A Method to support search string building in Systematic Literature Reviews through Visual Text Mining

Germano Duarte Mergel
PUCRS – Pontifícia Universidade Católica do Rio Grande do Sul
Av. Ipiranga, 6681, Prédio 32
Porto Alegre, RS, Brazil
gmergel@gmail.com

Milene Selbach Silveira
PUCRS – Pontifícia Universidade Católica do Rio Grande do Sul
Av. Ipiranga, 6681, Prédio 32
Porto Alegre, RS, Brazil
milene.silveira@pucrs.br

Tiago Silva da Silva
UNIFESP – Universidade Federal de São Paulo
Av. Cesare Mansueto G. Lattes, 1201
São José dos Campos, SP, Brazil
silvadasilva@unifesp.br

ABSTRACT
Despite the increasing popularity of systematic literature reviews in Software Engineering, several researchers still indicate it as a costly and challenging process. Aiming at alleviating this costly process, we propose an iterative method to support the process of building the search string for a systematic review. This method uses Visual Text Mining techniques to support the researcher by suggesting new terms for the string. In order to do so, the method extracts relevant terms from studies selected by the researcher and displays them in a way that facilitate their visualization and supports building and refining the search string. In order to check the feasibility of this approach, we developed a tool that implements the proposed method. Interviews with researchers identified their difficulties in performing systematic reviews and captured their feedback with regards the use of the proposed method in a user study. The researchers indicated that this approach could be used to improve the process of building the search strings for systematic reviews. The study indicates that our approach can be used to facilitate the construction of the systematic literature review search string.

Categories and Subject Descriptors
D.2. [**Software Engineering**]

Keywords
Systematic literature review, Visual Text Mining, Information Visualization.

1. INTRODUCTION
The volume of empirical research in software engineering is constantly expanding and reviews have become essential tools for any researcher wanting to keep him updated [4]. However, it is necessary to adopt a systematic approach to assess and aggregate the results of the review in order to provide an objective and balanced synthesis of the research evidence [9]. As a solution, the systematic literature review (SLR) evaluates and interprets available research relevant to a particular research question, topic area, or phenomenon of interest. SLR plays an important role, synthesizing existing research using an unbiased approach [9], using a predefined search strategy and identifying relevant researches. Moreover, the strategy adopted in a review must be clear enough to allow its repetition by other researchers.

Nevertheless, SLRs require considerably more effort from researchers than traditional reviews [10]. According to studies on the subject [2] [4] [16], amongst the major difficulties of conducting this type of review are building the search string and selecting resulting studies. Moreover, these are common problems in performing systematic reviews faced by both beginners and experienced researchers, and a poorly built search string can result in returning too many non-relevant studies, or even excluding relevant ones [16].

In order to assist the researcher in the SLR conduction, this paper presents a method that supports the process of building the search string. By applying Visual Text Mining techniques, it extracts and recommends relevant terms based on the studies returned from using the current search string. The researcher then can refine the query and re-apply it iteratively until the search string can be considered complete.

Researchers participated on a preliminary study that involved using a tool that implements the proposed method, providing material to analyze its applicability. At the end of the study, participants indicated that this method could help researchers conducting SLRs by recommending relevant terms.

The remainder of this paper is organized as follows: Section 2 provides background information on systematic review and Visual Text Mining; Section 3 presents the method proposal; Section 4 details the analysis performed on the proposed method; Section 5 discusses the results of the performed analysis; Section 6 draws conclusions and future directions to this research.

2. BACKGROUND
This section presents the necessary basis for understanding the techniques applied in the proposed approach as well as research related to this topic.

2.1 Systematic Literature Review
Defined as a form of identification, evaluation and interpretation of relevant papers to a research question [9], the systematic literature review provides mechanisms to identify and aggregate research evidence in Software Engineering. It is a review that

SAC'15, April 13 - 17 2015, Salamanca, Spain

http://dx.doi.org/10.1145/2695664.2695902

supports a full and fair assessment of the evidence of a topic of interest.

The SLR can be seen as a specific research methodology, following a strict set of methodological steps in sequence and in accordance to a predefined protocol [2]. Such protocol contains central review question, the focus of the research, knowledge bases and filtering criteria used in the selection of papers, among other definitions.

Biolchini et al [2] define the process of systematic literature review comprehending three phases: planning, implementation and analysis of results (Figure 1). The planning phase encompasses listing the research objectives and defining a review protocol. Such protocol specifies the central research question and the methods that will be used to execute the review. During the implementation of the review, primary studies are identified, selected and evaluated according to pre-defined inclusion and exclusion criteria. Once the primary studies are selected, all information obtained is analyzed and synthetized, as part of the last phase. In the process, the authors propose iterations between the phases of a systematic review, verifying and addressing potential problems. *"Before performing a systematic review, it is necessary to ensure that the protocol is feasible"* [2].

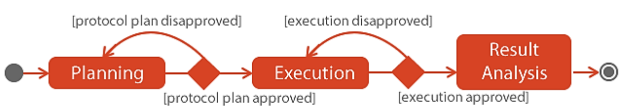

Figure 1. SLR process as defined by Biolchini et al [2]

2.2 Visual Text Mining

The Visual Text Mining (VTM) field is an intersection between the fields of Text Mining and Information Visualization [15]. Applications in Visual Text Mining go beyond visualizing the outcome of mining documents, using techniques of Information Visualization in the process of Text Mining to obtain a visual representation and support data exploration. VTM combines text mining techniques and interactive visualizations, supporting term extraction and recommendation from a collection of primary studies.

For a better understanding of the concept of Visual Text Mining, it is necessary to understand the fields of Text Mining and Information Visualization, discussed in this section.

2.2.1 Text Mining

Text Mining is the process of discovering new and unknown information, using algorithms to automatically extract it from one or different text documents [8]. Sometimes it is applied to collections of texts, supporting the discovery of patterns and relationships between text documents [7].

One of the Text Mining techniques is to extract relevant terms from a text document. A common way to do that is to apply algorithms that run frequency-based calculations and give each term in a document a value or relevance [13]. To consider term relevance in a document that belongs to a collection, one can apply techniques that involve the term's frequency in all documents.

Tf-idf (term frequency - inverse document frequency) [13] is a statistical calculation that considers the occurrence of a term in a document and the number of documents in a collection where the term is present. For that, *tf-idf* value is calculated using the following formula: $tfidf(t,d) = tf(t,d) * idf(t)$, where *d* is a document containing term *t*, *tf(t,d)* is how many times *t* appears in *d*, and *idf(t)* is how many documents in the collection contain the term *t*.

A normalization of the frequency values of the terms in a document is recommended, since terms in big text documents might have a greater frequency value than in smaller text documents, leading to a higher *tf-idf* value without necessarily indicating a greater relevance [13]. To apply normalization, the *tf* portion of *tf-idf* calculation is divided by the frequency of the most frequent term in that document (*tfmax*). In other words, a big text document may repeat a term more times than the smaller one but using the number of times the most-frequent term appears in that document normalizes this. The final formula for *tf-idf* then becomes:

$tfidf(t,d) = (tf(t,d)/tfmax(d)) * idf(t)$, where *tfmax(d)* contains the number of times the most-frequent term appears in document *d*. For purposes of this paper, the *tfidf* value mentioned in the following sections will always be normalized.

2.2.2 Information Visualization

Information Visualization is considered an emerging multidisciplinary field that comprehends visual representations of abstract data, with the goals of facilitating the communication of and assisting in exploration and analysis of information [17].

The benefits of applications from Information Visualization are within the exploration of human vision. One of the strongest points in visualization is the human ability to process visual information much faster than, for example, verbally. It is possible to condense a larger amount of data in a single view and the visualization process involves the human sense with greater capacity to capture information per unit of time [6].

A good application of Information Visualization should be unusual, informative, efficient and attractive. Aiming to create visualization with such features, some technique must be applied. One of the most popular techniques when visualizing differences is the heatmap [18].

Using heatmap for visualizing differences facilitates the action of comparing items, where the shades of colors represent the distance between values. Without having to look at accurate values, the heatmap is helpful to efficiently compare and spot differences [18].

2.3 Related Work

In a recent systematic mapping carried out in order to identify and classify tools that seek to support all or part of the process of SLRs conduction [14], it is possible to observe the lack of research that addresses activities on the planning stage of the review, as well as the search string construction.

The systematic mapping found 14 studies presenting tools that support SLRs [14]. Regarding the phases of the systematic literature review process that are addressed by the tools found in the mapping, from all studies found (14), 11 studies discuss tools that address activities in the conduction phase of the SLR, and 3 studies present tools that address the process of systematic literature review as a whole.

Among the studies mapped by Marshal et al [14], the study selection activity is the most commonly supported. There are 5 studies that support this activity, 3 of which use VTM techniques. Noteworthy that in this recent mapping there are no studies on supporting the search string construction during the development of the review protocol. Therefore, this gap is an opportunity for the development of a study as the presented in this paper, exploring techniques of Visual Text Mining targeting the process of building the systematic review search string.

3. THE PROPOSED METHOD

In this section, we present the results from a set of interviews with researchers and the proposed method.

3.1 Understanding the difficulties

Recent studies mention difficulties that researchers face when performing a systematic literature review [1] [3] [5] [16]. In order to deepen the understanding about the mentioned difficulties we applied a semi-structured interview with a small group of researchers. We choose to go deep instead of broad, as presented by Lazar, Feng and Hochheiser [12], in order to gather freely and detailed responses that help our reflection about the problem. The interviews were scheduled and conducted in person or remotely via communication tools such as Skype[1] and Google Hangout[2].

3.2 Interviews

In the following subsections, we present the profile of the participants that answered the interview and the analysis of the results.

3.2.1 *Participants profile*

We interviewed seven researchers, herein identified as P1, P2, P3, P4, P5, P6 and P7. All participants have completed a systematic literature review at least once. Concerning the experience of the interviewed researchers, two had performed more than one systematic literature review (P6 and P7) while the vast majority applied the technique only once. Three respondents (P1, P6 and P7) stated that they had undertaken a SLR without being able to finish it for different reasons, which were explored in detail in the continuation of the interview. Regarding the evaluation and publication of SLR, four respondents (P3, P5, P6 and P7) reported having the results of their SLRs and two (P6 and P7) had their SLRs published.

3.2.2 *Interview approach*

We conducted standardized open-ended interviews with a pre-defined set of questions that facilitated collecting data in a qualitative fashion. Table 1 presents the applied questions.

Table 1. Interview Questions

ID	Question
Q1.1	Have you ever performed a systematic review? How many times?
Q1.2	Have you ever started a systematic review but have not finished?
Q1.3	Have you ever had a systematic review evaluated in some way?
Q1.4	Have you ever published a systematic review?
Q1.5	How did you learn to perform a systematic review?
Q1.6	What is the greatest difficulty in conducting a systematic review? At what stage of the process?

3.2.3 *Result analysis*

Analyzing the interview answers, five of the seven participants referenced issues during the construction of the search string, as indicated by some studies on the subject [1] [4]. Specifically asking about systematic review that they could not complete, a respondent (P7) said it was due to a poorly built search string, resulting in too many studies.

Considering the difficulties to conduct a SLR, one of the respondents (P1) states that his greatest difficulty is the definition of the keywords for the review search string, highlighting the importance of this step: *"[...] If you choose a non-representative group of words, you won't get relevant results. So, the toughest part to me is finding the right words for what I'm searching for [...]"*. He further states that he once had to define the search string *"[...] by trial and error, going directly to the digital library, applying the search string and observing the results"*. According to him, because of that he had to revisit the planning phase of his SLR in order to update the review protocol with a new search string.

A similar experience was reported by another respondent (P6), mentioning that he lost time due to poorly defined search string: *"[...] I lost three weeks of work once because of a wrong term in the query. [...] I defined the query, returned items and saw that very few studies were related to the review [...]"*. Other two respondents cited challenges when defining the search string due to not knowing the proper terminology or it not being standard. The use of non-standardized terms is actually identified as one of the problems that lead to difficulties in performing systematic literature reviews [1].

Looking at the most mentioned difficulties, building the search string is clearly a top challenge for the interviewed researchers, as shown in the Figure 2.

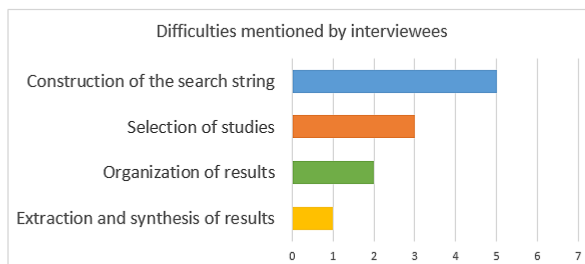

Figure 2. Difficulties most mentioned by interviewees

Based on the difficulties pointed out by the literature [1] [3] [5] [16] and the results of the interviews, we identified a need for a method that supports the construction of the search string in the systematic review process. Thus, we propose a method to address this need by using Visual Text Mining, which is detailed in the next section.

3.3 A Method to support SLRs' search string building

The method proposed in this research works iteratively between the planning and execution phases of a systematic literature review, applying Visual Text Mining techniques to support the researcher in building and validating the search string by extracting and recommending relevant terms.

[1] http://www.skype.com/

[2] https://www.google.com/hangouts/

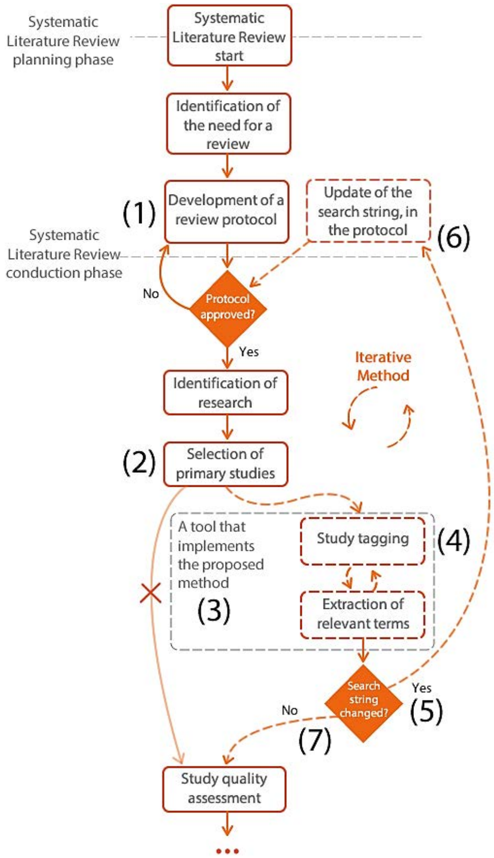

Figure 3. Proposed method workflow

As indicated in Figure 3, this method is an adaptation from Biolchini et al. [2] proposed approach for applying systematic literature reviews. Once the planning phase is complete and the protocol is created – including the defined search strategy – then an expert on the subject is responsible for the approval (1). Once approved, the protocol is put into practice and the searching commences. During the selection of primary studies (2), we recommend applying the method proposed on this paper.

Using the method, a tool implementing it would act as a search string verifier (3). Giving the researcher the ability to mark resulting studies as relevant or non-relevant to the current review (4), Text Mining techniques extract terms and expose the key ones in a visualization that indicates the relevance of those terms and supports the researcher on the decision of updating the search string. Once the researcher is done marking documents and updating the search string to incorporate new terms or remove former ones (5), the proposed method takes the researcher back to planning phase, updating the review protocol to reflect the new search string (6). The protocol then is taken into review for approval before entering execution phase again, repeating the process that iteratively refines the systematic review search string based on the studies being returned. If no changes were made, the researcher might proceed applying the systematic literature review by moving to the next step (7), continuing on Biolchini et al. [2] process.

Term recommendation is one of the key features of the method proposed on this paper. The following subsection clarifies on the approach and techniques applied with the objective of identifying the most relevant terms for recommendation.

3.3.1 Term Recommendation

Using the predefined inclusion and exclusion criteria, the researcher can select a study as relevant or non-relevant to the current review. This action triggers the term recommendation activity, which uses *tf-idf* weighting to find key terms from the marked documents. To support that, a text document is extracted from each study in a format that allows text-mining techniques to read from it. This weighting technique then calculates a global *tf-idf* value for each term on the collection of all extracted documents.

There are three steps involved in the process of calculating a term's *tf-idf* weight: (1) getting a frequency value for that term in studies marked as relevant (from R group), (2) getting the frequency value for that term in studies marked as non-relevant (NR group) and (3) calculating the final *tf-idf* by subtracting the non-relevant frequency value from the relevant frequency value, and then dividing the resulting value by the number of documents marked as relevant (|R| or the cardinality of R group).

Studies selected as relevant are grouped in a way that facilitates calculating intermediary frequency values, identified as the R group. A frequency value for each term t in R group documents is calculated by summing all its *tf-idf* values, using:

$\sum_{di \in R} tf - idf(t, di)$, where di is a document contained by R and t is the term for which *tf-idf* weight is calculated. Studies marked as non-relevant by the researcher are grouped separately, in a group identified as NR. Term frequency is then calculated the same way, by summing all *tf-idf* values for a given term t:

$\sum_{dj \in NR} tf - idf(t,\ dj)$, where dj is a document contained by NR and t is the term for which *tf-idf* weight is calculated. The final *tf-idf* value for a given term t in the body of documents resulting from the literature search D can then be found by the following formula:

$$tf - idf(t, D) = \left(\left(\sum_{di \in R} tf - idf(t, di)\right) - \left(\sum_{dj \in NR} tf - idf(t,\ dj)\right)\right) / |R|$$

where $|R|$ is the number of documents marked as relevant (or the cardinality of R group). The outcome of that is a list of global *tf-idf* values for every term identified in the studies. Sorting that list by the highest relevance values results in exposing the key relevant terms at the top. Those are the terms to be recommended by the method. Including them in the search string implies in returning to the planning phase, updating the review protocol and searching again, refining the search strategy of the SLR.

3.4 Implementing the proposal: SLR.qub tool

In order to allow a discussion about the proposed method, we developed a tool that implements it. The SLR.qub tool (Systematic Literature Review – query builder) applies Text Mining algorithms to the collection of abstracts from the primary studies returned from searching in IEEExplore digital library with the current SLR search string. It is worthwhile to mention that this tool was built as a proof of concept, and not a commercial use – its limitations are discussed in the final considerations.

Once done mining and extracting useful information, it displays a screen on top of the results page with a visual representation of all returned studies in a carousel. A screen captured from the tool is in Figure 4.

Besides the retrieved studies, the SLR.qub tool displays the search string applied with its terms highlighted in blue and logical operators in gray. In line with the search string, a button labeled "*Search!*" applies a new search in IEEExplore using the updated search string.

The tool allows the researcher to update terms or logical operators on the search string, manually adding or removing terms. Just below the search string, there is an area reserved for listing recommended terms. In order to display recommend terms based on the researcher's indication of relevant and non-relevant studies, this area is dynamically populated and varies according to the user interaction. When clicked, the terms recommended are added to the current search string, concatenated with a logical operator *OR*. It would be up to the user then to edit the search string and tailor the term newly added to the search string.

As shown in Figure 4-a, each term suggested by the tool carries a global value of *tf-idf*, calculated by the applied mining techniques. This value indicates the relevance of a term in relation to the collection of marked documents, and facilitates comparison between terms (to verify, for instance, the distance between two terms). Still in Figure 4-a, the first recommended term (*graphs*) has a relevance value (*tf-idf*) of *8.05*.

To select a study as relevant, the researcher must click on the representation of the study in the carousel area. On the first click, the color on the bottom of the document becomes green, indicating that the study is relevant for the systematic literature review in progress. As an effect of this interaction, the recommended terms are updated to reflect the changed list of key terms, after the re-calculating *tf-idf* values.

To indicate a document as non-relevant to the SLR, the researcher must click again on a study that had been selected as relevant in the carousel. Right after the click, the color changes from green to red, indicating it as non-relevant. Figure 4-b illustrates the flow of selecting studies. Initially, the study is unselected. Selecting and unselecting studies is a cyclical action and state changing is triggered by clicking on the study in the carousel. In the unselected state, a study is neutral to the algorithm, not interfering in the recommended terms.

At the bottom, SLR.qub displays a heatmap visualization of the recommended terms and all the returned studies. This visualization has a number of functions in the SLR.qub tool. The first one is to show which studies are visible in the carousel, positioning the researcher on the list of studies. The second function of the heatmap visualization is to identify if a study is selected as relevant, non-relevant or if it is unselected, using the colors green and red or leaving it uncolored, respectively. Finally, it helps to determine what could be a potential set of control papers to validate candidate search strings in a topic of expertise to the researchers, considering a given set of terms and papers.

As its main feature, the heatmap in SLR.qub is also responsible for representing the relevance of the recommended terms in relation to the studies. In order to build this visualization, shades of colors from yellow to red are applied based on the relevance (*tf-idf(t,d)*) values of a given term (*t*) in a particular study extracted abstract (referred to as document *d*). In the format of a table, the heatmap represents the recommended terms in rows and studies in columns. As noted in Figure 4-c, reddish cells indicate a higher relevance value. For instance, in Figure 4-c, the term *graphs* has a greater relevance in study 2, whereas the term *beginners* finds more relevance in study 1. Visualizing the relevancies this way supports the researcher in both looking for studies based on a term known to be relevant to the SLR or selecting terms based on studies.

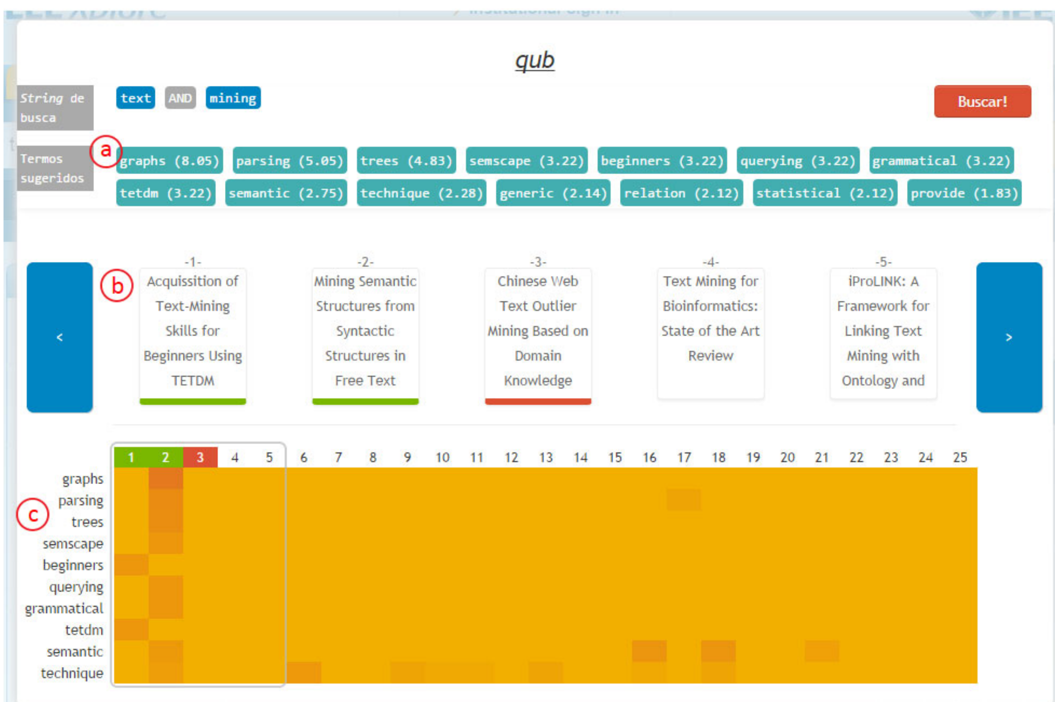

Figure 4. SLR.qub screenshot

4. ANALYSIS OF THE PROPOSED METHOD

In order to discuss the applicability of the proposed method, we conducted a preliminary study using the tool that implements it (*SLR.qub*). This section discusses this study and its results.

4.1 Study description

The preliminary study consisted of executing the construction of the search string for a proposed systematic literature review. At the beginning of the study, participants were briefed with information regarding a proposed SLR, comprehending the inclusion and exclusion criteria and a basic search string to start with. Its' objective was to understand the current state of systematic literature reviews in the area of Software Engineering and finding the challenges faced by researches that applied it. The basic search question provided for that SLR was "*systematic review in Software Engineering*". We conducted the study individually with each participant, in a mixed present and remote fashion but always observing the participant's actions.

Given the brief description on the proposed SLR, the next step consisted on pulling up IEEExplore digital library, applying the basic search string and activating the SLR.qub tool. Using the tool, the participants had to mark the returned studies they considered relevant or non-relevant for the hypothetical SLR.

4.2 Participants profile

We chose to perform the study with the same seven participants from the interviews discussed in section 3.2, given our familiarity to their level of experience with SLRs. Participants had to be familiar with Software Engineering researches and had already completed at least one systematic literature review. That last prerequisite regarding the researcher's experience with SLRs was crucial to the understanding and realization of the proposed task.

4.3 Results Analysis

Figure 5 represents the documents found during the preliminary study (from D1 to D25) and the respective tagging by the researchers. Looking at it, we can see how many times each document has been tagged as relevant, non-relevant or left unselected.

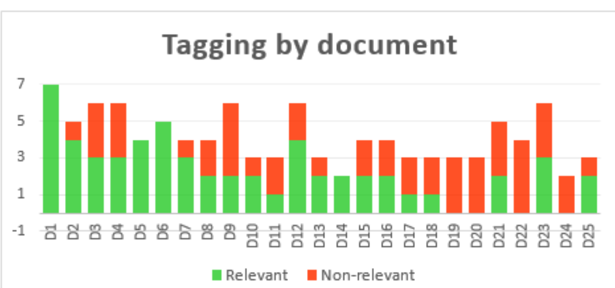

Figure 5: the 25 documents found and their tagging

In Table 2, document *D1* shows as selected by all participants in the study as relevant. *D1* is the first study resulting in IEEExplore search, and it provides a discussion on the challenges faced by researchers that conducted systematic literature reviews. That paper is highly relevant to the SLR proposed in the study and was correctly selected as relevant by all participants.

There are also documents like *D5* and *D6*, which were never selected as non-relevant. On the other hand, documents *D19, D20, D22* and *D24* were not selected relevant by any of the participants. Document *D5* is a key reference on systematic literature reviews in SE [9] [10] [11]. Reading the transcript of the execution of the study, three of the four participants selected the study after noticing that.

Regarding documents *D19, D20, D22* and *D24*, one of them is related to the area of Software Testing (*D19*), two of them discuss certifications in project management (*D20* and *D22*) and the latter is a meta-analysis on the relationship between software developers' productivity and personality (*D24*). At the end of each execution, we captured the search string built by each participant.

Table 2 shows the incidence of terms included by the participants during the study, in addition to the original terms from the proposed SLR. There are terms like *SLR* (included by three of the seven participants) and *method*, *replication*, *study* and *tertiary* (each included by two of the seven participants). Note: the same participant (P3) applied the term search twice on his search string (S3).

Table 2. Frequency of terms present in final search strings

	term	frequency	S1	S2	S3	S4	S5	S6	S7
t1	*review*	10	■	■	■	■	■	■	■
t2	*systematic*	9	■	■	■	■	■	■	■
t3	*engineering*	7	■	■	■	■	■	■	■
t4	*software*	7	■	■	■	■	■	■	■
t5	*slr*	3	■				■		■
t6	*method*	2	■					■	
t7	*replication*	2		■				■	
t8	*search*	2			■				
t9	*study*	2					■		■
t10	*tertiary*	2					■		■
t11	*approach*	1			■				
t12	*contribute*	1			■				
t13	*findings*	1			■				
t14	*groups*	1	■						
t15	*knowledge*	1			■				
t16	*lessons*	1	■						
t17	*literature*	1							■
t18	*methodology*	1						■	
t19	*motivation*	1		■					
t20	*outcomes*	1			■				
t21	*search*	1			■				
t22	*statistical*	1	■						
t23	*string*	1			■				

4.4 Post-study Interviews

For capturing participants' opinion on the conducted preliminary study and the proposed method, we conducted interviews right after each execution. This section presents the results.

Seven researchers indicated they have not found a search string they considered relevant to the proposed systematic review. Three of the five researchers (*P5, P6* and *P7*) indicated the lack of relevant terms in the recommendations. The other two researchers (*P1* and *P2*) indicated lack of time[3] to work on

[3] *Lack of time,* as indicated by the participants might be related to the limited time each participant had on this preliminary study.

selecting and marking the studies, analyzing the recommended terms and editing the search string.

When asked about the way researchers often build the search string of a SLR, four researchers indicated that they start by using terms they are familiar with (*P2, P4, P5* and *P6*). Through a simple search with these terms, they capture synonymous by reading some of the returned studies. One of the interviewees (*P6*) describes the process as follows: *"I first include the terms I know. Use them to search, revise and update the search string based on the returning papers. Then I repeat the process until I find an interesting search string [...]"*.

The other three researchers (*P1, P3* and *P7*) reported using the most popular and referenced studies (cited by more papers). *"[...] Then try extracting terms. Would be somewhat similar to your tool but using a predefined set of studies",* noted one of the researchers (*P3*).

The researchers were unanimous in their responses when asked if the method would help a researcher conducting a systematic literature review in an unfamiliar area, by recommending relevant terms. One of the researchers (*P3*) answered: *"Yes it would help, as it would aid directly in the issue of finding adequate terms. If you don't know a particular area, it is difficult to find some terms, and the method would help exactly with that"*.

When asked about researchers using the proposed method to perform SLRs in a field they are already experienced in, most participants claimed that the tool would help in building the search string with some caveats. Researchers *P2, P3, P5* and *P7* said it is more likely that it would be used in cases where there is an outline of the search string but needing verification. One of the researchers (*P5*) also indicated that he could imagine himself using the tool to update his familiarity with the latest terminology used in the area.

5. DISCUSSION

From the participants of this study, the most experienced researchers (*P6* and *P7*) – who had carried out more than one systematic review and had some reviews evaluated and published – mentioned difficulties in building the search string. This finding is in agreement with authors claiming that problems with the construction of the search string and selection of primary studies are within the most frequent in performing systematic reviews, faced by both beginners and experienced researchers [16]. The experience gained by performing SLRs also made the task of tagging studies easier by them, using their known references on the systematic review area. One of the most experienced researchers (*P6*) states that it accelerates the usage of the method and facilitates its integration in the SLR process.

As mentioned by researchers *P2, P3, P5* and *P7* when asked about the usage of such method by experienced researchers, it is possible to notice that *P6* tried to verify terms he already knew, which could be considered a valid use of the proposed method. However, for the method to work that way, it must be used iteratively, running new searches as the search string is modified. Thus, refining the resulting list of studies to the point of finding a set of studies familiar to the researcher and known to be relevant, confirming the relevance of the terms in the search string.

6. CONCLUSION AND FUTURE WORK

As observed in the interviews, researchers face difficulties in building the SLR's search string regardless of their experience with it, as mentioned in the literature [1] [4]. Yet, methods or tools addressing this specific activity on the SLR are still uncommon.

In order to assist the construction of the search string of a SLR, this paper presented an iterative method that supports the researcher by suggesting relevant terms based on manually selected studies. Through techniques of Visual Text Mining, terms are extracted and presented through a heatmap visualization that facilitates identifying relevant terms on returned studies. Once the search string is modified to accommodate any new terms, the method supports the update of the review protocol and re-execution of the search activity, allowing the researcher to repeat the process as often as necessary for the construction and refinement of the search string.

For the analysis of the applicability of this proposal, we developed a tool that implements the proposed method and tested it with researchers previously interviewed, as part of a preliminary study. At the end of each execution, the researcher was subjected to a post-study interview that captured his opinion, relating to difficulties he might have experienced in performing systematic literature reviews. All participants indicated that the tool could help researchers conducting a systematic literature review in an area they are not familiar with. Most participants also believe it would help researchers who have some experience in the area they are conducting a SLR but in a slightly different fashion, by verifying a built search string or updating the researcher with the latest terminology in the area.

The *SLR.qub* tool requires some refinement to address its limitations, before submitting it for use by a larger and more diverse sample of researchers – with different experiences in performing systematic literature reviews. Therefore, we will be able to capture more information and feed more depth discussions in this promising research field.

6.1 Limitations

One of the main limitations of this research is technical and refers to the Text Mining techniques applied. The SLR.qub tool reads the results page in IEEExplore but accessing and extracting the full text of each study showed to be technically unfeasible for this preliminary study. This is because the results page in IEEExplore shows only the abstract for each study (vs the full study text). Thus, *SLR.qub* could only extract and analyze terms the abstract of each study.

Due to that limitation, the relevance of the extracted terms is calculated over a body of text that, on average, has around fifty words. With recent studies questioning the quality of abstracts of studies in the area of Software Engineering and indicating that the abstracts of publications contains, in general, less than half of the information contained in a study [10], this limitation introduces a bias to the search result. Another limitation is the low number – seven – of participants in the discussed study. There is a need for a more extensive and rigorous evaluation. It is important though to notice that the *SLR.qub* tool is a proof-of-concept created to support the preliminary study conducted as part of this research to explore the proposed method.

6.2 Future Work

A further research lies in rewriting the mining algorithm in a more robust way, extracting content from the full studies text (vs. reading from only the abstracts). This approach could influence significantly the relevance of the recommended terms, and a new research would be required to verify that hypothesis. Another possible further research involves enhancing the tool that implements the method by making it broader, so it would read studies not only from IEEExplore but also from other main digital libraries. By doing that, the mining techniques would be performed on top of a larger set of studies, possibly increasing relevance of the recommended terms. It is important, though, to account for study duplication (studies that can be found in more than one digital library) and how each search engine in those digital libraries handle logical sentences.

7. ACKNOWLEDGMENTS

Our thanks to the participants of the study and to FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo).

8. REFERENCES

[1] Babar, M. A., and Zhang, H. (2009). Systematic literature reviews in software engineering: Preliminary results from interviews with researchers. In Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement (pp. 346-355). IEEE Computer Society.

[2] J. Biolchini, P. G. Mian, A. C. C. Natali, and G. H. Travassos. Systematic review in software engineering. *System Engineering and Computer Science Department COPPE/UFRJ, Technical Report ES*, 679(05):45, 2005.

[3] T. Dybå, T. Dingsøyr, and G. K. Hanssen. Applying systematic reviews to diverse study types: An experience report. In *Empirical Software Engineering and Measurement, 2007. ESEM 2007. First International Symposium on*, pages 225–234. IEEE, 2007.

[4] T. Dybå and T. Dingsøyr. Strength of evidence in systematic reviews in software engineering. In *Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement*, pages 178–187. ACM, 2008.

[5] K. R. Felizardo, S. G. MacDonell, E. Mendes, and J. C. Maldonado. A systematic mapping on the use of visual data mining to support the conduct of systematic literature reviews. *Journal of Software*, 7(2):450–461, 2012.

[6] S. Few. *Now you see it: simple visualization techniques for quantitative analysis*. Analytics Press, 2009.

[7] R. Garg et al. Study of text based mining. In *Proceedings of the International Conference on Advances in Computing and Artificial Intelligence*, pages 5–8. ACM, 2011.

[8] M. A. Hearst. Untangling text data mining. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 3–10. Association for Computational Linguistics, 1999.

[9] B. Kitchenham. *Procedures for performing systematic reviews. Keele, UK, Keele University*, 33:2004, 2004.

[10] B. Kitchenham and S. Charters. Guidelines for performing systematic literature reviews in software engineering. *School of Computer Science and Mathematics, Keele University, EBSE Technical Report*, EBSE-2007-01, 2007.

[11] B. Kitchenham, P. Brereton, and D. Budgen. Mapping study completeness and reliability - a case study. 2012.

[12] J. Lazar, J. H. Feng, and H. Hochheiser. *Research methods in human-computer interaction*. John Wiley & Sons, 2010.

[13] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to information retrieval*, volume 1. Cambridge university press Cambridge, 2008.

[14] C. Marshall and P. Brereton. Tools to support systematic literature reviews in software engineering: A mapping study. In *Empirical Software Engineering and Measurement, 2013 ACM/IEEE International Symposium on*, pages 296–299. IEEE, 2013.

[15] M. C. F. De Oliveira and H. Levkowitz. From visual data exploration to visual data mining: A survey. *Visualization and Computer Graphics, IEEE Transactions on*, 9(3):378–394, 2003.

[16] M. Riaz, M. Sulayman, N. Salleh, and E. Mendes. *Experiences conducting systematic reviews from novices' perspective*. In Proc. of EASE, volume 10, pages 1–10, 2010.

[17] J. Steele and N. Iliinsky. *Beautiful Visualization: Looking at Data through the Eyes of Experts.*" O'Reilly Media, Inc.", 2010.

[18] N. Yau. *Visualize this: the FlowingData guide to design, visualization, and statistics*. John Wiley & Sons, 2011.