

Using GQM and TAM to evaluate StArt – a tool that supports Systematic Review

Elis Hernandez, Augusto Zamboni, Sandra Fabbri

Universidade Federal de São Carlos, Computing Department

São Carlos, Brazil, 13565-905

{elis_hernandes, sfabbri}@dc.ufscar.br, augusto_zamboni@comp.ufscar.br

and

André Di Thommazo

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP)

São Carlos, Brazil, 13565-905

andredt@ifsp.edu.br

Abstract

Background: Although Systematic Literature Review (SLR) is a reliable way of conducting literature review, its process is laborious and composed of repetitive activities. Hence, aiming to facilitate and support the conduction of such a process, the StArt tool was developed. *Objective:* As any new technology should be evaluated before its use, the objective of this paper is to present an overview of this tool and describe an evaluation that was carried out aiming at characterizing its usefulness and its ease of use. *Method:* The evaluation, applied twice, was designed through GQM paradigm and TAM model. The participants were graduate students who had a previous knowledge on SLR and have already applied the SLR process manually. *Results:* In both the evaluations the results were concentrated on the answers “extremely agree” or “quite agree” both for the usefulness and for the ease of use. *Conclusion:* Based on the results the further actions are: improvements related to the “quite agree” answers and the conduction of an experiment for evaluating the StArt in a deeper way. Despite these needed improvements, the results provide insights that StArt indeed helps the conduction of SLR and facilitates the application of its process.

Keywords: systematic literature review, SLR, evidence-based software engineering, tool, literature review.

1 Introduction

Evidence-based software engineering (EBSE) has received great attention nowadays. It focuses on the identification of the best research evidences, aiming at integrating them with practical experience and human value. In addition, EBSE focuses on the application of this knowledge in the decision making process regarding software development and maintenance [1] [2] [3].

According to Kitchenham et al. [1], the term evidence corresponds to the synthesis of the best studies on a research topic provided in primary studies of literature. A way to compose this synthesis is applying Systematic Literature Review (SLR), which is a type of secondary study, and makes use of a process that is reliable, rigorous and that allows auditing [4]. If on one hand this process provides advantages to SLRs in relation to the traditional literature review, on the other hand, it is laborious and error prone when applied only manually. Therefore, the support of a tool is essential to achieve the expected results of a SLR.

Based on this context, the objective of this paper is to present a tool that supports the SLR process and comment two viability studies carried out aiming to evaluate its perceived usefulness and ease of use. This tool is named StArt - State of the Art through Systematic Review and has been developed at Federal University of São Carlos (UFSCar), in the Software Engineering Research Laboratory (LaPES).

The paper is organized as follows: Section 2 presents a summary on Systematic Literature Review and the characteristics that differentiate it from traditional literature review. Section 3 presents the tool StArt by exploring its functionalities and the way it supports the SLR process, as well as how it facilitates some tasks that should be

executed during that process. Section 4 provides an overview of related tools found in literature, Section 5 presents some preliminary studies that were carried out to evaluate StArt and finally, Section 6 presents the final remarks and further work.

2 Systematic Literature Review

According [5], the Systematic Literature Review is supported by a well-defined process that makes it different from the traditional literature review. Some characteristics of the SLR are: starts by defining a Protocol which must contain a set of information used during the process execution, including the question being addressed; is based on a search strategy carefully defined for identifying as much of the relevant literature related to the research question; documents its search strategy such that it can be followed rigorously; requires that the inclusion and exclusion criteria used to evaluate each potential primary study are explicitly defined in the Protocol; requires the specification of the quality criteria that should be used to evaluate the content of each primary study; and must always be conducted when a quantitative meta-analysis is required.

Despite the advantages of a SLR, as good coverage, replicability and reliability, its process is more laborious than the one related to an informal literature research [5]. Thus, considering that there are several stages to be executed and several documents to be managed, computational support can facilitate the work and enable higher quality in the execution process. Although there are slight differences among the SLR processes commented in literature, they all involve planning, execution, analysis and dissemination of results [5] [6].

In the *Planning* stage, the aim is to define a Protocol which contains all the information and the necessary procedures for the execution of the following stages. Examples of the information and the procedures needed are: the research question, the keywords, the search engines and the studies inclusion and exclusion criteria.

In the *Execution* stage, three steps must be conducted: the Studies Identification on the search engines defined in the Protocol, the Selection of these studies, based on the inclusion and the exclusion criteria, and the Extraction of data from the selected studies.

In the *Summarization* stage, the data extracted from the studies are analysed and summarized aiming at answering the research question defined in the Protocol.

After the conclusion of these three stages, it is important to report the results through technical reports or scientific papers to show the state of the art of the topic in focus.

3 The Tool StArt

Some activities in the SLR process are repetitive and require discipline and systematic practice by the researcher. The information must be registered in an organized way so that the SLR provides the expected results, is replicable, and allows that all the information can be packed. StArt provides support to the SLR process activities, except to the automated search of primary studies in electronic databases, since this is considered as a robot action, which is blocked by these mechanisms. Therefore, the researcher must do the search manually through the search engines registered in the *Protocol* every time a search is necessary. The search result must be exported from the search engine as a BibTex file which must be imported into StArt.

Figure 1 presents a screen of the tool. In the left side a hierarchical tree shows the process stages to be followed. Some pieces of information in this tree are filled out dynamically, as the researcher defines the Protocol or as the process steps are carried out. This resource of StArt helps the researcher in keeping the information updated and consistent.

The following subsections describe the way each SLR stage is supported by the tool.

3.1 Planning

In this stage, the researcher must define the Protocol that will support the other SLR stages. The *Protocol* fields available in the tool are the ones suggested by [5]. StArt has a help icon that provides the description and example of each field. As some fields have influence over other process stages, the tool assists in keeping these relations controlled. For instance:

- *Source List*: this field contains the list of all the search engines which will be used to gather the studies. Meanwhile they are inserted in the Protocol, the name of the search engines are automatically added in the side-tree of the main screen (Figure 1). The separation of the search engines allows a better organization of the studies as well as the information control in the studies identification step;

- *Keywords*: this field contains the keywords that will be used to compose the search strings. When the studies are uploaded into the StArt, it uses the keywords to score the studies according to the number of occurrences of these words in their title, abstract and keywords. This score, showed in Figure 2, suggests the studies relevance order;

- *Studies Inclusion and Exclusion Criteria Definition*: this field, as showed in Figure 1, contains the criteria that will be used to accept or reject each study during the selection step. StArt makes them available in this step and allows the researcher to register the ones were applied to each of the studies, as showed in Figure 2;

- *Information Extraction Form Attributes*: this field contains the attributes that will compose the form which must be filled in by the researcher in the Extraction stage, as explained in subsection 3.2.3.

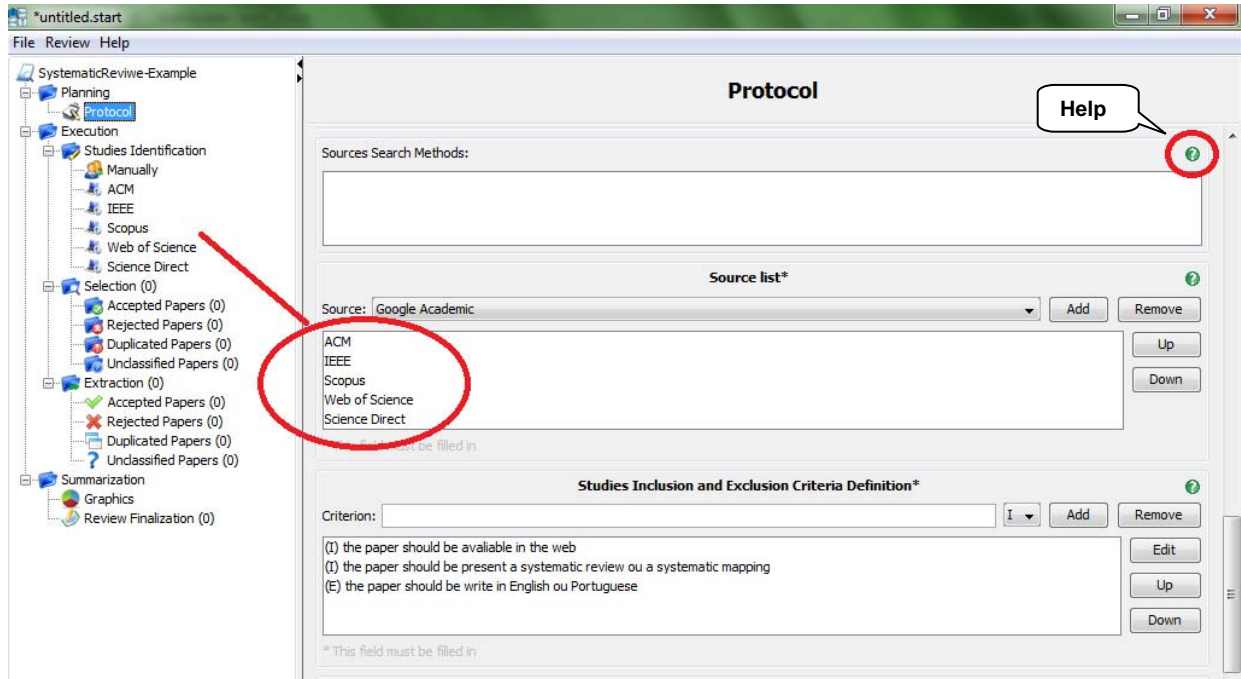


Figure 1. Part of the *Protocol* highlighting the source list that is added dynamically to the side-tree

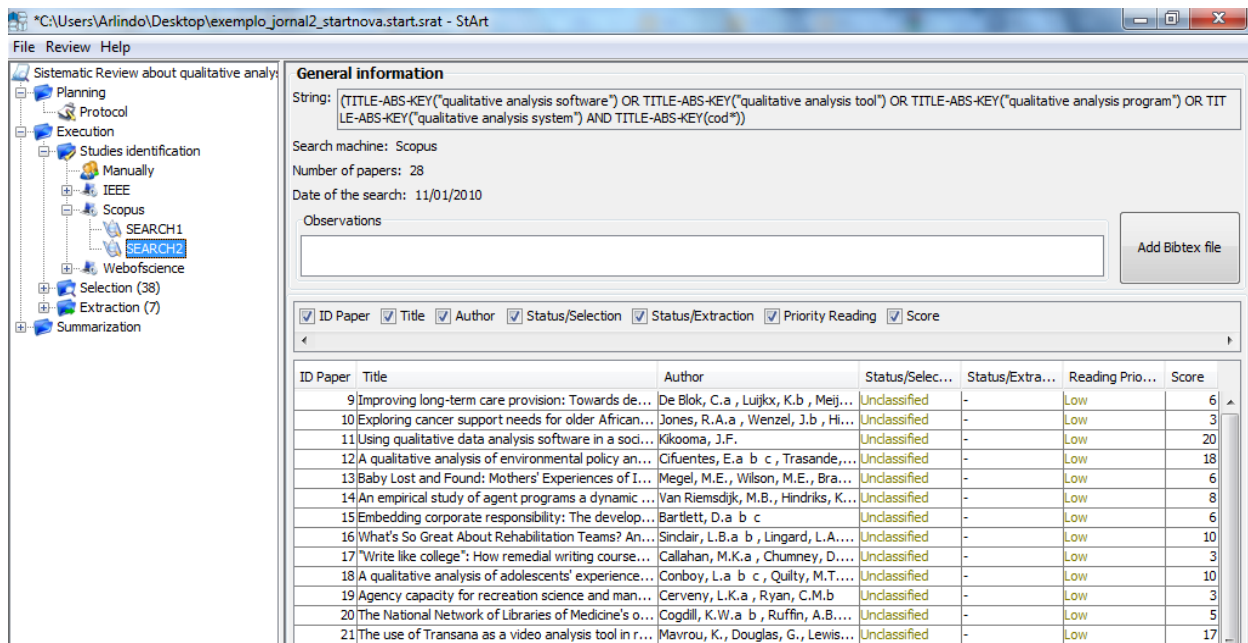


Figure 2. Information available in the StArt when the studies are uploaded into the tool

3.2 Execution

Once the Protocol is concluded, the researcher is able to perform the *Execution* stage that is composed of three steps: Studies Identification, Selection and Extraction.

3.2.1 Studies Identification

In this step the objective is to gather a set of studies that is related to the research question. Thus, the researcher should: (i) apply the search strings to each of the search engines specified in the Protocol and export the results in a BibTex format, and (ii) import into StArt the BibTex and store the search string used by the search engine, since the search string that allows a faithful replication of the SLR. The tool also allows manually insertion of studies.

Once the BibTex file was imported, all the information presented in Figure 2 is available in the StArt. This screen shows the search string used in this session, the number of studies identified, and a table with some attributes of each study, as its identification, title, author(s), status at the Selection step, status at the Extraction step, *Reading Priority* and *Score*. The score, as mentioned before, is automatically calculated according to the number of times the keywords defined in the *Protocol*. Duplicated studies are also automatically identified by the tool. The two fields *Status* must be filled by the researcher, according to the process step.

3.2.2 Selection

In this step the primary studies uploaded into the Start must be accepted or rejected according to the inclusion and exclusion criteria defined in the *Protocol*. Figure 3 illustrates the facility provided by the tool for doing this activity. The decision should be made after reading the title, abstract and keywords of the study, which are available for each study, as shown in Figure 4. At the end of this step all the accepted studies are automatically transferred to the *Extraction* step. Figure 5 exemplifies this fact: see that there are seven papers as *Accepted Papers* in the *Selection* step and a total of seven papers in the *Extraction* step.

3.2.3 Extraction

In this step, all the studies that have been accepted in the *Selection* step should be read in full and be analyzed again a study is rejected if they are not relevant to answer the main question defined in the *Protocol* it must be rejected in this step. The *Reading Priority* field that can be filled during the Selection step may help the researcher with the reading order. Although the full studies must be downloaded by the researcher, they can be linked to the SLR, which facilitates the access to the document. For the papers classified as *Accepted* in this step, the researcher must extract the information correspondent to the attributes of the *Information Extraction Form*, defined in the *Protocol*. This form is available in this step as shown in Figure 5. This facility promotes a systematic way for extracting information.

3.3 Summarization

In this stage the researcher should describe the state of the art of the topic in focus. StArt facilitates the access to the information extracted during the *Extraction* step and provides a text editor to help in a first version of the summarization document when this stage is reached some data on the whole SLR are available, as shown in Figure 6. In addition, Start provides some reports that also facilitate the conduction of a SLR.

Figure 3. Application of the Inclusion and exclusion criteria

Paper

Study Data Selection Data Information Extraction Form

Author: Cogdill, K.W.a b , Ruffin, A.B.c , Stavri, P.Z.d

Title: The National Network of Libraries of Medicine's outreach to the public health workforce: 2001-2006

Keywords: article; computer program; health care access; health care personnel; health promotion; information processing; interview; medical information; medical staff; methodology; outcome assessment; qualitative analysis; training, Access to Information; Community-Institutional Relations; Cooperative Behavior; Humans; Libraries, Medical; Library Associations; Models, Organizational;

Journal: Journal of the Medical Library Association

Font: Tahoma Size: 11

Abstract: Objective: The paper provides an overview of the National Network of Libraries of Medicine's (NN/LM's) outreach to the public health workforce from 2001 to 2006. Description: NN/LM conducts outreach through the activities of the Regional Medical Library (RML) staff and RML-sponsored projects led by NN/LM members. Between 2001 and 2006, RML staff provided training on information resources and information management for public health personnel at national, state, and local levels. The RMLs also contributed significantly to the Partners in Information Access for the Public Health Workforce collaboration. Methods: Data were extracted from telephone interviews with directors of thirty-seven NN/LM-sponsored outreach projects directed at the public health sector. A review of project reports informed the interviews, which were transcribed and subsequently coded for emergent themes using qualitative analysis software. Results: Analysis of interview data led to the identification of four major themes: training, collaboration, evaluation of outcomes, and challenges. Sixteen subthemes represented specific lessons learned from NN/LM members' outreach to the public health sector. Conclusions: NN/LM conducted

Year: 2007 Type: Article

Comment:

Status: Unclassified Search session: SEARCH2

Reading Priority: Low Score: 5

Save Cancel

Figure 4. General data of each study

Systematic Review about qualitative analysis ar

Accepted Papers (Extraction)

ID	SS	ID Paper	Title	Author	Reading Priority	Score
0		1	CD4: a system for understanding the dynamic properties of data processing prog...	Howden, W.E. and Guangming Shi	Very high	0
0		2	An FPGA Based Quench Detection and Protection System for Superconducting Ac...	Caricagno, R.H. and Feher, S. and Lamm, M. and Makulski, ...	Very high	3
0		3	User perception of seat functionality	Salindez-Novoa, C.V. and Goyal-Merida, D.A. and Perez-Za...	High	0
0		4	Support for quality-based design and inspection	Tervonen, I.	High	0
0		6	The ICT-Supported Unified Process Model of Offshore Outsourcing of Software P...	Yalaho, A. and Nahar, N.	Low	0

Paper

Study Data Selection Data Information Extraction Form

2007

Article Date

NVivo 7

Coding Tools

Computer Science

Research area

Proprietary Software

Tool type

Management of Engineering and Technology, Portland International Center for

Published

Semi-Structured Interview.

Status: Accepted Search session: SEARCH0 This paper is in Summarization step

Reading Priority: Low Score: 0

Full text

Save Cancel

Path: 'do\Desktop\jornal\exemplo\arquivos pdf\The ICT-Supported Unified Process Model of Offshore Outsourcing of Software Production Exploratory Examination and Validation.pdf

Name: The ICT-Supported Unified Process Model of Offshore Outsourcing of Software Production Exploratory Examination and Validation.pdf

Size: 10964 KB

Systematic Review opened successfully

Figure 5. Information Extraction Form

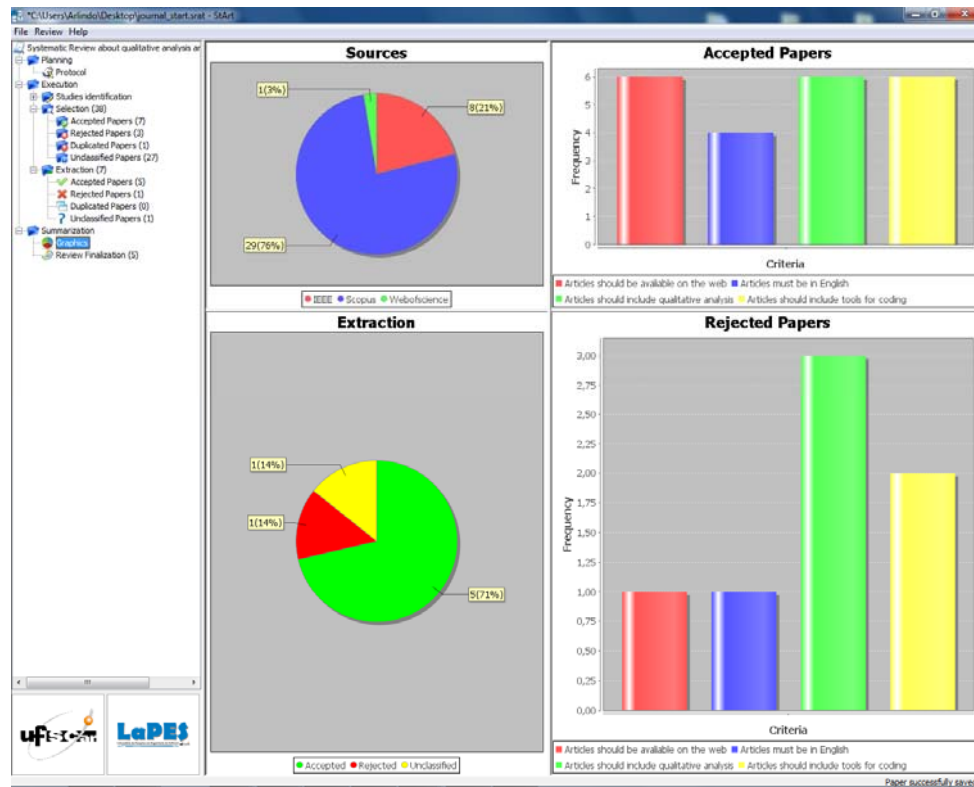


Figure 6. Some final data provided at *Summarization* stage

4 Related tools

In the literature, there are some tools to support the management of bibliographic references, which are commonly used by researchers to aid in the SLR process. The purpose and the coverage of these tools are different and they are not related to the SLR process proposed by [5], except for SLR Tool [7].

Only SLR Tool [7] focuses on Systematic Literature Review. However, its installation requires the availability of a specific database management system and a pre-configuration of the environment, which can restrict its use, mainly by researchers of other research areas such as Medicine and Nursing, who are also users of the SLR process. Another characteristic of the SLR Tool is that it only works with the English and the Spanish versions of the Windows operating system. On the other hand, StArt does not have this restriction and can be easily installed through a wizard interface. Table 1 presents the main characteristics of tools that are being used in the context of literature review.

Table 1. Characterization of related tools

Tool name	Free	Protocol definition	Management references	Export references	Customization of attributes	Automatic classification of papers
JabRef	Y	No	Yes	Yes	Yes	No
EndNote	No	No	Yes	Yes	Yes	No
ProCite	No	No	Yes	No	No	No
Reference Manager	No	No	Yes	No	No	No
RefWorks	No	No	Yes	Yes	No	No
BibEdt	Yes	No	Yes	No	No	No
Zotero	Yes	No	Yes	Yes	Yes	No
Biblioscape	No	No	Yes	No	No	No
Bookends	Yes	No	Yes	Yes	Yes	No
Library Master	No	No	Yes	Yes	Yes	No
Mendeley	No	No	Yes	Yes	Yes	No
Mekentosj	No	No	Yes	Yes	No	No
SLR Tool	Yes	Yes	Yes	Yes	No	No
Review Manager	Y	Y	Yes	No	Yes	No
StArt	Y	Y	Yes	No	Yes	Yes

5 StArt evaluation: preliminary data on the Usefulness and Ease of Use

According to [8], all proposed technology (method, technique, tool, etc.) should be evaluated before being made available for use. The objective of the evaluation described below was to characterize the two aspects of the TAM model (Technology Acceptance Model) [9], to get preliminary data on the tool viability of use.

The evaluation was applied twice. In both occurrences the participants were graduate students in Computer Science (MSc. and PhD.) who had applied the SLR process, manually, during the Research Methodology course. Fourteen students participated of the first evaluation and thirty five participated of the second one.

The evaluation was planned through the GQM (Goal, Question, Metric) paradigm [10][11], which is composed of four steps: Planning; Definition; Data Collection; and Interpretation that are described below.

5.1. Planning and Definition

The GQM model constructed for planning the evaluation consists of four goals, thirteen questions (Table 2) and fourteen metrics (Table 3), according to details presented in Figure 7. Based on that model, two questionnaires were used in the evaluation: Questionnaire1 (Q1 to Q4) for collecting data on student's opinion and their current contact with systematic review; and Questionnaire2 (Q5 to Q13) for characterizing the usefulness and ease of use of the tool, according to TAM. The questions related to TAM were inspired on the study presented in [12] and were evaluated according to the Likert scale [13]. Both questionnaires contained blank fields for comments. Table 6 presents the interpretation model of the GQM, which should be read as follows: "If Expression then Interpretation". Taking line 9 as an example where the question is Q4, "If $M9 + M10 + M11 \geq M12 + M13 + M14$ then the SR is seen as a key resource for the quality of academic research."

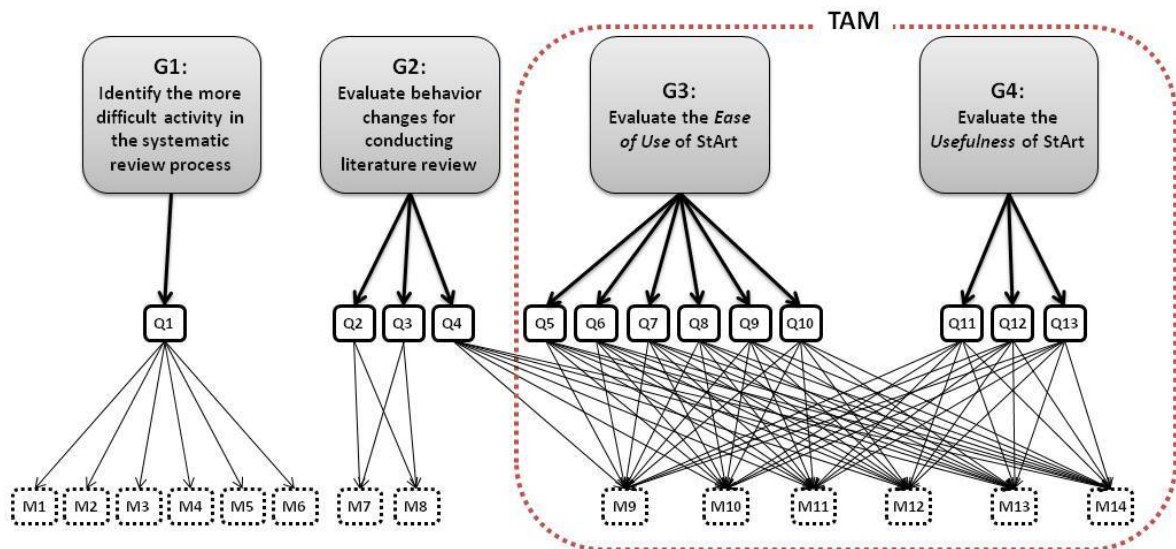


Figure 7. GQM to evaluate the StArt

Table 2. Questions used in the GQM

Question	Description	Question	Description
Q1	Which activity required in the process of systematic review do you judge as the most difficult?	Q9	It was easy to remember how to execute a systematic review through the use of StArt
Q2	After experiencing the process of systematic review, have you changed your way of doing a literature review?	Q10	I think StArt is easy to use
Q3	Have you tried to reapply the process of systematic review after the course was conducted in the second semester of 2009?	Q11	StArt has allowed me to perform a systematic review in accordance with the process defined in the literature
Q4	What is your level of agreement with the statement: "The systematic review is critical to the quality of academic research"	Q12	Considering what I have learnt about RS, using StArt facilitated the execution of the activities that compose the systematic review process
Q5	It was easy to learn how to use StArt?	Q13	I consider StArt useful to execute my systematic review(s)
Q6	I could use StArt the way I wanted.		
Q7	I could understand what was happening during my interaction with StArt		
Q8	It was easy for me to become skilful at using StArt		

Table 3. Metrics used in the GQM

Metric	Description	Metric	Description
M1	Number of people who chose "designing the protocol"	M8	Number of people who chose "NO"
M2	Number of people who chose "creation of search strings"	M9	Number of people who chose "extremely agree"
M3	Number of people who chose "search for articles on search engines"	M10	Number of people who chose "quite agree"
M4	Number of people who chose "select items that will be read in full"	M11	Number of people who chose "slightly agree"
M5	Number of people who chose "extract information from articles read in full"	M12	Number of people who chose "extremely disagree"
M6	Number of people who chose "summarize the systematic review's results"	M13	Number of people who chose "quite disagree"
M7	Number of people who chose "YES"	M14	Number of people who chose "slightly disagree"

Table 4. Model interpretation of the GQM

N°	Expression	Interpretation (next actions of the team)
1	$M1 > M_i, i=2, 3, 4, 5 \text{ and } 6$	The activity of the greatest difficulty is designing the protocol
2	$M2 > M_i, i=1, 3, 4, 5 \text{ and } 6$	The activity of the greatest difficulty is the creation of the search string
3	$M3 > M_i, i=1, 2, 4, 5 \text{ and } 6$	The most difficult activity is to search for articles on search engines
4	$M4 > M_i, i=1, 2, 3, 5 \text{ and } 6$	The most difficult activity is to select items that will be read in full
5	$M5 > M_i, i=1, 2, 3, 4 \text{ and } 6$	The most difficult activity is to extract information from articles read in full
6	$M6 > M_i, i=1, 2, 3, 4 \text{ and } 5$	The most difficult activity is to summarise the results of the systematic review
7	For Q2: $M7 \geq M8$	Changes have occurred in how to conduct literature review after learning the process of systematic review
8	For Q3: $M7 \geq M8$	Most of the students conducted another SLR after the course
9	For Q4: $M9 + M10 + M11 \geq M12 + M13 + M14$	SR is seen as a key resource for the quality of an academic research
10	For $Q_i, i=5 \text{ to } 10$: $M9 \geq M10 + M11$	The StArt is easy to use, and the next step is to conduct an experimental study with a new group of participants to test the result
11	For $Q_i, i=5 \text{ to } 10$: $M9 \leq M10 + M11$ and $M10 \geq M11$	The StArt is easy to use; however, the next step is to analyse the data sent by participants through comments in order to identify improvements needed to facilitate the use of the tool. Once improvements are made, it could be carried out with a new group of participants
12	For $Q_i, i=5 \text{ to } 10$: $M9 \leq M10 + M11$ and $M11 \geq M10$	The StArt does not have the 'ease of use' and so the next step is to study heuristics and usability standards defined in the literature to perform a self-assessment on the tool's interface; the comments submitted by participants should be analysed and, based on that, the StArt project should be reviewed and improvements implemented. The assessment should be conducted again with the same group of participants to see if there was any improvement in the results
13	For $Q_i, i=11 \text{ to } 13$: $M9 \geq M10 + M11$	the StArt is useful to perform a systematic review, and the next step is to conduct an experimental study to confirm the result
14	For $Q_i, i=11 \text{ to } 13$: $M9 \leq M10 + M11$ and $M10 \geq M11$	The StArt is useful, but analyse of the data sent by participants through comments must be done in order to identify improvements needed to make the tool useful. Once these improvements made, the same assessment could be carried out with a new group of participants
15	For $Q_i, i=11 \text{ to } 13$: $M9 \leq M10 + M11$ and $M11 \geq M10$	The StArt is not very useful, and therefore the team's next step is to review the design of the tool in order to identify whether there are features that can be implemented to provide greater utility to the StArt. The assessment should be conducted again with the same group of participants to see if there was any improvement in the results
16	For $Q_i, i=5 \text{ to } 10$: $M14 + M13 + M12 \geq M9 + M10 + M11$	The implementation of new features in the StArt must be aborted, the comments submitted by participants should be analysed and the design of the tool must be fully reviewed by the team, especially the user interface. Once the tool is modified, the same study should be conducted until the participants judge the tool easy to use
17	For $Q_i, i=11 \text{ to } 13$: $M14 + M13 + M12 \geq M9 + M10 + M11$	The implementation of new features in the StArt must be aborted, the comments submitted by participants should be analysed and the design of the tool must be fully reviewed by the team. Once the tool is modified, the same study should be conducted until the participants judge the tool useful

5.2. Data Collection

The data were collected as follows: firstly, Questionnaire1 was sent by e-mail to the participants who, after answered it, had access to two training videos about StArt and had permission to download the tool. Hence, the students were asked to explore the tool as they have done manually during the course.

Secondly, the students who have finished Questionnaire1, received Questionnaire2 that was, sent by e-mail. By using electronic questionnaires, at the end of the evaluation process all the answers were available on spreadsheets, which facilitated the data collection.

The summary of the data collected is showed in the next tables and figures. Tables 5, 6 and 7 present the questions and answers of the first questionnaire and Figures 8, 9, 10 and 11 present charts that show the questions and the answers of the second one. In the charts, the order of the bars obeys the order of the questions.

Table 5. Data collected in questionnaire 1 (question 1)

Q1) Which activity required in the process of systematic review do you judge as the hardest?	Protocol filling M1	Creation of search strings M2	Search for articles on search engines M3	Selection of articles that will be fully read M4	Extract information / data from items read completely M5	Summarize the systematic review's results M6
1 st Evaluation	6	5	1	0	2	0
2 nd Evaluation	9	15	2	4	2	3

Table 6. Data collected in questionnaire 1 (questions 2 and 3)

Q2) After knowing the process of systematic review, have you changed the way you perform literature review?	Yes M7	No M8
1 st Evaluation	13	1
2 nd Evaluation	32	3
Q3) Have you tried to reapply the process of systematic review after the course held in the second half of 2009?		
1 st Evaluation	6	8
2 nd Evaluation	12	23

Table 7. Data collected in questionnaire 1 (question 4)

Q4) What is your level of agreement with the statement: "The systematic review is critical to the quality of academic research"	Extremely agree M9	Quite agree M10	Slightly agree M11	Slightly disagree M12	Quite disagree M13	Extremely disagree M14
1 st Evaluation	7	3	4	0	0	0
2 nd Evaluation	21	6	8	0	0	0

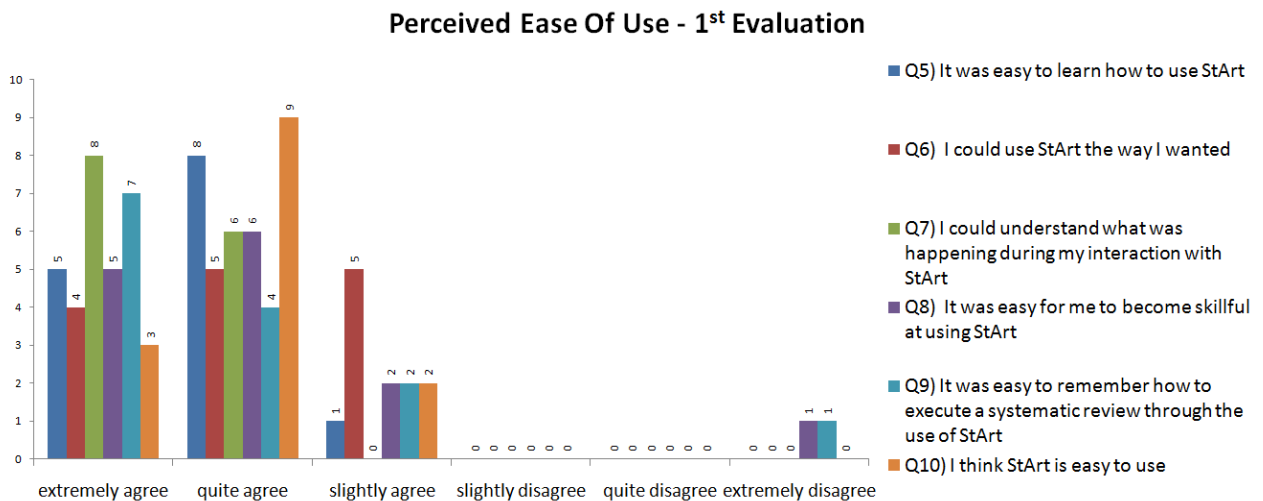


Figure 8. Questions and answers related to the ease of use – 1st Evaluation

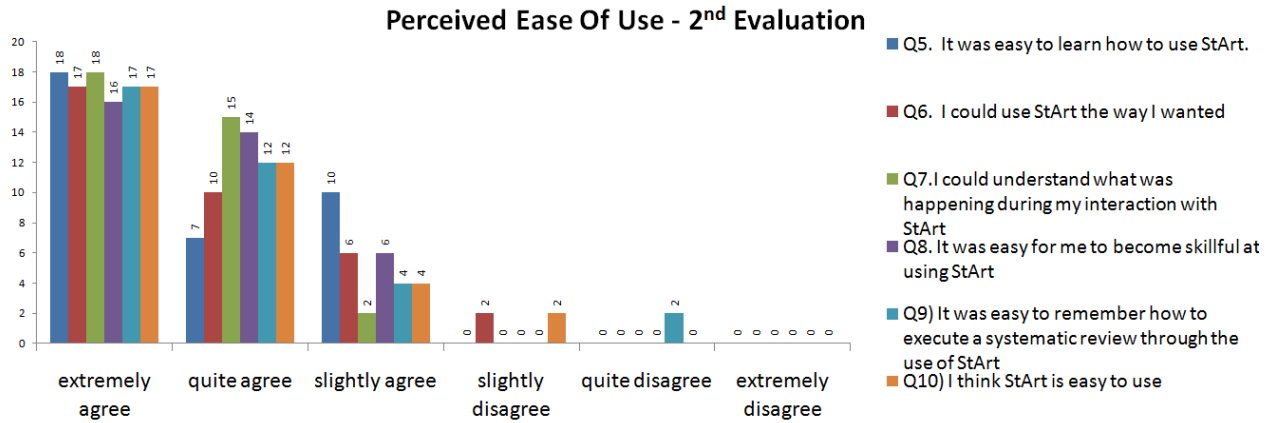


Figure 9. Questions and answers related to the ease of use – 2nd Evaluation

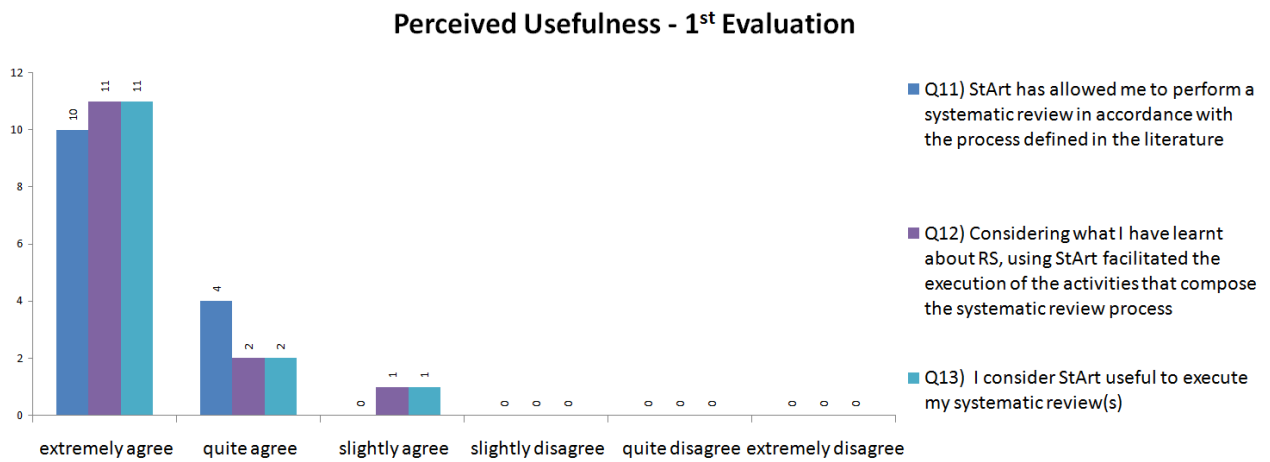


Figure 10. Questions and answers related to the usefulness – 1st Evaluation

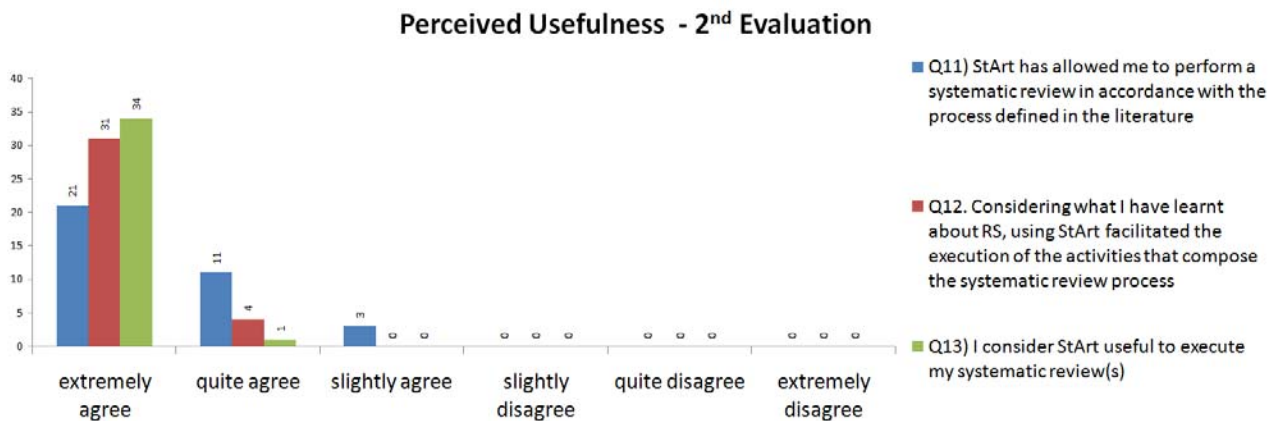


Figure 11. Questions and answers related to the usefulness – 2nd Evaluation

5.3 Interpretation

Applying the interpretation model presented in Table 6 on the data collected and presented in the previous section, we can assume the following about the goals:

- G1: in relation to this goal, the results showed that in the first evaluation the protocol filling was selected as the most difficult activity of the SLR process, since six participants (42%) have selected this option (Table 1) and this value makes the expression 1 true (Table 6). However, in the second evaluation the construction of search strings was selected as the most difficult activity of the SLR process, since fifteen participants (42%) have selected this option (Table 1) and this value makes the expression 2 true (Table 6). From now, we are planning

to address this issue in a deeper way, aiming to provide facilities that can help the researcher in doing these activities.

- G2: in relation to this goal, the results showed that the participants change their behaviour for conducting literature review. Although most of the participants have not tried to apply the SLR process after the course, they consider the SLR a key for the quality of academic research. This conclusion is supported by the following expressions (Table 6):
 - expression 7: Q2: $M7 \geq M8$, that is true for the evaluation 1 ($13 \geq 1$) and for the evaluation 2 ($32 \geq 3$), according to the values of the Table 2;
 - expression 8: Q3: $M7 \geq M8$, that is false for the evaluations 1 ($6 \geq 8$) and for the evaluation 2 ($12 \geq 23$), according to the values of the Table 2;
 - expression 9: Q4: $M9 + M10 + M11 \geq M12 + M13 + M14$, that is true for the evaluation 1 ($7+3+4 \geq 0+0+0$) and for the evaluation 2 ($21+6+8 \geq 0+0+0$), according to the values of the Table 3.
- G3: in relation to this goal, there are four expressions in the interpretation model (Table 6): 10, 11, 12 and 16. The results showed that in both the evaluations, most of the participants agree with the ease of use of the StArt. In the evaluation 1 the answers of the participants were concentrated in “quite agree” (Figure 8). Considering the answers for questions 5 up to 10, the expression 11: $Q_i, i=5$ to 10: $M9 \leq M10 + M11$ and $M10 \geq M11$ is true, since $32 \leq 34 + 12$ and $34 \geq 12$. Hence, the next step should be the analysis of the comments submitted by the participants in order to identify improvements needed to facilitate the use of the tool. Once the improvements are made, the evaluation could be carried out with a new group of participants. In the evaluation 2 the answers of the participants were concentrated in “extremely agree” (Figure 9). Considering the answers for questions 5 up to 10, the expression 10: $Q_i, i = 5$ to 10: $M9 \geq M10 + M11$ is true, since $103 \geq 70 + 32$. Therefore, the development team should conduct an experimental study with a new group of participants to test the result of the evaluation 2.
- G4: in relation to this goal, there are four expressions in the interpretation model (Table 6): 13, 14, 15 and 17. The results showed that in both the evaluations, the majority of the participants agree with the usefulness of the StArt (Figures 10 and 11). Considering the answers for questions 11 up to 13, the expression 13: $Q_i, i=11$ to 13: $M9 \geq M10 + M11$ is true, since for the evaluation 1 ($50 \geq 17 + 2$) and for the evaluation 2 ($86 \geq 16 + 3$). According to the interpretation model (Table 6) the next step is to conduct an experimental study to confirm this result.

6 Final remarks and future work

This paper presented the StArt tool that supports the conduction of the systematic literature review process [14] providing facilities for minimizing this laborious process. It has been developed in an iterative and interactive way, with constant feedback from users. For directing the next steps, an evaluation was carried out twice, aiming to explore the support of the tool for conducting all the stages of the SLR process. This evaluation involved students who had already applied the SLR process manually.

The evaluation was planned using the GQM model and established four goals: two of them, related to the aspects addressed by the Technology Acceptance Model (TAM) – ease of use and usefulness; one related to the identification of the activity considered the most difficult among the SLR activities; and another one related to the investigation on user’s behaviour change in conducting literature review.

The use of the TAM has turned the evaluation quick and objective, and disseminated the model among the participants. The use of the GQM led objectivity to the evaluation and to the definition and elaboration of forms for data collection. One of the limitations of the evaluation is the number of participants, which does not allow generalizing the results.

As our main objective was to explore the TAM aspects, in relation to these issues, the evaluation indicated that the StArt is useful, since 71.42% of the participants of the evaluation 1 and 81.90% of the participants of the evaluation 2 extremely agreed with the usefulness of the tool. For the ease of use, in the evaluation 1 the answers were concentrated on quite agree (45.23%) and extremely agree (38.09%) and in the evaluation 2 the answers were concentrated on extremely agree (49.04%) and quite agree (33.33%).

According to the GQM interpretation model, the actions that should be taken are: conducting an experimental study to confirm the results of this evaluation and analyse the qualitative data sent by the participants in order to identify improvements needed to facilitate the use of the tool. In summary, the evaluation has provided evidence that the tool will be accepted by users to support the conduction of SLRs.

In addition to the actions defined in the GQM interpretation model, arose directly from the evaluation, some functionalities are already being implemented such as systematic mapping process support [15] and mechanisms of communication with other tools.

Acknowledgements

The authors thank the students who participated in the evaluation and CNPq, CAPES and Observatório da Educação Project for financial support.

References

- [1] B. A. Kitchenham, T. Dyba, M. Jørgensen, "Evidence-based software engineering", in Proc. *International Conference on Software Engineering (ICSE'04)*, Edinburgh, Scotland, May. 2004, pp. 273-281.
- [2] T. Dyba, B. A. Kitchenham, M. Jørgensen, "Evidence-based software engineering for practitioners", *IEEE Software*, vol. 22, pp. 58-65, Jan. Feb. 2005.
- [3] M. Jørgensen, T. Dyba, B. Kitchenham, "Teaching evidence-based software engineering to university students", in Proc: *IEEE International Software Metrics Symposium (METRICS 2005)*, Como, Italy, Sep. 2005, pp. 24.
- [4] B. Kitchenham, P. Brereton, D. Budgen, M. Turner, J. Bailey, S. G. Linkman, "Systematic literature reviews in software engineering - a systematic literature review": *Information & Software Technology*, vol. 51, pp. 7-15, Nov. 2009.
- [5] B. A. Kitchenham, "Procedures for Performing Systematic Reviews", Software Engineering Group , Keele University, Keele, UK, Tech. Rep. TR/SE 0401, Jul. 2004.
- [6] J. Biolchini, P. G. Mian, A. C. C. Natali, G. H. Travassos, "Systematic Review in Software Engineering", UFRJ, Rio de Janeiro, Brazil, Tech. Rep. RT-ES (679/05), May. 2005.
- [7] A. M. Fernández-Sáez, M. G. Bocco, F.P. Romero, "SLR-tool - a tool for performing systematic literature reviews", in Proc: *International Conference on Software and Data Technologies (ICSOFT' 10)*, Athens, Greece, Jul. 2010, pp. 144.
- [8] V. R. Basili, S. Green, O. Laitenberger, F. Lanubile, F. Shull, S. Sorumgard, M. Zelkowitz, "Packaging researcher experience to assist replication of experiments", in Proc. *International Software Engineering Research Network Meeting (ISERN'96)*, Sydney, Australia, Aug. 1996, pp. 3-6.
- [9] F. D. Davis, "User acceptance of information technology: system characteristics, user perceptions and behavioral impacts", *International Journal of Man-Machine Studies*, vol 38, pp. 475-487. Jan. 1993.
- [10] V. R. Basili, C. Caldiera, H.D. Rombach, "Goal Question Metric Paradigm" in *Encyclopedia of Software Engineering*, New York: John Wiley & Sons, 1994, pp. 528-532.
- [11] R. Soligen, E. Berghout, *The Goal/Question/Metric Method: a practical guide for quality improvement of software development*. London, UK: McGraw W-Hill Companies, 1999.
- [12] O. Laitenberger, H. M. Dreyer, "Evaluating the Usefulness and the Ease of Use of a Web-based Inspection Data Collection Tool", in Proc.: *International Symposium on Software Metrics (METRICS'98)*, Bethesda, USA, March, 1998, pp. 122-135.
- [13] J. P. McIver, E. G. Carmines, "Unidimensional Scaling". London, UK: Sage Publications, 1981.
- [14] A. B. Zamboni, A. Di Thommazo, E. C. M. Hernandez, S. C. P. F. Fabbri, "StArt Uma Ferramenta Computacional de Apoio à Revisão Sistemática", in Proc.: *Congresso Brasileiro de Software (CBSOFT'10)*, Salvador, Brazil, Sep. 2010.
- [15] K. Petersen, R. Feldt, S. Mujtaba, M. Mattsson, "Systematic Mapping Studies in Software Engineering", in Proc.: *Inter. Conf. on Evaluation and Assessment in Software Engineering (EASE'08)*, Bari, Italy, Jun. 2008, pp.26-27

