What do researchers mean by “the right requirements elicitation techniques”?

¿Qué significa para los investigadores la “técnica de educción correcta”?

Dante Carrizo¹  Cristián Ortiz¹  Luis Aguirre¹

Recibido 6 de octubre de 2014, aceptado 22 de julio de 2015
Received: October 6, 2014  Accepted: July 22, 2015

ABSTRACT

Researchers often claim that the most appropriate elicitation techniques should be chosen to capture software requirements, but: what does it mean for them? This work carried out a systematic mapping on comparisons among elicitation techniques for knowing the constructs utilized to represent their appropriateness. This study identified 43 works that utilized 58 ways of measuring the goodness of the techniques. These metrics were classified, mainly, by types of constructs such as quality, adequacy, effectiveness, quantity and efficiency. The results show the large dispersion occurring between researchers on how to select the most appropriate technique for each elicitation session and, therefore, more convergent proposals are required.

Keywords: Elicitation requirements, techniques appropriateness, systematic review, elicitation techniques requirements, systematic mapping.

RESUMEN

Para capturar los requisitos de software los investigadores frecuentemente pregonan que se debe seleccionar la técnica de educción más adecuada, pero ¿qué significa para ellos esto? En este trabajo se realiza un mapeo sistemático de investigaciones acerca de comparaciones de técnicas para captura de requisitos con el fin de conocer los constructos que se utilizan para representar su adecuación. El estudio identificó 43 trabajos que utilizaron 58 formas de medir la bondad de las técnicas. Estas métricas se clasificaron principalmente en tipos de constructos de calidad, adecuación, efectividad, cantidad y eficiencia. El estudio demuestra la gran dispersión que existe entre los investigadores respecto de cómo seleccionar la técnica más apropiada para cada sesión de educción y que, por tanto, se requiere más propuestas convergentes.

Palabras clave: Educación de requisitos, adecuación de técnicas, revisión sistemática, técnicas de educación de requisitos, mapeo sistemático.

INTRODUCTION

The software requirements process, regardless the theoretical model considered (eg: [1-3]), has as first phase, the capture of relevant information from the problem domain and the stakeholders’ needs to specify the requirements that the software product must meet. To do this, there are a myriad of techniques, some of them from other disciplines such as psychology or social sciences [4].

Because of this diversity of origins, and the intrinsic nature of each technique, its performance in capturing information from stakeholders is obviously different.
Therefore, practitioners need to have methods and tools to select the most appropriate technique to be used at any time of the software development project [5].

Most of these methods [6-7] considered influential aspects of the context of the process that, in opinion of their authors, modulate the outcome of each elicitation session. However, in order to compare and select the most appropriate elicitation technique you need to use a metric system that faithfully represents this construct [8].

In order to move in this direction, this study aims to gather how researchers represent the appropriateness of the techniques. For this, we have conducted a systematic mapping of the publications on elicitation techniques, used in various domain areas. If the researchers visualize the elicitation techniques in a convergent and valid way then such vision may guide future experimental work on comparisons of techniques and thus form a body of knowledge about their performance. Conversely, if there is divergence of views, there will be the need for more research proposals to determine the construct of appropriateness. Once the researchers’ view is established, it is intended by the authors to continue capturing the practitioners’ view proceeding to compare them and then to induce a model that represents this construct.

For the presentation of this work, the paper is structured as follows: in next section backgrounds on systematic mapping and related work are discussed, then, the research methodology is described and the results of this work are presented, finally, the conclusions and future work are shown.

BACKGROUND AND RELATED WORK

A systematic mapping study is a methodology used in other areas such as medical research, but it has been suitable for being used in areas of software engineering [9]. It requires less effort than a systematic review of literature as it provides a broader view in order to identify more and less treated areas of research. A systematic review considers obtaining more detailed information than can be processed to form a body of knowledge based on empirical evidence with certain degrees of reliability.

The main goal of the systematic mapping performed in this study is to obtain an overall quantitative and qualitative overview of existing research on the performance evaluation of elicitation techniques.

This review has noted a significant number of studies on the use of the techniques, however, there are not any researches about the techniques with a focus on establishing a unique way to measure their performance or that show the great existing diversity. Only one work, Dieste and Juristo [10], deals with the dependent variables that represent the effectiveness of elicitation techniques in empirical studies. However, there is no further discussion of this diversity and these variables are only treated in order to generalize and obtain empirical evidence.

METHODOLOGY

Research question

The systematic mapping begins with the specification of the research questions to be answered. In this case, as it is an exploratory study, a main and three secondary questions are considered.

The main question is: What do researchers mean by appropriateness of elicitation techniques? This question relates directly to the constructs that represent the performance of a technique in a given contextual situation.

The other questions are:

- What techniques are used in studies about adequateness constructs?
- In what application domains certain constructs for appropriateness of the techniques are more commonly used?
- What kinds of studies use certain constructs for appropriateness of techniques?

Selection of studies

The identification stage of primary studies was performed by searching in the following databases: IEEEXPLORE, ACM DL, SCIENCE DIRECT. Eventual Internet searches were also conducted between the references of the selected articles and other papers already identified. The searching period included from the year 1984 to November 2013.

The searching string was:

(Requirements) AND (elicitation OR gathering OR capture OR acquisition) AND (techniques OR methods)
This search string was adjusted to the formats of each database and focused on the publications abstracts. The following inclusion/exclusion criteria were considered for selecting studies:

- Scientific publications related to elicitation techniques and requirements related books were considered.
- Studies may focus on one or more techniques. Thus, candidates could be publications on the application of a technique to a case study, descriptive and prescriptive techniques comparisons, and empirical evaluations of techniques, among others.
- No restriction on the application area of the literature was performed, as long as elicitation techniques that are used in software engineering are used.

To identify primary studies the following filters were performed:

First Filter (1F)
- Title: Each author reviewed the titles of publications contained on each database. One of them also made the opportunistic searching.
- Abstract: The abstract of those publications that were selected by their title, were reviewed by each assigned author.

Second Filter (2F)
- Full Text: Finally, publications that passed the previous filter were subjected to a revision of their contents. Those articles, in which an assigned author had doubts, were reviewed by the three authors assigned.

Classification scheme
In order to answer the research questions, the literature was reviewed focused on the following aspects of the classification scheme: Adequateness metric, elicitation techniques used, application domain, type of study, degree of reliability.

The adequateness metric refers to the definition and/or formula showing how the authors of the publication state the performance of the technique. The techniques used are those considered in each publication. The application domain is the area in which one or several elicitation techniques are used. The type of study refers to the main method that supports the study. The degree of reliability is a validity assessment of the authors’ proposal. In this last case, the following values were considered:

- Low: expert opinions of books or articles, or empirical studies without statistical validation results
- Media: empirical studies with results without statistical validation
- High: empirical studies with conclusive results and statistical validation

ANALYSIS OF THE RESULTS

The results of the literature searching are summarized in Figure 1 by DCM format. After applying the filters presented above, 43 primary articles were identified. The eventual searching was the most productive, since much of the related literature belongs to conferences and in specialized journals on software requirements but which are not indexed in the selected databases.

Of these primary studies [6, 11-52], 58 constructs or ways of evaluating the performance of the techniques were identified. This means that some of the reviewed studies proposed more than one independent metric. For reasons of space it is not possible to show the detail of the information obtained from the review of the identified bibliography. However, it can be reviewed at https://docs.google.com/file/d/0BzV_sIkoJTGAxy1oSzVvOU9ZX0k/edit. A summary of the primary papers found, with information about their origin, is shown in Table 1.

This information was analyzed considering four related aspects with the research questions: constructs or metric by which the appropriateness of the techniques is measured, the types of studies and their relationship with these constructs, types of application domains and their relationship with the constructs, and finally, the techniques used in these studies. Each of these results is reviewed in the following sections.

Constructs of appropriateness
When the identified papers were analyzed, 58 different ways to measure the appropriateness of the techniques used were obtained. Many of these measures are quantitative and have metric shape, such as: Number of generated concepts per time unit,
Table 1. Identified works.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Ref.</th>
<th>Application Domain</th>
<th>Study Type</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen, Khoo and Yan 2002</td>
<td>[12]</td>
<td>Marketing</td>
<td>Case Study</td>
<td>Low</td>
</tr>
<tr>
<td>Laguna, Marqués and García 2003</td>
<td>[14]</td>
<td>Software Eng</td>
<td>Case Study</td>
<td>Low</td>
</tr>
<tr>
<td>Laporti, Borges and Braganholo 2009</td>
<td>[16]</td>
<td>Software Eng</td>
<td>Experiment</td>
<td>Medium</td>
</tr>
<tr>
<td>Saiedian and Dale 2000</td>
<td>[17]</td>
<td>Software Eng</td>
<td>Theoretical</td>
<td>Low</td>
</tr>
<tr>
<td>Moore and Shipman 2000</td>
<td>[18]</td>
<td>Software Eng</td>
<td>Experiment</td>
<td>High</td>
</tr>
<tr>
<td>Sutcliffe 1997</td>
<td>[19]</td>
<td>Software Eng</td>
<td>Case Study</td>
<td>Medium</td>
</tr>
<tr>
<td>Sadiq, Ghafir and Shahid 2009</td>
<td>[20]</td>
<td>Software Eng</td>
<td>Experiment</td>
<td>Low</td>
</tr>
<tr>
<td>Kausar, Tariq, Riaz and Khanum 2010</td>
<td>[22]</td>
<td>Software Eng</td>
<td>Theoretical</td>
<td>Low</td>
</tr>
<tr>
<td>Sen and Hemachandran 2010</td>
<td>[23]</td>
<td>Software Eng</td>
<td>Case study</td>
<td>Low</td>
</tr>
<tr>
<td>Boulila, Hoffmann and Herrmann 2011</td>
<td>[24]</td>
<td>Software Eng</td>
<td>Experiment</td>
<td>Medium</td>
</tr>
<tr>
<td>Authors</td>
<td>Ref.</td>
<td>Application Domain</td>
<td>Study Type</td>
<td>Reliability</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>------</td>
<td>--------------------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Hickey and Davis 2003</td>
<td>[26]</td>
<td>Software Eng</td>
<td>Theoretical</td>
<td>Medium</td>
</tr>
<tr>
<td>Mustafa 2010</td>
<td>[27]</td>
<td>Software Eng</td>
<td>Experiment</td>
<td>Medium</td>
</tr>
<tr>
<td>Browne and Rogich 2001</td>
<td>[28]</td>
<td>Information Syst</td>
<td>Experiment</td>
<td>Medium</td>
</tr>
<tr>
<td>Duggan and Thachenkary 2003</td>
<td>[29]</td>
<td>Information Syst</td>
<td>Experiment</td>
<td>High</td>
</tr>
<tr>
<td>Agarwal and Tanniru 1990</td>
<td>[31]</td>
<td>Knowledge Eng</td>
<td>Experiment</td>
<td>High</td>
</tr>
<tr>
<td>Batista and Carvalho 2003</td>
<td>[32]</td>
<td>Software Eng</td>
<td>Theoretical</td>
<td>Low</td>
</tr>
<tr>
<td>Tsumaki and Tamai 2005</td>
<td>[33]</td>
<td>Software Eng</td>
<td>Theoretical</td>
<td>Low</td>
</tr>
<tr>
<td>Maiden and Rugg 1996</td>
<td>[34]</td>
<td>Software Eng</td>
<td>Theoretical</td>
<td>Low</td>
</tr>
<tr>
<td>Zhang 2007</td>
<td>[36]</td>
<td>Software Eng</td>
<td>Theoretical</td>
<td>Low</td>
</tr>
<tr>
<td>Aranda, Vizcaíno, Cechich and Piattini 2005</td>
<td>[38]</td>
<td>Software Eng</td>
<td>Theoretical</td>
<td>Low</td>
</tr>
<tr>
<td>Cordbrige, Rugg, Burton and Shadbolt 1994</td>
<td>[39]</td>
<td>Knowledge Eng</td>
<td>Experiment</td>
<td>High</td>
</tr>
<tr>
<td>Chao and Salvendy 1995</td>
<td>[40]</td>
<td>Knowledge Eng</td>
<td>Experiment</td>
<td>High</td>
</tr>
<tr>
<td>Lauesen 2002</td>
<td>[41]</td>
<td>Software Eng</td>
<td>Theoretical</td>
<td>Low</td>
</tr>
<tr>
<td>Hua 2008</td>
<td>[42]</td>
<td>Knowledge Eng</td>
<td>Theoretical</td>
<td>Low</td>
</tr>
<tr>
<td>Scapolo and Miles 2006</td>
<td>[43]</td>
<td>Information Syst</td>
<td>Survey</td>
<td>Low</td>
</tr>
<tr>
<td>Wagner, Chung and Najdawi 2003</td>
<td>[45]</td>
<td>Knowledge Eng</td>
<td>Survey</td>
<td>High</td>
</tr>
<tr>
<td>Holsapple, Raj and Wagner 2008</td>
<td>[46]</td>
<td>Knowledge Eng</td>
<td>Experiment</td>
<td>Medium</td>
</tr>
<tr>
<td>Rugg, Cordbrige, Burton and Shadbolt 1992</td>
<td>[48]</td>
<td>Knowledge Eng</td>
<td>Experiment</td>
<td>Low</td>
</tr>
<tr>
<td>Sauer, Schramme and Rüttinger 2000</td>
<td>[49]</td>
<td>Product design</td>
<td>Experiment</td>
<td>High</td>
</tr>
<tr>
<td>Fowlkes, Salas and Baker 2000</td>
<td>[50]</td>
<td>Knowledge Eng</td>
<td>Experiment</td>
<td>High</td>
</tr>
<tr>
<td>Vásquez, Sánchez, Medina and Amescua 2013</td>
<td>[51]</td>
<td>Knowledge Eng</td>
<td>Experiment</td>
<td>High</td>
</tr>
<tr>
<td>Dhaliwal and Benbazat 1990</td>
<td>[52]</td>
<td>Knowledge Eng</td>
<td>Theoretical</td>
<td>Low</td>
</tr>
</tbody>
</table>

Number of identified use cases, Numbers of goals elicited, Number of requirements, among others. Many others are qualitative, such as: Different levels and types of knowledge and information, Types of requirements, etc. And finally others are quite subjective and do not have defined metrics, such as: Fit with problem types (Analysis problems, Synthesis problems, Combination problems), Fit properties of requirements elicitation techniques with contextual facets, Suitability of elicitation techniques according to stakeholders’ preferences, among others. These representations were grouped according to nature in 13 main constructs, which are shown in Figure 2 together with the number of times that are used by researchers to define the adequacy of elicitation techniques.

As can be clearly seen, Quality of information elicited is the most used construct to represent the concept of appropriateness, followed by Adequacy as such, Quantity of Elicited Information and Effectiveness.

Due to the constructs Quality level of understanding and Quality of elicited information belong to the
same concept, were joined in one construct called Quality, the same was done with the constructs Quantity of information elicited and Quantity of knowledge elicited forming the construct Quantity. On the other hand, the constructs with low incidence: Productivity, Utility, Performance, Specification and Usability, were named Others because their low use. Figure 3 shows those 6 types of constructs as final result of the characterization table and the percentage in which they are used to define the adequacy. These types will be taken into account in the rest of this article as a basis for various analyses.

**Type of study**
The identified papers use the constructs, in its proposals, in form of expert opinions or empirical studies. Figure 4 shows the distribution of the 43 items according to the type of study presented, being classified in 4 categories: Experimental, Theoretical, Case Study and Survey.

Figure 5 illustrates the detail of the distribution showing at the same time that nearly half of the constructs are utilized in experimental articles. Theoretical proposals also have a significant proportion (33%). In this case, a preference to use the Adequacy construct by researchers is observed, that is to say, researchers claim that this or that technique is the most appropriate based in many cases on their experience and in others in the literature. They also use the Quality construct, but in a descriptive or qualitative form.

The case of the research conducted by experiments is different. In them, researchers use mostly constructs Quality and Quantity to define which technique is most appropriate based on empirical evidence from controlled experiments. In these studies, an increase in the presence of other “quantitative” constructs like Effectiveness and Efficiency is also observed, and opposite to the case of the Theoretical studies, the Adequacy construct is the least used.
The articles that deal with a Case Study show a similar trend of the Experiment ones, but with very few articles. Following this approach, conclusive information cannot be extracted. Survey based studies presented the same situation, with a few of the articles. These latter types of studies mainly focus on constructs of Quantity and Quality of elicited information.

Application Domains

Among the selected articles six application domains were identified: Information Systems, Knowledge engineering, Software Engineering, Marketing Analysis, Discrete Cell Control Systems (Manufacturing), and Product design. The last three joined a group called Others, because they were not found in more than one article. This distribution is presented by Figure 6.

As expected, most of the selected studies apply the elicitation techniques in Software Engineering domain. A significant amount of articles uses techniques in Knowledge Engineering area, since, as it is known, many elicitation techniques were first used in this discipline.

Figure 7 shows the distribution of the application domain related with the constructs. In Software Engineering domain, the most considered construct is Adequacy followed by the Quality construct. Instead, the Quantity and Efficiency constructs are mainly used in Knowledge Engineering domain. In the area of Information Systems, with much less

Figure 5. Distribution of type of study with respect to the main constructs.

Figure 6. Distribution of articles according to the application domain.

Figure 7. Distribution of the application domain with respect to the main constructs.
number of studies; focus on measuring the Quality and Quantity of information elicited.

It is noteworthy also that six studies the elicitation techniques were focused in areas of marketing, design and manufacturing.

**Techniques considered**

Finally, Figure 8 shows a graph of word cloud with all of the elicitation techniques studied by researchers in the 43 primary articles. On the map, the most named techniques by researchers are highlighted in larger font size. A convergence of most articles for studying the best-known techniques such as the various forms of Interviews, Brainstorming, Questionnaire, Prototyping, JAD, amongst others, is appreciated.

This studied aspect also highlights the fact that there are more than fifty elicitation techniques that have been studied and applied in areas such as software engineering, knowledge engineering and information systems areas.

**CONCLUSIONS AND FUTURE WORK**

Systematic mapping performed in this work showed that, despite there is a significant amount of research related to the use of elicitation techniques, there is no common way to assess their performance yet. Moreover, a great variability among metrics and / or constructs proposed in the primary studies was found. Only few cases less significant, whose authors have worked together in prior projects, use similar definitions.

This evidence is relevant because it has implications on the results of empirical research. Mainly, it is relevant for the purpose of shaping a body of knowledge to generate guidelines for requirements engineers. In particular, in branch of research such as Evidence-based Software Engineering, where aggregated evidence. The greater the diversity of constructs is used the greater the generalization is required, which means a significant loss of prescriptive information.

For this reason, it is necessary to carry out more proposals to help converging towards a unique way to measure the performance of requirement elicitation techniques.

In this direction, our planned future work aims to contrast the views of practitioners and to raise a proposal on how to find a faithful representation of the performance of the techniques.

**ACKNOWLEDGMENTS**

The proofreading of this work was done by Marcia Poblete Ríos of the Language Department, Universidad de Atacama.

**REFERENCES**


