Information and Software Technology 112 (2019) 68–82

Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof

Towards a reduction in architectural knowledge vaporization during agile global software development

Gilberto Borrego[a,*], Alberto L. Morán[a], Ramón R. Palacio[b], Aurora Vizcaíno[c], Félix O. García[c]

[a] Universidad Autónoma de Baja California, Mexico
[b] Instituto Tecnológico de Sonora, Mexico
[c] Universidad Castilla-La Mancha, Spain

ARTICLE INFO

Keywords:
Agile global software development
Architectural knowledge
Knowledge vaporization
Documentation debt

ABSTRACT

Context: The adoption of agile methods is a trend in global software development (GSD), but may result in many challenges. One important challenge is architectural knowledge (AK) management, since agile developers prefer sharing knowledge through face-to-face interactions, while in GSD the preferred manner is documents. Agile knowledge-sharing practices tend to predominate in GSD companies that practice agile development (AGSD), leading to a lack of documents, such as architectural designs, data models, deployment specifications, etc., resulting in the loss of AK over time, i.e., it vaporizes.

Objective: In a previous study, we found that there is important AK in the log files of unstructured textual electronic media (UTEM), such as instant messengers, emails, forums, etc., which are the preferred means employed in AGSD to contact remote teammates. The objective of this paper is to present and evaluate a proposal with which to recover AK from UTEM logs. We developed and evaluated a prototype that implements our proposal in order to determine its feasibility.

Method: The evaluation was performed by conducting a study with agile/global developers and students, who used the prototype and different UTEM to execute tasks that emulate common situations concerning AGSD teams' lack of documentation during development phases.

Results: Our prototype was considered a useful, usable and unobtrusive tool when retrieving AK from UTEM logs. The participants also preferred our prototype when searching for AK and found AK faster with the prototype than with UTEM when the origin of the AK required was unknown.

Conclusion: The participants' performance and perceptions when using our prototype provided evidence that our proposal could reduce AK vaporization in AGSD environments. These results encourage us to evaluate our proposal in a long-term test as future work.

1. Introduction

Agile and global software development (AGSD) is currently an important trend [1]. In fact, VersionOne of the 11th annual state of agile report[1] states that 86% of the respondents had distributed teams practicing agile software development (ASD). AGSD leads to many challenges, given the inherent nature of both paradigms: ASD and global software development (GSD). On the one hand, GSD communication is commonly based on documents, i.e.,explicit knowledge, that decrease the effect of the four distances of this paradigm (physical, temporal, linguistic and cultural) [2]. On the other, the agile manifesto [3] states that in ASD, face-to-face interactions are preferable to following a strict communication processes, and working software is preferable to comprehensive documentation, leaving the interpretation of the term "comprehensive" to each agile team [4]. In fact, ASD suggests that most documentation can be replaced by enhancing informal communication, i.e., a stronger emphasis on tacit knowledge rather than explicit knowledge [5]. However, prioritizing communication in ASD does not mean disregarding formal documentation [6]. This shows an internal antagonism within AGSD, since tacit knowledge is preferred in ASD (face-to-face interaction) and explicit knowledge (based on documents) is preferred in GSD.

In AGSD teams, tacit knowledge tends to predominate over explicit knowledge [6–8], leading to a lack of documents concerning architectural design, user manuals, data models, updated requirements specification, etc., known as documentation debt [9]. AGSD teams are affected by documentation debt, particularly when there is insufficient explicit

* Correspondence author.
E-mail address: gilberto.borrego@uabc.edu.mx (G. Borrego).

[1] https://explore.versionone.com/state-of-agile/versionone-11th-annual-state-of-agile-report-2.

https://doi.org/10.1016/j.infsof.2019.04.008
Received 23 June 2018; Received in revised form 22 March 2019; Accepted 17 April 2019
Available online 23 April 2019

architectural knowledge (AK). AK is composed of architectural design (including fundamental system concepts in its environment, embodied in its elements, relationships, and in the principles of its design and evolution [10]) and of the design decisions and rationale used to attain architectural solutions [11].

One of the main problems in GSD is generally the lack of explicit knowledge (including AK) when stakeholders attempt to resolve previously presented problems, especially when this occurs in small and medium companies [12,13]. The most significant causes of a lack of explicit AK in AGSD teams are: (1) the most popular agile methods,[2] Scrum and XP [14,15], specify AK in a very lax manner, leading to documents with informal notations [4]; (2) the inherent time pressures of ASD cause the omission of appropriate documentation [16], and (3) agile developers consider that documentation is a secondary and non-creative activity [17].

In co-located ASD, the lack of AK documentation is mitigated by developers' daily face-to-face interactions. In AGSD teams, however, the lack of AK documentation is often mitigated by communicating with remote teammates using unstructured textual electronic media (UTEM), such as emails, forums, comments boards, instant messenger, etc., mainly because UTEM reduce the language gap [18]. If remote teammates are unavailable or are unable to answer their questions, agile/global developers usually attempt to obtain answers by analyzing source code [19], which is time consuming. Furthermore, the knowledge obtained is generally unstructured, incomplete and inconsistent [20], which does not guarantee that the software will evolve as planned at design time.

Furthermore, literature reports that UTEM logs contain important AK for agile/global developers [19,21], but that is unstructured, inaccessible, dispersed and prone to be lost over time, i.e., prone to be vaporized [22]. Moreover, in AGSD teams, requirements and user stories are usually the only documented knowledge referring to software development tasks [19]; there are also informal diagrams, but they are created only as an aid to problem understanding and are, therefore, considered as disposable documents. Furthermore, agile/global developers usually attempt to obtain AK from UTEM logs to mitigate this lack of documentation [19]. However, the problem is that UTEM are not designed to search AK, and developers usually use more than one UTEM to share knowledge, signifying there is no single point at which to find AK. It is, therefore, important that agile/global developers have efficient means to access the AK in UTEM logs to reduce AK vaporization.

In this paper, we present the AK Condensation concept, conceived as a means to reduce AK vaporization in AGSD by taking advantage of the knowledge stored in UTEM logs, and by giving agile/global developers the means to search for the AK contained in the aforementioned logs at a single point. This concept was implemented in a tool evaluated by agile/global developers and students to determine its feasibility. The remainder of this paper is organized as follows: Section 2 presents the related works, while the concept of AK Condensation and its implementation are presented in Section 3. Section 4 describes the evaluation method, while Section 5 presents the evaluation results and Section 6 shows the threats to validity. Finally, a discussion of the results and our conclusion are presented in Sections 7 and 8, respectively.

2. Related work

2.1. Architectural knowledge management in agile and global software development

Knowledge management is currently an important part of any software development process. Dalkir [23] proposed that KM consists of creating/capturing, sharing/disseminating and acquiring/applying knowledge assets, where: creating/capturing refers to developing new knowledge from experience and/or explicit knowledge, and then coding the knowledge in an agreed format; sharing/disseminating refers to storing knowledge in a common repository, sending it to the appropriate people or sharing it during a training session, and acquiring/applying refers to the learning process and using new knowledge in practice, with the possibility of creating knowledge to start the cycle again. This definition could, therefore, be adapted to define AKM as the discipline of creating/capturing, sharing/disseminating and acquiring/applying a software process's AK assets. This adaptation is very close to Farenhorst and de Boer [24] AKM's definition, which states that the aim of AKM is to codify software architects' tacit knowledge explicitly in either structured or semi-structured knowledge bases.

KM is a challenge in AGSD [25–27], signifying that AKM is also a challenge. A critical part of AKM is the process of knowledge capturing, because the AGSD environment [16] and the agile developers' attitudes [17] cause documentation debt [9]. Since an AGSD environment leads to a lack of captured AK, then the AKM phases of sharing/disseminating and acquiring/applying are also affected, because AK is shared and acquired on the basis of inappropriate documentation or even tacit knowledge.

Several works address AKM in software engineering [28–32], in GSD [33–37], and even in ASD [6,38,39]; however, these works do not cover AGSD environments. We, therefore, conducted a systematic mapping review (reported elsewhere [40]), in which we identified nine approaches used to manage AK that were grouped into three areas: (1) artifact-based, (2) communication-based, and (3) methodology-based. The artifact-based documentation area refers to the use of software development support (repositories, wikis and groupware) to share AK, auto-generated documentation based on communication analysis (relating to emails and code repositories), and lightweight approaches to register architecture designs and decisions. The communication-based area refers to the use of videoconference and UTEM to discuss and share knowledge, and the use of smartboards or electronic displays to show information about project architecture. The methodology-based area refers to agile method modification by introducing an architecting phase or an architect role to manage projects' AK.

We additionally observed that the papers reviewed evenly support the three phases of the integrated KM cycle [23] (Capture/Creation – 35%, Sharing/Dissemination – 33%, and Acquisition/Application – 32%). We analyzed the cases of the Capture/Creation phase using the states of knowledge [41]: tacit knowledge, which is in the stakeholders' minds; documented knowledge, which is codified in an informal/ad hoc manner, and formalized knowledge, which is codified in a standardized structure. We observed that only 7% of all the papers report a formalized means of coding AK, 11% report a documented means, 4% report a tacit means, and 13% do not specify how AK is captured. Most of the papers reporting a way in which to capture AK employ a volatile means to do so, since AK remains tacit or is informally codified. AK could, therefore, lose meaning over time or in another context, and there is consequently a lack of adequate means to capture AK in AGSD environments that ensure the duration of AK.

2.2. Architectural knowledge management solutions based on social tagging

As stated previously, this paper proposes the concept of AK Condensation (see Section 3), implemented using a prototype based on tagging personal interactions using UTEM in real time, as a means to classify AK so as to ease its subsequent retrieval. Researchers and software companies have chosen social tagging as a lightweight and unobtrusive manner to organize unstructured or dispersed data or to add meaning and metadata to software development environments, to recover knowledge that is generally hard to find. To the best of our knowledge, seven

[2] https://explore.versionone.com/state-of-agile/versionone-11th-annual-state-of-agile-report-2.

Table 1
Tools that use social tagging in software development reported in literature (A=Analysis, D=Design, I=Implementation, T=Testing, M=Maintenance, Full=Cover all the phases). ArchiKCo refers to the prototype presented in this paper, which was designed for AGSD to cover any development phase and was focused on tagging UTEM interactions, where valuable AK is located. ArchiKCo tags are linked to base tags and include an auto-complete mechanism to ease the tagging action during the developers' interactions in order to avoid tag explosion. Finally, ArchiKCo has different parameters to perform AK retrieval, which other tools do not consider. See Section 3.2 for more ArchiKco details. *Parsed by TagSEA.

Tool name	Coverage			Tagging		Knowledge retrieval	Tool type
	Environment	Phases	Items to tag	Mechanism	Type		
IBM® Rational® Jazz® [44]	Agile distributed	Full	Artifacts and workitems	Auto-completed	Free	Free text and tags	Commercial
TagSEA [45]	Distributed	I	Sourcecode	Auto-completed	Free	Based on waypoints and tags	Research open source
Trac [69]	Distributed	I, T	Version control and Tickets (bugs)	Free	Free	Free text	Open source
eMoose [65]	Not defined	I	Source code	Auto-completed*	Free	Contextual "Push"	Research
CodeSnippets [70]	Not defined	I	Code Snippets in Source code	Free	Free	Based on tags	Open source
Paul et al. tool [71]	Open source	D, I	Software components	Free	Free	Free text associated with tags	Research
TAGGER [42]	Distributed	A	UTEM interactions	Free	Linked to base tags	Not reported	Research
ArchiKCo	Agile distributed	Full	UTEM interactions	Auto-completed	Linked to base tags	Free text, tags, dates, remittent, recipient, UTEM source	Research prototype

tools use social tagging (see Table 1). Most of these tagging tools are designed to support the implementation phase and are focused on tagging source code, software components or version control entries, i.e., they help manage AK. Only TAGGER [42] was designed to tag personal interactions in UTEM, but is focused on capturing domain knowledge during the analysis phase. Moreover, most of these tools are oriented toward a distributed development environment, and only IBM® Rational® Jazz® was evaluated in ASD. Our prototype, ArchiKCo (explained in Section 3.2), is shown in Table 1 in order to contrast its characteristics.

Regarding the implementation of tagging, only three tools have an auto-complete mechanism to aid during tag assignment. Moreover, most of them use free tags, i.e., there are no fixed or predefined tags to assign, and users are, therefore, free to write or compose any tag. Free tagging and unassisted tag assignment could lead to tagging difficulties and information retrieval problems caused by: (1) a huge number of tags, known as tag explosion; (2) differences in the interpretation of a tag's meaning; (3) an incomplete context in which to understand a tag; (4) the locality of tags, i.e., tags based on a team's jargon; (5) tags that only make sense when used together, known as composite tags, and (6) tags with the same meaning but written differently, known as obscure similarity [43].

Despite the above problems, literature reports that developers prefer using free tags because of their low cognitive load for everyday work [44,45]. Some efforts have been made to develop auto-tagging mechanisms [46] or tag-based recommender systems [43,47] to reduce developers' cognitive loads to an even greater extent. Sohan et al. [46] auto-tagging mechanism relates email messages to user stories in ASD projects with an accuracy of 70%; however, the remaining 30% of error could cause knowledge retrieval problems. Moreover, the tag recommender systems TagRec [47] and LS^{3}AutoTagger [43] are promising means to complement tag assignment and enhance the basic auto-complete mechanism.

Most of the knowledge retrieval mechanisms shown in Table 1 are based on tags and/or free text, except eMoose, which "pushes" AK to the users depending on the coding context. No tools except ArchiKCo base their knowledge retrieval mechanisms on dates, people and origins (i.e., a ticket, source code, or any other artifact). This is relevant, since agile/global developers struggle to find AK because they do not usually remember who originally provided it, or where and when a certain piece of AK was posted [19].

2.3. Architectural knowledge vaporization consequences in agile and global software development

The major challenges in AGSD are related to communication, culture, trusted relationships and KM [1,48]. A key success factor in any software development project is the correct appliance of KM [49], and consequently of AKM. As stated in Section 2.1, AKM is still a challenge in AGSD because most AK remains tacit or documented and could, therefore, lose meaning over time or in another context, i.e., it is prone to vaporization. It could be argued that AK vaporization in ASD is mitigated by practicing shared source code ownership [50], thus making developers aware of the project's AK. However, the four distances inherent in GSD cause inefficient AK sharing, since there are less opportunities for casual interaction [51] and informal awareness [52]. Shared source code ownership does not, therefore, have the same effect in AGSD.

AK vaporization could cause the following problems in AGSD [53,54]: (1) poorly understood requirements and technical solutions; (2) a lack of knowledge transfer between teams; (3) defects in software evolution and maintenance, i.e., architectural technical debt [55]; (4) a lack of visibility in project monitoring, and (5) time wasted by experts answering the same questions on certain issues and attempting to find solutions to problems that have already been solved. A consequence of the last point is that team members could annoy experts: constant questions could lead to an erosion in interpersonal relationships, which could affect the knowledge flow [56]. Interpersonal relationship erosion could be critical when building trusting relationships, which is important for any agile team. All the aforementioned problems and situations show the importance of addressing AK vaporization in AGSD. Our approach to mitigate this phenomenon is presented as follows.

3. A proposal of architectural knowledge condensation

As stated above, AK vaporization hinders the KM cycle because there is documentation debt. In a previous study [19], we found that UTEM logs contain valuable documented AK, and that developers attempt to recover it from those sources. Finding AK in UTEM logs is difficult because these media are not designed to find knowledge and because AK storage is unstructured. Moreover, it is difficult to retrieve AK because it is dispersed among different UTEM, which developers use to communicate. In this section, we present our contribution to AKM in AGSD, proposing a means of structuring and retrieving the AK shared using

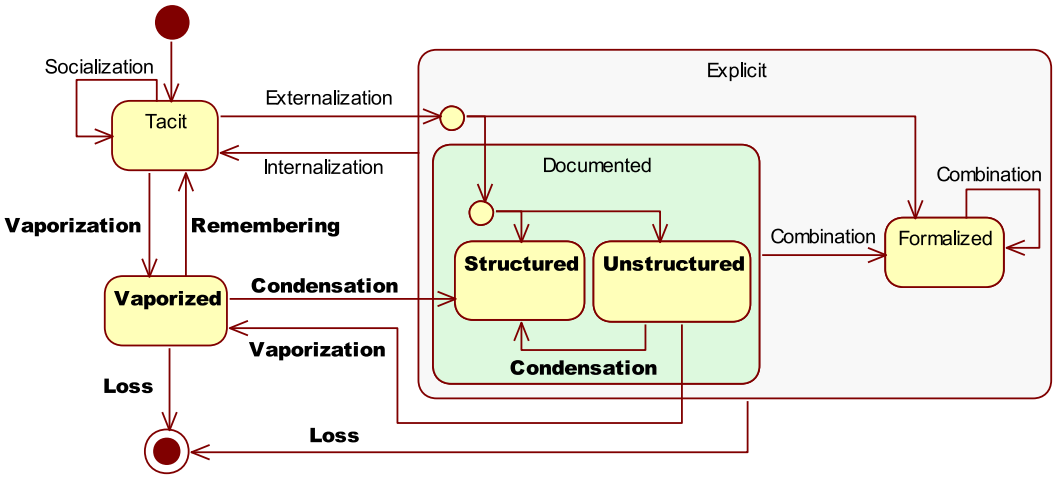

Fig. 1. UML state diagram representing SECI model with documented and formalized sub-states, and extended with the vaporization and condensation concepts. States and transitions in bold are product of our SECI model extension.

UTEM, in which AK structuring is based on a lightweight classification mechanism. This proposal is called AK Condensation – the opposite of AK vaporization. We additionally present a prototype that implements AK Condensation in order to evaluate the proposal's feasibility.

3.1. Conceptual definition

Explaining the AK Condensation concept implies exploring the vaporization concept in greater depth. Various authors define AK vaporization as the disappearance of AK owing to documentation debt [22,57,58]. However, we propose that vaporization could be a state just before AK disappears. In AGSD, developers attempt to find AK in UTEM logs when AK has evaporated from their minds. AK is, therefore, still recoverable because UTEM logs contain AK traces [19], which could help them infer/remember AK. In order to show our concept of AK vaporization and condensation, we extend the SECI model [41]. Fig. 1 shows the AK vaporized state, which may occur when developers forget tacit AK or cannot find AK in any kind of unstructured repository (e.g. UTEM logs). We propose that vaporized AK can be recovered when teammates help developers remember a piece of AK, or when unstructured AK or vaporized AK are gathered and structured to ease their retrieval (AK Condensation). Fig. 1 also shows that condensation is not a means to convert vaporized AK into AK in formal notation, but simply a step forward to ease AK formalization and reduce AK loss; we consider there still is a big gap between documented AK in UTEM and formalized AK. AK Condensation could, therefore, be implemented by considering the following elements:

1. **Accessible UTEM logs**. All stakeholders must be able to access all information from UTEM logs and thus be able to access all the AK being shared among them.
2. **UTEM log classification mechanism**. There must be a classification mechanism to structure the UTEM log information in order to ease AK retrieval. UTEM include features to find information in their logs. However, these features do not find AK efficiently [19]. In addition, developers do not usually remember the exact terms/concepts in which AK was shared and consequently need a semantic structure associated with the UTEM log to help them find AK without knowing the exact term to search for. In addition, this semantic structure would ease the transition from documented AK and formalized AK in later stages.
3. **AK searching mechanism**. All stakeholders could use the classification scheme to find valuable AK with less effort in the structured UTEM logs. The searching mechanism could include any other search parameter to ease AK retrieval, e.g., date period, message sender or message author.

Since AK Condensation is an abstract definition, there could be different ways to develop concrete instances. The following sub-section explains how we implemented a technological solution based on this concept.

3.2. Prototype of architectural knowledge condenser

In order to prove the feasibility of AK Condensation, we instantiated this concept using a technological solution called ArchiKCo, evaluated by professional developers and students (see Section 4). We based the ArchiKCo classification mechanism on social tagging, which can be applied during UTEM interactions. Social tagging is a lightweight and popular means to classify knowledge, which has been successfully used by other authors (see Section 2.2). Furthermore, in [59] we observed that social tagging is not a great effort for agile/distributed developers, and that they are interested in tagging UTEM messages in order to retrieve AK in the future.

We based ArchiKCo on Windows and Skype[3] (as the UTEM log source), since most of the subjects that were able to evaluate it use them both in their daily work. Fig. 2 depicts the ArchiKCo operation, showing the activities that implement the three elements of AK Condensation, along with the common situations described by agile/global developers (depicted as dialog clouds in Fig. 2); we explain how ArchiKCo implements each element below.

3.2.1. Accessible UTEM log information

We implemented this part using a Gatherer Service (see Fig. 2, part A) to periodically extract and send the Skype interaction logs to a shared repository in the cloud (depicted as a UTEM Messages database in Fig. 2). We used Algolia[4] server as a shared repository, since it provides robust indexing functions that ease the development of a searcher.

3.2.2. UTEM log classification mechanism

In order to avoid the problems related to free tagging [43], we implemented a semi-fixed tagging mechanism (successfully evaluated in [59]) that allows developers to add user tags with a web application called Tags Administrator (see Fig. 2, part B). We propose that developers perform this activity during a development cycle planning meeting,

[3] https://www.skype.com/.
[4] https://www.algolia.com/.

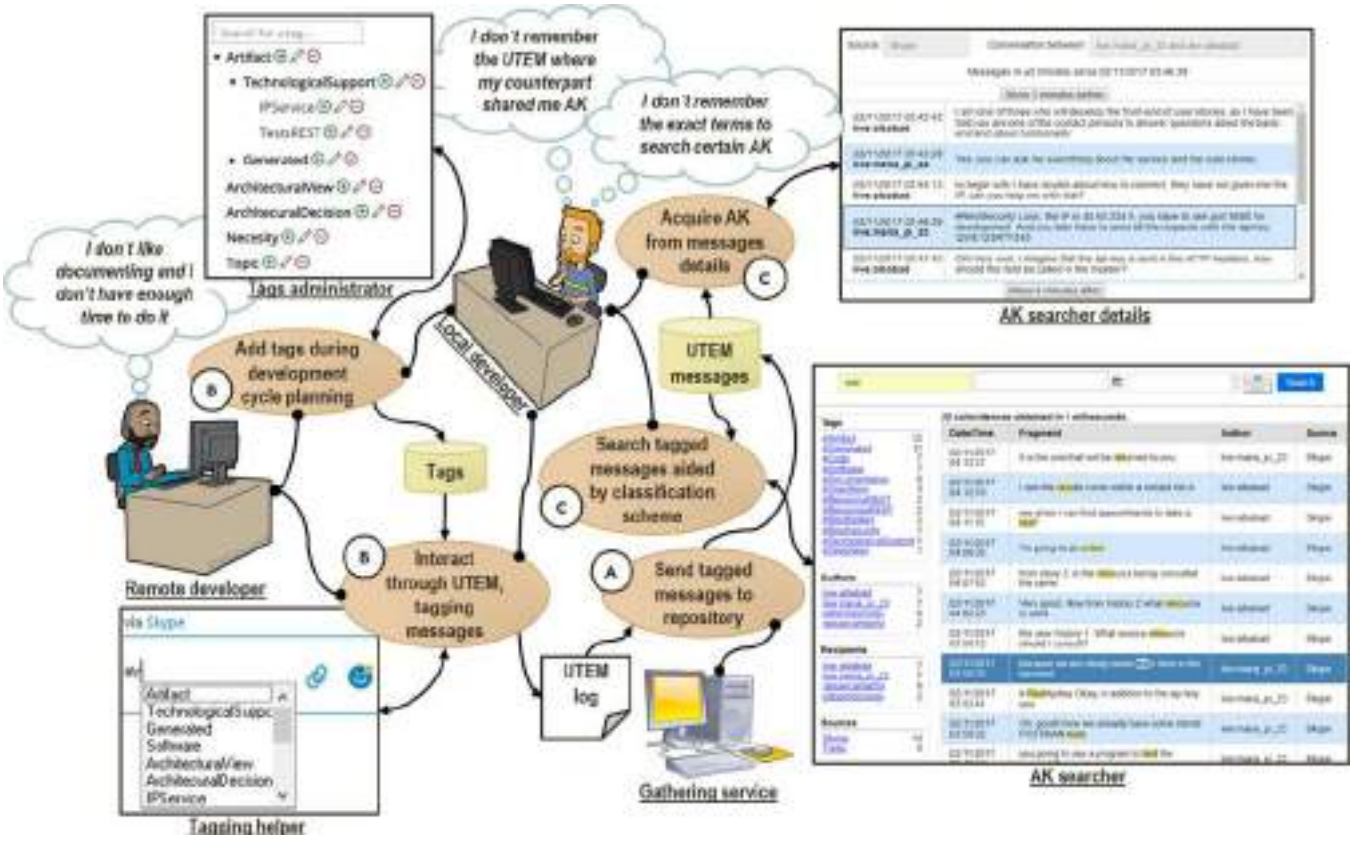

Fig. 2. ArchiKCo rich picture with activities (ovals) corresponding to the three elements of AK Condensation concept, A = Accessible UTEM logs information, B = UTEM logs Classification mechanism, C = AK searching mechanism. Bulleted lines represent links between activities and who performs them. Arrowed lines represent links between activities and artifacts. There are three types of artifacts: resulting artifacts (activities' outgoing arrows), source artifacts (activities' ingoing arrows), and interacting artifacts (linked with double arrow lines).

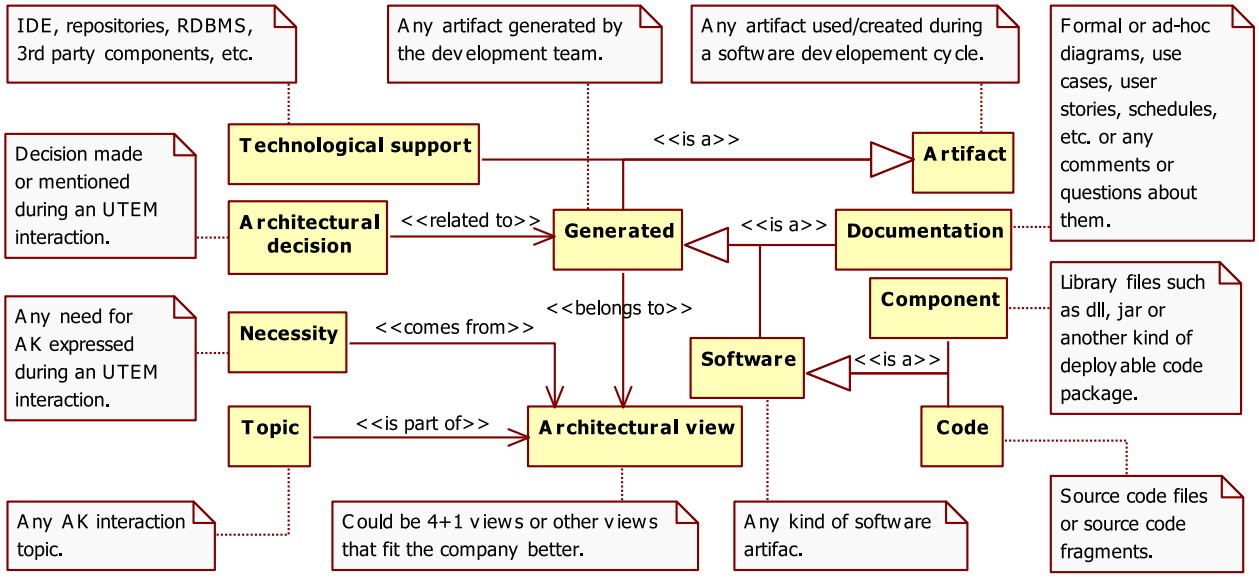

Fig. 3. Conceptual model based on UML, representing the aspects involved in AK articulation through use of UTEM by AGSD teams (reproduced from [19]).

since they would already know the key terms to be used during the next cycle. Every custom tag must be related to a meta-tag from an AK Model, which represents the AK that developers share; this model was empirically obtained in [19] (see Fig. 3). The aim of relating user tags and meta-tags is to provide an abstract means to store and find AK, signifying that when developers wish to recover AK and they do not remember the exact name of a tag, they can use a meta-tag to conduct an initial search.

Once the development cycle has started, remote and local developers could interact using UTEM, tagging messages that contain important AK. We are aware that developers sometimes make typing errors, or forget the exact way in which each tag was registered, or even the existence of certain tags, and have, therefore, developed a tagging helper component (see Fig. 2, part B) that auto-completes tags while developers are typing in conversations, whose source was the tag repository (depicted as a

Tags database in Fig. 2), which was updated using Tags Administrator. We thus aim to reduce the number of typing errors, ensure that users are using the exact tag writing and reduce tagging problems [43].

3.2.3. Architectural knowledge searching mechanism

We developed a web-based searcher, called AK searcher (see Fig. 2, part C), which has three search parameters: (1) free text, in which users can input any text to be searched for in the message content, along with the names of the author and recipient of the message; (2) Date range, from which users can select the period when the knowledge was shared; and (3) Tag filter, from which users can select the tags from a tag tree, on which meta-tags and user tags are hierarchically organized. This enables users to remember which tags are available and then add any number of them to the tag filter.

Having executed a search, a list of the coincident messages is shown, along with four panels (see Fig. 2, part C), which show all the related authors, recipients, sources (UTEM from which the messages originated) and tags of the resulting messages. Developers could apply extra filters to narrow the results, by clicking onto the panels' elements. The tag panel also shows each tag's parents; for example, if there is a tag called #nodeMongoDB, which is related to the #Code meta-tag, the tag panel shows both. We are aware that obtaining AK from a single tagged message could be difficult. AK searcher includes a feature to obtain the interaction context of a selected message, in which developers can read messages that were sent five minutes before and five minutes after a certain message (see Fig. 2, part C); they can even load messages from an additional five minutes before or after, if necessary.

3.3. Architectural knowledge condensation in agile and global software development

Instantiating the AK Condensation concept would give agile/global developers a lightweight means to structure AK (with a low cognitive load) while they are interacting with UTEM (maintaining agility), and an easier means to retrieve dispersed AK from UTEM logs. Our proposal could consequently reduce AK vaporization in AGSD environments. However, before implementing AK Condensation in a real scenario, we must first determine its concept feasibility in a controlled environment. We determined this concept feasibility by observing two key phases: AK structuring and AK retrieval. The following research questions arose from the latter:

- **RQ1.** Is an assisted semi-fixed tagging mechanism suitable to structure AK and avoid tagging explosion during UTEM interactions in AGSD environments?
- **RQ2.** Is it better to search for AK using the ArchiKCo searcher than by directly using the AK sources (UTEM) in AGSD environments?
- **RQ2.1.** Is ArchiKCo searcher trustworthy as regards finding correct AK?
- **RQ2.2.** Do developers find the correct knowledge faster using the ArchiKCo searcher than directly using the AK sources (UTEM)?
- **RQ2.3.** Is the ArchiKCo searcher preferable when searching for AK rather than directly searching in the UTEM source in AGSD environments?

The method employed to determine the feasibility of the AK Condensation concept using the ArchiKCo prototype is presented below.

4. Method to evaluate architectural knowledge condensation feasibility

4.1. Scoping

As stated above, AK structuring and AK retrieval are the key phases of the AK Condensation concept. We, therefore, designed a two-part evaluation to determine the feasibility of the concept presented. We define these parts below.

In [59], we observed that social tagging could be a lightweight manner to structure AK, using a semi-fixed and assisted tagging mechanism. In that study, we added a tagging helper to a web-based messenger, developed ex-professo, thus giving us full control over the tagging environment, allowing us to ensure that participants only used registered tags. The ArchiKCo tagging helper has now been added to Skype, signifying that we do not have sufficient control to avoid unregistered tags. In order to answer RQ1, the first part of evaluating AK Condensation comprised an observation study focused on tagging behavior in terms of registered (valid) tags and obtaining participants' perceived usability of the tagging mechanism.

We believe AK Condensation could enhance AK searching in UTEM logs during AK retrieval. We, therefore, designed the second part as a quasi-experiment to compare participants' searching performance when using ArchiKCo and two UTEM: Skype and Trello (RQ2). This section presents the method employed to determine the feasibility of AK Condensation, which was structured following Wohlin et al. [60] specification.

The objective of the whole study is to *Analyze* the use of the implementation of the concept of AK Condensation, *for the purpose of* determining its feasibility *with respect* to tagging behavior and AK retrieval, *from the point of view of* professional developers and students, *in the context of* AGSD.

4.2. Planning

4.2.1. Context selection

We had two experimental contexts: industrial AGSD and academia (replica). The participants in the industrial context were professional developers from seven Mexican companies: four small[5] (<100 employees) and three medium[5] (100–999 employees). The participant companies develop software for diverse areas (transportation, health care, internet of things, administration in general, etc.), have worked in a distributed or global environment, and all work in an agile manner. The main objective of working in an industrial context was to attain richer qualitative feedback about the AK Condensation concept. The academia replica took place at Castilla-La Mancha University in Spain (Spanish acronym, UCLM) with undergraduate and graduate students from the Superior School of Computer Science, all of whom had knowledge of AGSD.

4.2.2. Selection of subjects

The subjects of both contexts were chosen for convenience. The academia subjects were 3rd year undergraduate students, who had already studied subjects regarding agile methods, programming and software design. There were also graduates researching topics related to software development, who consequently also know about agile methods and software design. The industry subjects were professionals who have worked on AGSD projects.

4.2.3. Study design

This study had a within-subject design, since all the treatments were applied to all the participants. It consisted of two parts: AK structuring and AK retrieval. The first part was an observational study during which all the participants were mentally situated in a context scenario to interact in pairs using Skype and the ArchiKCo tagging helper (with predefined user tags) and following a chatting script containing seven marks suggesting what to tag. Since the participants did not register the user tags in the catalog, they were free to assign unregistered tags if they did not find one that fitted a certain message.

In the AK retrieval part, the participants had to answer a 12-question survey concerning the context scenario in an attempt to emulate AK needs. The survey answers were stored in the Skype log generated in the previous part, and on a Trello board that contains user stories and

[5] https://www.gartner.com/it-glossary/smbs-small-and-midsize-businesses/.

comments related to the same context scenario. We chose Trello as a second UTEM because it is easy to use and commonly used by the participants. Eight survey questions indicated which media to use to search for the answer: Skype, Trello or AK Searcher, which contains the logs of both UTEM. The participants were free to choose the media they preferred to search for answers to the last four questions, which were designed so that the answers to questions 9 and 10 could be found using Skype or AK Searcher, and the answers to questions 11 and 12 could be found using Trello or AK Searcher. We consider that this part was a crossover quasi-experiment with one factor and three treatments, in which only two comparisons are relevant: AK Searcher/Skype and AK Searcher/Trello.

4.2.4. Variables selection

In the AK retrieval part, the independent variable was represented by the different media used to search for AK. There is no independent variable for the AK structuring part, since it is an observational study. The dependent variables were: tag validity, i.e., number of registered and unregistered tag instances used by participants during the chatting session; media preference, i.e., percentage of time a participant used a certain media to search for AK; correctness of answers, percentage of correct AK found per media; and time required to find correct knowledge.

4.2.5. Hypotheses formulation

In this part, we present the four study hypotheses, which are directly related to the dependent variables defined above.

- $\mathbf{H_{0TagsValidity}}$: There is no difference between the number of valid and invalid tag instances used during the UTEM interaction.
- $\mathbf{H_{0Preference}}$: There is no difference as regards to the preference to search for knowledge using any of the media provided.
- $\mathbf{H_{0Correcteness}}$: There is no difference in the percentage of correct answers found using any of the media provided.
- $\mathbf{H_{0Time}}$: There is no difference in the time required to find correct answers using any of the media provided.

4.2.6. Instrumentation

Below, we present the instruments developed in order to conduct this study as it was designed.

- **Context scenario**. This scenario concerned two agile developers from different companies and locations working on the same project (medical appointments system), one of whom required information about a RESTful service that the other was developing. They had documentation debt, and consequently had to acquire the project AK by asking each other questions.
- **Chatting scripts**. Each pair of participants had to follow 2 scripts (one per scenario role) to simulate a technical conversation taking place using Skype regarding the context scenario.
- **SUS questionnaire**. We prepared a questionnaire based on the System Usability Scale [61] (SUS)using a Likert-7 scale and focused on the Tagging Helper. We added two SUS-style questions (one positive and the other negative) to explore the participants' perceptions of the helper's unobtrusiveness. This questionnaire also included an open-ended question to request suggestions regarding the Tagging Helper.
- **Extended TAM questionnaire**. We prepared a questionnaire based on the Technology Acceptance Model [62] (TAM) using a Likert-7 scale and focused on the AK Searcher. We added questions concerning reductions in interruptions (one question), finding relevant AK easily and in a timely manner (three questions) and the participants' overall impression of the whole ArchiKCo prototype (two questions).
- **12-questions survey**. This survey was uploaded onto LimeSurvey.[6] To compare the participants´ performance using AK Searcher versus Skype and Trello, we created two survey versions (one for each pair member), in which we varied the indicated media per question. For instance, while one pair member was required to answer a question using Trello, the other pair member was required to answer the same question using AK Searcher; and the same between Skype and AK Searcher.

4.3. Operation

4.3.1. Preparation

We deployed an ArchiKCo instance for each pair of participants and registered 10 user tags linked to different meta-tags related to the context scenario. The user tags were: IPService, TestsREST (related to TechnologicalSupport meta-tag); RestApikey, RestSecurity, Encryption, TestData, RESTResource, RESTResponse (related to Code meta-tag); AngularEncryption (related to Component meta-tag); and UserStory (related to Documentation). We pre-defined tags because each participant pair could define a different set of tags, which hindered the tagging behavior analysis. We additionally carried out a pilot test, which showed that the pre-defined tags really accorded with the context scenario. We also added five cards to a Trello public board on which fictitious members of the development team provided user story clarifications. Finally, we activated the two versions of the 12-question survey.

4.3.2. Execution

We first carried out the study in the industrial context with 30 professionals (average age = 28, SD = 3.9), 10 from the medium-sized companies, and the rest from the small ones. The industry participants had experience in ASD (average of 3.1 years' experience, SD = 1.9) and in GSD (average of 2 years' experience, SD = 1.4). The study in UCLM was carried out three months after the industry study, with 30 students (average age = 24.1, SD = 3.5): four graduates and 26 undergraduates. The experiment took place in three sessions per week for both contexts, so not all participant began on the same day. These sessions are explained below.

- **Installation session** (duration ≈ 10 min). The participants were given an overall explanation of the study sessions along with their objectives. We organized the participants into pairs and then helped them configure the tagging helper (to work with Skype) and the Skype extractor (to send each pair's conversations to a shared server).
- **Solving scenario session** (duration ≈ 25 min). We gave the participants a short training session regarding how to use the tagging helper (three minutes, approx.), and they quickly explored the available tags (two minutes, approx.). We then described the scenario in which they would be located to carry out the tasks, and assigned a role to each pair member: either the developer working on the RESTful service or the developer who wished to use it. Each member of each pair sat in a different part of the session room, ensuring they had no visual contact, as if they were geographically distributed. We asked them to avoid talking to each other to better emulate an environment of geographic distribution. They then used Skype to chat, following the corresponding script, and tagging aided by the Tagging helper. We also told them that they could tag any message as they considered necessary, and that they could write a new tag (unregistered/invalid tag) if they could not find one that fitted a certain message on the options shown by Tagging helper. After the participants had finished following the chat script, they answered the SUS-based questionnaire.

[6] https://www.limesurvey.org/.

- **Searching session** (duration ≈ 20 min). This session took place two days after the chatting session to prevent the participants from being able to remember the chat topics, thus mitigating the learning effect. The participants were required to answer the 12-question survey easily. The version of the electronic survey was assigned to each pair member randomly. The participants were trained to search in the three different media, after which we explained that each question indicated where to search for the answer, along with the fact that they could answer "I don't know" if they were unable to find any information. They then responded to the corresponding electronic survey and finally to the extended TAM questionnaire.

4.3.3. Data collection

The tags' validity was determined by obtaining all the tagged messages and comparing them with the tags catalog (user tags and meta-tags) to obtain the number of valid and invalid tags per participant. Regarding media preference, each 12-question survey had a field for the last four questions in which the participants indicated the media used to obtain the answer. We, therefore, counted only the number of answers for each media. Correctness of answers was obtaining by manually checking each one. The time required to find the correct knowledge was measured using LimeSurvey, which registers the time that has elapsed between the presentation of a question and that at which the participant clicks onto the next button to pass to the next question. Finally, we obtained the qualitative perception about Tagging Helper and AK Searcher using the results of the SUS and TAM questionnaires, respectively.

4.3.4. Data validation

The AK retrieval part of the industry context was conducted in the respective participant companies. We could not avoid interruptions during the session and the time required to find the correct knowledge was consequently affected. We, therefore, take into account only the time data from the academic context.

5. Results

The results obtained are presented in three parts: (1) the results of the AK retrieval part including the Tagging helper usability perception; (2) those of the AK retrieval part including the participants' perceptions of AK Searcher; and (3) the participants' overall perceptions of ArchiKCo.

5.1. AK structuring part

In this section, we analyze how the participants tagged messages using the tagging helper and its usability perception. During the tag analysis, we also identified tag instances that were used correctly in terms of their semantics. The results are consequently presented in terms of tag validity, tag correctness and Tagging Helper usability and unobtrusiveness.

5.1.1. Tags validity

Fig. 4 shows that most of the professionals (80% approx.) used between six and nine tag instances during the chatting session, while the UCLM students (80% approx.) used between three and eight tag instances. A considerable percentage of the participants (50% approx.), therefore, used valid tags as required, or more, i.e., at least seven tagged messages. Moreover, some participants used 12 or more valid tag instances (13% approx.), i.e., at least five instances more than expected. We can, therefore, interpret that they had the initiative to tag messages when this was not suggested in the script.

Upon considering invalid tag instances, 23% of the professionals did not use invalid tags, while only 6% of UCLM students did not do so (see Fig. 4). Furthermore, around 73% of the professionals used between one and three invalid tag instances, while 83% of the UCLM students used between one and six invalid tag instances. Invalid tags also represent new tags, and in this respect, 53% of the UCLM students used between one and three instances of new tags, while only 23% of the professionals used one new tag instance (see Fig. 4). However, 33% of the UCLM students used between four and eight new tag instances, indicating their disposition to tag messages or their inexperience with the script topics, signifying that they had to create new tags that would fit their knowledge. Fig. 4 also shows that around 50% of all the participants made no typing errors in the tag instances, and that 46% of the participants had only one or two "typos". This could mean that the Tagging Helper really helped them obtain a low error rate.

5.1.2. Tagging correctness

Around 13% of all the participants had no semantic errors when using tag instances, and 40% used only one or two tag instances erroneously (see Fig. 4). This is significant, because the participants did not know the exact semantics of the tags beforehand. However, 50% of the professionals had between three and five instances of incorrect use, while only 23% of the UCLM students did so in the same range. Regarding the correctly used tag instances, 78% of all the participants had between two and six correct instances, and the professionals registered more variability than the UCLM students (see Fig. 4). Finally, around 13% of the participants registered between eight and ten correctly used instances, again highlighting that they did not know the tags in advance.

5.1.3. Overall tagging behavior

To summarize the tagging behavior (see Fig. 5), the participants used more valid tag instances (75%) than invalid ones (25%), and more tag instances were used correctly (47%) than incorrectly (28%). Both differences were confirmed statistically using the Wilcoxon Signed-Rank Test ($\alpha = 0.05$): valid vs. invalid instances – p-value = 0; correct vs. incorrect instances – p-value = 0.000008. Moreover, there were 18% of new tag instances and only 7% of typing errors, which was also statistically significant using the same test (p-value = 0.0056). This signifies that the invalid instances resulted more from the need for new tags than from errors caused by the tagging mechanism.

We also noticed that 38% of the correct tag instances were unexpected (see Fig. 5), i.e., tag instances that fitted the message's semantics but that the participants were not expected to use, or that instances were even used in messages in which we did not suggest tagging. Unexpected tags comprise 26% of user tags and 12% of meta-tags. These meta-tags were: Technological Support, Component, Software, Code, and Necessity. These meta-tags would, therefore, appear to be intuitive, since we did not explain their meaning.

5.1.4. Tagging helper usability and unobtrusiveness

We obtained the scores from all the SUS questionnaires and transformed them according to the curved grading scale [63], including the two questions regarding unobtrusiveness. Tagging helper obtained an averaged SUS score of 77 (SD = 13, med. = 78 = B+), corresponding to a B grade, and represents a good usability perception [64]. Both sets of participants had similar usability perceptions. However, around 20% of the professionals had lower usability perceptions (grades D and F). This was mainly owing to the mechanism employed to select a tag and navigate through the suggestion list; it was necessary to press the <Alt> + arrows to navigate, and the <Alt> + <Enter> to select a tag. Moreover, both sets of participants had similar unobtrusiveness perceptions; we obtained an average unobtrusiveness score of 72 (SD = 20, med. = 83 = A), which corresponds to a C+ grade. In this case, 30% of the UCLM students and 17% of the professionals considered that Tagging Helper would be obtrusive in their daily work. Upon analyzing the participants' comments, we concluded that this was caused by the navigation and selection mechanism and also by the low availability of tags, since the participants did not create/register the tags.

In summary, we observed a significant rate of valid tag instances used correctly during the chat session, with a low rate of typing errors. This behavior could lead to a reduction in tag explosion and its

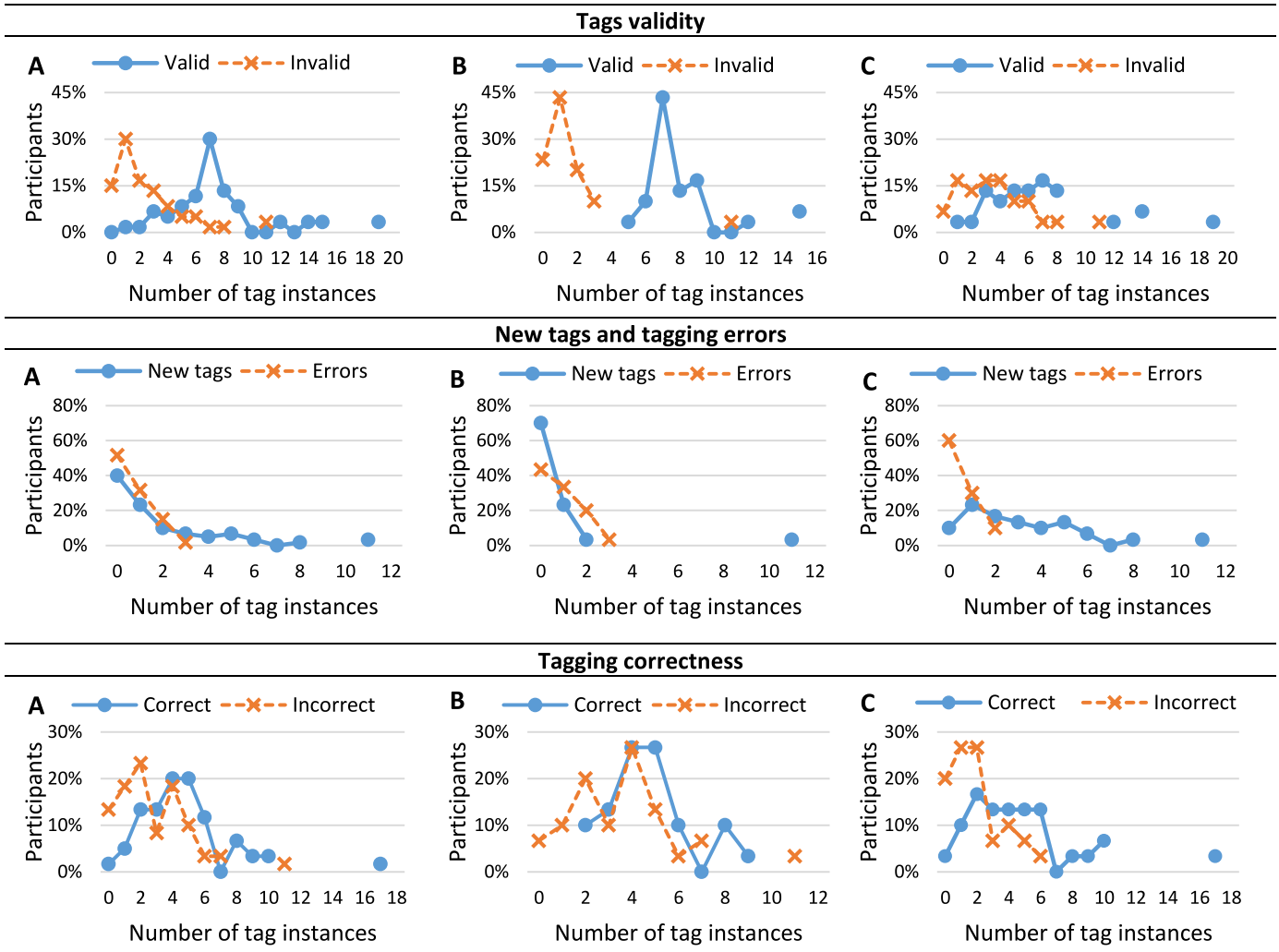

Fig. 4. Tagging behavior expressed by tag instances per participant.A = Mexican developers and UCLM students, B = Mexican developers only, C = UCLM students only. Outliers correspond to identified participants (< 2) who used many valid/invalid tag instances, or many new tags instances, or many correct/incorrect tag instances.

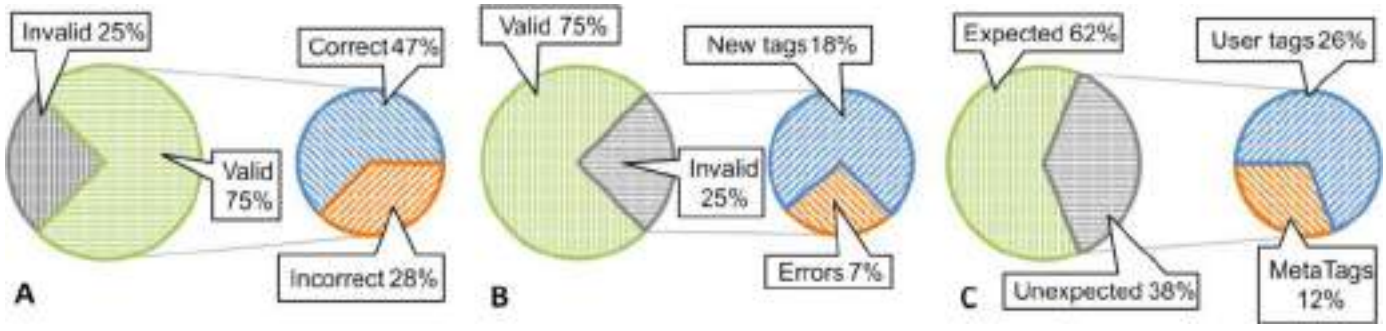

Fig. 5. A = Distribution of all tag instances (valid and invalid), detailing valid tags. B = Distribution of all tag instances (valid and invalid), detailing invalid tags. C = Distribution of correct tag instances (expected and unexpected), detailing unexpected.

related problems. Moreover, Tagging Helper has a great chance of being adopted to structure AK, since it is perceived as usable and unobtrusive.

5.2. AK retrieval

5.2.1. Media preference results

The participants preferred AK Searcher to Skype and Trello when answering all the questions (see Fig. 6, part A). There were questions to which the participants could not find the answers and did not, therefore, register a preferred media. Focusing only on the answered questions, and considering both participant profiles (professionals and students), AK Searcher and Skype obtained preferences of 69% and 31%, respectively, by joining the preferences attained for questions 9 and 10. AK Searcher and Trello similarly obtained preferences of 71% and 29%, respectively, when joining the preferences attained for questions 11 and 12. Since using the data obtained for each question was insufficient for the use of a paired test such as the Wilcoxon Signed-Rank Test, we applied goodness of fit tests based

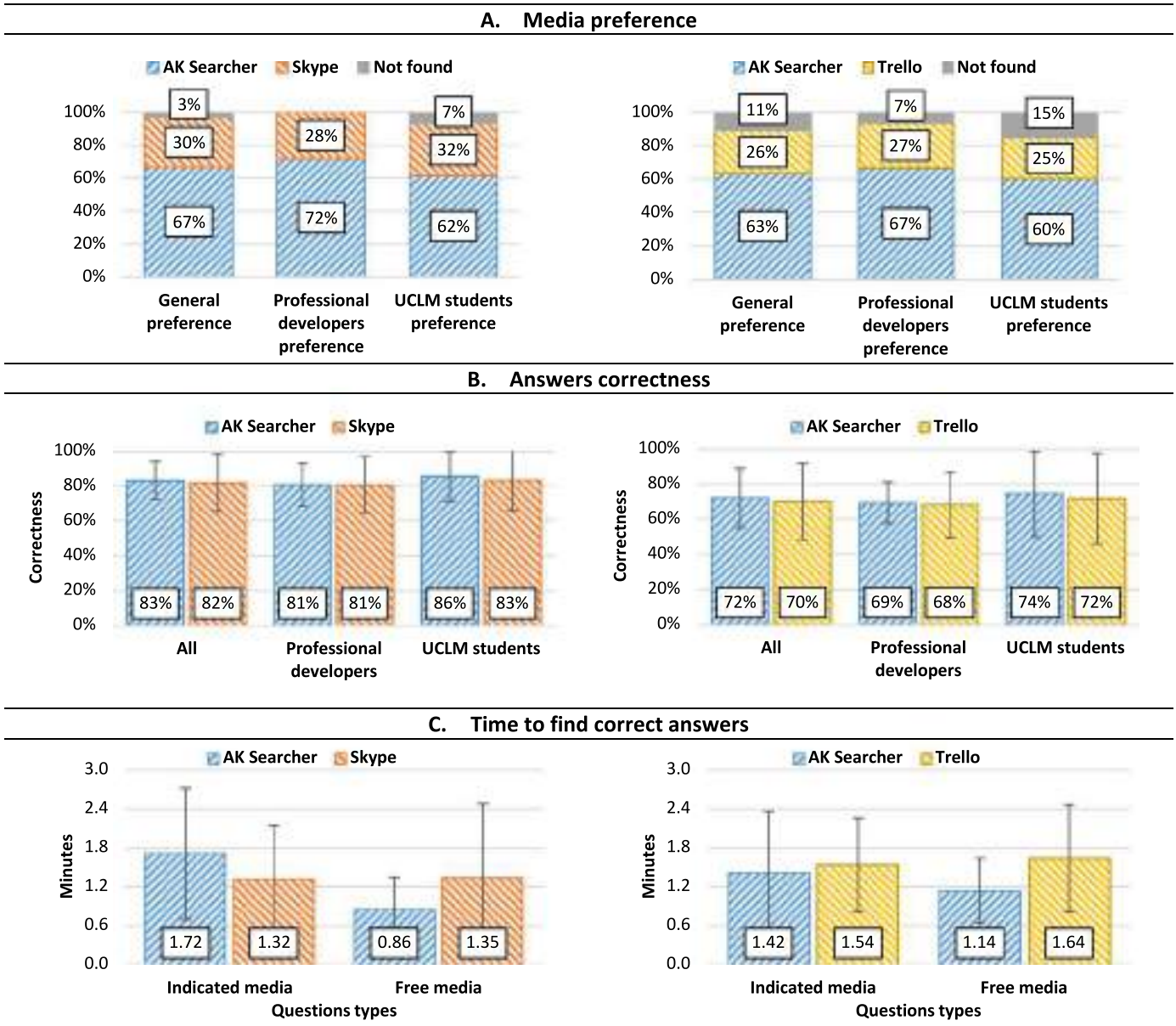

Fig. 6. A = Media preference of the participants as regards questions in which they had free choice. B = Comparison between correctness of answeres obtained using AK Searcher and Skype. C= Answering average times per question type. Vertical lines in B and C represent standard deviation.

on $X^2(\alpha = 0.05)$, supposing a uniform distribution for media preference. There is sufficient evidence to state that the participants' preferences are not uniformly distributed (AK Searcher/Skype questions: $X^2 = 16.69$, p-value < 0.001; AK Searcher/Trello questions: $X^2 = 18.925$, p-value < 0.001) and that there is, therefore, a tendency to prefer AK Searcher in all cases. We can, therefore, reject the null hypothesis $\mathbf{H_{0Preference}}$ since there was a significant difference in the preferred media.

5.2.2. Correctness of the results

Fig. 6 (part B) shows there were cases in which participants using AK Searcher obtained a higher correctness of answers than when using Skype, while there were others in which they obtained a lower correctness of answers than when using Skype, and yet others in which the correctness of the answers was the same for both media. Furthermore, both sets of participants behaved in a similar way when using either AK Searcher or Trello (see Fig. 6, part B). However, in this case, AK Searcher appears to have obtained a higher correctness of answers in more cases. We grouped the correctness data regarding questions 1–4, 9 and 10 (AK Searcher vs Skype), and questions 5–8, 11 and 12 (AK Searcher vs. Trello) of both participants' profiles to obtain two paired sets of data (AK Searcher vs. Skype and AK Searcher vs. Trello), each of which contained 12 elements, i.e., the results obtained for six of the students' questions and six of the professionals' questions. We then applied the Wilcoxon Signed-Rank Test ($\alpha = 0.05$) to both sets, and obtained that there is no difference among the correctness of answers when using AK Searcher or Trello ($W = 15 > W_{\alpha=0.05} = 3$), or using AK Searcher or Skype ($W = 30 > W_{\alpha=0.05} = 10$). It is not, therefore, possible to reject the null hypothesis $\mathbf{H_{0Correcteness}}$.

5.2.3. Time to find correct answers

Fig. 6 (part C) shows the time required to obtain answers with AK Searcher was higher than that required when using Skype for most of the questions in which the media required to search for the answer was indicated (questions 1–8). However, when the participants were free to choose any media (questions 9–12), less time was required to

obtain answers with AK Searcher than with Skype and Trello. We grouped the data by media and type of questions (free media choice or indicated media) such that we obtained four groups: (1) indicated media – Skype vs. AK Searcher, (2) indicated media – Trello vs. AK Searcher, (3) free media choice – Skype vs. AK Searcher, and (4) free media choice – Trello vs. AK Searcher. We determined whether there was a statistical difference within these groups by applying a Mann–Whitney U Test ($\alpha = 0.05$), because the samples are not paired, since we consider only the time required to find the correct answers. In the case of questions with an indicated media, there is a large difference in the answering time (p-value = 0.009) between AK Searcher and Skype, indicating that the participants found the correct answers faster when using Skype. In the same case, there is no considerable difference between AK Searcher and Trello (p-value = 0.153), although the participants were, on average, faster when using AK Searcher. In the case of questions with a free media choice, there is a large difference in the answering time (p-value = 0.045) between AK Searcher and Trello, indicating that the participants found the correct answers faster when using AK Searcher. However, there is no considerable difference between AK Searcher and Skype (p-value = 0.254), although the participants were, on average, faster when using AK Searcher. We can, therefore, reject the null hypothesis $\mathbf{H_{0Time}}$, since there were cases in which the participants were faster when using AK Searcher and others in which they were faster when using Skype.

5.2.4. Extended TAM results

The results of the extended TAM questionnaire showed that AK Searcher is extremely useful (median = mode = 6) and easy to use (median = mode = 6). The participants perceived that AK Searcher could help them find important AK in a timely manner (median = mode = 6). They also perceived that interruptions could be reduced using AK Searcher (median = mode = 6), since they would have a source of AK other than that of their teammates. It was also perceived that AK Searcher could ease the discovery of AK during development cycles (median = mode = 6).

5.2.5. AK retrieval results summary

AK Searcher was greatly preferred by the participants when searching for AK, which is supported by the perceptions presented above. Our results also provide evidence that professionals may perform better when locating knowledge if they use AK Searcher rather than Trello or Skype when they do not know the knowledge source. Finally, AK Searcher could be trustworthy when searching for knowledge, since the participants obtained a high percentage of correct answers and there was no significant difference between this percentage and that obtained with Skype or Trello.

5.2.6. Lessons learned by using Archikco prototype

We included a question in the TAM questionnaire to obtain a rating for the ArchiKCo prototype: a median of 8 (mode = 8) on a scale of 10. We were additionally able to learn some lessons from this study, which we grouped into AK structuring, AK retrieval and AK Condensation.

5.2.7. AK structuring

The participants stated that they have to get used to tagging their conversations. However, we believe that getting used to tagging could be easy because people are currently used to tagging in social networks. Furthermore, the participants suggested some enhancements to Tagging Helper: an automatic tagging or smart tag suggestion depending on the conversation topics, adding tags during the conversation, and selecting tags from a trending topic list. Finally, some participants struggled with the mechanism employed to browse tags in Tagging Helper, which was affected by Skype´s keyboard functions. This mechanism could, therefore, vary regarding the UTEM selected.

5.2.8. AK retrieval

Some participants commented that tags were not relevant during AK retrieval; one participant stated, "*I would end up not using labels, since I would look for words, not for labels*". In fact, while the participants were using AK Searcher, we observed that most of them preferred to search using free text, but some of them used tags to refine the results.

Regarding detail browsing interactions, this should change depending on the type of media: synchronous or asynchronous media; for instance, in asynchronous media, the time that elapses between messages could be more than five minutes, which is the time window that searchers have configured by default. We discuss this topic at greater length in the discussion section.

We also noticed that there should be a means to exclude informal messages from the repository. In this respect, one participant said AK Searcher "*could cause a lot of distraction because really important conversations are mixed with personal conversations, jokes, etc.*"; this situation could be problematic for AK retrieval.

5.2.9. AK condensation

Most of the participants commented that AK Condensation could speed up AK retrieval, reduce interruptions and reduce the repetition of information among teammates. The participants also appreciated that AK Searcher could be a single point of reference, rather than searching in multiple sources multiple times. In this respect, one participant stated: "*…notes and requirements are not discussed in the same place… there are conflicts because not all the parties have access to this information all the time.*"

6. Threats to validity

In order to understand to what extent the results are valid and how they can be used, a discussion regarding the validity threats is presented below according to Wohlin et al. [60] specification.

6.1. Conclusion validity

This study comprised participants from different backgrounds, but all of them were familiar with Skype and Trello. However, in order to balance the participants' knowledge, they received a brief amount of training regarding how to search for information in these media. We are aware that the participants could have been biased toward AK Searcher, because it was introduced as a new tool, or because they might have wanted to please us; however, their participation was anonymous. Moreover, we did not have previous contact with them before or after the evaluation and there was, therefore, no reason to try to please us. Moreover, the researchers were not close to the participants during the tasks in which they were free to choose one of the three available media. Furthermore, we are aware that the evaluation period was, perhaps, short. However, the results indicate an initial trend. It is also significant that, despite the short time and the use of a new tool, (people do not generally like to change their way of working) the subjects preferred AK Searcher when they did not know the location of a certain piece of knowledge.

6.2. Internal validity

All the participants were volunteers and showed a great interest in collaborating in this study. In addition, the study sessions were short to prevent them from getting bored or tired. We attempted to avoid learning effects by using a counterbalancing technique, i.e., we placed the participants in groups and presented the conditions (indicated media to search) to each group in a different order (see Section 4.3.2). Regarding persistence effects, the study was run with subjects who had never taken part in a similar study. Moreover, the participants did not have any previous knowledge of the context scenario, since it was fictitious. Furthermore, we conducted the searching session two days after the

chatting session to avoid the situation of the participants remembering all the details about the script topics. In order to clearly observe the results of the treatments on the participants' performances, they received the same set of questions to be searched for in the three media. All the questions could be answered using the indicated media, thus reducing the risk of the participants not being able to find the correct answer. The Wilcoxon Signed-Rank Test confirmed that there was no significant difference in the correct answers according to the media used, and the results are consequently independent of the study package.

6.3. Construct validity

We measured the time required to answer a question using a Limesurvey feature, which registers the time between a question being shown and the participant clicking onto the button to show the next one. This time could have been affected by the participants' reading speed and by the time needed to understand the question. We considered that the participants had similar abilities and this threat could, therefore, have been reduced by the arrangement of the sets. In order to discover the perceptions of Tagging Helper as regards unobtrusiveness, we extended the standard SUS questionnaire by adding two questions (one positive and the other negative), which were also processed by following the steps required to obtain the SUS score. This extension provided us with a structured means to obtain the participants' perceptions of topics that the conventional SUS questionnaire does not include.

6.4. External validity

We identified two main threats to external validity: subjects and tasks/materials. Regarding subjects, we also included students in order to have more controlled conditions in an academic context. Unfortunately, these students had no experience of real AGSD projects, but they had taken courses concerning ASD and GSD during their university education. We included professional developers with experience in AGSD to enforce external validity. Concerning tasks/materials, the chatting scripts were based on a fictitious scenario, but with real world characteristics. Although, tagging was suggested in the scripts, the participants also tagged messages using their own initiative. Moreover, the need for AK was motivated by a questionnaire, not by a project necessity. Real scenarios should, therefore, be considered, as they are supposedly more complex and articulated.

7. Discussion

In this paper, we propose the concept of AK Condensation and present its implementation (ArchiKCo), which was evaluated to determine the feasibility of this concept. These evaluation results are discussed in three parts: (1) AK classification mechanism, (2) AK searching mechanism, and (3) Feasibility of AK Condensation.

7.1. AK classification mechanism

Literature reports that developers prefer using free tags in tagging systems, given their low cognitive load in everyday work [44,45]. We based our AK classification mechanism on an assisted tagging mechanism (Tagging Helper), as IBM® Rational® Jazz® [44], TagSEA [45] and eMoose [65] do. However, we included predefined user tags, which are in turn associated with a fixed set of meta-tags, rather than just allowing free tagging. Our results do not reflect that the participants disliked using predefined tags, and they merely stated that it would have been better if they could have defined the tags that they used during the evaluation. However, we should explore the participants' perceptions in more long-term studies. The participants also stated that it might be appropriate to include a mechanism by which to add tags on the fly or a smarter tag suggestion (based on the context of the topics), but none of them mentioned free tagging.

The participants did not show any sign of disliking tag conversation messages and around 50% of them used the expected number of tag instances (seven instances). In a previous study [59], around 90% of the participants used at least the number of tag instances provided, although in that case, only three suggestions were marked in the scripts, which were also shorter. Both results show evidence that developers may need to tag messages during conversations, thus indicating that an AK classification mechanism based on social tagging could be successful.

We decided to use an assisted tagging mechanism to avoid problems such as tag explosion, which could ruin the AK classification mechanism. Our results do not show any signs of tag explosion. Despite the fact that we allowed the participants to use new tags if they considered it necessary, there were only 18% of instances of unregistered tags. This tagging mechanism could also help reduce the problem of obscure similarity [43], since the participants selected a tag from a suggestion list and messages were, therefore, tagged with correctly written tag instances. In that respect, there were only 7% of tag typing errors during the chatting sessions. This low error rate also contributed to keeping the AK classification mechanism functional.

Although the participants did not know the registered tags beforehand, significantly more tag instances were correctly (47%) rather than incorrectly used. This means the tags' semantics corresponded to the message topics. However, this tagging accuracy is lower than that obtained by Sohan et al. [46], who used an intelligent auto-tagging mechanism (70% accuracy). We believe that the accuracy of Tagging Helper could be increased in future evaluations if the participants define their own tags.

In our previous study [59], 30% of the participants used meta-tags correctly, while in the present study this percentage increased to 43%. This could, therefore, be considered as evidence that the conceptual model entities on which meta-tags are based are expressive and are related to AGSD-type situations in terms of AK. However, we should conduct studies focused only on the refinement of the model.

These results lead us to believe that Tagging Helper could be part of a good AK classification mechanism for use during UTEM conversations. Although the participants perceived Tagging Helper to be usable (SUS score 77 = B grade), this perception was lower than that attained in our previous study [59] (SUS score 87 = A+ grade = Excellent according to Bangor et al. [64]), in which we evaluated another implementation of Tagging Helper that was integrated into a custom Web instant messenger. In that study, we had absolute control over the autocomplete features and the participants, therefore, reported fewer problems with tag selection and navigation. However, the implementation employed in this study was integrated into Skype, signifying that we had to adapt the selection and navigation features so as not to interfere with the Skype features. Despite this difference in usability, the Skype-based Tagging Helper was perceived as unobtrusive (score 72 = C+ grade), but we believe that this perception could be improved if tag navigation and selection are also improved. Nevertheless, some participants stated that it was just a matter of getting used to the helper's features.

7.2. AK searching mechanism

Our results indicated that AK Searcher is a trustworthy application that is preferred by developers when searching for AK, because users tend to find AK faster than in UTEM when they do not know the knowledge source. However, AK Searcher was slower than Skype when we indicated a media in which to search. This could have been caused by the user interface design. While searching in Skype consists of at least three steps: (1) Ctrl+F to show the search textbox, (2) write text to search, (3) press <Enter>, AK Searcher requires two more steps to show a specific result: (1) write text to search, (2) press <Enter>, (3) look for an interesting result item, (4) open corresponding conversation, (5) look for knowledge in conversation. In a real situation, therefore, if a developer remembers the source of a certain item of knowledge, it might be better to search directly in that source. However, if a developer does not

remember the AK source and needs more than free text searching to find a specific item of knowledge, AK Searcher would be the best option, since it offers more parameters in which to carry out a search and refine the result set.

During the searching session, we observed that participants barely used the proposed classification mechanism to find knowledge, although they commented that tagging conversations is an interesting way to search for knowledge later. We believe that tags are more useful to search for AK if a developer wishes to attain knowledge from a general view to a detailed view. For example, when a new team member needs to acquire AK of the team's project, a good starting point might be to explore the existing tags and then explore the comments of a particular tag in greater depth.

The participants' comments and our observations led us to realize that improvements should be made to AK Searcher, one of which is to present the results differently depending on whether the AK source is a synchronous or an asynchronous UTEM. In our study, Trello could be considered as asynchronous, since card comments are not received as frequently as Skype messages in a conversation. The AK Searcher feature used to show a range of messages 5 min before and after a selected message may, therefore, be useless, since Trello comments could have a greater time difference. The same problem would occur with other asynchronous media (e.g. email). Another improvement would be to include a feature to show only tagged messages in the first result set. The problem of showing irrelevant messages (e.g. personal interactions, jokes, etc.) could, therefore, be reduced because these kinds of messages would not be tagged. Another way to reduce this problem would be to add exclusion tags that tell the Gatherer Service not to send certain messages to the repository.

To the best of our knowledge, there is no similar AKM research tool to ArchiKCo. The important differences between the reported tools (see Section 2.2) and our approach are: (1) they are not focused on searching for AK sourced in electronic interactions; (2) almost all of them are based solely on free text searching, and only TagSEA [45] and IBM® Rational® Jazz® [44] include extra parameters(e.g. tags or waypoints); (3) they do not have features to refine a result set; (4) they do not integrate AK from different sources (IBM® Rational® Jazz® could be configured to do so, but the paper [44] that refers to this does not present any integration), and (5) they do not present empirical results regarding searching in their respective papers. All these differences prevent us from making a direct comparison with our results.

Despite the improvement opportunities, AK Searcher was perceived to be a usable and useful media that eases the discovery of AK during an AGSD cycle, which could reduce interruptions among teammates when they have questions about the project architecture. Moreover, the participants perceived that AK Searcher would allow them to find AK in a timely manner, as required in an agile environment.

7.3. Feasibility of AK condensation

Agile/global developers know that UTEM logs contain important AK and they, therefore, need to search for architectural topics in those logs. However, they often spend too much time searching for AK because it is dispersed throughout different UTEM. The results obtained provide sufficient evidence to state that it could be feasible to implement the AK Condensation concept in AGSD for the following reasons.

- UTEM logs lack a structure that eases AK searching. We have, therefore, proposed a classification mechanism based on assisted social tagging. The participants used this mechanism well and it was perceived to be useful and unobtrusive. The results showed that agile/global developers could tag UTEM interactions accurately, even if they do not know the tags' meaning beforehand. In a real situation, developers should be careful to define useful tags, and to tag in a correct manner, since they are the most interested in retrieving AK from UTEM logs, because documentation debt is often present in AGSD environments. Furthermore, when developers tag during UTEM interactions, they are able to tag coherently because they are aware of the interaction topic. The evaluation results, therefore, allow us to state that (**RQ1**) social tagging is suitable and feasible to classify AK in AGSD because: (1) it is usable and unobtrusive, (2) agile/global developers could classify AK correctly, and (3) it is integrated into the agile work style.
- The AK retrieval mechanism was well received by the participants, since it integrates different AK sources, and they thus preferred searching for AK using this mechanism than searching directly in each UTEM. What is more, the participants tended to discover AK faster than when doing so directly in UTEM, particularly when they did not know which UTEM contained the AK required. The participants perceived that the AK retrieval mechanism was useful and usable. Furthermore, since AK is previously structured by the classification mechanism, AK retrieval could be easier and quicker than code analysis, even if the developers do not know the terms required to search for AK, because tags are linked to meta-tags that have a fixed meaning. Meta-tags could, therefore, be a guide to find AK, since they would not lose meaning overtime. However, we must conduct a long-term evaluation to better assert this. The evaluation results, therefore, show that the AK retrieval mechanism is (**RQ2**) suitable and feasible to help agile/global developers obtain AK from UTEM logs, because the mechanism is (1) useful and usable, and (2) the AK retrieval performance was better using the proposed mechanism when developers ignored which UTEM log contained the required knowledge.

Implementing the AK Condensation concept could help reduce AK vaporization in AGSD environments, taking into account the global and distribution aspects, without affecting the teams' agility. Furthermore, by reducing AK vaporization in AGSD, problems related to wasted time, software defects, and software projects' technical understanding [53,54] could be also reduced. It is worth recalling that we are presenting a single means to implement the concept of AK Condensation. Different implementations could be created solely by considering the basic items of this concept. It might be interesting to evaluate another implementation to confirm the feasibility of AK Condensation.

8. Conclusions and future work

In this paper, we present the concept of AK Condensation, which consists of structuring and retrieving AK shared by means of UTEM to reduce its vaporization in an AGSD environment. We also present an implementation of this concept, which was evaluated to determine AK Condensation feasibility. The evaluation results allowed us to determine that this concept could be feasible in AGSD environments.

On the one hand, these results could be attractive for AGSD practitioners, since an implementation of AK Condensation could reduce the amount of time wasted trying to find solutions to past problems, along with reducing the number of interruptions among teammates, since an additional source of AK would be available, i.e., an AK Condenser. In addition, AGSD practitioners might be interested in an implementation of AK Condensation because it could be a workaround to documentation debt, a means to alleviate architectural technical debt, and thus reduce AK vaporization. However, we are aware that there are cases of AGSD teams in which there is even a team of architects in charge of all the projects' architectural issues, as this is also reported in the empirical studies conducted by Clerc et al. [7], Razzak & Smite [66], and Alzoubi & Gill [67]. AK vaporization could, therefore, occur less frequently than in companies in which there is not a role or team that has these responsibilities. This leads us to believe that AK Condensation may be more appropriate for small and medium-sized AGSD companies,[7] which do not have sufficient resources and infrastructure to have an architect role.

[7] https://www.gartner.com/it-glossary/smbs-small-and-midsize-businesses.

On the other hand, our results may be interesting for software engineering researchers, since AK Condensation represents a convenient way in which to manage AK without much obstruction to the developers' work in AGSD. It may, therefore, be worth continuing exploring the UTEM logs as an AK source. Furthermore, AK Condensation represents a first step toward converting AK from tacit to explicit in a formalized manner [41] (e.g. UML notation), since AK would be structured by a classification scheme, which could ease the formal representation of this knowledge. Moreover, the impact of AK Condensation on the transitions between tacit and explicit knowledge and vice-versa (expressed using the SECI model [68]) could be explored in the future, owing to the close relation between AK Condensation and these two types of knowledge.

As future work, we shall improve ArchiKCo by using artificial intelligence to ease tag selection, adding a context aware suggestion feature to the Tagging Helper component. We shall also develop new versions of this component that will run with other UTEM, such as Trello, Slack, Jabber or Outlook, to be able to conduct studies in other contexts. AK Searcher must be adapted to these UTEM, since there will be synchronous and asynchronous media, and the frequencies of messages are different. Another improvement is the inclusion of a feature to exclude personal messages, such that only work-related messages would be considered during the search. In order to conduct evaluations in real scenarios, we must also include a strategy to motivate developers to tag. This strategy could be the following: when a developer finds a useful AK, s/he could qualify the message author, which could incentivize the best-qualified tagger. Finally, evaluations in real scenarios will allow us to refine the meta-tag model, and to observe how AK Condensation is conducted during pressure scenarios, thus enabling us to observe the implications as regards adopting the concept in AGSD.

Acknowledgments

The authors are grateful to the participating companies: EMCOR, Sahuaro Labs, Tufesa, Grupo SMI, Trapishar, Ubilogix and Softtek, for the support provided to carry out the present study, and for their willingness to continue working with us in future projects.

Funding

This work was supported by the National Council of Science and Technology (whose acronym in Spanish is Conacyt) of Mexico, with scholarship number 394125 for the first author. This work is also partially supported by: GINSENG (TIN2015-70259-C2-1-R, Ministerio de Economía y Competitividad y Fondo Europeo de Desarrollo Regional FEDER); and G3Softproject (SBPLY/17/180501/000150) funded by "Consejería de Educación, Cultura y Deportes de la Dirección General de Universidades, Investigación e Innovación de la JCCM" of Spain.

Conflict of interest

We have no conflict of interest to declare.

References

[1] B. Ramesh, L. Cao, K. Mohan, P. Xu, Can distributed software development be agile? Commun. ACM 49 (2006) 41, doi:10.1145/1164394.1164418.
[2] H. Holmstrom, E.O. Conchuir, P.J. Agerfalk, B. Fitzgerald, Global software development challenges: a case study on temporal, geographical and socio-cultural distance, Glob. Softw. Eng. (2006) 3–11 ICGSE '06. Int. Conf. (2006), doi:10.1109/ICGSE.2006.261210.
[3] K. Beck, M. Beedle, A. van Bennekum, A. Cockburn, W. Cunningham, M. Fowler, The Agile Manifesto, (2001). http://agilemanifesto.org/.
[4] R. Hoda, J. Noble, S. Marshall, How much is just enough? in: Proc. 15th Eur. Conf. Pattern Lang. Programs - Eur. '10, New York, USA, ACM Press, 2010, p. 13, doi:10.1145/2328909.2328926.
[5] A. Cockburn, J. Highsmith, Agile Software Development:the People Factor, Computer 34 (2001) 131–133, doi:10.1109/2.963450.
[6] A.R. Yanzer Cabral, M.B. Ribeiro, R.P. Noll, Knowledge Management in Agile Software Projects: a Systematic Review, J. Inf. Knowl. Manag. 13 (2014) 1450010, doi:10.1142/S0219649214500105.
[7] V. Clerc, P. Lago, H. Van Vliet, Architectural knowledge management practices in agile global software development, in: 2011 IEEE Sixth Int. Conf. Glob. Softw. Eng. Work., IEEE, 2011, pp. 1–8, doi:10.1109/ICGSE-W.2011.17.
[8] M.A. Razzak, R. Ahmed, Knowledge sharing in distributed agile projects: techniques, strategies and challenges, in: 2014 Fed. Conf. Comput. Sci. Inf. Syst., Warsaw, Poland, IEEE, 2014, pp. 1431–1440, doi:10.15439/2014F280.
[9] E. Tom, A. Aurum, R. Vidgen, An exploration of technical debt, J. Syst. Softw. 86 (2013) 1498–1516, doi:10.1016/j.jss.2012.12.052.
[10] ISO/IEC/IEEE, Systems and software engineering – – Architecture architecture description, ISO/IEC/IEEE 420102011(E) (Revision ISO/IEC 420102007 IEEE Std 1471-2000). (2011) 1–46. doi:10.1109/IEEESTD.2011.6129467.
[11] P. Kruchten, P. Lago, H. van Vliet, in: Building Up and Reasoning About Architectgural Knowledge, 4214, 2006, pp. 95–110, doi:10.1007/11921998. Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics).
[12] A. Moraes, E. Silva, C. da Trindade, Y. Barbosa, S. Meira, Recommending experts using communication history, in: 2nd Int. Work. Recomm. Syst. Softw. Eng. - RSSE '10, 2010, pp. 41–45, doi:10.1145/1808920.1808929.
[13] M. Aniche, M.A. Gerosa, C. Treude, Developers' perceptions on object-oriented design and architectural roles, in: Proc. 30th Brazilian Symp. Softw. Eng., New York, NY, USA, ACM, 2016, pp. 63–72, doi:10.1145/2973839.2973846.
[14] I. Ghani, M. Bello, Agile adoption in IT organizations, KSII Trans. Internet Inf. Syst. 9 (2015) 3231–3248, doi:10.3837/tiis.2015.08.029.
[15] A.M.M. Hamed, H. Abushama, Popular agile approaches in software development: review and analysis, in: 2013 Int. Conf. Comput. Electr. Electron. Eng., 2013, pp. 160–166, doi:10.1109/ICCEEE.2013.6633925.
[16] H.M. Sneed, Dealing with Technical Debt in agile development projects, Lect. Notes Bus. Inf. Process. 166 (2014) 48–62 .LNBIP, doi:10.1007/978-3-319-03602-1.
[17] T. Clear, Documentation and Agile Methods: striking a Balance, SIGCSE Bull. 35 (2003) 12–13, doi:10.1145/782941.782949.
[18] H.-C. Estler, M. Nordio, C.A. Furia, B. Meyer, J. Schneider, Agile vs. structured distributed software development: a case study, in: Proc. IEEE Int. Conf. Glob. Softw. Eng., 2012, doi:10.1109/ICGSE.2012.22.
[19] G. Borrego, A.L. Morán, R. Palacio, O.M. Rodríguez, Understanding architectural knowledge sharing in AGSD teams: an empirical study, in: 2016 IEEE 11th Int. Conf. Glob. Softw. Eng., Los Alamitos, CA, USA, IEEE Computer Society, 2016, pp. 109–118. doi:doi.ieeecomputersociety.org/10.1109/ICGSE.2016.29.
[20] B. Selic, Agile documentation, anyone? IEEE Softw. 26 (2009) 11–12, doi:10.1146/annurev.ps.29.020178.002001.
[21] M.L.I. Gervigny, S.D. Nagowah, Knowledge sharing for agile distributed teams: a case study of Mauritius, in: 2017 Int. Conf. Infocom Technol. Unmanned Syst., Dubay , UAE, IEEE Computer Society, 2017, pp. 413–419, doi:10.1109/ICTUS.2017.8286043.
[22] J. Bosch, Software architecture: the next step, in: F. Oquendo, B. Warboys, R. Morrison (Eds.), EWSA, Springer, 2004, pp. 194–199.
[23] K. Dalkir, Knowledge Management in Theory and Practice, Second ed., The MIT Press, 2011.
[24] R. Farenhorst, R.C. de Boer, Knowledge management in software architecture: state of the art, in: M. Ali Babar, T. Dingsøyr, P. Lago, H. van Vliet (Eds.), Softw. Archit. Knowl. Manag. Theory Pract., Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 21–38, doi:10.1007/978-3-642-02374-3_2.
[25] S. Dorairaj, J. Noble, P. Malik, Knowledge management in distributed agile software development, in: 2012 Agil. Conf., Dallas, USA, IEEE, 2012, pp. 64–73, doi:10.1109/Agile.2012.17.
[26] M. Jiménez, M. Piattini, A. Vizcaíno, Challenges and improvements in distributed software development: a systematic review, Adv. Softw. Eng. 2009 (2009) 14, doi:10.1155/2009/710971.
[27] K.B. Awar, M.S.I. Sameem, Y. Hafeez, A model for applying Agile practices in Distributed environment: a case of local software industry, in: Proc. 2017 Int. Conf. Commun. Comput. Digit. Syst. C-CODE 2017, 2017, pp. 228–232, doi:10.1109/C–CODE.2017.7918933.
[28] T. Dingsøyr, Strategies and approaches for managing architectural knowledge, in: M.A. Babar, T. Dingsøyr, P. Lago, H. van Vliet (Eds.), Softw. Archit. Knowl. Manag., Springer, Berlin Heidelberg, 2009, pp. 59–68, doi:10.1109/ASWEC.2008.4483186.
[29] R. Capilla, A. Jansen, A. Tang, P. Avgeriou, M.A. Babar, 10 years of software architecture knowledge management: practice and future, J. Syst. Softw. 116 (2016) 191–205, doi:10.1016/j.jss.2015.08.054.
[30] A. Tang, P. Avgeriou, A. Jansen, R. Capilla, M. Ali, A comparative study of architecture knowledge management tools, J. Syst. Softw. 83 (2010) 352–370, doi:10.1016/j.jss.2009.08.032.
[31] M. Galster, M.A. Babar, Empirical study of architectural knowledge management practices, in: 2014 IEEE/IFIP Conf. Softw. Archit., 2014, pp. 239–242, doi:10.1109/WICSA.2014.28.
[32] R. Farenhorst, R.C. De Boer, Knowledge management in software architecture: sState of the art, (n.d.) 21–39. doi:10.1007/978-3-642-02374-3.
[33] N. Ali, S. Beecham, I. Mistrik, Architectural knowledge management in global software development: a review, in: Glob. Softw. Eng. (ICGSE), 2010 5th IEEE Int. Conf., 2010, pp. 347–352, doi:10.1109/ICGSE.2010.48.
[34] S. Beecham, J. Noll, I. Richardson, N. Ali, Crafting a global teaming model for architectural knowledge, in: Proc. – 5th Int. Conf. Glob. Softw. Eng. ICGSE 2010, 2010, pp. 55–63, doi:10.1109/ICGSE.2010.15.

[35] P. Gaubatz, I. Lytra, U. Zdun, Automatic enforcement of constraints in real-time collaborative architectural decision making, J. Syst. Softw. 103 (2015) 128–149, doi:10.1016/j.jss.2015.01.056.
[36] S. Sherman, I. Hadar, M. Levy, N. Unkelos-Shpigel, Enhancing software architecture via a knowledge management and collaboration tool, in: S. Kunifuji, G.A. Papadopoulos, A.M.J. Skulimowski, K. Janusz (Eds.), Knowledge, Inf. Creat. Support Syst. Sel. Pap. from KICSS'2014 – 9th Int. Conf. Held Limassol, Cyprus, Novemb. 6-8, 2014, Cham, Springer International Publishing, 2016, pp. 537–545, doi:10.1007/978-3-319-27478-2_41.
[37] M. Che, D.E. Perry, G. Yang, Evaluating architectural design decision paradigms in global software development, Int. J. Softw. Eng. Knowl. Eng. 25 (2015) 1677–1692, doi:10.1142/S0218194015400380.
[38] C. Yang, P. Liang, P. Avgeriou, A systematic mapping study on the combination of software architecture and agile development, J. Syst. Softw. 111 (2016) 157–184, doi:10.1016/j.jss.2015.09.028.
[39] M.A. Babar, A.W. Brown, I. Mistrik, Agile Software Architecture: Aligning Agile Processes and Software Architectures, Elsevier Inc., 2013, doi:10.1016/C2012-0-01208-2.
[40] G. Borrego, A.L. Morán, R. Palacio, O.M. Rodríguez-Elias, E. García-Canseco, Review of approaches to manage architectural knowledge in Agile Global Software Development, IET Softw. 11 (2017) 77–88, doi:10.1049/iet-sen.2016.0197.
[41] I. Nonaka, H. Takeuchi, The Knowledge-Creating company: How Japanese Companies Create the Dynamics of Innovation, Oxford University Press, New York, 1995.
[42] H. Richter, G. Abowd, C. Miller, H. Funk, Tagging knowledge acquisition sessions to facilitate knowledge traceability, Int. J. Softw. Eng. Knowl. Eng. 14 (2004) 3–19, doi:10.1142/S0218194004001543.
[43] E. Bagheri, F. Ensan, Semantic tagging and linking of software engineering social content, Autom. Softw. Eng. 23 (2016) 147–190, doi:10.1007/s10515-014-0146-2.
[44] C. Treude, M.A. Storey, How tagging helps bridge the gap between social and technical aspects in software development, in: Proc. – Int. Conf. Softw. Eng., 2009, pp. 12–22, doi:10.1109/ICSE.2009.5070504.
[45] M.A. Storey, J. Ryall, J. Singer, D. Myers, L.T. Cheng, M. Muller, How software developers use tagging to support reminding and refinding, IEEE Trans. Softw. Eng. 35 (2009) 470–483, doi:10.1109/TSE.2009.15.
[46] S.M. Sohan, M.M. Richter, F. Maurer, Auto-tagging emails with user stories using project context, in: Lect. Notes Bus. Inf. Process. 48 LNBIP, 2010, pp. 103–116, doi:10.1007/978-3-642-13054-0_8.
[47] J.M. Al-kofahi, A. Tamrawi, T.T. Nguyen, H.A. Nguyen, T.N. Nguyen, Fuzzy set approach for automatic tagging in evolving software, Softw. Maint. (ICSM), 2010 IEEE Int. Conf., IEEE, 2010, doi:10.1109/ICSM.2010.5609751.
[48] S. Jalali, C. Wohlin, Global software engineering and agile practices: a systematic review, J. Softw. Evol. Process. 24 (2012) 643–659, doi:10.1002/smr.561.
[49] M. Levy, O. Hazzan, Knowledge management in practice: the case of agile software development, in: 2009 ICSE Work. Coop. Hum. Asp. Softw. Eng., 2009, pp. 60–65, doi:10.1109/CHASE.2009.5071412.
[50] J. Shore, S. Warden, The Art of Agile Development, First, O'Reilly, 2007.
[51] E. Isaacs, S. Whittaker, D. Frohlich, ... B. O'Conaill, Informal communication re-examined: new functions for video in supporting opportunistic encounters, Mediat. Commun. (1997) 1–30.
[52] C. Gutwin, S. Greenberg, M. Roseman, Workspace awareness in real-time distributed groupware: framework, widgets, and evaluation, in: M.A. Sasse, R.J. Cunningham, R.L. Winder (Eds.), People Comput. XI, Springer London, London, 1996, pp. 281–298, doi:10.1007/978-1-4471-3588-3_18.
[53] H. Holz, G. Melnik, M. Schaaf, Knowledge management for distributed agile processes: models, techniques, and infrastructure, in: Enabling Technol. Infrastruct. Collab. Enterp. 2003. WET ICE 2003, IEEE, 2003, pp. 291–294, doi:10.1109/ENABL.2003.1231423.
[54] N. Uikey, U. Suman, ... A. Ramani, A documented approach in agile software development, Int. J. Softw. 2 (2011) 13–22.
[55] Z. Li, P. Liang, P. Avgeriou, Architectural debt management in value-oriented architecting, Econ. Softw. Archit. (2014) 183–204, doi:10.1016/B978-0-12-410464-8.00009-X.
[56] S. Ryan, R.V. O'Connor, Acquiring and sharing tacit knowledge in software development teams: an empirical study, Inf. Softw. Technol. 55 (2013) 1614–1624, doi:10.1016/j.infsof.2013.02.013.
[57] M.A. Babar, Supporting the software architecture process with knowledge management, Softw. Archit. Knowl. Manag. Theory Pract. (2009) 69–86, doi:10.1007/978-3-642-02374-3_5.
[58] T. Zernadji, C. Tibermacine, F. Cherif, Processing the evolution of quality requirements of web service orchestrations: a pattern-based approach, in: Proc. – Work. IEEE/IFIP Conf. Softw. Archit, 2014, pp. 139–142, doi:10.1109/WICSA.2014.35. WICSA 2014. (2014).
[59] G. Borrego, A.L. Morán, R. Palacio, Preliminary evaluation of a tag-based knowledge condensation tool in agile and distributed teams, in: 2017 IEEE 12th Int. Conf. Glob. Softw. Eng., 2017, pp. 51–55, doi:10.1109/ICGSE.2017.14.
[60] C. Wohlin, P. Runeson, M. Hst, M.C. Ohlsson, B. Regnell, A. Wessln, Experimentation in Software Engineering, Springer Publishing Company, Incorporated, 2012.
[61] J. Brooke, SUS-A quick and dirty usability scale, Usability Eval. Ind. 189 (1996) 194, doi:10.1002/hbm.20701.
[62] F.D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology, MIS Q. 13 (1989) 319–340, doi:10.2307/249008.
[63] J. Sauro, J.R. Lewis, Quantifying the User Experience: Practical Statistics for User Research, first ed., Morgan Kaufmann Publishers Inc, San Francisco, CA, USA, 2012.
[64] A. Bangor, P.T. Kortum, J.T. Miller, An empirical evaluation of the system usability scale, Int. J. Hum. Comput. Interact. 24 (2008) 574–594, doi:10.1080/10447310802205776.
[65] U. Dekel, J.D. Herbsleb, Pushing relevant artifact annotations in collaborative software development, in: Proc. ACM 2008 Conf. Comput. Support. Coop. Work – CSCW '08, 2008, p. 1, doi:10.1145/1460563.1460565.
[66] M.A. Razzak, D. Smite, Knowledge management in globally distributed agile projects – lesson learned, in: Glob. Softw. Eng. (ICGSE), 2015 IEEE 10th Int. Conf., Ciudad del Real, Spain, 2015, pp. 81–89, doi:10.1109/ICGSE.2015.22.
[67] X. Shen, Y. Li, Y. Sun, An agile enterprise architecture-driven model for geographically distributed agile development, transform, Healthc. Through Inf. Syst. 17 (2016) 185–197, doi:10.1007/978-3-319-30133-4.
[68] I. Nonaka, R. Toyama, N. Konno, SECI, Ba and leadership: a unified model of dynamic knowledge creation, Long Range Plan. 33 (2000) 5–34, doi:10.1016/S0024-6301(99)00115-6.
[69] F. Calefato, D. Gendarmi, F. Lanubile, Investigating the use of tags in collaborative development environments: a replicated study, in: Proc. 2010 ACM-IEEE Int. Symp. Empir. Softw. Eng. Meas, 24, 2010, pp. 1–24, doi:10.1145/1852786.1852818. 9.
[70] A. Forward, T. Lethbridge, D. Deugo, CodeSnippets plug-in to eclipse: introducing web 2.0 tagging to improve software developer recall, in: Proc. – SERA 2007 Fifth ACIS Int. Conf. Softw. Eng. Res. Manag. Appl., 2007, pp. 498–502, doi:10.1109/SERA.2007.81.
[71] S. Paul, T. Makkar, K. Chandrasekaran, Software development using context aware searching of components in large repositories, in: Int. Conf. Comput. Commun. Autom., IEEE, 2015, pp. 765–772, doi:10.1109/CCAA.2015.7148513.

www.komputersapiens.org

KOMPUTER SAPIENS

Revista de Divulgación de la Sociedad Mexicana de Inteligencia Artificial

Año 10
Volumen 1
Ene-Abril 2018
$50.00

SEMBLANZA DE JOSÉ NEGRETE

HERRAMIENTA PARA EL LIGADO DE OBJETOS

ENFOQUES BASADOS EN LEXICÓN Y LSA

VERIFICACIÓN DE TUITS USANDO ALGORITMO DE CLASIFICACIÓN

SMARTPHONES COMO RECOLECTORES DE DATOS

CLASIFICANDO CONOCIMIENTO ARQUITECTÓNICO

TWITTER PARA VIGILAR LA INCIDENCIA DE ENFERMEDADES

ISSN 0187-9693

©Komputer Sapiens, Año X Volumen I, enero-abril 2018, es una publicación cuatrimestral de la Sociedad Mexicana de Inteligencia Artificial, A.C., con domicilio en Ezequiel Montes 56 s/n, Fracc. los Pilares, Metepec, Edo. de México, C.P. 52159, México, http://www.komputersapiens.org, correo electrónico: editorial@komputersapiens.org, tel. +52 (833)357.48.20 ext. 3024, fax +52 (833) 215.85.44. Impresa por Sistemas y Diseños de México S.A. de C.V., calle Aragón No. 190, colonia Álamos, delegación Benito Juárez, México D.F., C.P. 03400, México, se terminó de imprimir el 30 de abril de 2018, este número consta de 1000 ejemplares.
Reserva de derechos al uso exclusivo número 04-2009-111110040200-102 otorgado por el Instituto Nacional de Derechos de Autor. ISSN 2007-0691.

Komputer Sapiens es una revista de divulgación en idioma español de temas relacionados con la inteligencia artificial.
Creada en LaTeX, con la clase **papertex** disponible en el repositorio *CTAN*: Comprehensive TeX Archive Network, http://www.ctan.org/
Indizada en el IRMDCT de CONACYT y en Latindex.

Contenido

Columnas

ARTÍCULO ACEPTADO

Clasificando conocimiento arquitectónico a través de técnicas de minería de texto

Samuel González-López, Gilberto Borrego Soto, Aurelio López-López y Alberto L. Morán y Solares

Introducción

En los últimos años, el desarrollo ágil de software (DAS) ha desplazado al desarrollo basado en planes (convencional), debido a que el DAS se adapta a situaciones de requerimientos cambiantes, al incremento de la productividad, a la reducción de defectos y costos de software y a que los productos ingresan rápidamente al mercado. La adopción del DAS ha sido tal, que empresas de desarrollo global de software (DGS), basado en equipos virtuales distribuidos geográficamente, ya trabajan bajo ese esquema; es decir, aplican el desarrollo ágil global de software (DAGS). Tradicionalmente, en el DGS la comunicación se basa en documentos (conocimiento explícito) para disminuir el efecto de las 4 distancias (física, temporal, lingüística y cultural) propias del DGS. En cambio, en el DAS se prefiere la interacción personal (cara a cara) sobre el seguimiento de procesos, así como el software funcionando sobre una documentación exhaustiva. Esto muestra un antagonismo interno en el DAGS: por un lado se prefiere el conocimiento tácito (interacción cara a cara) y por el otro, se prefiere el conocimiento explícito.

En los equipos DAGS predomina el conocimiento tácito sobre el explícito; es decir, predominan las prácticas ágiles sobre las prácticas tradicionales basadas en documentos. Esto se debe a que las inherentes presiones de tiempo del DAS conducen a la reducción de la documentación; además, los desarrolladores ágiles consideran la documentación como una actividad secundaria y no creativa [1,2].

Todo lo anterior produce la llamada "deuda de documentación" [1]; es decir, poca claridad o ausencia de documentos donde se expresen elementos de diseño arquitectónico, manuales de usuario, modelos de datos, especificación de requisitos, etc.. Frecuentemente, esta deuda conduce a la vaporización del conocimiento o pérdida del conocimiento a través del tiempo. Los desarrolladores son unos de los principales roles que se afectan por la vaporización, particularmente cuando se trata del conocimiento arquitectónico (CA), el cual se refiere a estructuras y componentes de software que conforman un sistema informático, y las decisiones que llevaron al estado actual del sistema. La vaporización del CA afecta a los desarrolladores provocando pérdidas de tiempo para encontrar soluciones a problemas que ya se habían presentado con anterioridad, provocando defectos en la evolución y mantenimiento del software, limitando la visibilidad para el seguimiento del proyecto, provocando requerimientos o soluciones técnicas mal entendidas, por mencionar algunos [3].

Se han implementado diversas soluciones para manejar el CA en el DAGS y disminuir su vaporización; algunas de ellas se basan en la facilitación de la comunicación entre los miembros remotos y locales mediante medios electrónicos textuales, como Wikis, repositorios, groupwares, mensajeros instantáneos, etc.. Usando Wikis, repositorios y groupware, la comunicación queda registrada en una estructura adecuada para su recuperación. Sin embargo, los desarrolladores prefieren comunicarse con sus pares remotos usando mensajeros instantáneos, correo electrónico, foros, tablones de comentarios, etc., a los que llamamos medios electrónicos textuales no estructurados (METNE), en los que la recuperación de conocimiento se dificulta.

En un estudio con empresas de software donde se aplica el DAGS [4] se observó que se comparte frecuentemente CA importante a través de los METNE, quedando este conocimiento almacenado en las bitácoras de estos medios. También se obtuvo que es importante para los desarrolladores tener una manera eficiente de acceder al CA en las bitácoras de los METNE, ya que sería una manera de solventar la deuda de documentación. Al respecto, se propone una manera de estructurar las interacciones a través de los METNE usando etiquetado social, pero con las etiquetas ligadas a una estructura semántica fija para facilitar la recuperación del conocimiento a través de herramientas de búsqueda [5]. En otro estudio se evaluó con desarrolladores un componente de ayuda para el etiquetado manual de interacciones en METNE en tiempo real [6]. Los resultados mostraron la necesidad de contar con un mecanismo que sugiera opciones de etiquetas o bien que etiquete interacciones basado en el contexto de la plática, lo cual se podría llevar a cabo con técnicas de Procesamiento de Lenguaje Natural (NLP por sus siglas en inglés) y de minería de textos.

En este artículo se presenta un método para clasificar CA aplicando técnicas de minería de textos con el objetivo de apoyar en el etiquetado de interacciones en METNE y así facilitar la recuperación de CA a través de herramientas de búsqueda. También se muestran los resultados obtenidos al experimentar con una colección de bitácoras de METNE, llevándolas a una representación Bolsa de Palabras (*BoW* por sus siglas en inglés).

 ISSN 2007-0691

Finalmente, se discuten estos resultados y se presentan nuestras conclusiones finales.

Trabajo relacionado

La clasificación en el área de NLP es una tarea ampliamente desarrollada. Se han realizado estudios sobre la clasificación de textos cortos, que no alcanzan a formar una oración, donde se enfatiza la complejidad de clasificarlos, ya que es difícil encontrar patrones de los textos al presentarse matrices con muchas frecuencias en cero. En nuestro trabajo, los mensajes intercambiados en un chat pueden ser cortos o, inclusive, pueden contener más de una sentencia. Para resolver la tarea de clasificación, Wang et al. [7] utilizaron un modelo de lenguaje de red neuronal profundo. Para su experimentación utilizaron 5 corpus diferentes con textos cortos. De forma similar, pero utilizando una representación de bolsa de palabras, nuestro trabajo plantea una primera solución para la clasificación de CA.

La categorización de textos cortos ha sido abordada desde el aspecto semántico. En el trabajo de Wang et al. [7]se aplica un agrupamiento semántico; es decir, palabras cercanas conformando un grupo. Esta cercanía puede ser obtenida a través de una representación de vectores, midiendo la distancia entre ellos; por ejemplo, con la similitud de coseno. En nuestro trabajo, la parte semántica no se contempla. Sin embargo, bajo el enfoque *BoW*, podemos obtener una categorización para que el desarrollador seleccione o deseche tal conocimiento con base en su necesidad.

Por otro lado, en el área de DAGS se han realizado esfuerzos para clasificar información con el fin de ayudar a la operación del desarrollo de software en estos ambientes. Por ejemplo, se desarrolló un sistema recomendador [8] que usa algoritmos de minería de datos para identificar temas en correos electrónicos y relacionarlos con código fuente y documentación. Con esta relación, se puede consultar sobre un tema específico y el sistema recomienda las personas expertas en ello. En otro trabajo [9] se reporta el uso de "machine learning" para crear relaciones entre correos electrónicos e historias de usuario. De tal manera que, cuando un desarrollador encuentra un mensaje de correo, se conozca la historia de usuario relacionada. Por el momento, esta técnica reporta el 70 % de la precisión de la asignación, por lo que requiere mejoras para una implementación industrial.

Estos trabajos relacionados con el área de DAGS no toman en cuenta la comunicación por mensajeros instantáneos y no hace énfasis en la identificación de CA, el cual es muy importante para la correcta operación de un equipo de desarrollo de software.

Materiales y método

En este trabajo se hace uso de técnicas del área de NLP para abordar la clasificación de CA en busca de organizar de forma inteligente la información no estructurada en bitácoras de METNE (almacenada en formato de texto). Para realizar la experimentación se planteó un método de clasificación de CA y se creó una colección de interacciones en METNE. A continuación, se describen ambos pasos.

Colección de interacciones en METNE

En un estudio anterior [4] se capturaron 216 interacciones en METNE de 20 programadores de una compañía mexicana de desarrollo software, quienes participan en proyectos ágiles junto con empresas latinoamericanas, españolas y de Estados Unidos. Particularmente las interacciones en METNE fueron capturadas de correos electrónicos y chats. Cabe mencionar que obtener este tipo de interacciones es complicado, ya que se considera información clasificada de las empresas.

Del total de 216 interacciones se extrajeron 150, las cuales fueron enviadas a etiquetar por un experto. El etiquetado se hizo con base en 11 categorías de CA identificadas en un estudio previo [4], donde se analizaron de manera manual las 216 interacciones mencionadas anteriormente. Abajo se presentan las 11 categorías de CA:

- Segmentos de código compartido.
- Configuración del entorno de pruebas o de despliegue.
- Petición de información sobre el flujo interno de sistemas.
- Comunicar información sobre clases.
- Referenciar reglas de negocio u operación interna que fue explicada por un tercero.
- Explicación técnica acerca de soluciones a errores.
- Aclaración de reglas de negocio, funcionalidad o historias de usuario.
- Proposiciones de nuevos proyectos.
- Dudas acerca del funcionamiento de un producto terminado.
- Clarificación de tópicos de bases de datos.

Para efectos de nuestro experimento, se eligieron 2 categorías principales, que fueron identificadas por el experto como las más frecuentes en la colección de interacciones. Estas categorías fueron: (1) Segmentos de código compartido y (2) Configuración del entorno de pruebas o de despliegue. Se decidió crear una tercera categoría que representa el resto de interacciones (es decir, aquellas diferentes a las categorías 1 y 2 con menor frecuencia), esto con el fin de validar el algoritmo de clasificación y poder diferenciar los textos que no forman parte de las categorías 1 y 2.

 ISSN 2007-0691

En la Figura 1 se muestra un ejemplo de una conversación en mensajero instantáneo de las que fueron etiquetadas por el experto. En este caso sólo se etiquetó el texto sombreado, el cual es un fragmento de una sentencia de SQL (*Structured Query Language*) compartida por uno de los desarrolladores y que corresponde a la categoría "Segmentos de Código Compartido" (categoría 1). El resto de la conversación no fue etiquetada, ya que no correspondía a ninguna de las categorías elegidas para el experimento. Cabe mencionar que en una sola sesión de mensajes se pudo identificar más de una categoría.

Figura 1. Segmento de conversación en mensajero instantáneo (código compartido sombreado).

Método Clasificador de Conocimiento Arquitectónico

En la Figura 2 se muestra el esquema del método diseñado para clasificar CA. A continuación, se describen sus componentes.

Figura 2. Método Clasificador de Conocimiento Arquitectónico.

Corpus de Interacciones en METNE: Las interacciones están almacenadas en una base de datos, donde se tiene el texto completo de las conversaciones y correos electrónicos en dato crudo.

Subcorpus: Es un subconjunto del Corpus creado de la siguiente manera: se realizó el preprocesamiento del texto, eliminando información propia de los METNE (por ejemplo, etiquetas HTML que dan formato a la conversación) para obtener sólo la conversación entre los desarrolladores. También se eliminó el nombre del desarrollador para cuidar la privacidad de la información.

Representación BoW-bigramas: Dentro de este componente el algoritmo *BoW* contabiliza la frecuencia de aparición de las palabras en el documento, creando una matriz de frecuencias. *BoW* [6] no considera el orden, la estructura o el significado; sin embargo, da buenos resultados en tareas de clasificación donde la aparición de términos es frecuente. Para desarrollar los experimentos se formaron bigramas (secuencia adyacente de dos elementos, letras, sílabas o palabras) para encontrar frecuencias, como se muestra en la Figura 3. Cada bigrama forma parte de los atributos dentro de la matriz de frecuencias. La representación obtuvo aproximadamente 1660 atributos. Se decidió usar bi-gramas ya que en el trabajo de Ogada et al. [10] se evalúan textos cortos utilizando n-gramas (de 1 a 7 gramas), empleando tres enfoques: NaiveBayes, el vecino más cercano KNN y máquina de vectores de soporte SVM, siendo bigramas los de mejor desempeño.

Figura 3. Ejemplo de formación de bigramas.

Algoritmo Clasificador: Para entrenar el algoritmo de clasificación se utilizó la representación BoW-bigramas. Cabe mencionar que ambos grupos, el de entrenamiento y el de prueba, tienen la misma representación. Utilizamos *WEKA* (herramienta computacional con algoritmos de aprendizaje automático para minería de datos) para ejecutar la experimentación con los clasificadores *NaiveBayes*, *Complement NaiveBayes* y *SMO* (*Sequential Minimal Optimization*). *NaiveBayes* es un algoritmo simple y eficiente para la clasificación de textos, ya que considera independencia entre las características o términos de cada texto, lo cual se ajusta al enfoque *BoW*.

Después de realizar la tarea de clasificación obtuvimos los resultados de las medidas de Precisión, Recuerdo y F-Measure, que nos permiten identificar el nivel de éxito de la representación y del clasificador para las categorías seleccionadas.

 ISSN 2007-0691

Presentamos los resultados preliminares de un método de clasificación de conocimiento arquitectónico en bitácoras de comunicación, para habilitar su posterior recuperación.

Experimentación y resultados

Para realizar la experimentación en la tarea de clasificación utilizamos la validación cruzada a 10 iteraciones o pliegues. Se formaron tres grupos: el de entrenamiento, el de validación y el de prueba. Este último corresponde a datos que no forman parte del entrenamiento. De esta forma se busca dar un resultado confiable. La representación *BoW* se realizó utilizando el *Toolkit FeatureSpaceTree*[1], que provee los archivos de entrenamiento y prueba formateados. En la Tabla 1 se muestran los resultados obtenidos con los diferentes clasificadores, donde destaca que el algoritmo *Complement NaiveBayes* obtiene una mejor efectividad en la mayoría de resultados. La medida de *precisión* da un panorama del desempeño del algoritmo respecto al conjunto de interacciones clasificadas. El *recuerdo* es una medida que permite conocer la capacidad del algoritmo para clasificar las interacciones relevantes de todo el conjunto de interacciones correctas.

Tabla 1. Resultados de la tarea de clasificación (todos los valores numéricos están expresados en porcentajes, excepto los de la columna de Categorías).

Categorías	Algoritmos		
	NaiveBayes	Complement NaiveBayes	SMO
F-measure			
1	0.658	0.94	0.867
2	0.55	0.86	0.834
3	0.425	0.613	0.687
Precisión			
1	0.52	0.9333	0.9
2	0.65	0.8334	0.9
3	0.5967	0.6667	0.616
Recuerdo			
1	0.95	0.95	0.85
2	0.5	0.95	0.85
3	0.35	0.6	0.8

Para nuestra tarea, tanto el *recuerdo* como la *precisión* son vitales, ya que buscamos que el sistema tenga buena cobertura al clasificar el CA, considerando los documentos relevantes de cada categoría, y al total de interacciones.

La medida *F-Measure* es una media armónica ponderada que contempla el recuerdo y la precisión, brindando un valor más homogéneo del comportamiento del clasificador. Las categorías 1 y 2 obtuvieron resultados de 0.94% y 0.86% de F-Measure. Durante el desarrollo del experimento se observó cualitativamente que la categoría 1 (segmentos de código compartido) es más homogénea, porque los comandos del lenguaje SQL son frecuentes en el corpus utilizado. Al comparar estos resultados con tareas similares (clasificación de textos cortos) y con enfoque *BoW*, encontramos que, para la tarea de clasificar textos en Twitter obtenidos por Wang et al. [7], los niveles alcanzados son de 59.8%. Para la tarea de clasificación de títulos de artículos en noticias alcanza un 72.7%. Esto muestra que nuestros resultados para la clasificación de textos de CA son alentadores; aunado a que para la tarea realizada, "clasificación de conocimiento arquitectónico", no se encontró algún punto de comparación directo en otros artículos.

Conclusiones

En este trabajo presentamos un método basado en técnicas de minería de textos para clasificar CA a partir de datos no estructurados. Además, presentamos los resultados de experimentación con nuestro método usando textos de bitácoras reales de interacciones en correo electrónico y chats entre desarrolladores de software. El experimento estuvo enfocado a identificar 2 de 11 categorías dadas de CA obtenidas en un estudio anterior [4]. Los resultados nos indican que el algoritmo *Complement NaiveBayes* obtuvo un alto porcentaje de identificación de conocimiento, lo cual nos alienta a seguir experimentando con el método propuesto.

Por otro lado, estos resultados incrementan la factibilidad de desarrollar un mecanismo que sugiera etiquetas adecuadas al estar interactuando mediante METNE entre desarrolladores en un ambiente ágil y distribuido; o bien, abre la posibilidad de que dicho mecanismo etiquete las interacciones automáticamente, en lugar de hacerlo manualmente tal como se implementó en un estudio anterior [19]. Esto ayudaría a la estructuración del CA presente en las bitácoras de METNE, y por lo tanto sería más fácil desarrollar herramientas de búsqueda de CA que aprovechen las interacciones etiquetadas. Con esto se reduciría la deuda de documentación y la eventual vaporización del CA en ambientes de DAGS, lo cual se reflejaría en la reducción de pérdidas de tiempo para encontrar soluciones a problemas anteriores y en la reducción de defectos en la evolución y mantenimiento del software, sólo por mencionar los más importantes.

Como trabajo futuro se planea incrementar la colección de interacciones anotadas en METNE de 150 a 300,

[1] https://github.com/lopez-monroy/FeatureSpaceTree

 ISSN 2007-0691

para evaluar las 11 categorías de CA y evaluar nuestro método de una manera más robusta. También se dispondrá de un sitio web para poder descargar los conjuntos de datos de entrenamiento y prueba. Además, evaluaremos otra representación del conocimiento diferente de *BoW*, donde se incorpore el aspecto semántico, como la técnica de palabras embebidas. Esta técnica busca identificar la similitud de palabras vinculadas a vectores numéricos, para ello se podría utilizar la herramienta *Word2Vec*. Este enfoque de representación del conocimiento considera la distribución de las palabras; es decir, la posición de aparición de las palabras en las interacciones.✳

REFERENCIAS

1. Sneed, H. M. (2014). "Dealing with Technical Debt in agile development projects" *Lect. Notes Bus. Inf. Process.*. Vol. 166, pp. 48–62.

2. Clear, T. (2003). "Documentation and Agile Methods: Striking a Balance" *SIGCSE Bull.* Vol. 2003, 35, No.(2), pp. 12–13.

3. Holz, H. y Maurer, F. (2003). "Knowledge management support for distributed agile software processes", *Adv. Learn. Softw. Organ.* Vol. 2640, pp. 60–80.

4. Borrego, G., Morán, A.L., Palacio, R. y Rodríguez, O. M. (2016). "Understanding architectural knowledge sharing in AGSD teams: An empirical study". En *Proc. 11th IEEE International Conference on Global Software Engineering*, pp. 109–118.

5. Borrego, G. (2016). "Condensing architectural knowledge from unstructured textual media in agile GSD teams". En *Proc. IEEE 11th International Conference on Global Software Engineering Workshops*. pp. 69–72.

6. Borrego, G., Morán, A. L. y Palacio, R. (2017). "Preliminary Evaluation of a Tag-Based Knowledge Condensation Tool in Agile and Distributed Teams". En *Proc. IEEE 12th International Conference on Global Software Engineering (ICGSE) (2017)*. pp. 51–55.

7. Wang, J., Wang, Z., Zhang, D. y Yan, J. (2017). "Combining Knowledge with Deep Convolutional Neural Networks for Short Text Classification". En *Proc. of the Twenty-Sixth International Joint Conference on Artificial Intelligence* pp. 2915–2921.

8. Moraes, A., Silva, E., da Trindade, C., Barbosa, Y. y Meira, S. (2010). "Recommending experts using communication history". En *2nd International Workshop on Recommendation Systems for Software Engineering*, pp. 41–45.

9. Sohan, S. M., Richter, M. M. y Maurer, F. (2010). "Auto-tagging emails with user stories using project context". *Lect. Notes Bus. Inf. Process.* pp. 103–118.

10. Ogada, K., Mwangi, W. y Wilson, C. (2015). "N-gram Based Text Categorization Method for Improved Data Mining". *J. Inf. Eng. Appl.* Vol. 2015, No. 8, pp. 35–44.

SOBRE LOS AUTORES

Samuel González López es Profesor-Investigador Titular A de la Universidad Tecnológica de Nogales (UTN), Sonora. Obtuvo su Doctorado en Ciencias Computacionales en el Instituto Nacional de Astrofísica, Óptica y Electrónica. Sus líneas de investigación son procesamiento de lenguaje natural, minería de textos, extracción de información en textos y tutores inteligentes.

Gilberto Borrego Soto es estudiante de doctorado en la Facultad de Ciencias en la Universidad Autónoma de Baja California (UABC) campus Ensenada. Obtuvo su grado de maestría en Ciencias Computacionales en el Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE). Sus intereses de investigación incluyen procesos de ingeniería de software, sistemas colaborativos e interacción humano-computadora.

Aurelio López López es Profesor-Investigador Titular B del Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE). Obtuvo su doctorado en Ciencias Computacionales y de la Información en Syracuse University, Syracuse Nueva York, E.U.A. Sus áreas de interés son representación del conocimiento, extracción y recuperación de información de textos, minería de textos, así como tratamiento de lenguaje natural. Forma parte del Laboratorio de Tecnologías de Lenguaje del INAOE.

Alberto L. Morán y Solares es Profesor-Investigador de la Facultad de Ciencias en la Universidad Autónoma de Baja California (UABC) campus Ensenada. Obtuvo su grado de doctor en el Instituto Nacional Politécnico de Grenoble, Francia. Sus intereses de investigación incluyen interacción humano-computadora, computación móvil y ubicua, sistemas colaborativos e ingeniería de software.

 ISSN 2007-0691

Universal Access in the Information Society
https://doi.org/10.1007/s10209-020-00746-3

LONG PAPER

Key factors that influence the UX of a dual-player game for the cognitive stimulation and motor rehabilitation of older adults

Gilberto Borrego[1] · Alberto L. Morán[2] · Victoria Meza[2] · Felipe Orihuela-Espina[3] · Luis Enrique Sucar[3]

Abstract
In this work, the results of usability and user experience (UX) evaluation of a serious video game for the cognitive stimulation and motor rehabilitation of the upper limb of the elderly are presented. The serious game includes features that allow (1) performing cooperative therapy exercises between two patients, (2) remote session configuration therapy, and (3) monitoring/analyzing the sessions' results by the therapist. An evaluation of the game with 16 older adults provides evidence about how the tool is perceived by participants, who embraced it as stimulating, useful, usable and even fun, and which impacts in therapy compliance and acceptability by the elderly. In addition, through an in depth analysis of the participants' performance and observed emotions, as well as their self-report, we determined which engagement attributes are present in the game. Finally, we also found evidence that suggests that the participants' skill level and the game difficulty level rather than just a good performance on the game are key factors that influence their enjoyment and frustration.

Keywords Serious game · Usability evaluation · User experience evaluation · Tele-rehabilitation · Dual-player video game

1 Introduction

Older adult population is receiving increasing importance due to its growth projections for the coming years. With an increasingly aged population, cognitive impairment is a major health and social issue. Cognitive decline is among the most feared aspects of growing old due to the high cost, in terms of the economic and social burdens, since cognitive decline is a precursor of dementia, illness and death [1]. Serious games are a promising approach to provide non-pharmacological treatments, aiming to meet, maintain, and in some cases restore, the cognitive state and physical health of the elderly through cognitive stimulation, physical activation and rehabilitation activities [2–4]. Particularly, multiplayer serious games have been used for this purpose and for addressing the potential social isolation and demotivation caused by their condition [5–7]. However, in order to be successful in the implementation of rehabilitation multiplayer games specially for the elderly, their engagement and accessibility must be considered as an important factor to keep in mind during game design [8]. In a previous work, a serious video game for cognitive stimulation, upper limb motor rehabilitation and physical activation of the elderly was proposed [9]. This video game, called Balloons Rescuer, has the following main features: remote configuration, monitoring, evaluation and adaptation of the therapy by the therapist as the patient performs it, and the realization of cooperative therapy exercises between two patients. Balloons Rescuer was designed to be accessible and to provide engagement for older adults. Regarding accessibility, the game was designed simple enough to facilitate its use and learning; in [9], the results of a pilot usability evaluation

✉ Gilberto Borrego
gilberto.borrego@itson.edu.mx

Alberto L. Morán
alberto.moran@uabc.edu.mx

Victoria Meza
mmeza@uabc.edu.mx

Felipe Orihuela-Espina
f.orihuela-espina@inaoe.mx

Luis Enrique Sucar
esucar@inaoe.mx

[1] Departamento de Computación y Diseño, Instituto Tecnológico de Sonora, Cd. Obregón, México

[2] Facultad de Ciencias, Universidad Autónoma de Baja California, Ensenada, México

[3] Instituto Nacional de Astrofísica, Óptica y Electrónica, Tonanzintla, Puebla, México

Published online: 14 July 2020

suggest that the video game is perceived as being fun, stimulating, useful and usable.

Further, engagement is important for a serious game as it is one of the factors that enable the underlying aims of the game to be fulfilled [8, 10, 11]. Engagement is a quality of user experience characterized by attributes such as challenge, positive effect, endurability (the probability to remember pleasant situations and the intention to perform again the activity that caused them), esthetics and sensory appeal, attention, feedback, variety/novelty, interactivity and perceived user control [12]. Also, in order to achieve players' engagement in a video game, a balance of factors is required, such as: emotions (e.g., interest, frustration and joy), task difficulty (i.e., task challenge), cognitive load and motivation, among others [13–16].

Furthermore, user engagement is related to user experience (UX). A good user engagement design makes interaction exciting and fun, while UX encompasses the engagement but also tries to answer why people adopt and continue to use a particular design over time [17]. This relation is important to take into account as a priority during the design of a rehabilitation video game [8], since UX and engagement could affect the game adoption and consequently therapy success.

In this work, we report the results of an inquiry-based and observation-based usability and UX evaluation of the Balloons Rescuer video game, considering the performance and observed emotions of a group of older adults during their use of the game, aiming at determining whether the Balloon Rescuer video game generates user engagement, as well as, to identify the key factors that influence the participants' UX.

The paper is organized as follows: In Sect. 2, we present related works about serious video games focused on physical and cognitive rehabilitation. In Sect. 3, we present a brief description of the Balloons Rescuer video game, followed by the evaluation method description and results in Sect. 4, which are discussed in Sect. 5. Finally, we present our conclusions in Sect. 6.

2 Related work

The use of serious games has been explored in health care and in other domains (e.g., education [10]). Particularly, multiplayer serious games have been studied for rehabilitation, either psychological or physical. For instance, [18] presents a psychological rehabilitation application, where a virtual world was proposed as a means for social phobias rehabilitation. They found that the use of a virtual avatar-motivated interactions among the patients. Further, digital games hold a significant promise for enhancing the lives of older adults, potentially improving their mental and physical wellbeing, enhancing their social connectedness, and generally offering an enjoyable way of spending time [19].

Regarding physical rehabilitation, Novak et al. [5] used a robotic arm to control shoulder movements, along with an air-hockey video game. After evaluating the game, regarding game scores, participants' personality and self-report questionnaires, they concluded that multiplayer video games have great potential as a rehabilitation means, because they can provide enjoyment in the patients, as well as being more exercise intensive due to a sense of competition and engagement. However, they also stated that in the future, they have to consider more objective ways of measuring motivation; for example, considering the engagement concept. Also, they found that the personality and the skills of the patients have to be considered for the design or selection of a game. Also, impairment or skill differences were identified in [6] as influencing factors of patient motivation and game engagement. In the same sense, Maier et al. [7] considered the skill differences of rehabilitation patients. They developed a multiplayer game (an air-hockey video game as Novak et al. [5]) that compensates the motor impairments of the more affected patients and allows them to interact with other participants on equal game conditions. They evaluated the game and the compensation mechanism at the patients' homes. Maier et al. concluded that social interaction among the patients was improved, and that they played more rounds than the expected ones, because the game was perceived as very entertaining, however, game engagement was not formally considered in this study. After the game rounds, researchers measured the patients' ability to perform activities of daily living, obtaining that they presented an improvement in this point.

In addition, Mace et al. [20] developed a collaborative game focused on upper limb rehabilitation, which was evaluated in single mode and in multiplayer mode with patients and healthy people. They reported a significant performance improvement in patients playing in multiplayer mode, i.e., patients are motivated to play better when they have a partner or counterpart. This work neither reports motivational/cognitive aspects as a result of the reported performance improvement. In this last case, collaboration was a factor to motivate patients to perform better, however, Goršič et al. [21] report that competition can motivate patients too. They designed and evaluated four games focused on arm rehabilitation (one competitive, two cooperative, and one in single player mode); 40% of the patients preferred the competitive game, and showed more motivation and physical intensity playing the game than the rest of the players. In another study with the same video games [22], Goršič et al. concluded that competitive games have a high potential for in-home rehabilitation, given the increase in enjoyment and intensity shown by patients during the video game rounds.

From these later works, it can be inferred that engagement is a very important factor in order to have a successful rehabilitation video game. However, as far as we know, there are very few evaluations of rehabilitation serious games in terms of performance and emotions of the patients to determine game engagement. The only work we found was that of d'Ornellas et al. [23], where a questionnaire (Game Experience Questionnaire) was used to determine the perceived game engagement of patients. Their performance in the game is important since it could affect the emotions that they may present during the game (e.g., positive effect when patients obtain high scores), and consequently both factors could affect not only the user experience, but also game engagement [13, 24]. On the one hand, if a patient achieves a high score in the game, it does not warrant that the patient will have pleasant emotions, because the game could be so easy that it results boring [25]. On the other hand, if the game presents a high challenge level to the patient (and this reflects into a low score), this situation could affect the patient emotions, making him/her feel sad or listless [25]. From a point of view of serious video games for physical rehabilitation, these situations could affect the continuity of the therapies, since patients could quit them because they are not engaged in the game or they have a bad UX during therapy sessions.

3 Environment for cognitive stimulation and physical rehabilitation

In this section, we present the main video game components, i.e., the dual-player game and the associated telerehabilitation platform, along with a use scenario of both components.

3.1 Balloons Rescuer: the patients' serious game

Balloons Rescuer is a dual-player environment for cognitive stimulation and upper limb motor rehabilitation. The main objective of the Balloons Rescuer game is to avoid the balloons from hitting the ground. Balloons fall from the sky every few seconds for a specified time (both parameters are configurable). In order to "rescue" a balloon, the user must "touch it" or "hit it" three times with his "fist" (i.e., GUI pointer). The fist moves from left to right and vice versa on the screen (see Fig. 1). With each touch or hit, the balloon color and gesture change. As a balloon falls from the sky, first it is green, with the 1st hit it turns yellow, with the 2nd hit it turns red, and with the 3rd one it turns blue and flies away back to the sky. If a balloon falls to the ground, it turns gray and explodes. If a balloon touches below the blue vortex, this balloon is "teleported" and appears just below the vortex of the other player (playmate). The game score shows the countdown time, the number of saved and fallen balloons and balloon hits. The game ends at the completion of the specified time.

The Balloons Rescuer serious game was developed using Unity 4.5 in 2D mode. It was designed based upon the characteristic Gripper of the Gesture Therapy platform as the game control in order to bring accessibility to the elderly [26]. However, this evaluation was conducted using a PC keyboard, since we were more interested in game engagement aspects, rather than on game control aspects, thus we tried to avoid effects on the game perception caused by the Gripper control.

3.2 Cognitive stimulation

Balloons Rescuer is useful for cognitive stimulation because the players have to be able to discern and select which balloon to hit first given its color, as the color represents a priority for hitting them. For instance, if a balloon is red, with the

Fig. 1 The Balloons Rescuer serious game. As a balloon falls from the sky, first it is green, with the 1st hit it turns yellow, with the 2nd hit it turns red, and with the 3rd one it turns blue and flies away back to the sky. Whenever a balloon is sent through the blue vortex, it disappears from the current user's play area (left pane) and reappears in the playmate's play area (right pane)

next hit it would be saved and this action would be reflected in the score; but if the balloon were of another color, the player would have to hit it more times (see Table 1). Thus, in a situation where a red balloon and a green balloon were falling almost at the same time, the red one should be given priority since it would represent a balloon rescued and registered in the score. Similar decisions and selections should be conducted in the case of the other colors.

Additionally, Balloons Rescuer is useful for cognitive stimulation given its competitive/cooperative feature, i.e., passing balloons to the playmate using the blue vortex. We consider that this feature could be competitive when a player has the possibility of hindering the task of his/her playmate, passing more balloons to s/he, and this action could be reflected in the score as a greater number of fallen balloons. Similarly, this feature could be considered cooperative when the number of balloons passed is not that much as to hinder the playmate's play, but enough as to give him/her the chance to hit more balloons than those that are present in his/her play area. In the same way, this competitive/cooperative feature could be a stimuli to perform better during the game and be motivated to keep playing longer, as reported by Goršič et al. [21, 22].

3.3 Tele-rehabilitation

Remote management of the platform is done through a Web application directed to the therapist. With this application, s/he can manage his/her patients and their files, and manage the configuration of their therapies (e.g., duration, type of movement, etc.). Also, the therapist can consult the results of past game sessions through tables and graphs; s/he can see the video of an ongoing session in (quasi) real time through a video streaming service while the two players are online and performing their therapy. Further details regarding the collaborative game and service architecture are provided elsewhere [9].

3.4 Use scenario

In order to illustrate how the collaborative tool is used, consider the following use case description. The use case starts when John, the therapist, configures the therapy for his patients Alice and Bob, using the Therapy Management Web Tool on his PC. Therapy configuration includes selecting the activity (i.e., the Balloons Rescuer game), and setting its parameters (e.g., duration and interval between balloons). Subsequently, Bob (or a relative) starts the virtual therapy platform from his home and communicates with Alice to ensure her presence in the collaborative game; either via audio call or chat (not included on the game environment). When Bob starts the video game, it receives the therapy configuration parameters from the server, and then it displays Bob's avatar (i.e., a fist). Bob has to wait for Alice to join him. Once Alice enters the game at her home PC, it shows Bob's configuration with which she will play and she could start the game for both. During the game, John, the therapist, can see in a Web page how both patients are playing. When the game finishes, it sends the session results (scores) to the server. Thus, John can see the patients' results and even reconfigure their therapy program based on their recent performance. This scenario is shown through a sequence diagram in Fig. 2.

Table 1 Description of the balloon's color changes in the balloons rescuer game

Balloon color	Color change with one hit	Hits needed to save it
Green	Yellow	3
Yellow	Red	2
Red	Blue	1

3.5 Usability and UX evaluation

In this section, we present the main aspects and results of this evaluation. Previously, we conducted a usability evaluation of the Balloons Rescuer game to obtain the perceived ease of use and intention to use it of the elderly [9]. In this work, we conducted a usability and UX evaluation of the Balloons Rescuer game. The main difference between both usability evaluations was the way that participants used the game; in the first case, the evaluation of the participants was made based upon a demonstration of the game and gripper, in order to know the participants' perception regarding the game usability, meanwhile in the actual study, participants played the game with an opponent, to know their perception of usability, as well as how they felt about using it.

Another difference with our previous study is that we considered engagement as a key aspect to have a good UX. Thus, we measured variables related to challenge, positive effect, endurability, esthetics and sensory appeal, attention, feedback, variety/novelty, interactivity and perceived user control. We describe this study in the following sections.

3.5.1 Objective

Our objective was to analyze the use of the Balloons Rescuer serious video game, aiming at identifying key factors that influence the players' UX regarding performance, emotions, usefulness and usability and how these aspects help to achieve game engagement from the point of view of healthy elders, in the context of a physical rehabilitation therapy.

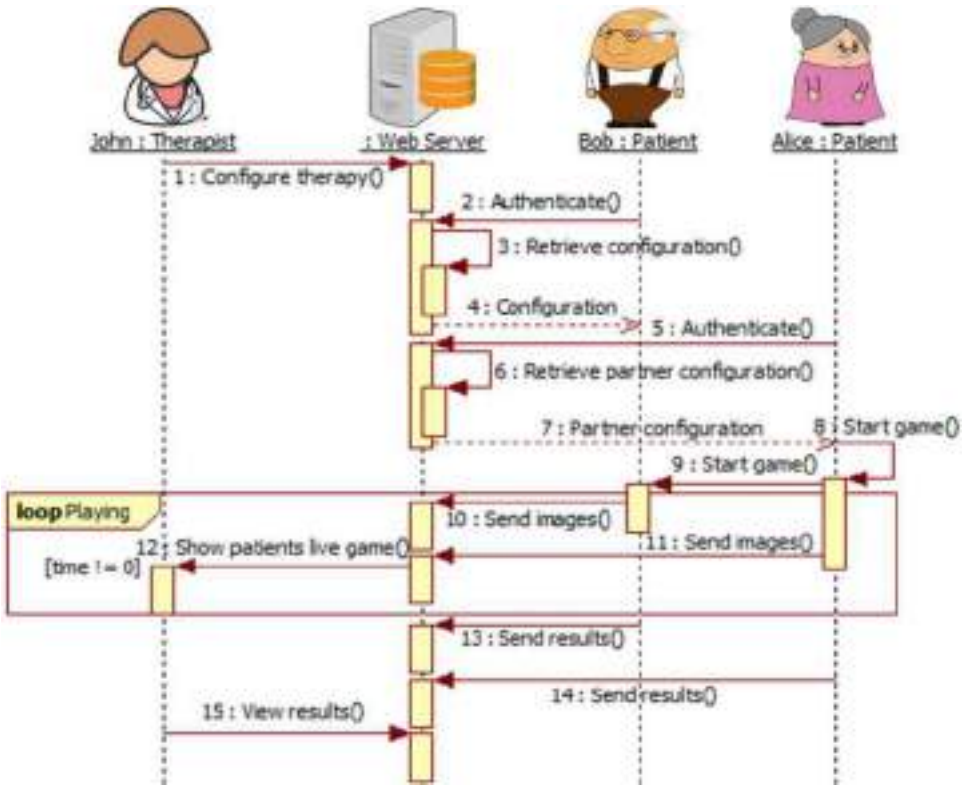

Fig. 2 Sequence diagram of the main scenario of use. The therapist configures the therapy for both patients. The game receives the configuration once the patients are authenticated. They play the game, and the therapist sees the game progression and its results

3.5.2 Methods and variables

The variables to measure/observe in this study were the following.

$$\text{Effectiveness} = \frac{\text{Amount of rescued balloons}}{\text{Total of session's balloons}} \quad (1)$$

- *Effectiveness*. We called "effectiveness" the performance of participants during their playing with Balloons Rescuer. Effectiveness was calculated dividing the number of rescued balloons by the total number of balloons during a game session (see Eq. 1). We decided to measure effectiveness because generally there is a relation between game challenge and game score. Also, players usually present positive emotions when they achieve high scores [13, 24]. Thus, effectiveness can be considered a factor that impacts the players' emotions and UX.
- *Emotions' prevalence during game sessions*. Participants' emotions were identified by indirect observation (using videos recorded during the game sessions), based on Tomkins' [27], proposals of basic emotions, and Eckman's et al. guidelines to recognize emotions in the human face [28] (see Table 2), in a similar way to the FACES coding system proposed by [29]. The coders extracted information about the frequency and duration of facial expressions; an expression is defined as any change in the face from neutral display. Then, the duration obtained for each emotion was divided by 60 s (the duration of each task), in order to obtain the emotions'

Table 2 Guide to identify emotions based on [27] and [28]

Type	Emotion	Distinctive expressions
Positive	Enjoyment joy	Smiling, lips wide and out
	Interest Excitement	Eyebrows down, eyes tracking, eyes looking, closer listening
Neutral	Surprise	Eyebrows up, eyes blinking
Negative	Anger frustration	Frowning, a clenched jaw, a red face
	Discomfort	The lower lip raised and protruded, head, forward and down, ask when finish the task, use both hands to finish the task
	Fear terror	A frozen stare, a pale face, coldness, sweat, erect hair
	Sadness	Eyes lowered, the head down and averted, blushing

prevalence. We decided to observe emotions given the above stated relation with challenge level and effectiveness, as well as their relation with positive effect and endurability of the game. We had two raters coding each participant's video as in [29].

- *Usefulness and intention to use*. These variables were obtained by applying the Technology Acceptance Model (TAM) [30] questionnaire (based on Likert-7 scale) to the participants after the game sessions. This instrument could be effective to obtain whether the game had a positive effect, endurability, esthetics and sensory appeal.
- *Ease of use /Usability*. We applied a TAM questionnaire and a System Usability Scale (SUS) questionnaire [31] (based on Likert-7 scale) to the participants after the game sessions to measure participants' ease of use/usability [32]. These questionnaires are complementary whenever the SUS includes additional questions to ask about learning factors (learnability), while TAM considers elements to obtain the usefulness of a specific technology system, in order to predict the intention to use of a certain population. Further, SUS and TAM questionnaires are useful to determinate feedback, interactivity and perceived control of the game.

It is important to mention that we added two questions to the SUS questionnaire in order to explore whether it was easy to send balloons through the vortex, due to the SUS sensitivity to include a wide variety of independent variables [33]. On the other hand, the TAM questionnaire was adapted from Davis [30] to include an intrinsic motivator (perceived enjoyment) using the model proposed by Lee [34]. We added six questions to the TAM questionnaire in order to explore whether the game was perceived as fun (and makes the participants feel happy), satisfying or boring, and whether the participants find it motivating to play in pairs.

3.5.3 Participants

This evaluation was carried out in the same research laboratory of a local public university where our previous usability evaluation [9] was conducted. In this study, participants were 16 older adults (avg. age of 64.12 years old, s.d. = 7.6–9 women and 7 men) who were part of an information technologies alphabetization course in the local University. All participants declared (1) not having any visual or motor impairment, (2) making daily use of IT and (3) owning a mobile device.

3.5.4 Procedure

As first activity, the participants were classified according to their ability to coordinate their hands and eyes to provide input using the PC screen and keyboard. The skill level was determined through an application developed *ex-professo* (see Fig. 3). In this application, the participants had ten attempts to press the indicated key on the PC keyboard. Every half second a button is highlighted in the screen to indicate which key has to be pressed (in this case, only the left and right arrow keys). The application records the failed attempts, the omissions and the time (in milliseconds) that the participant took to press the indicated key. Using the data collected in the application, participants were divided into skillful and non-skillful. We used as first criterion the number of omissions, as a second criterion the number of failed attempts and as a third criterion the average time that the participant took to press the correct key. In particular, for this sample, it was only necessary to consider the first criterion, given the big differences among participants in the number of omissions. Particularly, the participants with less than 5 omissions were considered as skillful. We had 10 skillful participants and 6 non-skillful participants. Then, elderly couples were randomly arranged following these patterns: skilful–skillful (3 couples), skilful–non-skillful (4 couples) and non-skillful–non-skillful (1 couple), aiming at observing their game performance and the behavior of these types of participants' couples. We were interested in observing whether the difference in participants' skills has an effect in their emotions; for instance, whether a non-skillful participant would feel discomfort or sadness when s/he played with a skillful participant who has better effectiveness than her/him.

Once all the participants were paired, the objective of the game was explained (using a slide show) to each couple, as well as the balloon color code and gestures, and the functionality of the vortex (passing balloons to the playmate). In addition, an on-entry survey was applied to the participants, where demographics were captured and a consent form signed.

Subsequently, each couple was asked to perform four one-minute tasks in the Balloons Rescuer game. The specification of each of the tasks is described in Table 3. Task 2 has

Fig. 3 Skill level detection application

Table 3 Specification of each of the tasks of the game evaluation (IBR = Inter Balloon Rate)

Task	Objective	IBR
1	Rescue as many balloons, without considering the functionality of passing balloons to the companion playmate	6 s
2	Send as many balloons as possible to the playmate, regardless of the balloons that fall down to the ground	2 s
3	Free play. Try to save as many balloons and try to send balloons to the playmate	8 s
4	Free play. Tactry to save as many balloons and try to send balloons to the playmate	2 s

Fig. 4 Screenshot of one of the videos taken with Camtasia Studio

Fig. 5 View of one of the videos taken using the environmental video camera

an interval between balloons (Inter Balloon Rate—IBR) of 2 s, aiming at increasing the probability of having balloons near the vortex. Further, in this second task, the effectiveness was not taken into account since its objective was not similar to the rest of the tasks; instead of effectiveness, we counted the number of balloons passed to the other player.

In order to measure the performance of participants, and considering that there were a group of non-skillful participants, no changes were made in the order of tasks. Furthermore, each task execution was video recorded using Camtasia Studio (see Fig. 4) and using an environmental video camera (see Fig. 5), in order to quantify the emotions shown by the participants during the tasks and between tasks. Two usability experts (two authors) performed an indirect observational analysis, using a spreadsheet to register the periods of the session in which the different emotions were presented on the recorded video (see Fig. 4). In the case of the emotions between tasks (we used the environmental video, see Fig. 5), emotions were just identified by observation not quantified, since the time between tasks was very variable.

4 Results

4.1 Task performance

In order to analyze task performance, we grouped the results of tasks 1, 3 and 4 because they share the same objective, i.e., to rescue as many balloons as possible, thus we could calculate the effectiveness of these tasks. Although participants were able to pass balloons through the vortex during tasks 3 and 4, this functionality was ignored by them for reasons which are discussed later (see Sect. 5.2.2). The results from task 2 were analyzed separately because its objective was not similar to the rest, thus the results of this task are not comparable to the results of the three others.

Regarding tasks 1, 3 and 4, skillful participants were more effective in tasks 1 and 4 (see Fig. 6). However, only in task 1 (IBR = 6 s.), the difference in effectiveness between skillful and non-skillful participants was significant ($t = 2.2546$, $p = 0.0203$, $\alpha = 0.05$). Surprisingly, in task 3 (IBR = 8 s.), non-skillful participants were more effective, but without being significant ($t = -0.1244$, $p = 0.4513$).

Regarding effectiveness between tasks, a significant difference (F-ratio = 26.87363, p-value < 0.00001,

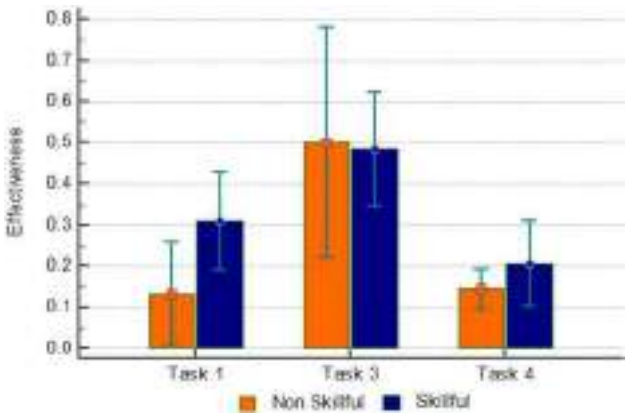

Fig. 6 Bar graph showing the effectiveness in tasks 1, 3 and 4, distinguishing between skillful and non-skillful participants (the vertical lines of each bar represent the standard error)

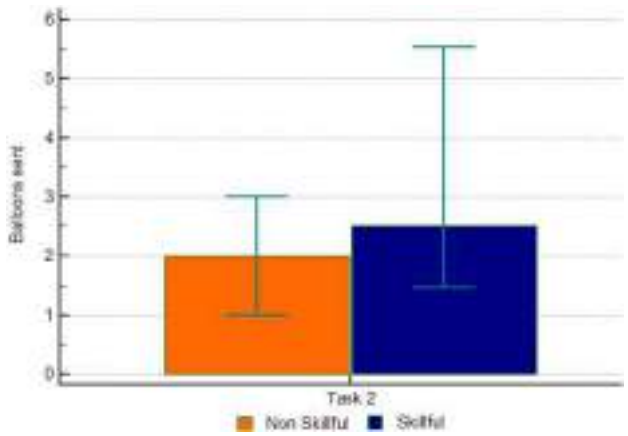

Fig. 7 Bar graph showing the performance in Task 2 of the skillful and non-skillful participants to pass balloons to the playmate (the vertical lines of each bar represent a median 95% confidence interval)

$\alpha = 0.05$) was found. In task 3, the one with the highest IBR, all participants scored a more significant effectiveness (avg = 0.4911, std. error = 0.0427) than that presented in the other two tasks (task 1: avg = 0.2438, std. error = 0.0537; task 4: avg = 0.1833, std. error = 0.0296). This same behavior was repeated when we compared the effectiveness between tasks, considering only the skillful participants (F-ratio = 14.61398, p-value < 0.00017, $\alpha = 0.05$) or only the non-skillful participants (F-ratio = 16.26511, p-value = 0.000719, $\alpha = 0.05$).

In task 2, as mentioned earlier, effectiveness was not taken into account to evaluate the performance of participants, but the number of balloons passed to the other player. In this task, the skillful participants (median = 2.5 balloons passed) performed better than non-skillful participants (median = 2 balloons passed), as shown in Fig. 7. However,

Table 4 Emotions detected during each task on all the participants (expressed on seconds)

	Skillful		Non-skillful	
	Avg	StD	Avg	StD
Enjoyment/joy	01.59	07.29	00.67	02.08
Interest/excitement	54.84	08.29	55.04	09.90
Surprise	00.46	01.04	00.00	00.00
Anger/frustration	03.03	04.32	04.21	09.84
Discomfort	00.08	00.49	00.08	00.29

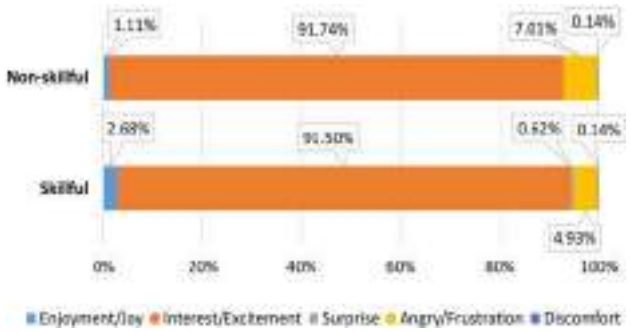

Fig. 8 Emotions shown during the execution of tasks by all participants

this difference was not statistically significant ($U = 22$, $U_{p<0.05} = 14$; Z-Score = 0.8135, $p = 0.2089$, $\alpha = 0.05$), according to the Mann–Whitney U test (applied because the data did not show a normal distribution).

4.2 Observed emotions

Table 4 presents the time (in seconds) in which each detected emotion was expressed by the participants (skillful and non-skillful) during each task. Interest/Excitement emotion was prominent over the rest of them. In fact, participants showed positive emotions most of the time (93.73%—See Fig. 8), where interest/excitement has 91.68% of prevalence with a significant difference according to Kruskal–Wallis test ($\chi^2 = 165.144 > \chi^2_U = 9.488$, p-value < 0, $\alpha = 0.05$). Among the negative emotions, anger/frustration (5.75% of prevalence) and discomfort (0.14% of prevalence) were also present. It should be noted that frustration occasionally arose in participants when balloons fell to the ground.

By looking at the prevalence of emotions by type of participant (skillful and non-skillful), some differences were observed (see Fig. 9). Skillful participants showed a higher prevalence of positive emotions (skillful—94.3%, non-skillful—92.85%), and a lower prevalence of negative emotions (skillful—5.08%, non-skillful—7.15%); in addition, skillful participants presented neutral emotions (surprise—0.62%).

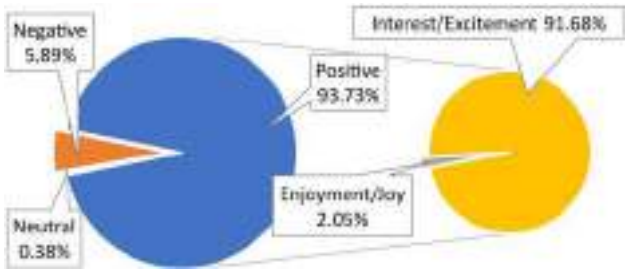

Fig. 9 Emotions shown during the execution of all the tasks by participant type

Fig. 10 Emotions shown during the execution of each of the tasks by all participants

Fig. 11 Emotions shown during the execution of each task by type of participant (S = skillful, NS = non-skillful)

It is emphasized that in positive emotions, skillful participants had a higher prevalence of enjoyment/joy (2.68%). However, none of these differences was significant, with the exception of the surprise emotion ($W = 516$, p-value = 0.041, $\alpha = 0.05$).

Now, if we look at the emotions presented by task (see Table 5 and Fig. 10), we see that task 4 presented a higher prevalence of enjoyment/joy (5%), even though it was the task that can be considered as the most difficult, as it had an IBR = 2 s., and with two goals to achieve: save balloons and pass balloons to the other player. However, there was no significant difference on enjoyment/joy prevalence between all tasks, according to the Kruskal Wallis test ($\chi^2 = 3.9004$, p-value = 0.2724).

Separating the results by task and by participant type (see Fig. 11), we observed that skillful participants presented more anger/frustration than the non-skillful ones, on task 2 ($U = 12$, $U_{p<0.05} = 12$; Z-Score = 1.70884, p-value = 0.04363, $\alpha = 0.05$). It is worth to remember that this task consisted on passing balloons to the other player, and there was no significant difference in the number of passed balloons between both types of participants; thus, this situation could be the cause of the higher prevalence of anger/frustration of skillful players. We also observed that only the skillful participants presented the enjoyment/joy emotion (8.33%) on task 4, the task with lower IBR, i.e., the most difficult. However, this difference was not significant ($U = 21$, $U_{p<0.05} = 12$; Z-Score = 0.64818, p-value = 0.25785, $\alpha = 0.05$). Furthermore, in this same task, the non-skillful participants presented a higher prevalence of anger/frustration (10.28%) than the skillful participants (2.22%), which is even the highest level for this emotion in all tasks (see Fig. 11), however, this difference was not significant ($U = 21.5$, $U_{p<0.05} = 12$; Z-Score = 0.58926, p-value = 0.2776, $\alpha = 0.05$).

Table 5 Emotions detected on skillful and non-skillful participants, expressed in seconds in which each emotion was present during each task

		Enjoyment joy	Interest excitement	Surprise	Anger frustration	Discomfort
Task 1	*Avg*	01.19	54.13	00.38	04.25	00.06
	StD	02.20	05.61	01.02	05.80	00.25
Task 2	*Avg*	00.67	56.00	00.27	03.07	00.00
	StD	02.32	05.87	00.80	05.43	00.00
Task 3	*Avg*	00.07	56.27	00.27	03.20	00.07
	StD	00.26	07.39	00.80	07.22	00.26
Task 4	*Avg*	03.00	53.33	00.20	03.40	00.20
	StD	11.35	14.44	00.77	09.48	00.77

In addition, it is also highlighted that the emotions of surprise and discomfort were the least shown in general. In the case of surprise, only the skillful participants presented it, particularly in cases where unexpected events (for them) happened during the game. For example, "*Ah look! I have sent one [balloon]*", "*hey! they are coming in bunches*." In the case of the discomfort emotion, it was mainly present in the non-skillful participants, particularly in tasks 1 and 3; and just one of the skillful participants presented it during task 4. This emotion was shown, both in skillful and non-skillful participants, in a gesture denoting their desire to finish as quickly as possible with the task.

To conclude the on-task analysis, it was evident in some of the participants that they were continuously looking at the keyboard. As shown in Fig. 12, the percentage of session time devoted by non-skillful participants to look at the keyboard in all tasks is visually greater. However, this difference was not significant either between tasks or between participant types, according to a Kruskal Wallis test ($\chi^2 = 1.3706$, p-value $= 0.7125$, $\alpha = 0.05$).

Concerning participants' behavior out of task, (i.e., before starting, in between, and after the tasks), it is noted that all couples expressed doubts about the interaction with the game (e.g., "*which keys are used to play the game?*", "*What is what I have to do to hit the balloons?*"). Also, there were confusion about identifying the playing area and the fist managed by each player. Only two couples (non-skillful–non-skillful, skilful–non-skillful) expressed doubts about the game in between tasks 1 and 2.

In most couples, there were interactions between participants which were motivated by the interest of knowing the score of the other. An additional motivation was to make jokes about their own results (e.g., "*How many balloons escaped from you?*", "*I had a lot of fallen balloons! [laughs]*"). In addition, relief attitudes were identified at the end of some tasks ("sigh or snorting"), usually by non-skillful participants. In between tasks 3 and 4 was the moment where fewer interactions among the participants were recorded, however, at the end of task 4 again empathic attitudes among couples were presented.

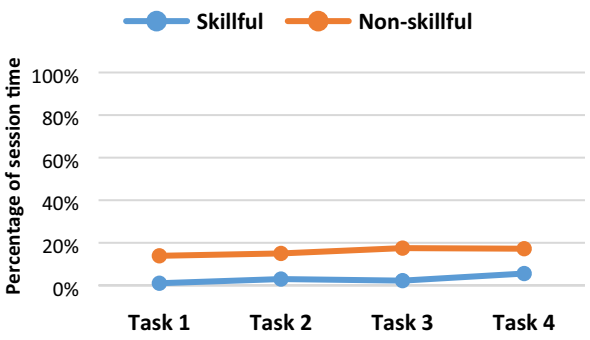

Fig. 12 Percentage of session time that skillful and non-skillful participants see the keyboard during each task

4.3 Emotion–effectiveness–skill level relationship

In order to explore the relationship between emotions, effectiveness and skill level, we discarded the interest emotion since it was the one with the lowest variability during the evaluation sessions. We also discarded task 2 from this analysis since its objective was different from the other three tasks.

In task 3, participants obtained the highest effectiveness, since the IBR was the highest of the remaining tasks (IBR = 8 s.), but it was the task where the least prevalence of enjoyment/joy was reported. Instead, in task 4 (where IBR = 2 s.), there generally was a lower effectiveness and more time dedicated to verify the keyboard. However, skillful participants reported more enjoyment/joy time, and non-skillful participants presented more frustration and no signs of enjoyment/joy. Finally, in task 1, which can be considered with a medium difficulty level (IBR = 6), all participants had a similar proportion of positive emotions. However, non-skillful participants had a higher frustration level than skillful participants (see Fig. 11), coinciding that in task 1 was when skillful participants obtained significantly more effectiveness than the non-skillful ones.

Regarding the stated above, we observed a relationship between the level of the emotions anger/frustration and enjoyment/joy, and the level of effectiveness. In order to present this relationship in a better manner, we grouped the prevalence per both emotions by participant type and by difficulty level regarding the IBR (low level = task 3, medium level = task 1, high level = task 4). Then, we ordered these data by the prevalence of each emotion, and we assigned a rank of level of emotion by thirds (see Table 6). In the same way, we grouped the effectiveness by participant type and difficulty level, and we assigned ranks by thirds (see Table 7).

Based on the data presented above, we created a radar graph for each participant type, in order to observe the relation between both emotions and effectiveness (see Figs. 13 and 14). On the one hand, the non-skillful participants (see Fig. 13) obtained a high rank of effectiveness when the game level was low, but also showed a high rank of anger/frustration and a low rank of enjoyment/joy. When the game was set in medium and high level, the non-skillful participants scored the lowest effectiveness, and presented a high rank of anger/frustration and a low rank of enjoyment/joy.

On the other hand, the skillful participants (see Fig. 14) scored a high rank of effectiveness when they played the game at the low level, and presented a medium rank of

Table 6 Emotions ranks by participant types and difficulty level

Rank	Participant type	Emotion	Prevalence	Difficulty level	Task
1	Non-skillful	Enjoyment joy	0.00%	Low	3
	Non-skillful	Enjoyment joy	0.00%	High	4
	Skillful	Enjoyment joy	0.20%	Low	3
	Non-skillful	Enjoyment joy	1.90%	Medium	1
2	Skillful	Enjoyment joy	2.00%	Medium	1
	Skillful	Anger frustration	2.20%	High	4
	Skillful	Anger frustration	3.90%	Low	3
	Skillful	Anger frustration	5.80%	Medium	1
3	Non-skillful	Anger frustration	8.10%	Low	3
	Skillful	Enjoyment joy	8.30%	High	4
	Non-skillful	Anger frustration	8.60%	Medium	1
	Non-skillful	Anger frustration	10.30%	High	4

Table 7 Emotions ranks grouped by participant types

Rank	Participant type	Effectiveness percentage	Difficulty level	Task
1	Non-Skillful	13.33%	Medium	1
	Non-Skillful	14.44%	High	4
2	Skillful	20.67%	High	4
	Skillful	31.00%	Medium	1
3	Skillful	48.57%	Low	3
	Non-Skillful	50.00%	Low	3

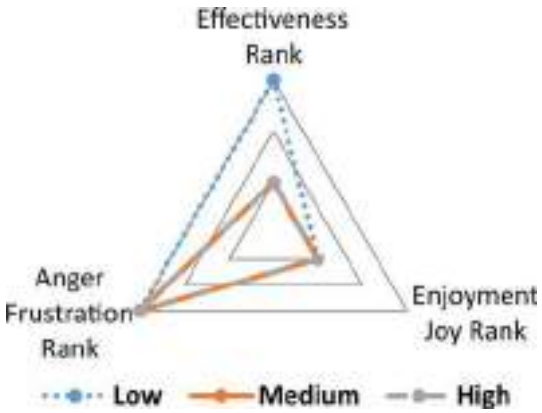

Fig. 13 Relation between effectiveness and the emotions anger/frustration and enjoyment/joy, regarding non-skillful participants. The different types of lines represent the rank of emotion or effectiveness

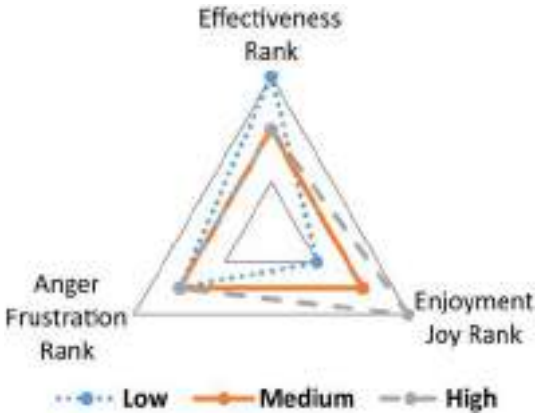

Fig. 14 Relation between effectiveness and the emotions anger/frustration and enjoyment/joy, regarding skillful participants. The different types of lines represent the rank of emotion or effectiveness

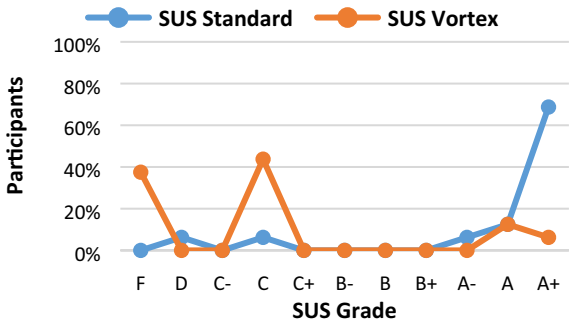

Fig. 15 SUS-based survey results (standard questions and vortex questions) by participants expressed on grades

anger/frustration and a low rank of enjoyment/joy. When they played on the medium level, they scored a medium rank of effectiveness, and presented a medium rank of anger/frustration and enjoyment/joy. We observed that anger/frustration did not increase its rank even when the effectiveness decreased one rank. Finally, when they played on the high level, they obtained the same rank of effectiveness and anger/frustration than the medium level, however, they also obtained a higher rank of enjoyment/joy.

4.4 Usability and intention to use

Regarding usability, in the SUS survey, a score of 88.19/100 was obtained, which indicates that overall the Balloons Rescuer game presents an acceptable degree of usability [35]. By applying the process proposed by Sauro and Lewis [33] to convert the SUS scores of each participant on grades, we obtained Fig. 15, where it can be observed that around 70% of the participants qualified the game as having an A + usability. Additionally, participants reported a low degree of usability regarding the functionality to send balloons to the other player, where a score of 50/100 was obtained. This score was calculated using the Likert-7 values obtained from the two questions added to the SUS survey to explore the vortex functionality, applying the same formula to obtain the SUS score. In Fig. 15, we observe that around 40% of the participants qualified the usability of the vortex functionality with an F or a C. These results suggest that this feature should be reviewed and redesigned.

Regarding the use of the TAM-based questionnaire (Likert-7 scale), we obtained a median of 7 for usefulness, a median of 7 the ease of use, and a median of 7 for the intention to use the game (see Table 8). These results indicate a high intention to use the video game by participants. It is particularly apparent that users find the game very entertaining (median = 7), easy to learn and operate (median = 7), as well as beneficial to exercise and increase their mental and physical agility (median = 7). On this regard, participants said: "*this game is cool! …*", "*it makes you exercise both your hands and eyes, right? …*", "*the task that follows will be even more difficult …?*", "*What a thrill!*", "*are we going to use only 2 keys to control the game?*". It is also noteworthy that participants reported on the extended TAM questions that it was motivating to play with a friend (median = 7), and that they felt happy (median = 7) and satisfied (median = 7) during the game play. Also, participants reported that the functionality to send balloons to the other was fun (median = 7).

5 Discussion

The results presented above are discussed in two parts: (1) comparison between usability evaluations of the Balloons Rescuer game; (2) analysis of the UX evaluation regarding the factors that characterize engagement; and (3) a summary about the key factors that affect the game's UX.

Table 8 TAM results based on Likert-7 scale (Q = Quartile)

	Median	Mode	$Q.25$	$Q.50$	$Q.75$
Ease of use	7	7	6	7	7
Usefulness	7	7	6	7	7
Intention to use	7	7	7	7	7

5.1 Comparison between usability evaluations of the balloons rescuer game

In general, we could say that the participants' perception of ease of use and intention to use the video game in this study were consistent with those previously reported [9]. The game was perceived as easy to operate and learn in both studies. There were also coincidences in coherence with the perception of reality of both evaluations, for example, participants considered that virtual balloons fall like actual balloons.

In addition, there was a consensus about the game being useful to improve cognitive and physical abilities, as well as in the motivation generated by the game when played with other players. Significantly, in our previous study [9], there was a person out of the 7 participants, who said that games were not appealing to him/her, meanwhile in the second study, none of the 16 participants reported anything similar.

All the above stated indicate that the ease of use perceived by participants who only saw the game in operation (previous study [9]) matches with that perceived by participants that actually played the game in specific tasks. This perception can be beneficial for the technology adoption process since it does not impose an idea of difficulty just by seeing it.

5.2 UX of the balloons rescuer game

In this part, we discuss the UX results analyzing the attributes which characterize engagement [12]: (1) positive effect and endurability, (2) feedback, interactivity and perceived control, (3) challenge level and attention, and (4) variety/novelty, esthetics and sensory appeal. Based upon the discussion of these points, we finalize the section determining whether Balloons Rescuer is an engaging game for the elderly.

5.2.1 Positive effect and endurability

Regarding positive effect, the Balloons Rescuer video game evoked joy and empathetic attitudes among participants, regardless of their ability condition (skillful or non-skillful), which is beneficial to improve self-esteem and increase the chances of engagement [5, 7, 23]. It should be noted that before starting the tasks, non-skillful participants generally showed nervousness, which was expressed through constant questioning about the operation of the game. However, during the execution of the tasks positive emotions were predominant in all participants, with some short episodes of frustration, usually caused by losing points (i.e., balloons falling to the ground).

In addition, it was found that some participants showed signs of relief at the end of a task, especially tasks 2 and 4 were those that had a higher output frequency of balloons (inter balloon rate IBR = 2 s.), which imposed a certain level of stress on them. In fact, the lowest effectiveness rate was obtained in task 4. However, the episodes of negative emotions were significantly lower than the proportion of positive emotions.

Concerning endurability, we consider that Balloons Rescuer meets this aspect; given that the TAM questionnaire results show a high usefulness and most participants said they would continue to use the game if they had it available (i.e., willingness to participate [15]), given that they found it fun, easy to use, beneficial to improve cognitive and physical aspects, as well as motivating to play with other players, i.e., they were satisfied to be part of a group [13, 20].

In addition, participants showed expressions of empathy and coexistence in between tasks, which were possible during the evaluation because the participants were co-located. Taking into consideration that the game could be played in a distributed mode (each player located on different places), an audio or video conference type functionality would be necessary, as to have a means through which these interactions could occur, considering that each player may not be in their home. This feature could help to increase the amount of time that patients spend playing, thus, doing therapeutic exercises [16]. Also, home rehabilitation would be possible, because its infrastructure is inexpensive, as in [8, 36], where such possibility is also planned.

5.2.2 Feedback, Interactivity and Perceived control

Regarding the control perceived by the user, although the game was considered easy to use (SUS Score = 88.19/100—70% of participant gave an A + of usability—Median = 7 Likert scale on ease of use), the functionality to send balloons to the other player was not considered as simple (median = 4.75 Likert 7). In fact, in task 2, where the only goal was to send balloons to the other player, skillful participants were the ones who showed a higher level of frustration due to their feeling of lack of control to send the balloons to the other player. On this matter, some of the comments expressed were: "*It does not enter [into the vortex]!*", "*it does not go there! [the balloon into the vortex]*" or "*but, how do I do that?*". In addition, it can be interpreted that the skillful participants showed more frustration in this task because, confident on their ability, they expected to send more balloons to the other player, but given that the balloons appear from different locations at random, it was not easy to guide them into the vortex. This observation is according to Whitlock et al. [14] work, where he stated that the mismatch between player expectation of difficulty, and actual difficulty experienced by the players lead to their frustration.

Further, concerning feedback during the observation, it was also identified that as there is no salient mechanism in the current implementation of the Balloons Rescuer game to indicate that one of the balloons has been sent to the other player, nor there is a score to indicate the number of balloons sent, some participants did not realize they had managed to send one or more balloons to his/her playmate. This omission in the current design and implementation of the game provides a possible explanation about why participants in general reported that it was not easy to send balloons to the other player, which may have caused stress in some participants, insomuch that at the end of the task there were those who sighed in relief. Further, the video game does not include any kind of reward to pass a balloon through the vortex, thus participants preferred to save balloons instead of passing them to their playmate during task 3 and 4. In fact, around 10% of the balloons of those tasks were passed through the vortex, and less than 5% of the passed balloons generated points.

The foregoing suggests the need to redesign the mechanism for sending balloons to the playmate, in order to make easier for the players to send balloons to their playmates, and make them feel more in control of the situation. Also, the presented situations suggest the inclusion of a notification mechanism with which the player is aware about his/her sending a balloon through the vortex, and about how many balloons s/he have sent, in order to inform the players in a better way about their achievements.

Regarding interactivity, understood as the degree of exchange of information between system and users [37], the current version of Balloons Rescuer is based on visual stimuli as output to the user and only two keys for input to the system, but there were no comments about a lack of an additional means of interaction. It will be necessary to further explore this in future work.

5.2.3 Challenge level and attention

In order to explore the level of challenge, we took into account both the relationship that we found between emotions, effectiveness and skill level. While none of the tasks showed an effectiveness greater than 50% (even the effectiveness of tasks 1 and 4 were below 25%), the participants remained focused/interested 91.68% of the duration of the tasks. This indicates that the game caught the attention of participants and that the level of challenge could be considered adequate in general terms. However, analyzing the effectiveness and the emotions of enjoyment/joy and anger/frustration (the ones with more variability) of each task, other aspects can be deduced.

We can argue that the level of challenge of task 3 was probably low for both types of participants, where they obtained the highest effectiveness, since the IBR was the largest of all tasks (IBR = 8 s.), but it was the task where the least prevalence of enjoyment/joy was reported. Instead, in

task 4 (where IBR = 2 s.), there was a lower effectiveness, however, skillful participants presented more enjoyment/joy, and non-skillful participants presented more frustration. This probably indicates that the level of challenge for skillful participants was adequate, but it was not for non-skillful participants. Finally, in task 1, (medium difficulty level—IBR = 6), non-skillful participants presented a higher frustration level than skillful participants, who obtained significantly more effectiveness than non-skillful participants.

Based on the evidence presented, we can deduce that there is not a linear relation between the challenge and the difficulty level of the video game, and the skill level of the player. Skillful participants presented a high rank of enjoyment in difficult levels, despite their not scoring high in effectiveness, understanding that a high score could be satisfying to the player [13, 24]. Conversely, skillful participants presented a low enjoyment on the easiest level of the game, despite their achieving a high score, as was also confirmed by Whitlock et al. [14]. On the other hand, non-skillful participants presented high ranks of anger/frustration and low ranks of enjoyment on any difficulty level, even though their achieving a high score rank on the easiest level.

The difference in the emotions shown by the participants and the effectiveness obtained regarding their ability, confirms the need for the existence of playing groups where there is a homogeneity in the degree of affectation of patients, or else, a mechanism in the game that dynamically levels the skill of the players, as was done in [7, 15, 38]. Even though the amount of time where participants show the anger/frustration emotion is very small relative to that of the interest/excitement emotion, in the case of participants with a real affectation and longer play sessions (greater than one minute), the degree of anger/frustration emotion could increase even more in the most affected patients.

5.2.4 Variety/novelty, esthetics and sensory appeal

Although these attributes of engagement were not explored in this study, it can be considered that participants had a neutral attitude about the esthetic and graphic design of the game, since neither negative nor positive comments were received on this respect. As for the variety/novelty, we consider that we should implement some rewarding mechanism for performance and/or perseverance which allow for novelty in the game. For example, if the user reaches a certain number of balloons saved or completed sessions, access to new functionality could be granted.

5.2.5 Does the balloons rescuer game cause engagement in participants?

Looking back at the results of this study and their relationship with the attributes that characterize engagement as discussed in the previous subsection, it can be said that the Balloons Rescuer game meets the following attributes: positive effect, endurability, level of challenge and attention. We can also say that it partially meets the attributes of feedback, interactivity and perceived control, given the design problems present while passing balloons to the other player, and that the current version of the game includes only visual stimuli. Also, it would be necessary to evaluate the attributes of variety/novelty, esthetic and sensory appeal. Therefore, considering that from 10 attributes, 4 are fully met and 3 are partially met, one could say that we are on the right track for the game to cause engagement on its users.

Although our results show that Balloons Rescuer meets certain attributes of engagement, we should conduct further studies in order to broaden the sample and consider not only healthy older adults, but also adults who suffer from cognitive disease or a motor affectation. Additionally, we should continue working on the development of the game to integrate the Gripper as the game control, so that we can conduct an assessment that may include clinical aspects, as the Gripper is required to perform the movements usually prescribed for rehabilitation therapy of the upper limb.

5.3 Key factors that influence the UX

This study provides evidence that the combination of players' skill and the game's challenge level is an important factor to take into account for the design and development of serious video games intended for the cognitive and physical rehabilitation of the elderly. As we presented above, skillful participants need to be challenged by the game (increasing its difficulty) so that they feel more joy/enjoyment, which is part of a good UX. Conversely, non-skillful participants need a moderated challenge in the game, and they could even require a support mechanism to ease their interaction with game. In the case of our study, non-skillful participants always presented a high rank of anger/frustration, even in the easiest level of the game. We consider important to avoid negative emotions in the players during the video game sessions, since they could be a factor to leave the sessions or play with reluctance, and consequently, this situation would affect the effectivity of the rehabilitation therapies based on serious games.

Additionally, we consider important to allow the players to interact among each other in-between game sessions or even during the game. In our study, we observed that participants interacted between them with good attitude, talking about how the game was, without judging their performance or score, i.e., they always presented positive emotions and a non-competitive behavior. In the case of tele-rehabilitation, we propose to add a feature to have a video-call during the game sessions in order to allow the players interaction, and consequently promote positive emotions in them.

Finally, game feedback is an important influencing factor on the participants' UX, particularly with elders. Feedback must be visual, auditory and even haptic so that the players could be aware of all the important game events. In our study, the vortex feature implements a visual feedback only, and for this reason, most of the participants did not realize when they achieved passing a balloon to the other player. We consider that better feedback about this feature would have a positive effect on the collaborative or competitive attitude of the participants, thus influencing the appearance and intensity of different emotions and even on the perceived UX of the participants.

6 Conclusions and future work

In this work, we present the results of a further usability and UX study of Balloons Rescuer, a serious video game for cognitive stimulation and motor rehabilitation of the upper limb of the elderly. These results were contrasted with those of an evaluation previously reported [9]. We found that the results of the first study were consistent with the results of second one, with a high level of usability and intention to use the video game, with the exception of the functionality of passing balloons to the playmate (send them through the vortex). At first glance, this functionality seemed to be easy for the participants of our previous study [9], but when it was used in a specific task during the second study, it was perceived as difficult to use. However, the participants claimed that it was fun to pass balloons to their playmates. This result leads us to propose a redesign to improve this feature as future work.

Regarding emotions, in the UX evaluation, we found that the interest level was about the same during the execution of the 4 tasks, regardless of their difficulty level. However, we also found that there were important differences in the enjoyment/joy and anger/frustration levels between user groups and tasks. For instance, let us consider the following cases: On the one hand, for the skillful participants 1) considering the most difficult task (i.e., task 4) they had their highest enjoyment/joy level even though their effectiveness was their lowest; and 2) considering the least difficult task (i.e. task 3) they had their lowest enjoyment/joy level even though their effectiveness was their highest.

These results provide evidence to suggest that performance and challenge in a serious game do not necessarily correlates linearly with enjoyment/joy or anger/frustration level, rather they are factors that depend on both the skill level of the participant and the difficulty/challenge level of the game. It has been stated in other works [13–16] that an engaging video game requires a balance of participants' emotions (negative and positive), perceived challenge and motivation. This evidence provides us with a line of future work, as it is necessary to keep working on the design and evaluation of this kind of serious games, in order to better understand the relationship among factors such as performance, enjoyment/joy, anger/frustration, difficulty level and skill level, so that we could improve the design of engagement/UX of these games, as this is a very important factor for successful cognitive and physical stimulation and rehabilitation of the elderly.

In addition, a clear understanding of the above mentioned factors, would allow us to explore as a future work the development of adaptive mechanisms for rehabilitation video games intended for the elderly (as in [8, 20, 38–41]), in order to fit the challenge level to each patient condition, and to achieve the patients' engagement, taking care of pushing them into their physical limits, without reaching high ranks of frustration [14], which could affect the motivation to continue playing/rehabilitating in the future.

Finally, although in the current version of Balloons Rescuer the Gripper control has not been integrated yet (and instead, the keyboard arrows were used), the aim of the presented evaluations was understanding the participants' perception of ease of use and intention to use the game, as well as identifying key factors that improve the game's engagement. As future work, the Gripper control will be included to ease the use of the game by people with motricity limitations, and to be used as a tool for upper limb rehabilitation.

Acknowledgements The authours would like to thank PhD. Arturo Jesús Laflor Hernández who helped in the development of the Balloons Rescuer game, and PhD. Cristina Ramírez Fernández for her support in data interpretation. We also thank Consejo Nacional de Ciencia y Tecnología of Mexico for scholarship No. 394125 provided to the first author during the study conducting. Finally, we thank the support from UABC for the realization of this research.

Compliance with ethical standards

Ethics approval All procedures followed were in accordance with the Ethical Code of the Research Ethical Comitee of Universidad Autónoma de Baja California.

Consent to participate Informed consent was obtained from all subjects for being included in the study.

References

1. Deary, I.J., Corley, J., Gow, A.J., et al.: Age-associated cognitive decline. Br. Med. Bull. **92**, 135–152 (2009). https://doi.org/10.1093/bmb/ldp033
2. Levin, M.F.: Can virtual reality offer enriched environments for rehabilitation? Expert Rev. Neurother. **11**, 153–155 (2011). https://doi.org/10.1586/ern.10.201
3. Tárraga, L., Boada, M., Modinos, G., et al.: A randomised pilot study to assess the efficacy of an interactive, multimedia tool of cognitive stimulation in Alzheimer's disease. J. Neurol.

Neurosurg. Psych. **77**, 1116–1121 (2006). https://doi.org/10.1136/jnnp.2005.086074
4. Gamberini, L., Alcaniz, M., Barresi, G., et al.: Cognition, technology and games for the elderly: an introduction to ELDERGAMES project. Psychnol. J. **4**, 285–308 (2006)
5. Novak, D., Nagle, A., Keller, U., Riener, R.: Increasing motivation in robot-aided arm rehabilitation with competitive and cooperative gameplay. J. Neuroeng. Rehabil. **11**, 64 (2014). https://doi.org/10.1186/1743-0003-11-64
6. Gerling, K.M., Mandryk, R.L.: Long-term use of motion-based video games in care home settings. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, Seoul, Republic of Korea, pp. 1573–1582 (2015)
7. Maier, M., Rubio Ballester, B., Duarte, E., et al.: Social integration of stroke patients through the multiplayer rehabilitation gaming system BT—games for training, education, health and sports. Lect. Notes Comput. Sci. Include. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinform. **8395 LNCS**, 100–114 (2014). https://doi.org/10.1007/978-3-319-05972-3_12
8. Burke, J.W., McNeill, M.D.J., Charles, D.K., et al.: Optimising engagement for stroke rehabilitation using serious games. Vis. Comput. **25**, 1085–1099 (2009). https://doi.org/10.1007/s00371-009-0387-4
9. Borrego, G., Morán, A.L., LaFlor, A., et al.: Pilot evaluation of a collaborative game for motor tele-rehabilitation and cognitive stimulation of the elderly. In: Cleland, I., Guerrero, L., Bravo, J. (eds.) Ambient Assisted Living, pp. 42–48. Springer International Publishing, Cham, ICT-based Solutions in Real Life Situations (2015)
10. Hamari, J., Shernoff, D.J., Rowe, E., et al.: Challenging games help students learn: an empirical study on engagement, flow and immersion in game-based learning. Comput. Hum. Behav. **54**, 170–179 (2016). https://doi.org/10.1016/j.chb.2015.07.045
11. Spek, E.D.: Towards designing for competence and engagement in serious games. In: Ma, M., Oliveira, M.F., Hauge, J.B., et al. (eds.) Serious Games Development and Applications, pp. 98–109. Springer, Berlin (2012)
12. O'Brien, H.L., Toms, E.G.: What is user engagement? A conceptual framework for defining user engagement with technology. J. Am. Soc. Inf. Sci. Technol. **59**, 938–955 (2008). https://doi.org/10.1002/asi.20801
13. Hoffman, B., Nadelson, L.: Motivational engagement and video gaming: a mixed methods study. Educ. Technol. Res. Dev. **58**, 245–270 (2010). https://doi.org/10.1007/s11423-009-9134-9
14. Whitlock, L.A., Collins-McLaughlin, A., Allaire, J.C.: Video game design for older adults : usability observations from an intervention study. Proc. Hum. Fact. Ergon. Soc. Annu. Meet. **55**, 187–191 (2013). https://doi.org/10.1177/1071181311551039
15. Sharek, D., Wiebe, E.: Investigating real-time predictors of engagement. Int. J. Gaming Comput. Simul. **7**, 20–37 (2015). https://doi.org/10.4018/ijgcms.2015010102
16. Lohse, K., Shirzad, N., Verster, A., et al.: Video games and rehabilitation. J. Neurol. Phys. Ther. **37**, 166–175 (2013). https://doi.org/10.1097/NPT.0000000000000017
17. Sutcliffe, A.: Designing for user experience and engagement. In: O'Brien, H., Cairns, P. (eds.) Why Engagement Matters: Cross-Disciplinary Perspectives of User Engagement in Digital Media, pp. 105–126. Springer, Cham (2016)
18. Mathur, A.: Augmented virtual identities: facilitating rehabilitation and societal reconnect. In: Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion. ACM, San Antonio, Texas, USA, pp. 219–224 (2013)
19. Ijsselsteijn, W., Nap, H.H., de Kort, Y., Poels, K.: Digital game design for elderly users. In: Proceedings of the 2007 Conference on Future Play. ACM, Toronto, Canada, pp. 17–22 (2007)
20. Mace, M., Kinany, N., Rinne, P., et al.: Balancing the playing field: Collaborative gaming for physical training. J. Neuroeng. Rehabil. **14**, 1–18 (2017). https://doi.org/10.1186/s12984-017-0319-x
21. Goršič, M., Cikajlo, I., Novak, D.: Competitive and cooperative arm rehabilitation games played by a patient and unimpaired person: effects on motivation and exercise intensity. J. Neuroeng. Rehabil. **14**, 1–18 (2017). https://doi.org/10.1186/s12984-017-0231-4
22. Goršič, M., Cikajlo, I., Goljar, N., Novak, D.: A multisession evaluation of an adaptive competitive arm rehabilitation game. J. Neuroeng. Rehabil. **14**, 1–15 (2017). https://doi.org/10.1186/s12984-017-0336-9
23. d'Ornellas, M.C., Cargnin, D.J., Prado, A.L.C.: Thoroughly approach to upper limb rehabilitation using serious games for intensive group physical therapy or individual biofeedback training. In: 2014 Brazilian Symposium on Computer Games and Digital Entertainment. pp. 140–147 (2014)
24. Wang, C.K.J., Khoo, A., Liu, W.C., Divaharan, S.: Passion and intrinsic motivation in digital gaming. CyberPsychol. Behav. **11**, 39–45 (2008). https://doi.org/10.1089/cpb.2007.0004
25. Pagulayan, R.J., Keeker, K., Wixon, D., et al.: User-centered design in games. In: Jacko, J.A., Sears, A. (eds.) The Human–Computer Interaction Handbook, pp. 883–906. L. Erlbaum Associates Inc., Hillsdale, NJ, USA (2003)
26. Morán, A.L., Orihuela-Espina, F., Meza-Kubo, V., et al.: Borrowing a virtual rehabilitation tool for the physical activation and cognitive stimulation of elders. In: Collazos, C., Liborio, A., Rusu, C. (eds.) Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 95–102. Springer, Cham (2013)
27. Tomkins, S.S.: Affect Imagery Consciousness: The, Complete edn. Springer, Two Volumes (2008)
28. Ekman, P., Hager, J.C., Oster, H.: Emotion in the Human Face, 2nd ed. Malor Books (2015)
29. Kring, A.M., Sloan, D.M.: The facial expression coding system (FACES): development, validation, and utility. Psychol. Assess. **19**, 210–224 (2007). https://doi.org/10.1037/1040-3590.19.2.210
30. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q **13**, 319–340 (1989). https://doi.org/10.2307/249008
31. Brooke, J.: SUS-A quick and dirty usability scale. In: Jordan, P.W., Thomas, B., McClelland, I.L., Weerdmeester, B. (eds) Usability Evaluation in Industry. CRC Press, pp. 189–194 (1996)
32. Assila, A., De, O.K.M., Ezzedine, H.: Standardized usability questionnaires : features and quality focus. Electron. J. Comput. Sci. Inf. Technol. **6**, 15–31 (2016)
33. Sauro, J., Lewis, J.R.: Quantifying the User Experience: Practical Statistics for User Research, 2nd edn. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2012)
34. Lee, M.K.O., Cheung, C.M.K., Chen, Z.: Acceptance of Internet-based learning medium: the role of extrinsic and intrinsic motivation. Inf. Manag. **42**, 1095–1104 (2005). https://doi.org/10.1016/j.im.2003.10.007
35. Bangor, A., Kortum, P.T., Miller, J.T.: An empirical evaluation of the system usability scale. Int. J. Hum. Comput. Interact. **24**, 574–594 (2008). https://doi.org/10.1080/10447310802205776
36. Palacios-Navarro, G., García-Magariño, I., Ramos-Lorente, P.: A kinect-based system for lower limb rehabilitation in parkinson's disease patients: a pilot study. J. Med. Syst. (2015). https://doi.org/10.1007/s10916-015-0289-0
37. Huang, M.-H.: Designing website attributes to induce experiential encounters. Comput. Human. Behav. **19**, 425–442 (2003). https://doi.org/10.1016/S0747-5632(02)00080-8
38. Gilleade, K.M., Dix, A.: Using frustration in the design of adaptive videogames. In: Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology. ACM, Singapore, pp. 228–232 (2004)

39. González Sánchez, J.L., Gutiérrez, F.L., Cabrera, M., Padilla Zea, N.: Design of adaptative video game interfaces: a practical case of use in special education. In: Lopez Jaquero, V., Montero Simarro, F., Molina Masso, J., Vanderdonckt, J. (eds.) Computer-Aided Design of User Interfaces VI, pp. 71–76. Springer, London (2009)
40. Montani, V., De Grazia, M.D.F., Zorzi, M.: A new adaptive videogame for training attention and executive functions: design principles and initial validation. Front. Psychol. **5**, 1–12 (2014). https://doi.org/10.3389/fpsyg.2014.00409
41. Bontchev, B., Vassileva, D.: Assessing engagement in an emotionally-adaptive applied game. In: Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality. ACM, Salamanca, Spain, pp. 747–754 (2016)

Revista Internacional de Investigación e Innovación Tecnológica

Página principal: www.riiit.com.mx

Aplicación del Modelo de Negocios Canvas en una planta de elaboración de alimento para cerdo

Application of the Canvas Business Model in a pig feed processing Company

Ramos-Solís, C.E.[a], Palacio-Cinco, R.R.[b], Córdova-Cárdenas, M.H.[b], Mendoza León J.G.[b]

[a] Unidad Nainari, Instituto Tecnológico de Sonora.
[b] Unidad Navojoa, Instituto Tecnológico de Sonora.
cramossolis91@gmail.com; ramon.palacio@itson.edu.mx; martin.cordova@itson.edu.mx; jorge.mendoza@itson.edu.mx

Aplicación empresarial: Generación de propuestas de valor para mejorar y hacer eficiente los canales y procesos de producción de alimento para cerdo.

Área de aplicación industrial: Tecnologías de la Información y Comunicación, Ingeniería Industrial e Ingeniería en Administración.

Enviado: 15 Abril 2019
Aceptado: 03 Julio 2019

Abstract

Currently, many organizations are unaware of the operation of their business, so they need to know how their business model is formed, this lets them to identify the key elements of how it is formed and how those elements are related in the organization. Therefore, the aim of this work is the design of a business model using the Canvas Business Model methodology, is a tool who permits finding the different opportunity areas in the product or business processes in an organization. For it, the tools of customer segment, empathy map, and creation of the value proposal were used, with which it was possible to obtain the Business Model of a pig feed processing plant. As a result, it was possible to infer the considerations and implications provided by the Business Model to the Plant through the elaboration of initiatives that help to achieve the aims established in the Plant's strategies. This Business Model is focused in an abstract way on how the plant has a profit, how the costs are composed, how its market segment is composed, how the Plant relates to its customers and how the value proposal makes them arrive. In addition, it includes what are its key activities, its resources, and its allies. This abstraction made it possible to show the overall operation of the plant from an agile perspective, thus achieving the sight of the current state of the plant, which in

turn allows us to project different scenarios of the operation of the model in each step of it. This capacity that has Canvas allows being flexible to the Plant to be able to project and analyze proposals and initiatives that allow them to prevent possible deviations that affect adversely interests of the Plant.

Key Words: Canvas, Processes, Production, Business Model.

Resumen

Actualmente, muchas de las organizaciones desconocen el funcionamiento de su empresa, por lo que requieren conocer su modelo de negocio, el cual les permita identificar los elementos clave de cómo está conformada y la forma en que se relacionan estos elementos en la organización. Por ello, el objetivo principal de este trabajo es diseñar un Modelo de Negocio utilizando la metodología del Modelo de Negocios Canvas, la cual es una herramienta que permite encontrar las diferentes áreas de oportunidad en los procesos productivos o de negocio en una organización. Para esto se utilizaron las herramientas de segmento de clientes, mapa de empatía y creación de la propuesta de valor con lo que fue posible obtener el Modelo de Negocio de una planta de elaboración de alimento para cerdo. Con esto fue posible inferir las consideraciones e implicaciones que proporciona el Modelo de Negocio a la planta mediante la elaboración de iniciativas que ayuden al logro de los objetivos establecidos en las estrategias de la misma. Este Modelo de Negocio está enfocado de manera abstracta en determinar la forma en la que la planta obtiene sus ganancias, cómo se componen los costos, cómo se compone su segmento de mercado, cómo se relaciona la planta con sus clientes y cómo les hace llegar la propuesta de valor. Además, incluye cuáles son sus actividades clave, sus recursos y sus aliados. Esta abstracción permitió visualizar la operación global de la planta desde una perspectiva ágil logrando de esta forma visualizar el estado actual de la planta lo que permite a su vez proyectar diferentes escenarios del funcionamiento del modelo en cada una de sus etapas que lo conforman. Esta capacidad que tiene Canvas permite ser flexible a la organización para poder proyectar y analizar propuestas e iniciativas que le permitan prevenir posibles desviaciones que afectan negativamente a los intereses de la planta.

Palabras clave: Canvas, Procesos, Producción, Modelo de Negocios.

1. Introducción

El modelo de negocio sirve como un plan de construcción que permite diseñar y organizar su estructura, además de ser la base para constituir la forma física y operativa de la empresa. Es importante reflexionar sobre su semántica, por lo que tanto el negocio, como el modelo por sí mismo, tienen un significado específico. Esta combinación de significados refleja muchas de las posibles aplicaciones del modelo de negocio [1]. El modelo es interpretado como "una descripción y representación simplificada de una entidad o proceso complejo", de manera que esta representación implica conceptualización, que puede describirse como “los objetos, conceptos y otras entidades que se asume que existen en algún área de interés y su interrelación”. Asimismo, también se interpreta la palabra negocio como "la

actividad de suministro de bienes y servicios que involucran aspectos industriales". Unificando estos elementos se establece que la reflexión sobre el concepto de modelo de negocio debe ir en la siguiente dirección: "un modelo de negocio es una herramienta conceptual que contiene un conjunto de objetos, conceptos y sus relaciones con el objetivo de expresar la lógica empresarial de una firma determinada".

Existen varios tipos de modelos de negocios o metamodelos que son genéricos, pero contienen características comunes. Los tipos de modelos se refieren a una categorización, mientras que los metamodelos dan referencia a diferentes modelos integrados, de forma que esta distinción refleja diferentes grados de conceptualización.

Por lo que podemos afirmar que un modelo de negocio es una plantilla que describe la forma en que la empresa lleva a cabo su operación, y es elaborado por los gerentes de una firma para satisfacer mejor las necesidades percibidas de sus clientes. Para abordar completamente la oportunidad de mercado, el modelo de negocio a menudo incluye a la firma misma y sus interacciones con la industria. En ese mismo sentido, se han realizado muchos intentos para describir y clasificar los modelos de negocio de forma taxonómica, es decir, desarrolladas al abstraer de observaciones típicamente de una sola industria [2]. Estos intentos rara vez tratan de manera completa y adecuada las dimensiones de clientes, organización interna y monetización; sin embargo, existe una tipología que considera cuatro elementos: a) identificar a los clientes (el número de grupos de clientes segmentados); b) compromiso del cliente (o la propuesta de valor); c) monetización, d) la cadena de valor y los vínculos (generalmente concierne a cómo funciona la empresa internamente) [3]. Cada una de estas dimensiones se relaciona con la definición del modelo de negocio, de creación de valor o captura de valor o ambas y se prestan para crear subcategorías y, por lo tanto, la oportunidad de un mapa significativo. Este mapa puede superponerse al mundo real de una industria, o la forma de pensar de un empresario, y al comparar el mapa con la tipología completa, podemos identificar la gama de modelos existentes. Por lo que el modelo de negocio es considerado como una plantilla que incorpora dentro de él un conjunto de relaciones causa-efecto.

Existen distintos tipos de modelo de negocio que la gran mayoría de las empresas utilizan y adoptan para encontrar una ventaja competitiva. Entre otros, se tiene: i) *Publicidad*, el fundamento de este modelo gira en torno a la creación de contenido que las personas desean leer o ver y luego mostrar publicidad a sus lectores o espectadores; ii) *Afiliado*, este modelo de negocio está relacionado con el modelo de publicidad, pero tiene algunas diferencias específicas, utiliza enlaces incrustados en el contenido en lugar de anuncios visuales que son fácilmente identificables; iii) *Desintermediación*, este se refiere cuando se desea hacer y vender algo en las tiendas, normalmente trabaja con una serie de intermediarios para obtener su producto de la fábrica a la estantería de la tienda; iv) *Franquicia*, es cuando se concede una licencia para utilizar una idea de negocio o marca; v) *Mercado*, este permite a los vendedores listar artículos en venta y proporcionar a los clientes herramientas fáciles para conectarse con los vendedores. Este conjunto de modelos, refieren a un concepto tradicionalmente utilizado por las empresas, y que hasta antes de la globalización de mercados, generaba una ventaja competitiva [4].

Un modelo de negocio describe las bases con las que una empresa crea, proporciona y capta valor, [5] de manera tal que como organización es posible identificar quién es, cómo hace las cosas, y cómo puede intervenir

con el fin modificar lo establecido para hacer las cosas de manera más eficiente. A su vez, el Modelo Canvas, busca que los proyectos se gestionen como unidades de negocio, además, el potencial emprendedor en la gestión de proyectos; en otras palabras, es una herramienta pertinente en un contexto de emprendimiento e innovación, que, si bien el plan de negocio es importante, debe ser flexible, pendiente a responder a las oportunidades y necesidades de la organización donde se aplique.

Lo antes expuesto nos lleva a considerar que existen diferentes tipos de modelos o planes de negocios que fueron diseñados para actividades industriales, mercantiles y de servicio específicos, de tal forma que estos modelos tienen la estructura que la organización les requirió en su momento; aun cuando este tipo de modelos o planes tienen como limitante que son complejos y poco flexibles, la tendencia advierte que en la actualidad se requiere de modelos agiles y flexibles que permitan adaptarse a las distintas variables que influyen en la empresa.

En el presente trabajo se propone el diseñar de un modelo de negocio en una planta procesadora de alimento para cerdo, ubicada en el sur del estado de Sonora, México; la cual produce para atender las diferentes etapas[1] de desarrollo del animal, tales como cuarentena, de gestación, maternidad, destete y engorda. Cuenta con antigüedad de 30 años y provee a 28 granjas, distribuidas entre los estados de Sonora y Sinaloa. La principal actividad de los clientes es la producción de cerdos en pie, los cuales son engordados en tallas de 120 kg a 125 kg de acuerdo con los requerimientos de calidad en gradeo, rendimiento magro y crecimiento del producto. La planta procesadora tiene como proveedores principales a distintas asociaciones productoras de grano, así como también de aceites y concentrados con vitaminas, los cuales aportan insumos que se agregan a la fórmula para complementar el alimento elaborado.

Los proveedores juegan un papel preponderante para la planta procesadora, debido a que proporcionan la materia prima necesaria para la elaboración de los productos requeridos por sus clientes. Entre los proveedores y la planta procesadora de alimentos existe una alianza estratégica para garantizar el abasto adecuado y mantener la calidad del producto.

Con base en el estado de arte descrito se presenta este trabajo, el cual tiene como objetivo diseñar el modelo de negocios para la planta de alimentos para cerdo, mediante el Modelo de Negocios Canvas con la finalidad de mejorar su propuesta de valor actual [6]. Esto es, debido a la necesidad de la planta para actualizar y mejorar sus actividades o procesos claves. Para dicha planta es de suma importancia que sus colaboradores puedan identificar como contribuyen al desarrollo del Modelo de Negocio Canvas y puedan observar las áreas de oportunidad por medio de esta herramienta ágil [7]. Todo esto para que mejore la calidad del servicio que se brinda de una mejor forma a sus clientes conforme a la demanda y exigencias de su producto.

2. Trabajos relacionados

En el siguiente apartado se mencionan diferentes trabajos que fueron consultados con el fin de conocer la utilidad del Modelo de Negocios Canvas en diferentes ámbitos [8] en los cuales se ha logrado abstraer de una

[1] Estas etapas caracterizan el desarrollo cronológico y físico del cerdo, desde el nacimiento hasta alcanzar el estándar de calidad requerido.

forma sencilla cómo los negocios u organismos funcionan.

En el trabajo elaborado en [9], se abordó la problemática de las Mipymes que se dedican a la seguridad y salud ocupacional, ya que no logran cumplir con los estándares de calidad en distintos ámbitos y que normalmente se encuentran reguladas por las organizaciones gubernamentales, debido al incumplimiento en las normativas y la mejora de los estándares de calidad. Para esto se aplicó el Modelo de Negocio Canvas con la información más importante acerca del tamaño del mercado, necesidades de las empresas, así como de oferta y demanda. Para esto se utilizaron fuentes de información primarias, recopilación de datos y encuestas. Como resultados de este análisis se diseñó el plan de negocio para la creación de una empresa de Servicios de Seguridad y Salud Ocupacional para Mipymes, orientada a mejorar la seguridad y la salud ocupacional de los colaboradores que prestan sus servicios personales y profesionales a estas empresas demostrando la factibilidad del negocio para satisfacer sus necesidades.

En el trabajo presentado en [10], se abordó la problemática social referente al desempleo, debido al incremento del trabajo poco digno y a su vez el incremento del comercio informal que afecta de manera negativa a la sociedad. Por ello, este trabajo tuvo como objetivo desarrollar un Modelo de Negocio Canvas para la implementación de una empresa de servicios de aseo y limpieza para hogares, oficinas y negocios con el fin de formalizar y dignificar este tipo de actividad económica. Para esto se hizo uso de un buen planteamiento del segmento de mercado, el cual estuvo enfocado a la propuesta de valor que determinó ofrecer servicios integrales con altos estándares de calidad con el personal debidamente capacitado y certificado mostrando compromiso con los clientes.

Asimismo, en el trabajo informado en [11] se presenta la necesidad de identificar el modelo de negocio de una empresa consultora, la cual desconocía su mercado objetivo, es decir no había logrado establecer al cliente como su objetivo principal en relación a los servicios ofrecidos, por lo que su propuesta de valor no hacía diferencia con respecto a su competencia, su principal objetivo fue diseñar un modelo de negocio bajo la metodología del Modelo de Negocio Canvas. El resultado fue la propuesta diferenciadora de establecer un modelo fundamentado en métodos de ejecución basados en procesos de negocio, así como las buenas prácticas para orientar a la organización logrando ofrecer a sus clientes asesoría, revisión y corrección, optimización de tiempo y dinero, mejoramiento de procesos satisfaciendo las necesidades de los clientes como su principal propuesta de valor.

En el trabajo presentado en [12], muestra la problemática de la empresa comercializadora D’ Perfect Color, la cual no había implementado una herramienta de gestión que le permitiera conocer su propuesta de valor para sus clientes. Por ello, este estudio se enfocó a implementar el Modelo de Negocios Canvas para poder crear un modelo con el cual la empresa se pueda posicionar y obtener una ventaja por medio de una propuesta de valor diferenciadora sobre sus competidores. Con la aplicación del Modelo de Negocios Canvas se logró abordar el negocio desde un punto de vista abstracto, teniendo en cuenta las áreas clave a desarrollar como la prioridad e importancia otorgada a los clientes para el mejor funcionamiento del negocio, además de lograr una aproximación más cercana a la realidad. Es claro que los objetivos propuestos al inicio de este proyecto se lograron con satisfacción y se logró que la empresa fuera más competitiva.

El trabajo presentado en [13], trata los problemas del crecimiento poblacional como un detonador de demanda a los servicios de salud, así como también del crecimiento de la clase media como una consecuencia en la demanda de los productos farmacéuticos. Para esto se realizó un análisis y descripción del modelo de negocio de la empresa farmacéutica similares y su forma de crear valor. Esto fue mediante un estudio de caso en el que se examinó el éxito de la industria farmacéutica en México, donde se observó que el planteamiento de la propuesta valor estuvo enfocada en el conocimiento de su segmento de mercado, el cual estuvo basada por el crecimiento demográfico y el aumento de la clase media. Con el uso de la herramienta de Modelo de Negocio Canvas se identificó que este negocio parte de un sentido social, y que su diversificación estuvo impulsada por adquisiciones y combinaciones de productos y servicios del mercado, que los caracterizó no sólo por diversificación dentro de la misma zona de especialización, sino también por la presencia en nuevas áreas como laboratorios de producción de fármacos, transportación para su propia distribución, ofrecimiento de franquicias y laboratorios de servicios médicos propios, las cuales fueron el factor de su éxito haciendo diferencia con su competencia.

Como se puede notar, los trabajos presentados en esta sección demuestran que el Modelo de Negocios Canvas es una herramienta que ha sido útil para identificar la propuesta de valor de empresas u organismos de distintos giros basados en problemas diversos. Además este modelo permite hacer análisis para conocer el estado actual de las empresas y permite solucionar los problemas relacionados con la interrelación de las diferentes áreas para satisfacción de los clientes internos y externos, que se puedan presentar mediante una metodología ágil, flexible y que permite establecer objetivos de una manera más rápida, lo que facilita adaptarse a los diferentes cambios que enfrentan las empresas u organismos.

3. Aplicación del Modelo de Negocio Canvas en la Planta comercializadora de alimentos para cerdo

El Modelo de Negocio Canvas [5] es una herramienta que facilita la captura, la visualización, el entendimiento y la lógica del negocio ofreciendo una visión distinta de la empresa que permite la comprensión de las relaciones entre las áreas que intervienen en las decisiones. Este modelo describe las bases sobre las que una empresa crea, proporciona y capta valor; está dividido en nueve módulos básicos que reflejan la lógica que sigue una empresa para conseguir ingresos, cubren las cuatro áreas principales de un negocio: clientes, oferta, infraestructura y viabilidad económica [14].

Para documentar el Modelo de Negocios Canvas para la planta de alimentos para cerdo se realizaron cuatro actividades, que se describen a continuación:

1. Entrevista inicial.- Se realizó una entrevista al responsable de la planta con el fin de conocer los diferentes procesos que constan desde la entrada del camión con materia prima hasta la salida del mismo y la carga del producto terminado.

2. Visita a planta. - se visitó la planta de elaboración de alimentos para cerdo con el objetivo de observar las actividades que se realizan de manera específica, por lo que se encontraron 9 técnicas clave que conforman el proceso de producción, las cuales son i) recibimiento de la materia prima, ii) pesaje de la materia prima, iii) descarga de materia prima, iv) proceso de molienda, v)

premezclados, vi) mezclados, vii) envío de producto terminado a tolvas, viii) carga del camión con el producto para su envío, ix) pesaje el camión con el producto. Además se realizó un inventario de recursos y de las instalaciones con las que se cuenta.

3. Segmentación de clientes. - se realizó una segmentación de los clientes de la planta de elaboración de alimentos para cerdo con el uso de formatos de diseño del segmento de clientes propuesto por [15] que fueron llenados por personal responsable de cada una de las 15 granjas que conforman los clientes de la planta, donde los encargados de ellas proporcionaron información relacionada con datos demográficos, geográficos, tipo de comunidad, capacidad de producción, necesidades como cuántas toneladas de alimento se requieren al mes, qué le ofrece cada granja a la planta de elaboración de alimentos para cerdo, tipos de granja, accesibilidad a las granjas, su modelo de comunicación, el número de granjas del mismo tipo, frecuencia de pedido de alimento, así como sus indicadores para el producto recibido.

4. Elaboración de Formatos "value proposition". - con la información obtenida en la segmentación y la información recolectada de los distintos procesos obtenidos de la visita a la planta de elaboración de alimentos para cerdo, se elaboraron las propuestas de valor y el mapa de empatía [16] que ayudaron a la elaboración del Modelo de Negocio Canvas.

Por lo anterior, y con el fin de establecer el conocimiento que la planta requiere de las diversas áreas que la conforman, se realizó el lienzo de Modelo de Negocio Canvas, que consta de 9 pasos, los cuales se explican a continuación:

1. Segmento de cliente: para este indicador la planta procesadora cuenta en su cartera de clientes con 28 empresas ubicadas en la región sur de Sonora, a las cuales provee en exclusiva el alimento. En el presente estudio participaron 15 granjas, entre las cuales se cuentan 5 denominadas de ciclo completo, 1 granja de destete, 8 de tipo engorda y 1 de engorda y destete.

2. Propuesta de valor: en este paso se tuvo como objetivo conocer qué iniciativa puede aportarles valor a los clientes de la planta, se identificó que ésta garantiza la calidad del alimento con estándares superiores respecto a la competencia, ofrece una exclusividad del producto elaborado, facilita cambios en las entregas del producto, mejora la accesibilidad y comunicación con los clientes.

3. Canales de distribución: esta se refiere a cómo se le hace llegar la propuesta de valor al cliente, el objetivo identificado es tener una buena distribución de los productos ofrecidos mediante entregas a través de una flotilla de camiones con salidas programadas gradualmente con cada uno de los clientes según lo especifiquen los pedidos solicitados.

4. Relación con el cliente: el objetivo fue conocer la relación de los clientes con las granjas, aquí se identificó que existe un intercambio de información con cada uno de los clientes vía telefónica. Esto con base a las especificaciones que el cliente requiera para posteriormente realizar la programación necesaria con respecto a la demanda.

5. Flujos de ingresos: en este paso se determina cómo se obtendrá la ganancia por medio de la propuesta de

valor, identificando que la venta de alimento a los distintos clientes establecidos genera cerdos de mayor calidad, así como también el aprovechamiento de los desperdicios para su venta.

Después de que se logró obtener la información necesaria sobre el entorno de la organización, se procedió a analizar los pasos que contribuyen a la propuesta de valor.

6. Recursos clave: en este punto se muestra lo que se requiere para llevar a cabo las actividades de la planta. Los recursos que se encontraron fueron la maquinaria para la elaboración del producto, las máquinas de molienda, bandas transportadoras, tolva bascula, mezcladoras, silos de almacenamiento, tolvas de producto terminado, edificios y las redes de distribución. También se considera como recurso clave los insumos (e.g. vitaminas, aceites, concentrados, granos, etc.) que son recursos esenciales para la elaboración del producto. También son clave los recursos intelectuales que incluyen las fórmulas de los alimentos y las bases de datos de los clientes. Finalmente se identificó como recursos financieros posibles nuevos empleados, sistemas a implementar, sistemas gráficos visuales, entre otros.

7. Actividades clave: este punto nos muestra los procesos que son parte fundamental para lograr el objetivo principal de la planta. Las actividades encontradas que inciden en la propuesta de valor son:

Pruebas de calidad: se analiza que la materia prima recibida cumpla con los estándares establecidos por la organización.

Almacenamiento: el producto recibido pasa a los silos de almacenamiento donde se resguarda para el momento de su uso.

Proceso de molienda: la materia prima recibida pasa a este proceso para poder aprovecharla de la mejor manera.

Mezclados: se combinan los diferentes ingredientes que se requieren para elaborar el alimento.

Carga del producto: se procede a llenar los camiones para la distribución del alimento a los diferentes clientes que lo solicitaron.

Envío de producto: el producto es enviado al cliente que lo solicitó en el camión de transporte.

8. Aliados clave: este paso tiene como objetivo conocer quiénes son los aliados principales para el funcionamiento de la planta de elaboración de alimentos para cerdo. Los principales socios clave que fueron encontrados son:

Vimifos: este proveedor abastece a la planta de medicamentos y vitaminas necesarias para la realización de sus productos, ofrece los mejores precios y a su vez la calidad requerida por los estándares establecidos, así como también es un proveedor de emergencia y provee de alimentos cuando se necesita.

Almacenadoras de grano: almacenadoras como San Rafael, Aríc, Pacsa y C. de Palma, son las que proveen de materia prima a la planta de elaboración de alimentos para cerdo y tienen abastecida a la planta de distintos granos que

se utilizan para la elaboración del alimento para cerdo.

Nutrikowi: este es un proveedor de emergencia al igual que Vimifos que abastece de alimento a la planta cuando se necesita.

9. Estructura de costos: en este punto se dividen los costos que impactan en la propuesta de valor, de los cuales se identificaron nómina, materia prima, producción, mantenimiento, flotilla de camiones de transporte del alimento y operación. Estos costos son con los que cuenta la planta de elaboración de alimentos para cerdo y que son de suma importancia en el logro de las metas y objetivos de negocio.

Con toda la información recopilada de la organización se documentó el lienzo del Modelo de Negocio Canvas con sus elementos y especificaciones correspondientes como se muestra en la Tabla 1.

Tabla 1. Lienzo modelo de negocio Canvas aplicado a la Planta de elaboración de alimento para cerdo.

<table>
<tr>
<td rowspan="2">Aliados Clave

Los socios clave que apoyan a la organización son:
- Vimifos
- Nutrikowi
- Almacenadoras de grano</td>
<td>Actividades Clave

- Pruebas de calidad.
- Almacenamiento.
- Molienda.
- Mezclados.
- Cargar alimento.
- Envío de producto.</td>
<td rowspan="2">Propuesta de Valor

- Estándares de calidad de alimento respecto a la competencia.
- Exclusividad del producto elaborado.
- Optimización y disponibilidad de cambios en la entrega del producto.</td>
<td>Relación con el Cliente

Hay un intercambio de información con cada uno de los clientes por medio de la vía telefónica con base a las especificaciones que este requiere y realizar la programación necesaria.</td>
<td rowspan="2">Segmentos de Clientes

Se analizaron como segmentos un total de 15 granjas porcinas de la organización, segmentando por tipo de granja las cuales fueron:

5 granjas ciclo completo
1 granja de destete
8 granjas de engorda
1 granja de engorda y destete.</td>
</tr>
<tr>
<td>Recursos Clave
- Máquina de molienda
- Tolva báscula
- Mezcladoras
- Silos de almacén
- Tolvas de producto terminado.
- Banda transportadora</td>
<td>Canales

La distribución se realiza a través de una flotilla de camiones previamente programada para la salida con cada uno de los clientes.</td>
</tr>
<tr>
<td colspan="2">Estructura de Costos

Los costos están divididos por:
- Nomina.
- Materia prima.
- Producción.
- Mantenimiento.
- Flotilla de camiones.
- Contabilidad.
- Servicios.</td>
<td colspan="3">Estructura de Ingresos

- Venta de alimentos a las distintas granjas.
- Aprovechamiento de los desperdicios salientes para su venta.
- Generar animales de calidad.</td>
</tr>
</table>

4. Consideraciones finales e implicaciones

En este artículo se diseña un modelo de negocio para una planta de elaboración de alimentos para cerdo utilizando la metodología del Modelo de Negocio Canvas, con la finalidad de que la empresa visualice su propuesta de valor actual y pueda reducir su riesgo detectando oportunamente las áreas de oportunidad, debido a que con la información resultante los directivos de la

planta pueden plantear estrategias y acciones coherentes a su misión y visión como empresa.

Esta procesadora de alimentos no dispone de una plantilla Canvas, aun cuando los tomadores de decisiones requieren información disponible de su situación actual, en este sentido el Modelo de Negocio Canvas permite que esta información sea generada de manera ágil y flexible, debido a que la alta dirección requiere diseñar y controlar no solo aspectos clave para satisfacer a los clientes actuales, sino también incluir aspectos que permitan la buena marcha y desarrollo del modelo de negocio [17]. Es aquí donde se estableció la necesidad aplicar este modelo para que de éste resulten iniciativas y objetivos enfocados a medir y controlar los aspectos relevantes de este negocio.

El Modelo de Negocios Canvas tienes varias limitantes que es importante tomar en cuenta, entre ellas la más común es el nivel de abstracción [8], lo cual dificulta la implantación de estrategias, pues no proporciona indicadores ni las relaciones que tienen los factores clave involucrados en el modelo, por lo que se debe considerar el identificar las fortalezas y debilidades de cada proceso del modelo de negocio y de esta forma poder elaborar iniciativas que ayudan a reducir las debilidades y consolidar las fortalezas.

Es por ello que se deben seguir las siguientes implicaciones que trae consigo el Modelo de Negocios Canvas [18] para la planta de elaboración de alimentos para cerdo:

1. Detectar fortalezas y debilidades de cada una de las áreas que conforma este Modelo de Negocio. Con esto se pueden establecer iniciativas que contribuyan a la consecución de los objetivos y a su vez reforzar las áreas débiles con base a las iniciativas establecidas.
2. La flexibilidad de este modelo facilita la toma de decisiones inmediatas sin esperar a identificar un indicador bajo, es decir detecta desviaciones que dificulten o impidan el logro de objetivos sin tener cuantitativamente la información del indicador. Esto con el fin de lograr identificar distintos indicadores de los procesos antes de que estos se encuentren en estado crítico.
3. Este modelo permite adaptabilidad, de tal manera que los procesos que no se encuentren en el lienzo (Figura 1) se pueden ir agregando y posteriormente pueden ser propuestos para la ayuda del cumplimiento de objetivos de la planta. Esto es de gran ayuda debido a que el modelo no se limita a ciertos objetivos y conforme la planta modifique objetivos y metas el Modelo de Negocio Canvas se acoplará a ello.
4. Los apartados de flujo de ingresos y costos del Modelo, facilitan a la planta la visualización y análisis de los objetivos trazados por los directivos, puesto que ayudan a las metas del negocio y de ser necesario mejoran algunos de ellos.
5. Los apartados de propuesta de valor, segmento de clientes y relaciones del cliente, facilitan el análisis de los objetivos de la propuesta de valor ya que dicha propuesta puede ser reforzada o modificada con base a lo que el cliente esté demandando, esto según la percepción y la calidad del producto ofrecido por la misma planta.
6. Los apartados de canales de distribución y actividades clave del modelo facilitan el análisis de cómo el producto es transportado de una manera eficiente y a su vez contempla contingencias que resulten de imprevisto.

7. Los apartados de recursos y aliados clave ayudan a identificar y reflexionar sobre los diferentes tipos de alimentos con los que dispone la planta y como estos pueden ser aprovechados de una forma más eficiente por nuestros clientes según sus especificaciones.

El Modelo de Negocios Canvas ha sido utilizado y probado en diversos giros a nivel internacional con éxito [19]. Esta Herramienta aun cuando fue diseñada para ayudar a definir propuestas de valor para startups ha sido utilizada para abstraer Modelos de Negocios Canvas de empresas con procesos complejos otorgando un planteamiento holístico, además ha comprobado efectividad para establecer iniciativas encaminadas a la consecución de objetivos, además su planteamiento abstracto facilita la flexibilidad en el acomodo de los elementos de sus distintas áreas, esto es, permite observar a nivel dirección que iniciativas funcionan y contribuyen a la consecución de objetivos.

Con este Modelo de Negocio Canvas se pretende inferir un mejor planteamiento de objetivos por medio de la identificación de los distintos elementos del modelo en la propuesta de valor, generando objetivos mínimos viables.

5. Conclusiones

Este trabajo suministró herramientas para el desarrollo y análisis de negocios que son indispensables en un mundo globalmente competitivo, el Modelo de Negocios Canvas permitió abstraer las actividades operativas y administrativas de la Planta de alimento para cerdo ofreciendo las herramientas necesarias para la definición de la propuesta de valor en base a la aplicación de utilidades del modelo que ayudaron a conocer las necesidades y deseos de sus clientes, esto es, se analizó la segmentación de clientes, también se realizó un estudio de la segmentación por medio del análisis del mapa de empatía, logrando consolidar la propuesta de valor actual de la planta, la identificación de las áreas restantes del Canvas ofrecen una visión ordenada de las actividades de la organización y la interrelación de las mismas ofreciendo la oportunidad de realizar modificaciones considerando distintos escenarios que se puedan presentar por medio de iniciativas que contribuyan a la consecución de los objetivos, estas iniciativas pueden establecerse para cada escenario planteado. Lo antes expuesto demuestra la flexibilidad del Modelo de Negocios Canvas y la capacidad de este en incidir en la toma de decisiones de la organización. De allí que toda organización que cuente con su modelo de negocios bien estructurado puede sortear las incertidumbres de las fuerzas tanto internas como externas contemplando desde una nueva perspectiva las oportunidades y amenazas en un escenario global y competitivo.

6. Referencias

[1] Osterwalder, Alexander, Pigneur Yves, and Tucci, Christopher L. 2005. *"Clarifying Business Models: Origins, Present, and Future of the Concept"*, Communications of the Association for Information Systems: Vol. 16, Article 1. DOI: 10.17705/1CAIS.01601 Available at: https://aisel.aisnet.org/cais/vol16/iss1/1.

[2] Zott, C., & Amit, R. 2013. *The business model: A theoretically anchored robust construct for strategic analysis*. Strategic Organization, 403-411.

[3] Baden-Fuller, C., & Mangematin V, 2013. *"Business models: A challenging agenda"* Volume: 11 issue: 4, page(s): 418-427 Available at: https://doi.org/10.1177/1476127013510112.

[4] Parson, Noah, 2012. *What Is a Business Model? Business Models Explained*; https://articles.bplans.com/what-is-a-business-model-business-models-explained/.

[5] Osterwalder, A., & Pigneur, Y, 2010. *Business model generation : A handbook for visionaries*. Clearance Center, Inc. USA.

[6] Cedillo Auquilla Diego Gerardo, 2010. *Generación de un modelo de negocios para la gestión e implementación de una empresa que solvente las necesidades de asesoría, dirigido de manera particular a las mipymes en temas de seguridad y salud ocupacional*. Universidad Politécnica SALESIANA. Ecuador.

[7] Osterwalder Alexander & Pigneur Yves, 2010. *business Model Generation*, Barcelona, España, Desusto.

[8] Osterwalder, Alexander, 2009. *Obstáculos de la innovación en los modelos de negocios.*

[9] Soto Nuñez Antonio, Sánchez Galván Fabiola, Bautista Santos Horacio, Purroy Vázquez Rubén, 2017. *Uso de la metodología business model canvas para comercializar un dispositivo electrónico para tratamiento de artritis reumatoide*, México, Revista de la Alta Tecnología y la Sociedad.

[10] Benavides Sánchez Paula, 2017. *Profundización para tecnologías gestión de marketing para el emprendimiento social*. Universidad Nacional Abierta y a Distancia. Pitalito-Huila, Colombia.

[11] Sierra Basantes Erika Johana, 2015. *Ventajas del modelo de negocio "Lean Canvas" en comparación al plan de negocios tradicional para emprendimientos dinámicos nacientes en el distrito Metropolitano de Quito*. Universidad Del Pacífico. Quito.

[12] Hernández Bedoya Estrella Maris, Silva Gómez Armando, 2016. *Modelo de negocios Canvas aplicado a la empresa D`Perfect color comercializadora de productos de lencería*. Universidad Nacional Abierta y a Distancia. Pitalito, Huila, Colombia.

[13] Sánchez Guerrero Marisol, Arellano González Alejandro, 2017. *Utilización del modelo de diagnóstico Canvas en el análisis de un caso de la industria farmacéutica en México*. Revista Ciencias Administrativas. México.

[14] Osterwalder, Alexander, Pigneur Yves, 2005. *Clarifying Business Models: origins, present, and future of the concept. Communication of the Association for Information Systems*, Clearance USA. Center, Inc.

[15] Mouriño Javier, 2018. *Authentic Leaders: Emprendimiento basado en innovación e impacto*, México, ilab.

[16] Sánchez Vásquez José Manuel, Vélez Elorza María Luisa, & Araújo Pinzón Pedro 2016. *Balanced scorecard para emprendedores: desde el Modelo Canvas al cuadro de mando*, España, Revista de la facultad de ciencias Económicas de la Universidad Militar Nueva Granada.

[17] Castillo S, 2017. *Aplicación de Teoría C-K y modelo de negocios Canvas para asistir el diseño de nuevos servicios*, Coloquio de Investigación Multidisciplinaria.

[18] Clark, T., Osterwalder, A., Pigneur, Y, 2012. *Business Model You: A One-Page Method For Reinventing Your Career*, London, Wiley.

[19] Osterwalder, Alexander, Pigneur Yves, 2003. *"Towards Strategy and Information Systems Alignment through a Business Model Ontology"*, USA, Proceedings of the Annual Conference of the Strategic Management Society.

El Examen de Grado se presentó el día 06 de septiembre de 2019

en Ciudad Obregón, Sonora, México

según consta en el acta número 2303

registrada por el Departamento de Registro Escolar.

Folio No. **44121**

CERTIFICACIÓN DE ANTECEDENTES ACADÉMICOS
POSGRADO MAESTRÍA

No. de control: 33265

Nombre: RAMOS SOLÍS CARLOS EDUARDO

Nacionalidad: MEXICANA

Grado: MAESTRO EN TECNOLOGÍAS DE LA INFORMACIÓN PARA LOS NEGOCIOS

CURP: RASC910806HSRMLR06

Estudios profesionales

Licenciatura: EN ADMINISTRACIÓN

Nº de Cédula: 042323 (CÉDULA DEL ESTADO DE SONORA)

Estudios de posgrado

Institución: INSTITUTO TECNOLÓGICO DE SONORA

Maestría: EN TECNOLOGÍAS DE LA INFORMACIÓN PARA LOS NEGOCIOS

Entidad Federativa: 26 Periodo: 2017-2019

Examen de grado, *fecha:* 06.SEP.19

Institución Educativa que otorga el grado

Mtra. Margarita Ross Ramírez
Jefa del Departamento de Registro Escolar

Nº 16773