Appendix

D

Paper: An Extension of the Systematic Literature Review process with Visual Text Mining: A case study on Software Engineering

This appendix presents the full contents of a paper submitted to the Automated Software Engineering Journal (ASE Journal – currently under evaluation), which is a synthesis of the works presented in Chapters 3 and 4 of this thesis.

An Extension of the Systematic Literature Review Process with Visual Text Mining: A Case Study on Software Engineering

Katia Romero Felizardo[a], José Carlos Maldonado[a], Rosane Minghim[a], Stephen G. MacDonell[b], and Emilia Mendes[c, *]

[a]Department of Department of Computing Science, University of São Paulo, Brazil.
[b]School of Computing and Mathematical Sciences, AUT University, New Zealand.
[c]College of Information Technology, Zayed University, Dubai, United Arab Emirates

Abstract

Background: The systematic literature review (SLR) is a methodology used to find and aggregate all relevant existing evidence about a specific research question of interest. One of the activities associated with the SLR process is the selection of primary studies. Prior research has shown that the quality of primary study selection impacts the overall quality of the SLR. Therefore, in order to ensure better quality outcomes of the SLR as a whole, it is important to conduct as completely and as reliably as possible the primary study selection activity. This is particularly relevant when the researcher faces large volumes of primary studies, where it can be challenging to select relevant articles for further analysis – the process used to select primary studies can be arduous, time consuming, and must often be conducted manually. **Objective:** This paper addresses these challenges validating an extension to the SLR process. Such extension uses Visual Text Mining (VTM) to automate and support the selection of primary studies and the review of decisions on whether to include or exclude studies from an SLR. **Method:** We have conducted a case study on a previously reported SLR to demonstrate the usefulness of the VTM techniques in the selection context. **Results:** The results show that the employed VTM techniques can successfully assist the SLR process. In particular, the visualizations are able to highlight inconsistencies in the treatment (inclusion or exclusion) of similar primary studies that might signal the need for closer attention when selecting studies. **Conclusions:** The use of VTM to support the study selection activity within the software engineering context requires minimal additional knowledge of visualization techniques and offers automated tool support to facilitate the execution of this activity. Different VTM techniques may be combined to support reviewers in their decisions regarding the inclusion or exclusion of primary studies from an SLR. This should result in a more reliable treatment of primary studies and, therefore, more reliable SLR outcomes.

Keywords: Systematic Literature Review (SLR), Visual Text Mining (VTM), Evidence-based Software Engineering (EBSE).

1. Introduction

The systematic literature review (SLR) is recognized as one of the key components of the evidence-based software engineering (EBSE) paradigm (Kitchenham and Charters, 2007). SLRs were introduced in the software engineering (SE) field in 2004 (Kitchenham, 2004) and have since gained increasing popularity among SE researchers (Dybå and Dingsøyr, 2008). The bibliographical review, also known as informal literature review, does not explicitly define the search process or the data extraction process; hence it may be vulnerable to bias in both conduct and outcome (Kitchenham and Charters, 2007). In contrast, an SLR employs a methodical process of identifying, assessing and interpreting all available and relevant research evidence in a thorough and unbiased manner. Moreover, the SLR can be used to identify publications related to a specific topic of interest using a predefined search strategy aimed at minimizing bias (Kitchenham, 2004). The term "primary study" refers to an individual piece of evidence, e. g. a case study or an experimental study, considered in an SLR. Therefore an SLR is itself referred to as a "secondary study". SLRs of high-quality avoid bias and are considered crucial to the progress of EBSE (Dybå et al., 2005).

Despite its importance, the SLR process is still largely carried out manually, which is both challenging and time-consuming. Riaz et al. (2010) have summarized the potentially problematic aspects of the process to be: (i) formulation of research questions (Brereton et al., 2007; Staples and Niazi, 2008); (ii) conducting searches (Dybå et al., 2007); (iii) primary study selection (Dybå and Dingsøyr, 2008); and (iv) primary studies' quality assessment (Dybå et al., 2007; Staples and Niazi, 2008), among others. Therefore, the problems faced by researchers while conducting SLRs and the approaches adopted to tackle these problems vary across studies (Riaz et al., 2010). One particular issue encountered by researchers involves the selection of primary studies, especially when many, primarily irrelevant, search results are returned. This issue not only makes the primary studies' selection process very cumbersome but could also introduce selection bias (Riaz et al, 2010). When the selection is performed, uncertainties about inclusion or exclusion should be investigated using sensitivity analysis, which involves revisiting the selection activity for the studies divergently classified by reviewers (Kitchenham and Charters, 2007). This implies an additional effort to re-read the studies in question. The quantity of papers to be read and

* Corresponding author. E-mail: katiarf, jcmaldon; stephen.macdonell@aut.ac.nz and Emilia.Mendes@zu.ac.ae

analyzed, and possibly re-read, may present a difficult task for any researcher or research team. The selection of primary studies for an SLR is an issue of central importance, given that one needs to aggregate studies that truly relate to the research subject focus of the review.

In recent years, there has been an increasing interest in the use of visualization techniques as supporting tools for SLRs (Malheiros et al., 2007; Emam et al., 2009; Ananiadou et al., 2009; Felizardo et al., 2010). This interest is motivated by the fact that humans possess strong visual processing abilities; visual-based techniques make use of these abilities, by exploiting the human system to support knowledge discovery (Keim, 2002). In the SLR context, this seems to be particularly beneficial in regard to the systematic discovery of relevant primary studies. The SLR process involves the analysis of an often large collection of documents (Dybå et al., 2007; Bereton07; Dybå and Dingsøyr, 2008). Visual representations may provide a novel and valuable way for a researcher to review and explore such a collection (Paulovich and Minghim, 2006; Paulovich and Minghim, 2008). VTM is an extension of Text Mining (TM), a well-established practice commonly used to extract patterns and non-trivial knowledge from unstructured documents or textual documents written in a natural language (Tan, 2005). Visual Text Mining (VTM) is the association of mining algorithms and information visualization techniques that support visualization and interactive data exploration (Oliveira and Levkowitz, 2003). Keim (2002) suggests that visualizing and exploring information using VTM might be beneficial when dealing with vast amounts of data. This is because the visualization supports user interaction with the mining algorithms and can direct the search towards a suitable solution to a given task. Moreover, VTM can be used to enhance user interpretation of mining tasks (Oliveira and Levkowitz, 2003).

Such a growing acceptance of VTM (Garcia et al., 2004; Malheiros et al., 2007; Emam et al., 2009; Felizardo et al., 2010) has motivated us to employ this technique in our current study, and also in our previous work (Felizardo et al., 2011). Therefore, the present study is an extension of our previous work (Felizardo et al., 2011), where we have investigate further the use of visualization techniques to help with carrying out SLRs by employing VTM techniques to aid the selection of primary studies when carrying out an SLR. The main motivation behind that work was to propose an approach based on VTM, the SLR-VTM (Systematic Literature Review based on Visual Text Mining) to support the selection of primary studies in an SLR, offering different perspectives, i.e., visualizations based on the content of the studies and based on their citation relationship. In the present study, the key contribution is to add empirical evidence regarding the effects of the VTM in the selection of primary studies context.

The remainder of this paper is structured as follows. Section 2 provides background information on the SLR process and an introduction to VTM concepts. Section 3 specifies in more detail the stages of the SLR process that can be supported by VTM techniques. Section 4 includes a case study in which we have applied VTM to a real SLR, the results of which are discussed in Section 5. Finally, Section 6 concludes our work, highlights the limitations of the study and discusses future research.

2. Background and Related Work

2.1. Use of SLRs in Software Engineering

The SLR provides reliable means and established methods to conduct a comprehensive and robust literature review based on three clearly defined phases: (i) planning; (ii) conducting; and (iii) reporting the review (Kitchenham, 2004). During the planning phase, the need for a review is identified and the review protocol is developed. The activities during the second phase include the identification of relevant research and selection of primary studies based on stated inclusion and exclusion criteria. It is important that the studies be reviewed (i.e., that there is an explicit selection review activity) to ensure, in particular, that relevant studies are not eliminated in error. Next, significant information is extracted and synthesized from the studies identified as appropriate for inclusion. Finally, the third phase includes dissemination or reporting of the SLR's results to interested parties including researchers and practitioners (Kitchenham and Charters, 2007).

Three studies (Kitchenham et al., 2009; Kitchenham et al., 2010; Silva et al., 2010) have assessed the impact of SLRs as an EBSE method for aggregating evidence in software engineering. In 2009, Kitchenham et al. concluded that the topic areas covered by SLRs were limited and that EBSE was principally supported by European-based researchers. The majority of topics were concerned with technical issues rather than research methods. They also found that the quality of SLRs was improving, but that many researchers still preferred to undertake informal literature reviews. In 2010, Kitchenham et al. repeated their assessment and the results of this study indicated that the number of SLRs published and the topic areas covered by these SLRs appeared to be increasing. The quality of these SLRs also appeared to be improving; however, relatively few SLRs evaluated the quality of the primary studies included. The findings also suggested that researchers based in the USA, which is the leading country in SE research, had conducted few SLRs. The purpose of EBSE is to inform researchers about empirical evidence that can be used to improve SE practice, however, Kitchenham et al. (2010) found out that relatively few SLRs provided advice oriented to practitioners.

The results of an independent assessment study conducted by Silva et al. (2010) showed that the main limitation constraining the use of SLRs in SE is that a large number of SLRs do not assess the quality of the underlying primary studies, confirming the previous findings of Kitchenham (Kitchenham et al., 2009; Kitchenham et al., 2010). In addition, they support the findings from the Kitchenham studies, showing that the number of SLRs providing guidelines to practitioners is small. Their findings also showed, that the number of SLRs being conducted was increasing, along with the number of researchers and organizations performing them.

In relation to the integration/synthesis of results from primary studies, Silva et al. (2010) found work was that they were poorly conducted in many SRLs. In this same context, Cruzes and Dybå (2010) performed a tertiary review to assess the types and methods of research synthesis evident in SLRs in SE. They included 31 studies in their review and found that almost half of those studies (13 of the 31 considered) did not contain any synthesis. This suggests that currently the attention given to research synthesis in SE may be limited. They reported that just over half of the studies analyzed used tables (i.e., the simplest type of graphic presentation) to show their findings, and that these tended to contain a lot of data from individual studies (e.g., title, authors, year, outline, strengths, among others). Other forms of visual representation were used in fewer than 20% of the studies. The adoption of approaches to aggregate research outcomes to provide a balanced, objective and more readily understood summary of research evidence appears to be an ongoing challenge in regard to SLRs (Brereton et al., 2007; Cruzes and Dybå, 2010).

While the number of SLRs on various topics within the SE discipline is increasing, related studies have also been carried out to report researcher experiences and consider the challenges encountered by those conducting SLRs. For a summary of the problems and experiences reported by various researchers, refer to Riaz et al. (2010). Due to the necessarily comprehensive and rigorous nature of an SLR, exhaustive searches for relevant primary studies are required. One particular issue involves the selection of primary studies, especially when many, primarily irrelevant, search results are returned. Consequently, this leads to difficulties in reading and evaluating the state of the art of a current topic of interest (Riaz et al., 2010). Recent studies provide evidence that unstructured and poorly written abstracts can complicate the study selection process (Brereton et al., 2007; Dybå et al., 2007; Kitchenham et al., 2008; Dybå and Dingsøyr, 2008). Riaz et al. (2010) suggest that one of the causes of selection bias is that titles and unstructured abstracts may not actually be sufficient as a basis for the initial selection of primary studies in SLRs. It appears, then, that one of the sources of difficulty in determining whether a study is relevant to be included in the SLR is the lack of clarity and incomplete information contained in unstructured abstracts; moreover, not only do unstructured abstracts often omit key information (e.g. background, aim, method, results and conclusions), they have been found to include irrelevant information (Kitchenham et al., 2008). One solution to minimize the difficulties encountered in SE SLR study selection is to promote the use of structured abstracts (Kitchenham et al., 2008).

2.2. Visualization and the SLR Process

Several studies have investigated the potential benefits of visualization in supporting the conduct of an SLR. El Emam et al. (2009) investigated the use of Electronic Data Capture (EDC) tools to support the identification of primary studies during an SLR process; however, their study selection activity was in general still manually conducted. Ananiadou et al. (2009) employed text mining tools to support three different activities of the SLR process: (i) search, (ii) study selection – using document classification and document clustering techniques – and (iii) syntheses of the data; however their focus was in the social sciences field, and it is unclear whether their findings would apply readily to SE, particularly given the relative immaturity of study reporting in this field (Kitchenham et al., 2008). Garcia et al. (2004) analyzed how graphical representations, such as parallel coordinates, may complement statistical data analysis, helping users to understand and treat data from empirical studies. This research was the first initiative towards introducing graphical representations in the analysis of data from empirical studies in SE; however the data analyzed came from only one experiment replication conducted in a specific scope (i.e., the application of several reading techniques, aimed at evaluating and comparing their efficacy and efficiency).

Visual analytics (or visual data mining (VDM)) is an interdisciplinary field of research supported by the combination of visualization techniques, human factors (e.g., interaction, cognition and perception) and data mining (DM) algorithms (Keim, 2006; Thomas and Cook, 2005). It combines traditional mining algorithms with information visualization techniques that exploit the advantages of both approaches (Keim, 2002). In addition, the DM approach employed within the context of VDM comprises the extraction of patterns or models from the data (Keim, 2002). Visual Text Mining (VTM), therefore, is simply VDM applied to textual data. The basic idea of VTM is to visually represent the information and use the capabilities of humans in terms of visual information exploration to interact with the information, gain insights and detect interesting knowledge (Keim, 2006). Visualization enables users to “see” and navigate through the data in multiple ways, with their optical abilities enhancing the knowledge acquisition process (Oliveira and Levkowitz, 2003). Edward Tufte, who developed foundational theories about the visual display of information, stated that graphical excellence consists of complex ideas communicated with clarity, precision, and efficiency (Tufte, 1983).

Two previous studies (involving some of the present authors) (Felizardo et al., 2010; Malheiros et al., 2007) have specifically investigated the use of VTM within the context of EBSE. Felizardo et al. (2010) employed VTM to support categorization and classification of studies when carrying out systematic mapping studies; and Malheiros et al. (2007) investigated the use of content-based VTM techniques (i.e. content maps), to help with the selection of primary studies, using a feasibility study. They compared the performance of reviewers in selecting primary studies using two different methods: (i) reading abstracts; and (ii) using VTM techniques. Their results were encouraging, and suggest that the use of VTM can both reduce the time required and improve the effectiveness of the selection of primary studies.

Similarly to the work proposed by Malheiros et al., our previous study (Felizardo et al., 2011) also makes use of VTM techniques to support the selection of primary studies in the process of SLR. We expanded the VTM techniques used by Malheiros et al. (2007) (i.e. content-based analysis of documents – content maps), suggesting the use of meta-data analysis of documents, via representations such as citation maps. In their study, Malheiros et al. (2007) used the content map as a single visualization technique to support the selection activity. Although the content maps are valuable tools to support the analysis of primary studies, the content (i.e. title, abstract and keywords) may not reflect the quality of a study precisely (Kitchenham and Charters, 2007). Therefore, additional support must be provided to the analysis of other properties that only the content does not reveal. Examples of these relationship-based visualizations are citation, co-citation and co-authoring (Chen et al., 2010). In our previous work we employed other visualization technique to indicate citation relationships as support to other types of association in the selection and review selection of primary studies activities. Furthermore, this additional visualization (i.e., the citation map) can be used together with the content map through the coordination technique, also proposed by us.

Other contribution of our previous work (Felizardo et al., 2011) is that we have also developed a VTM automated tool, named Revis – Systematic Literature Review Supported by Visual Analytics – to support and substantially automate the study selection and study selection review activities of the SLR process. In the current paper the use of VTM (i.e. content and citation maps) and the Revis tool are exemplified in an EBSE context through a case study using a real, previously published, SLR.

3. The Application of VTM to Support SLRs

In our previous work (Felizardo et al., 2011) we created the SLR-VTM (Systematic Literature Review based on Visual Text Mining – see Figure 1), an approach to support primary studies' selection during the SLR process. The entire VTM-based SLR process comprises seven stages: i) planning; ii) search process; iii) visualization; iv) VTM selection; v) VTM selection review; vi) data extraction/data synthesis; and vii) reporting the review.

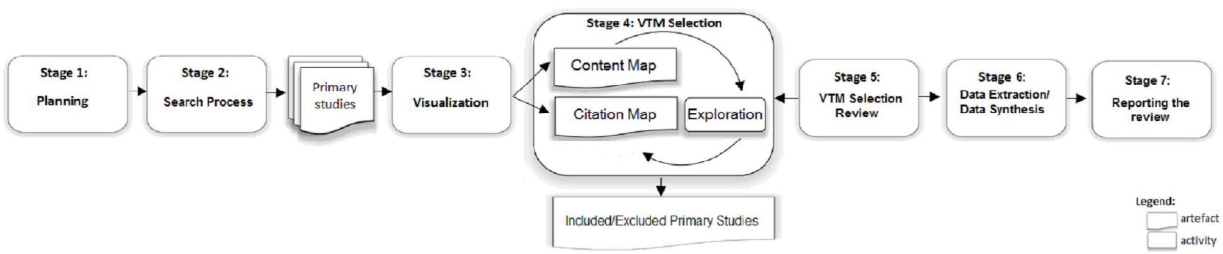

Fig. 1. VTM-based Systematic literature review process . Extended from (Kitchenham, 2004) and adapted from (Felizardo et al., 2011).

Stages 1, 2, 6 and 7 remain unchanged, as per the SLR description detailed in (Kitchenham, 2004); however, stages 3, 4 and 5 were revisited, as within the context of our previous work they include VTM support. In short, the revisited process becomes as follows: i) Stage 1 – the SLR protocol is defined containing, for instance, the research question, population, source search methods, keywords, paper inclusion and exclusion criteria, and the primary study selection process, among other elements; ii) Stage 2 – primary studies are searched in different sources according to the protocol; iii) Stage 3 – visual representations of the potentially relevant primary studies are generated; iv) Stage 4 – based on the set of primary studies found in Stage 3, the relevant primary studies are selected applying inclusion and exclusion criteria; v) Stage 5 – the studies classified as included or excluded are reviewed to ensure that all relevant studies were indeed classified as included; vi) Stage 6 – during this stage, data extraction and data synthesis activities are executed; vii) Stage 7 – the results of the review are reported. We have investigated the application of VTM in stages 3, 4 and 5 of the SLR process, briefly discussed below.

● Stage 3: Visualization

During stage 3, visual representations of the primary studies that were previously selected are generated in different ways, like: (i) content map; and (ii) citation map. Each is explained below:

(i) A **content map** (see Figure 3(a)) is a two-dimensional (2D) visual representation, where each document (i.e., a primary study in an SLR) is graphically represented as an element on the plane, normally shown graphically as a circle or point.

The documents' positions in a map reflect the similarity relationships between their content. Therefore, the more similar the documents (exhibiting content similarity) the closer together they are shown; conversely, the more dissimilar the documents, the further apart they are shown. Clusters of documents, i.e., documents in close proximity to each other, indicate that the content of these documents is similar. Therefore, the larger the distance between documents, the more their content varies. Details about the stages to create a content map (i.e., pre-processing; similarity calculation; and projection) (Paulovich and Minghim, 2006), can be found in Felizardo et al., 2011.

(ii) The most common way to visually represent citation maps is by means of graphs, which are composed of a set of vertices and edges, representing respectively objects and the relationships between them. The **citation map** shows the primary studies (central point – circle) along with their cited references (circles around the central point, connected by edges). Through this depiction it is possible to see citations between the primary studies with their own references and also citations between primary studies and references of other primary studies (references shared – see Figure 2). The citation map visualization uses a force-based algorithm (Eades, 1984) to position the points on the layout. This means that studies attract or repel one another depending on how strong their connections (references to each other) are. More details about citation maps can be found in Felizardo et al., 2011.

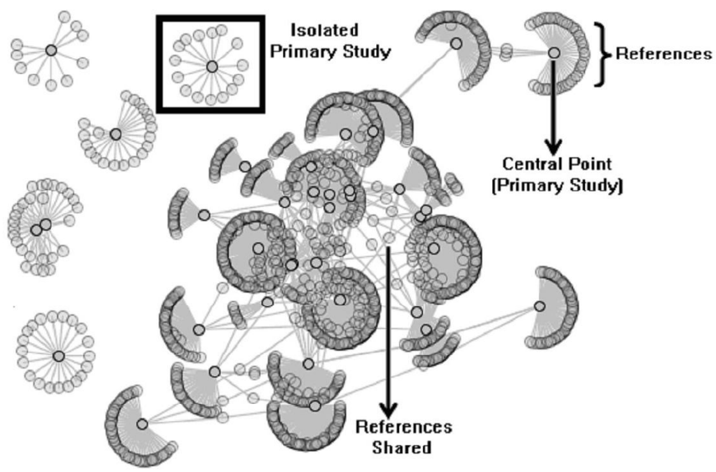

Fig. 2. Citation map: Primary studies that do not share references (isolated primary studies) are disconnected from the other studies in the network (Felizardo et al., 2011).

● **Stage 4: VTM Selection**

During stage 4 the primary studies' selection activity takes place. VTM can support the selection activity using different visualization methods (i.e., content and citation maps). Their respective strategies of exploration are detailed next through different examples those used in our previous work (Felizardo et al., 2011).

There are at least two VTM techniques (see Figure 3*) that can be applied to a **content map**:

1) **Expression Occurrence:** This technique changes the colour of each point in the content map in order to represent the frequency of occurrence of specific user-defined expressions in the primary studies. In this case, the colour scale varies from red (i.e., no occurrence) to blue (i.e., many occurrences). A user can then prioritize their reading towards documents coloured in blue (or closer to this colour) in order to consider whether these documents should be included in the SLR. Conversely, a user can read the documents coloured in red (or closer to this colour) to determine whether those documents should indeed be excluded from the SLR (assuming, of course, that the user-defined expression is relevant to the inclusion/exclusion decision). Figure 3(b) illustrates the colouring of a content map using this technique, where the blue points indicate the maximum occurrence of an expression. Probably the studies approaching inclusion (e.g. approaching blue colour) need most attention, as they are on the cusp of being relevant. Those that are red are most likely to be irrelevant (in terms of the expression) so could be given less attention.

---

*In general, visualization techniques employ colour in order to add extra information on a visual representation. We suggest the reading of a colour print version of this paper, where possible.

2) **Clusters and Topics:** One strategy to classify primary studies is to identify the regions (clusters) of documents with similar content in terms of their titles, abstracts and keywords. Using this technique, clusters are created automatically followed by the formation of their associated topics. These topics are labels representing the content of the documents contained within clusters. In order to efficiently include groups of primary studies, a user can focus their reading on documents belonging to clusters labeled with the topics that match most closely the SLR's research questions. Similarly, in order to exclude studies, a user can review (perhaps less thoroughly) the documents belonging to clusters labeled with topics that do not match their SLR's research questions. Figure 3(c) shows a content map to which the clusters and topics technique has been applied to. The topics appear inside boxes.

The **citation map** offers important additional information beyond that related to the initial set of documents, in particular the primary studies' references and the connections between papers via the set of references that they share. Reference lists from relevant primary studies could be other sources of evidence to be searched (Kitchenham and Charters, 2007). Hence, papers that share references with a relevant paper could be more appropriate for inclusion in the SLR. On the other hand, primary studies that are not connected to any other studies (i.e., do not share citations or references, referred to as isolated primary studies), are more likely to be irrelevant documents in terms of a research question, and may therefore be more readily excluded from the SLR.

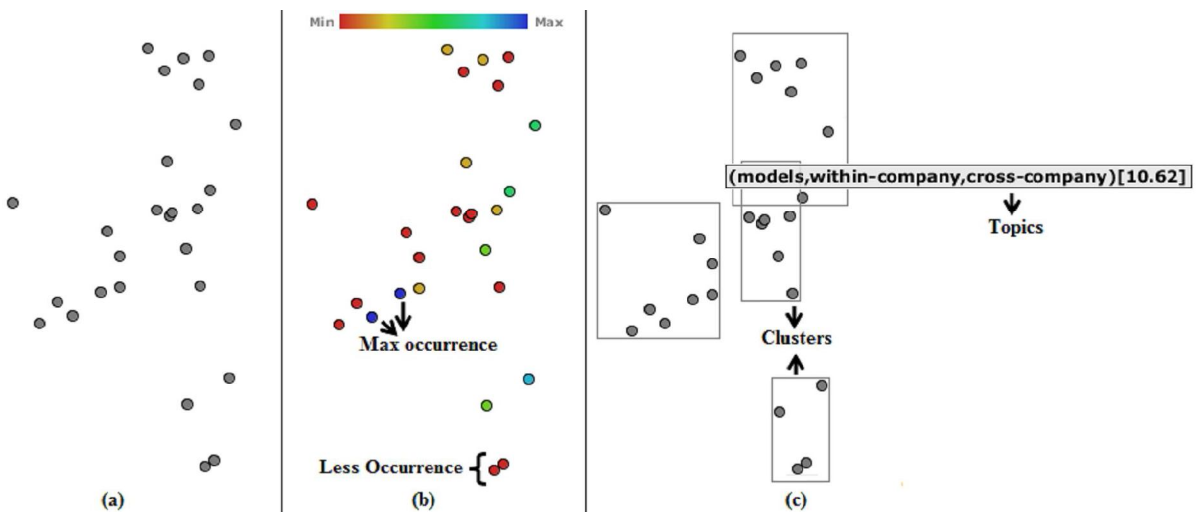

Fig. 3. The two different VTM techniques (b and c) that can be applied to the content map (a).

The strategies just mentioned can be applied iteratively in order to adequately seek primary studies, as depicted in the various visual representations. The number of iterations is determined by the user and should continue until all primary studies have been considered. Moreover, the user can combine the techniques using **coordination**, which represents an interaction among the two different views (i.e., content and citation maps). Using coordination, once a point or a group of documents in a view is selected, the corresponding point (or points) is then highlighted in the other view. Figure 4 illustrates the coordination of the two views. In Figure 4(a) – content map, the documents numbered as 1, 2 and 3 are similar in terms of content. These same documents also share their references (see Figure 4(b) – citation map). Hence, the user has additional information about these documents that could support their decision on whether to include the studies in the SLR. Thus, if document 1, for example, is considered appropriate to be included in the SLR, then it is likely that documents 2 and 3 are also to be included. Various other combinations can be used all aiming at supporting inclusion/exclusion decisions and users should employ the views to obtain additional information in relation to the document set or to individual studies. Note that at any point a user can refer to the abstract/full text of any study. At the end of this stage all primary studies are classified as included or excluded.

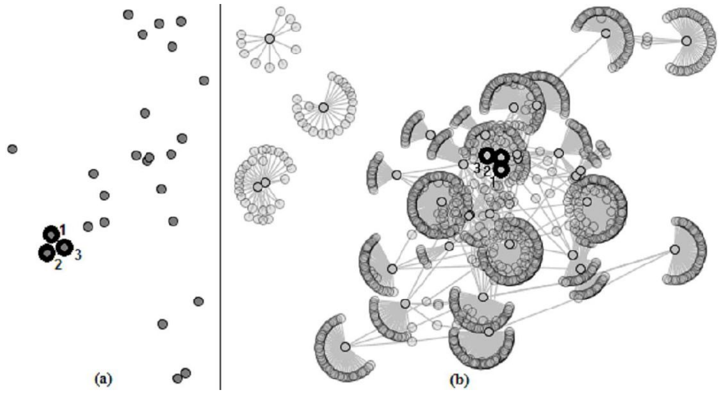

Fig. 4. Coordination between the two visualizations: (a) content map and (b) citation map (Felizardo et al., 2011).

● **Stage 5: VTM Selection Review**

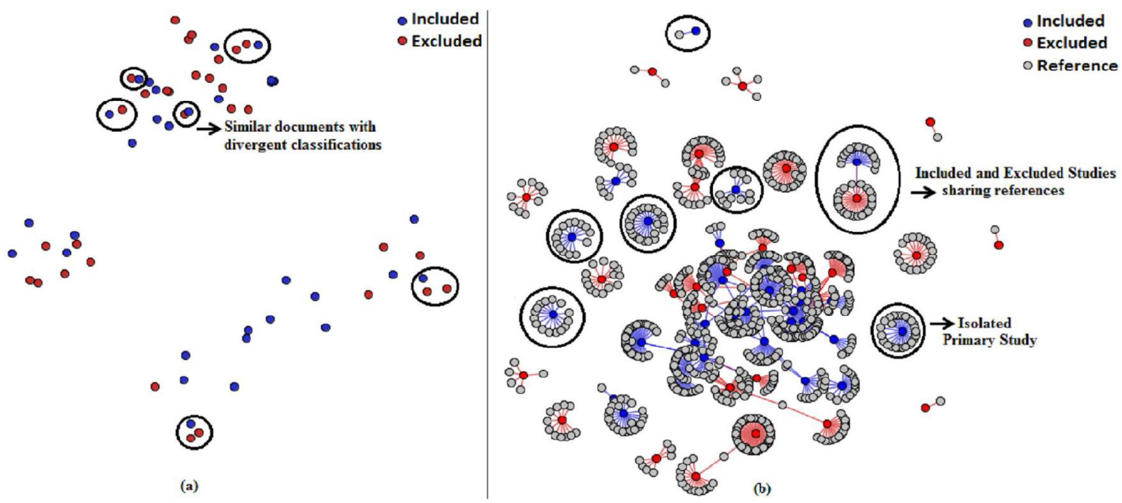

Fig. 5. The highlighted points in the content map (a) and citation map (b) are candidates to be reviewed.

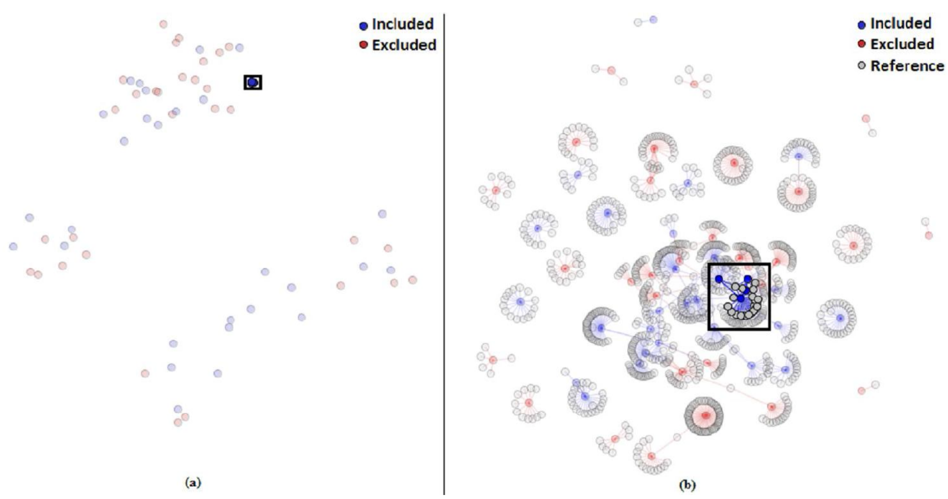

Fig. 6. Coordination between the content map (a) and the citation map (b). The same documents are highlighted on both maps, showing documents that are similar in content and also share citations between themselves, a strong indication that they should receive the same decision.

The strategy suggested here to review the studies' selection is to analyze the content map in terms of the included and excluded documents to find inconsistencies, i.e., documents with similar content (meaning that they are positioned close together on the map) with different classifications: included and excluded. These cases are hints to the reviewer, and the studies should be reviewed following the traditional method (reading the full text). It is important to be clear that finding papers that are similar with respect to their content (i.e., title, abstract and keywords) with some included and some excluded is not necessarily a sign of a problematic selection. It can occur if authors write several different papers about the same study (e.g. a conference paper followed by a more complete journal paper) and the reviewers decide to include only the most recent version. To avoid this scenario we recommend to include in the content map only the most recent version of papers about the same study, in other words, to create maps without "duplicate" papers (two or more papers about the same study). In addition, this adheres to the SLR convention that only one of multiple primary studies relating to the same underlying analysis should be included.

In a citation map, the vertices are the studies, and an edge indicates if a study cites another one. Employing this representation it is possible to identify, for instance, studies that are not connected to any other, that is, that do not share citations. These studies, which are isolated in terms of references, deserve attention from experts (reviewers) if they are included in the review. Another scenario which needs attention occurs when an excluded study, sharing citations with included studies, is not included in the review. In this case, it can indicate that important studies may be missing in the SLR.

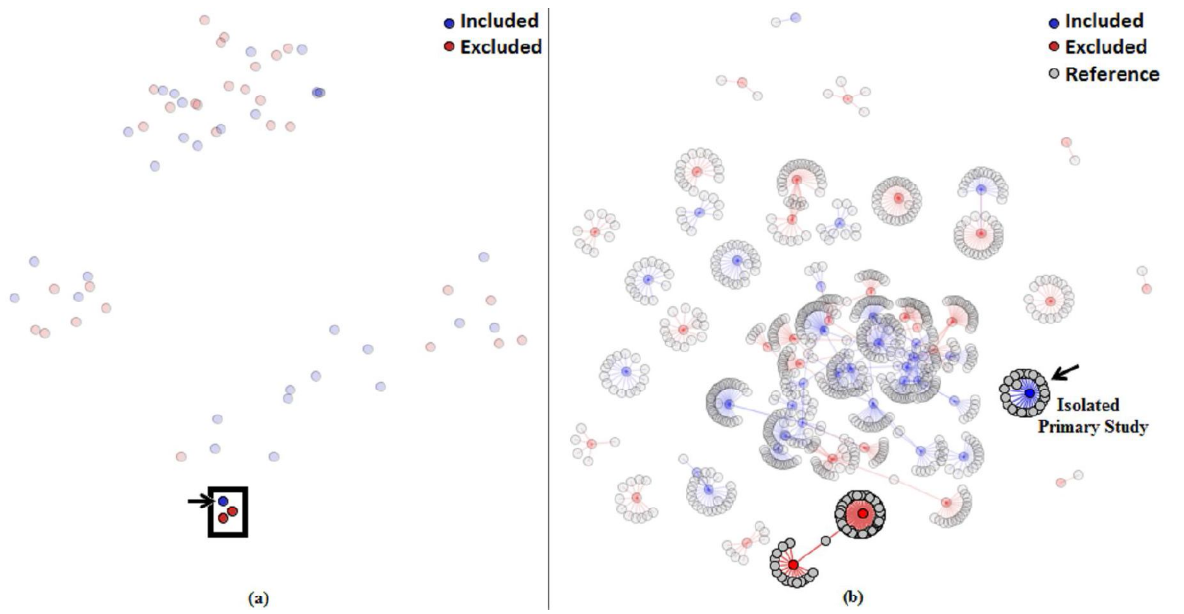

Fig. 7. Coordination between the content map (a) and the citation map (b). The included primary study highlighted in the citation map is an isolated point in terms of references, a strong indication that it could in fact be a candidate for exclusion, and should at least be given further attention.

Figures 5, 6 and 7 present content and citation maps composed of 63 primary studies of an unpublished SLR on SE for computer games, conducted by a collaborator of our research group (a master's student without experience in conducting SLRs). The SLR was conducted using the traditional strategy to select papers i.e., reading abstracts and full text. The colours identify the two possible classes of studies: red points are studies excluded from the review, and blue points represent included studies. Figure 5(a) illustrates the content map containing included studies (blue points) that are similar in terms of content to other, excluded studies (red points). In the citation map (Figure 5(b)) it is possible to observe one included study that has an excluded study as a reference, and included studies isolated in terms of references. All these situations point to inclusion/exclusion decisions that should be inspected.

The content and citation maps can be analyzed together, by coordinating their use. Figure 6 illustrates the coordination between the content map (a) and the citation map (b), with the same documents highlighted on both. Selected documents (points) remain with their original opacity and the other documents, not selected, become semi-transparent. It is possible to see that the included studies are similar in content (they are closely clustered) and share citations. These properties give support to their similar treatment – in this case, inclusion. Figure 7(a) shows the opposite situation, an included study (blue point) that is similar in content to two excluded studies and that is isolated in terms of references (see Figure 7(b)), an indication that its inclusion should be reviewed.

Revis - Systematic Literature Review Supported by Visual Analytics – is a flexible visualization tool developed by us in our previous work (Felizardo et al., 2011) that enables a user to leverage several VTM capabilities to explore a collection of documents, which are in our case primary studies. Figure 8 shows the Revis tool's main window. Revis takes as input a set of primary studies selected during stage 2 of the SLR process. These studies are organized according to the *bibtex* format, which includes their title, abstract, keywords, and references. Revis then executes the activities performed during stage 3 and presents the content and citation maps for the document set. The main VTM functionalities offered by Revis are described as follows: – it creates the views: content and citation maps; – it applies clustering algorithms in order to create clusters and their respective topics; – it allows changing of visual attributes (colour) of the points in the document map to represent the frequency of occurrence of an expression in the documents, or the status of the document (i.e., blue circles represent the included studies and red the excluded); – it supports coordination between different views, content and citation maps, among others. In the next section we present a case study that further illustrates and validates the utility of the Revis tool and the VTM techniques.

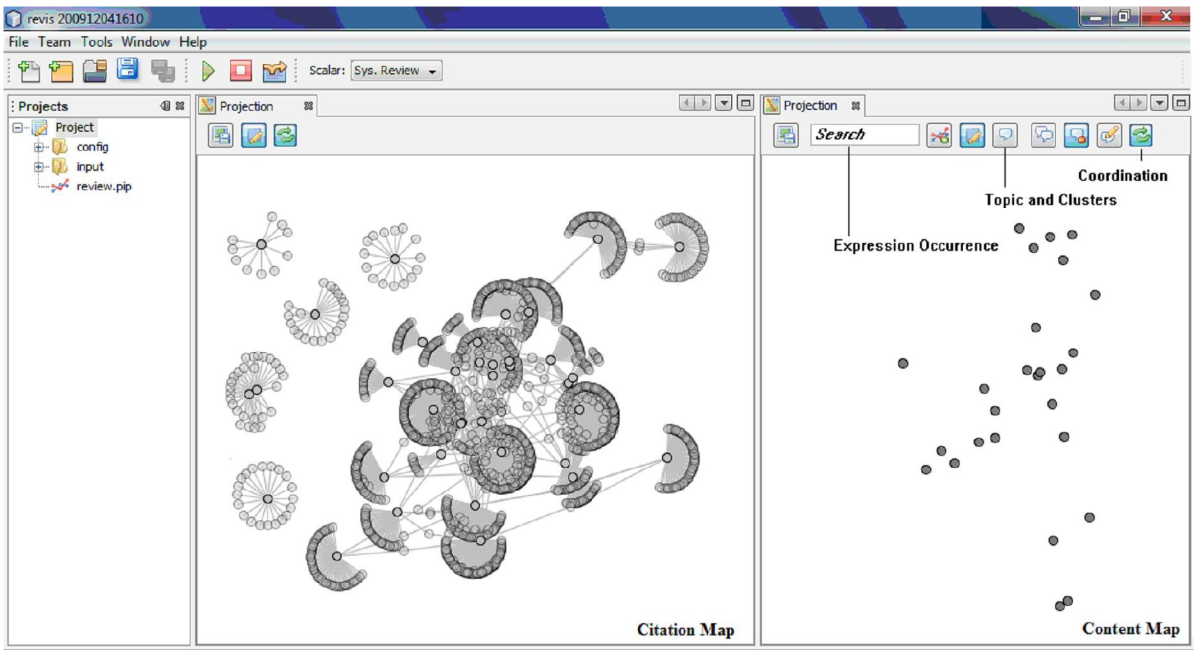

Fig. 8. Main Window of Revis. Adapted from (Felizardo et al., 2011).

4. Case Study: Using VTM in a Real SLR

In order to demonstrate the usefulness of the various VTM techniques a case study was conducted, using a real, previously published SLR on "Comparing Local and Global Software Effort Estimation Models" (MacDonell and Shepperd, 2007). This SLR contains 185 potentially relevant primary studies, including duplicates. In total, 147 papers were excluded and 38 were included (i.e., 28 repeated and 10 distinct).

● Stage 3: Visualization

As described above, stage 3 of the VTM-supported SLR process results in the production of two visual representations of the primary studies – the content and citation maps. Figure 9 shows the results of this stage. It is important to note that the activities undertaken during this stage were completely automated by Revis, which takes only some seconds to present the maps. The other two stages of the process, VTM selection and VTM selection review, which specifically explore VTM in the SLR study selection process, are presented in more detail in the next sections.

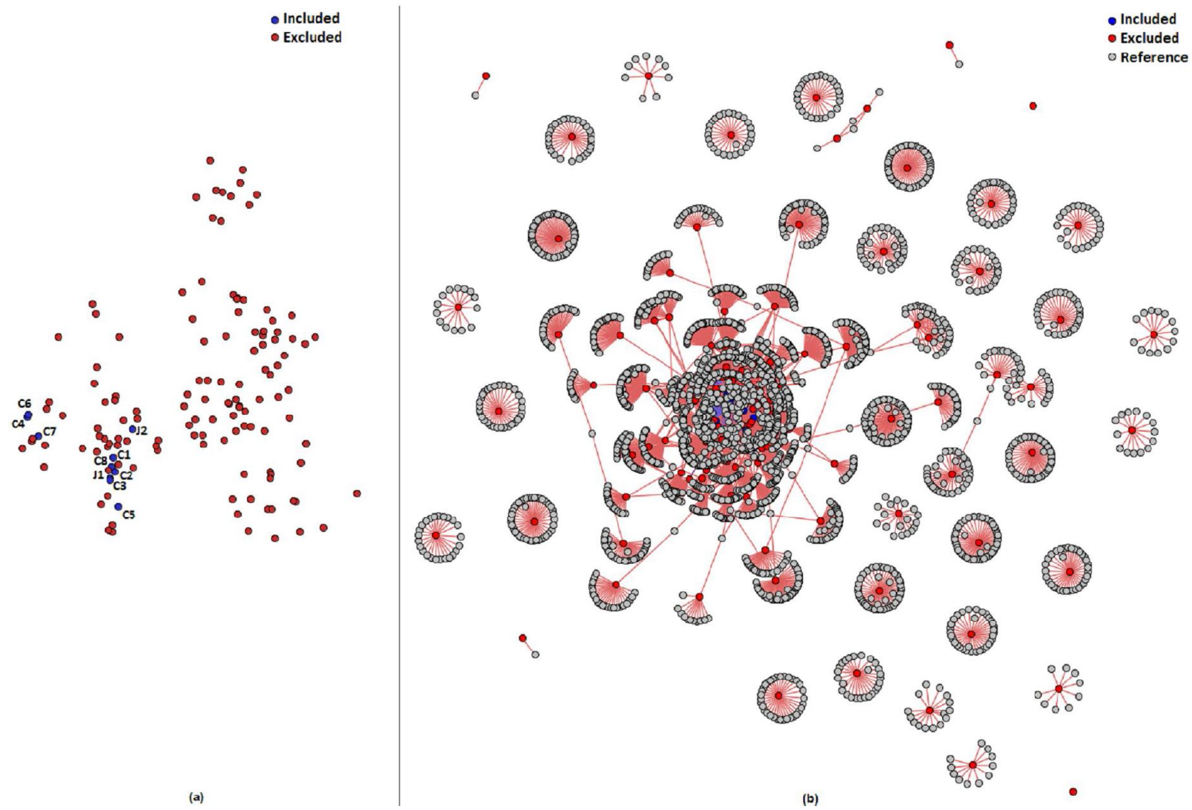

Fig. 9. Example of maps: (a) Content map and (b) Citation map. The colour of the nodes was changed to represent the status of each document i.e. blue circles represent included studies and red the excluded ones, grey circles are cited references.

● **Stage 4: VTM Selection**

During this stage, the selection of primary studies activity was carried out using the associated content map and citation map produced using Revis. The strategies suggested by us to explore the content map, taking into account expression occurrence and clusters and topics, are exemplified next.

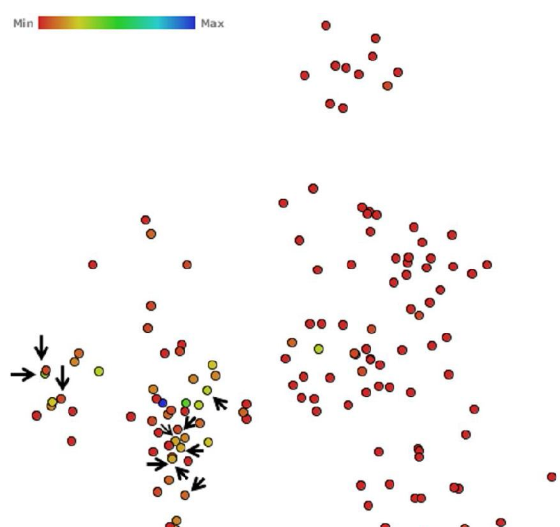

Fig. 10. Content map coloured to represent the frequency of occurrence of the expression “estimation” in the case study SLR. The colour scale varies from red to blue - no occurrence to many occurrences.

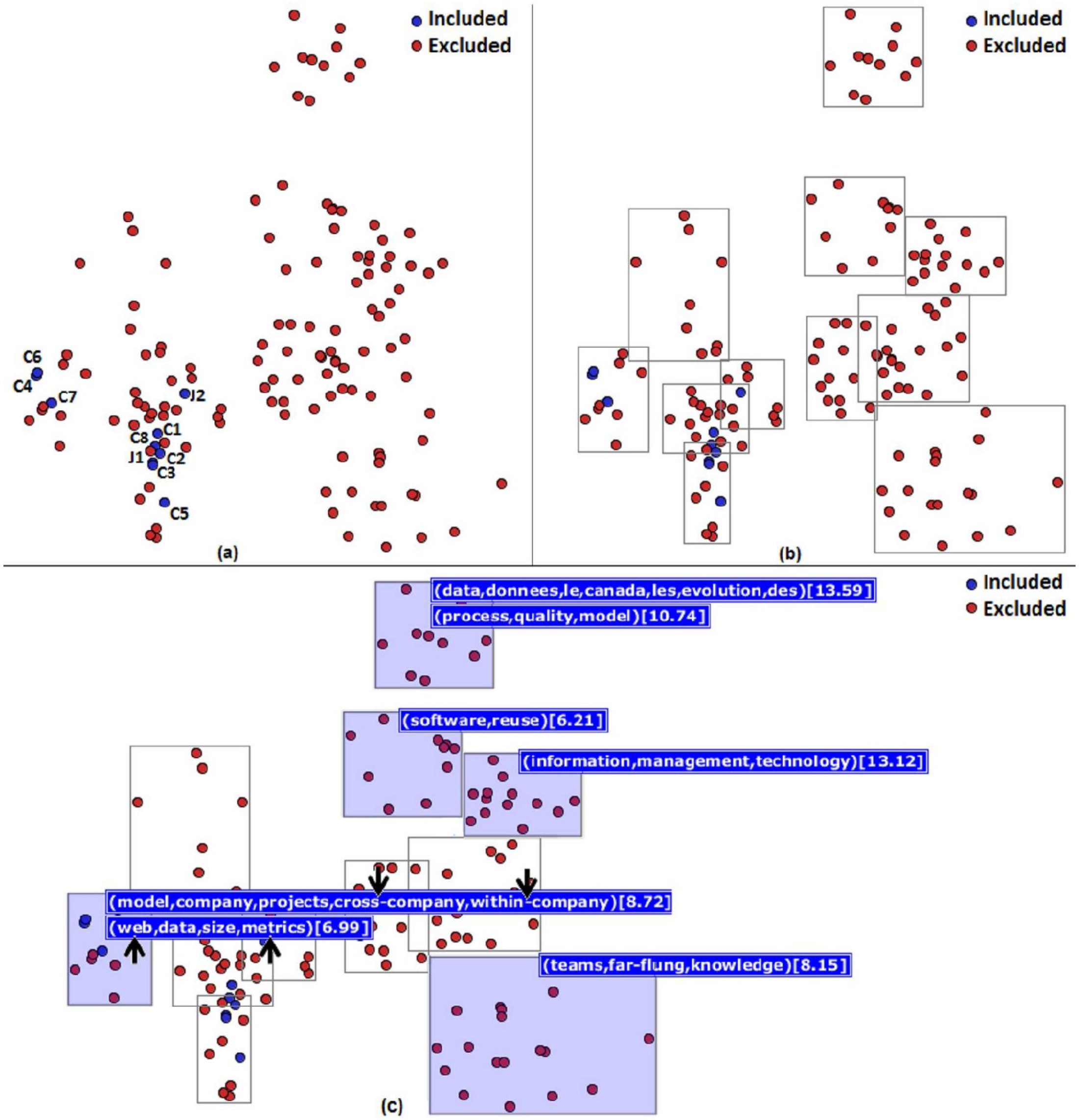

Fig. 11. (a) Content Map. (b) Content map after the application of the k-means clustering algorithm. (c) After the generation of clusters, topics, which are labels that represent the content of the documents contained in the clusters, were created automatically by Revis.

Analyzing the expression occurrence strategy as applied to the content map (see Figure 10), the points were coloured to represent the number of occurrences of the relevant expression "estimation" in the context of the SLR research question (i.e., *What evidence is there that cross-company estimation models are at least as good as within-company estimation models for predicting effort for software projects?*). As mentioned above, we recommend that a researcher could review the documents coloured in red (or closer to this colour) to determine whether those documents should indeed be excluded from the SLR. It is possible to observe in Figure 10 that all included papers (highlighted by arrows) were not coloured in red, while the bulk of the excluded studies (see the right side of the figure) were coloured in red (i.e., there is no occurrence of the expression "estimation" in these documents).

To identify groups of documents with high content similarity, a clustering algorithm can be applied to the content map. The *k-means* algorithm (MacQueen, 1967), which is known for its efficiency in clustering large data sets (Cios, 1998), is one of the classical clustering algorithms available in the Revis tool. Using this algorithm, we provided as input the number of clusters into which the collection of documents should be classified. Based on the sample size used in this study (i.e. 157 distinct papers), the Revis tool suggested that the documents on the content map could be divided into eleven clusters (i.e. the regions that concentrate similar documents), so we initially chose this number, but any other value could be chosen. An example of the content map after the application of the *k-means* algorithm is presented in Figure 11(b). As mentioned above, we recommend that a user should concentrate their reading on documents that belong to clusters labeled with topics that most closely match the SLR's research questions. We can observe in Figure 11(c) that clusters containing included studies had topics representative of the SLR context, such as "web", "metrics", "cross-company" and "within-company". These topics are highlighted in Figure 11(c) by arrows. However, in clusters containing excluded studies, the topics were less representative of the SLR's question, resulting in more general descriptive terms such as "data", "process", "software", "information" and "teams" (see Figure 11(c)).

The citation map visualization presented in Figure 12 shows that all 10 included papers (located in the middle of the map) share the same references, confirming that papers sharing references with a relevant paper could be more appropriate for inclusion in the SLR. All isolated studies (i.e., studies that do not share references) are classified as excluded.

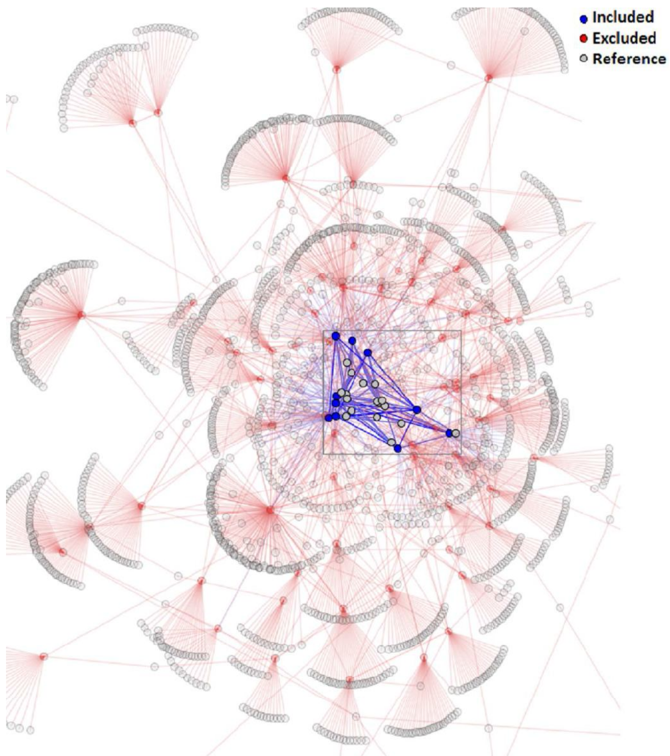

Fig. 12. Citation map: the 10 distinct papers included in the SLR are highlighted in the map and it is evident from this representation that they share citations among themselves.

● Stage 5: VTM Selection Review

During this stage, the selection of primary studies activity is reviewed, to lend reassurance that, in particular, important studies are not missing in the SLR.

Figure 13 illustrates the use of the coordination strategy. It is possible to see that the included studies C4 and C6 are similar in content and have citations between themselves, a clue to support their similar treatment in terms of inclusion or exclusion.

In a subsequent paper (MacDonell et al., 2010) the authors of the SLR noted that they did not include one relevant study in their SLR because this paper had not yet been published, although it was in press. We can use it here to create new content and citation maps (see Figures 14 and 15), to test whether it would indeed be positioned by the VTM techniques near other included studies. The use of VTM to analyze papers published after an SLR has been completed would also allow reviewers to "see" how the new papers are related to existing papers, giving some insights into the stability of the literature around a particular research question.

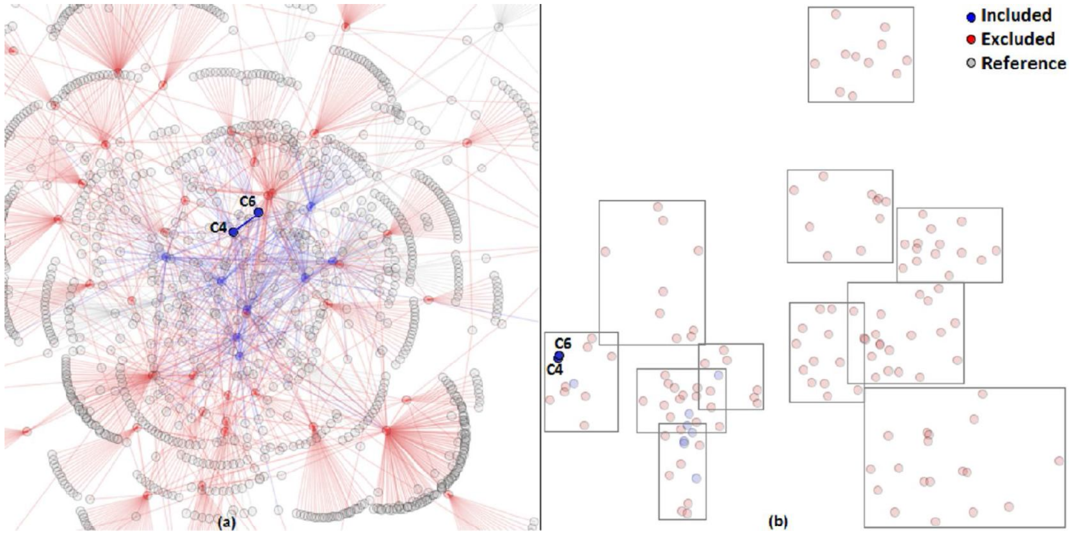

Fig. 13. Coordination between the citation map (a) and the content map (b). The same documents, papers C4 and C6, are highlighted on both maps, showing documents which are similar in content and also share citations between themselves, a strong indication that they should be considered in the same light regarding the inclusion/exclusion decision.

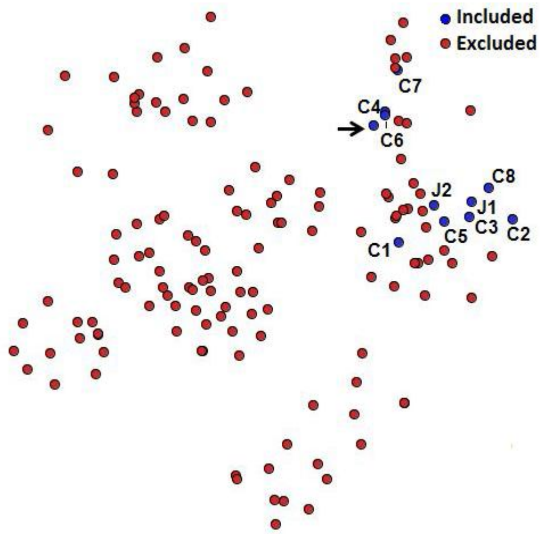

Fig. 14. Content map after the inclusion of the "missed" paper, which is a strong candidate to be considered as an included paper because it is similar in content with other already included papers (C4 and C6).

Figure 14 shows that the "missed" paper (highlighted with an arrow) was allocated in the content map near to other included papers (C4 and C6), a strong indication that it should receive the same decision. Figure 15 shows that the "missed" paper also shares citations only with included papers (blue points), illustrating quite effectively the utility of the VTM representations. In both visualizations and using the coordination between them, the reviewer has clues that the "missed" paper is relevant in the context of the SLR question and that it should be included.

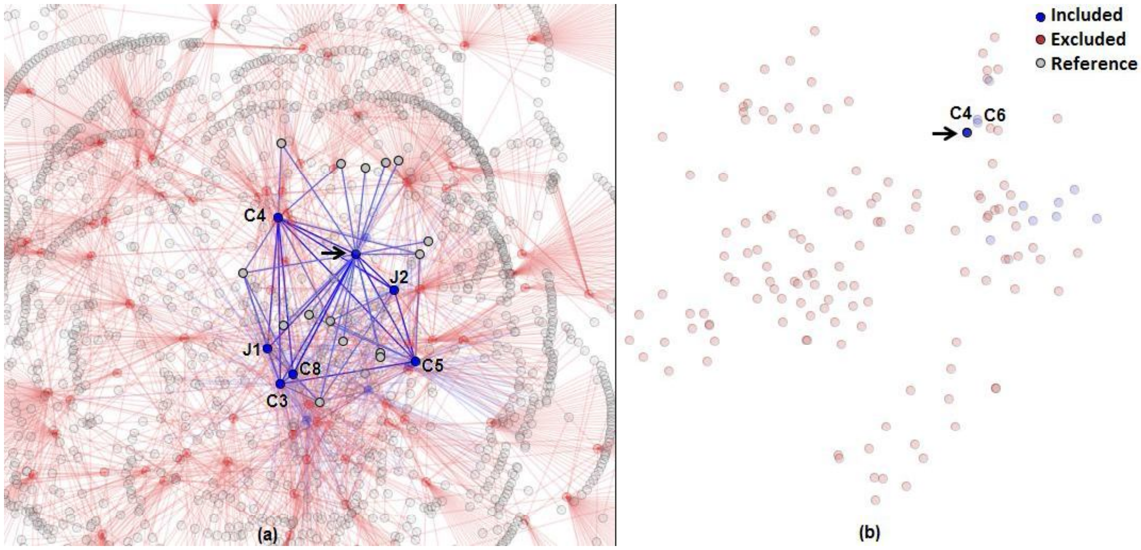

Fig. 15. Coordination between the content map (a) and the citation map (b). The "missed" paper also share citations with other included papers, a strong indication that it should be considered in the same light regarding the inclusion/exclusion decision.

5. Discussion

Our previous paper represented an initiative towards introducing visual techniques in the SLR process in SE. We have presented several VTM techniques, which we believe can support primary studies' selection and selection review activities. In the current paper we present evidence from a case study that confirm our contend that VTM techniques are important because they can help to address the challenges that arise in the exploration of large data sets. Visual data exploration has been used in many applications (e.g. fraud detection and data mining) (Keim, 2002), we are exploring information visualization technology to support improved data selection in SLR context of SE research. The results of our case study support the assertion that VTM techniques facilitate exploration, interpretation, and decision-making in regard to the inclusion or exclusion of primary studies.

Note, however, that it is not our intent to eliminate the traditional method, reading the abstracts and full texts, to select primary studies. Rather, it is our view that exploratory visualization techniques may augment the traditional selection approach, helping users to better understand the primary studies. In particular, reviewers can use the VTM techniques to judge their inclusion and exclusion decisions. In other words, the employed visual representations can be used to complement the decisions made by reviewers, given support to guide the researchers to a consistent treatment regarding inclusions and exclusions. Although visual techniques do not substitute the traditional selection process, our case study illustrates that visualizations may assist not only in the selection activity, but also the review of the selection. In both situations it may support prioritization of effort in of the consideration of primary studies. Using VTM techniques users can explore different visual representations of the primary studies to have additional and complementary information/details that are not readily available directly from reading the study abstracts (e.g. similarity relationships, citations between primary studies).

One of the potential threats to the internal validity of our study relates to the illustration of only case study, consequently, the use of VTM was investigated in a limited context. However, we believe our study is still useful, because the example used is a real SLR. To ensure internal validity, the results of our case study can be compared with other case studies have been conducted by our research group to demonstrate and validate the use of VTM in the research domain of software testing. This has also enabled us to apply VTM to SLRs of varying sizes (in terms of the number of primary studies considered), for example: (i) SLR1: 49 papers on testing distributed software (38 excluded and 11 included); (ii) SLR2: 34 papers on agile testing (16 excluded and 18 included); (iii) SLR3: 37 papers on software component testing (14 excluded and 23 included) and; (iv) SLR4: 100 papers on model-based testing (51 excluded and 49 included). In summary, the results obtained from the case study presented in this paper support our previous findings and suggest that the VTM approach can lend useful support to the primary study selection and selection review activities of SLRs. Furthermore, the VTM approach suggested here can be used in real SLRs, where a large number of candidate studies are considered – hundreds and even thousands. In fact, according to the VTM specialists, VTM techniques work best when applied to a large number of articles (Malheiros et al., 2007).

6. Conclusions and Further Work

In conclusion, the work presented herein extends our previous where we have proposed an extension of the traditional SLR process, comprising a novel and interactive visual approach that seeks to help reviewers to comprehend primary studies analyzed in SLRs. In doing so we are making use of specific VTM techniques (e.g. content map, citation map and coordination) to ensure that the SLR process also has more automated support. The key contribution of our paper is to illustrate the use of VTM to facilitate the execution of SLRs in the SE context. The case study showed that different VTM techniques may be combined to support reviewers in the selection and selection review activities, enabling them to (re-) evaluate their decisions regarding the inclusion or exclusion of primary studies from an SLR and, in particular, to help reviewers to ensure that important studies are not missed. In addition, the findings of our case study showed that the use of VTM to extend the traditional SLR approach provides additional and complementary information that is not readily available directly from reading study abstracts or full text e.g. similarity relationships, shared citations among and between primary studies. Given the apparent benefits of using VTM, we suggest that it would be useful as an extension within the SLR process, to assist researchers in selecting and verifying the results of their primary study selection.

The key disadvantage of introducing VTM in the SLR process is the additional knowledge required – it would be beneficial for the reviewer to have a degree of understanding of text mining and visualization, and they will need to become familiar with the Revis tool if used. Another current limitation is that the documents (papers) under analysis need to be in *bibtex* format in order to be loaded into Revis. Thereby, if the studies are in any other format (e.g. PDF), it is necessary to convert them prior to the analysis. However, many electronic databases (e.g. ACM Digital Library; IEEE Xplore; Web of Science; Scopus and Springer Link) have the facility to export citations from datasets to the *bibtex* format, hopefully making this process as automated as possible.

We believe that an interesting direction for future research relates to applying our process, based on VTM, to different topics and in other domains of interest (e.g. medicine domain), and to do so using researchers with different levels of experience in conducting SLRs (e.g., master's students, PhD students and experts in SLRs).

Acknowledgments
This research is supported by Brazilian funding agencies: CNPq (Processes n. 141972/2008-4; 201622/2009-2) and CAPES.

References

[1] B.A. Kitchenham, "Procedures for Performing Systematic Reviews". Tech. Rep., Keele University and NICTA, 2004.

[2] B.A. Kitchenham, O.P. Brereton, D. Budgen, M. Turner, J. Bailey and S. Linkman, "Systematic literature reviews in software engineering – A systematic literature review", Information and Software Technology, vol. 51, no. 1, pp. 7–15, 2009.

[3] B.A. Kitchenham, O.P. Brereton, S. Owen, J. Butcher and C. Jefferies, "Length and readability of structured software engineering abstracts", IET Software, vol. 2, no. 1, pp. 37–45, 2008.

[4] B.A. Kitchenham, R. Pretorius, D. Budgen, O. P. Brereton, M. Turner, M. Niazi and S. Linkman, "Systematic literature reviews in software engineering – A tertiary study", Information and Software Technology, vol. 52, no. 8, pp. 792–805, 2010.

[5] B.A. Kitchenham and S. Charters, "Guidelines for Performing Systematic Literature Reviews in Software Engineering". Tech. Rep. EBSE 2007-001, Keele University and Durham University Joint Report, 2007.

[6] C. Chen, F. Ibekwe-SanJuan, J. Hou, J. "The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis". Journal of the American Society for Information Science and Technology, vol. 61, no 7, pp. 1386–1409, 2010.

[7] D.A. Keim, "Information Visualization and Visual Data Mining", IEEE Transactions on Visualization and Computer Graphics, vol. 8, no. 1, 8 pages, 2002.

[8] D.A. Keim, F. Mansmann, J. Schneidewind and H. Ziegler, "Challenges in visual data analysis", Proc. Information Visualization (IV' 06), Washington, DC, USA, IEEE Computer Society, p. 9–16, 2006.

[9] D.S. Cruzes and T, Dybå, "Synthesizing evidence in software engineering research", Proc. ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '10). ACM, New York, 2010, 10 pages.

[10] E. Tufte. "The visual display of quantitative information", Graphics Press, Cheshire, Connecticut, 1983.

[11] F.Q.B. da Silva, A.L.M. Santos, S. Soares, A.C.C. França and C.V.F. Monteiro, "Six Years of Systematic Literature Reviews in Software Engineering: an Extended Tertiary Study", Proc. International Conference on Software (ICSE' 10), IEEE Computer Society, Cape Town, South Africa, 10 pages, 2010.

[12] F.V. Paulovich and R. Minghim, "HiPP: A Novel Hierarchical Point Placement Strategy and its Application to the Exploration of Document Collections", IEEE Transactions on Visualization and Computer Graphics, vol. 14, no. 4, pp. 1229–1236, 2008.

[13] F.V. Paulovich and R. Minghim, "Text Map Explorer: a Tool to Create and Explore Document Maps", Proc. International Conference on Information Visualisation (IV' 06), London, UK, pp. 245–251, 2006.

[14] J.B. MacQueen, "Some Methods for Classification and Analysis of MultiVariate Observations", Proc. Berkeley Symposium on Mathematical Statistics and Probability (University of California Press, Statistical Laboratory of the University of California, Berkeley, pp. 281–297, 1967.

[15] J.J. Thomas, K.A. Cook, "Illuminating the path: The research and development agenda for visual analytics", National Visualization and Analytics Center, 190 pages, 2005.

[16] K. Cios, W. Pedrycz and R.W. Swiniarski, "Data Mining Methods for Knowledge Discovery", Kluwer Academic Publishers, 1998.

[17] K.E. Emam, E.Jonker, M. Sampson, K. Krleza-Jeric and A.Neisa, "The Use of Electronic Data Capture Tools in Clinical Trials: Web-Survey of 259 Canadian Trials", Journal of Medical Internet Research, vol. 11, no. 1, 8 pages, 2009.

[18] K.R. Felizardo, E.Y. Nakwgawa, D. Feitosa, R. Minghim and J.C. Maldonado, "An Approach Based on Visual Text Mining to Support Categorization and Classification in the Systematic Mapping", Proc. International Conference on Evaluation and Assessment in Software Engineering (EASE '10) BCS-eWiC, Keele University, UK, 10 pages, 2010.

[19] K.R. Felizardo, N. Salleh, R.M. Martins, S.G. MacDonell, J.C. Maldonado, "Using visual text mining to support the study selection activity in systematic literature reviews", Proc. International Software Engineering and Measurement (ESEM' 11), Canada, ACM, p.1-10, 2011.

[20] M.C.F. Oliveira and H. Levkowitz, "From Visual Data Exploration to Visual Data Mining: A Survey", IEEE Transactions on Visualization and Computer Graphics, vol. 9, no. 3, pp. 378–394, 2003.

[21] M. Riaz, N. Sulayman, M. Salleh and E. Mendes, "Experiences Conducting Systematic Reviews from Novices' Perspective", Proc. International Conference on Evaluation and Assessment in Software Engineering (EASE' 10), BCS-eWiC, Keele University, UK, 10 pages, 2010.

[22] M. Staples and M. Niazi, "Systematic Review of Organizational Motivations for Adopting CMM-Based SPI", Information and Software Technology, vol. 50, no. 78, pp. 605–620, 2008.

[23] P.A. Eades, "A Heuristic for Graph Drawing", Congressus Numerantium, vol. 42, no. 1, pp. 149–160, 1984.

[24] P.N. Tan, M. Steinbach and V. Kumar, "Introduction to Data Mining", 1st ed., Addison Wesley, 2005.

[25] P.O. Brereton, B.A. Kitchenham, D. Budgen, M. Turner and M. Khalil, "Lessons from Applying the Systematic Literature Review Process within the Software Engineering Domain", Journal of Systems and Software, vol. 80, no. 1, pp. 571–583, 2007.

[26] R.E. Garcia, M.C.F. de Oliveira, J.C. Maldonado and M.G. Mendonça Neto, "Visual Analysis of Data from Empirical Studies". Proc. International Conference of Distributed Multimedia Systems (DMS' 04), San Francisco, pp. 225–230, 2004.

[27] S. Ananiadou, B. Rea, N. Okazaki, R. Procter and J. Thomas, "Supporting Systematic Reviews Using Text Mining", Social Science Computer Review, vol. 27, no. 4, pp. 509–523, 2009.

[28] S.G. MacDonell and M.J. Shepperd, "Comparing Local and Global Software Effort Estimation Models – Reflections on a Systematic Review", Proc. International Symposium on Empirical Software Engineering and Measurement (ESEM' 07), BCS-eWiC, Keele University, UK, 2007.

[29] S.G. MacDonell, M.J. Shepperd, B.A. Kitchenham and E. Mendes, "How Reliable Are Systematic Reviews in Empirical Software Engineering?", IEEE Transactions on Software Engineering, IEEE Computer Society, vol. 36, no. 5, pp. 676–687, 2010.

[30] T. Dybå, B.A. Kitchenham, and M. Jørgensen, "Evidence-Based Software Engineering for Practitioners", IEEE Software, vol. 22, no. 1, pp. 58–65, 2005.

[31] T. Dybå and T. Dingsøyr, "Strength of Evidence in Systematic Reviews in Software Engineering". Proc. ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM' 08), New York, NY, USA, pp. 178–187, 2008.

[32] T. Dybå, T. Dingsøyr and G.K. Hanssen, "Applying Systematic Reviews to Diverse Study Types: An Experience Report", Proc. International Symposium on Empirical Software Engineering and Measurement (ESEM'07), IEEE Computer Society, Washington, DC, USA, pp. 225–234, 2007.

[33] V. Malheiros, E.N. Höhn, R. Pinho, M. Mendonca and J.C. Maldonado, "A Visual Text Mining Approach for Systematic Reviews", Proc. International Symposium on Empirical Software Engineering and Measurement (ESEM' 07), IEEE Computer Society, Washington, DC, USA, pp. 245–254, 2007.