Information and Software Technology 54 (2012) 1079–1091

Contents lists available at SciVerse ScienceDirect

Information and Software Technology

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

A visual analysis approach to validate the selection review of primary studies in systematic reviews

Katia R. Felizardo *, Gabriel F. Andery, Fernando V. Paulovich, Rosane Minghim, José C. Maldonado

Instituto de Ciências Matemáticas e de Computação (ICMC), Universidade de São Paulo, P.O. Box 668, 13560-970 São Carlos, SP, Brazil

ARTICLE INFO

Article history:
Received 30 May 2011
Received in revised form 10 March 2012
Accepted 24 April 2012
Available online 7 May 2012

Keywords:
Systematic Literature Review (SLR)
Visual Text Mining (VTM)
Information visualization
Content document map
Citation document map

ABSTRACT

Context: Systematic Literature Reviews (SLRs) are an important component to identify and aggregate research evidence from different empirical studies. Despite its relevance, most of the process is conducted manually, implying additional effort when the *Selection Review* task is performed and leading to reading all studies under analysis more than once.

Objective: We propose an approach based on Visual Text Mining (VTM) techniques to assist the *Selection Review* task in SLR. It is implemented into a VTM tool (Revis), which is freely available for use.

Method: We have selected and implemented appropriate visualization techniques into our approach and validated and demonstrated its usefulness in performing real SLRs.

Results: The results have shown that employment of VTM techniques can successfully assist in the *Selection Review* task, speeding up the entire SLR process in comparison to the conventional approach.

Conclusions: VTM techniques are valuable tools to be used in the context of selecting studies in the SLR process, prone to speed up some stages of SLRs.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

In general, Evidence-Based Software Engineering (EBSE) is related to research methods used to build a body of knowledge about when, how, and in what context methods or tools are more appropriate to be used for Software Engineering (SE) practices. In this context, Systematic Literature Reviews (SLRs), which have been attracting the attention of the SE community, provide a comprehensive and methodical evaluation of research using a predefined strategy of search, extraction and aggregation aiming to maximize coverage and minimize bias [22,25].

The process of conducting an SLR, adapted for SE, was suggested by Kitchenham [22] and involves three phases: *Planning*, *Execution* and *Result Analysis*. These phases and their respective stages and tasks are summarized in Fig. 1.

In the *Planning* phase, the need for a new SLR is identified, the research objectives are defined and the protocol is created, including items such as sources selection, search methods and keywords, inclusion, exclusion and quality criteria of primary studies [22]. Controlled experiments, case studies and surveys are examples of primary studies which compound the information source of SLRs. These empirical studies are then grouped and summarized by SLRs, composing the secondary studies [22,36].

The *Execution* phase is divided into two stages:

1. The *Selection Execution* is the acquisition and analysis of primary studies and involves three tasks: (i) in the *Initial Selection* the studies are identified, collected and organized in a list; (ii) in the *Studies Evaluation*, studies are marked as included or excluded in two steps, first by reading the title, abstract and conclusions, and then by reading the full text. If necessary, reviewers perform a third step based on a quality criterion; (iii) in the *Selection Review*, focus of this work, the studies list should be reviewed to ensure that relevant studies are not initially eliminated.
2. During the *Information Extraction* stage relevant information is extracted from the studies identified as included.

Finally, in the *Result Analysis* phase the results of the primary studies that meet the SLR purpose are summarized. This synthesis can be descriptive, but a quantitative summary obtained through a statistical calculation can complement the description [22].

As previously mentioned, the *Selection Review* task seeks to prevent the exclusion of relevant studies; it may also reset a particular study as excluded. This step is very important, since excluding information that should be evaluated and synthesized in the *Result Analysis* phase may impair the whole SLR process.

* Corresponding author.
E-mail addresses: katiarf@icmc.usp.br (K.R. Felizardo), gfandery@icmc.usp.br (G.F. Andery), paulovich@icmc.usp.br (F.V. Paulovich), rminghim@icmc.usp.br (R. Minghim), jcmaldon@icmc.usp.br (J.C. Maldonado).

0950-5849/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.infsof.2012.04.003

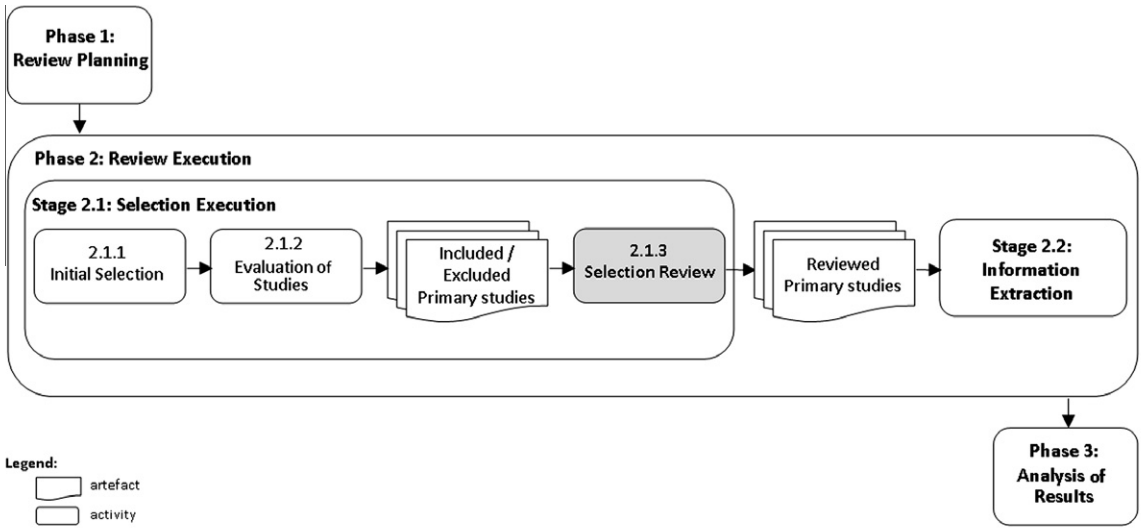


Fig. 1. SLR Process. Adapted from [22], highlighting the *Selection Review* task, which is the focus of this work.

When the selection is performed by two or more reviewers, uncertainties about the inclusion or exclusion should be investigated by sensitivity analysis, which involves repeating the *Evaluation* task in the studies divergently classified by reviewers [25]. When the selection is performed by an individual, the researcher should consider discussing his or her decisions with other researchers or, alternatively, the researcher can re-evaluate a random sample of primary studies to determine the consistency of his or her decisions [25]. Consequently, it implies additional effort to re-read the studies, mainly if more than one reviewer is considered. The task is time-consuming and currently performed with minimal support of text analysis tools. The need to find mechanisms to assist and speed up this task of the SLR process is therefore evident.

A highly successful approach to support tasks involving interpretation of a large amount of textual data suitable to be applied to the *Selection Review* stage is known as Visual Text Mining (VTM) [21,27]. VTM is the integration of Text Mining, which is the process of extracting patterns and non-trivial knowledge from textual documents [38], and Information Visualization techniques, which enable a user to visually interact in the knowledge acquisition process [31].

We believe that the use of VTM can assist the *Selection Review* of primary studies not only speeding up the SLR process, but also improving the reliability of the results. In this sense, the main contribution of this paper is the definition of an approach which employs VTM techniques in the SLR process, specially in the *Selection Review* stage. This approach gives support to resolve divergences more consistently when an SLR is conducted by more than one reviewer. When conducted by only one reviewer, the approach offers clues about what studies should be doubly reviewed for inclusion or exclusion, replacing the random choice strategy. Our results offer evidences of these claims.

The remainder of this paper is organized as follows: Section 2 presents the background and related work on VTM inside and outside of the domain of SLR; Section 3 describes our approach and gives examples to demonstrate its use; Section 4 shows a complete pilot case study, using the approach in a real SLR, and its validation through users' tests; finally, Section 5 draws the conclusions of our work and suggests future developments.

2. Background and related work

This section describes relevant background knowledge and related work for readers to easily understand the proposed approach to be presented and discussed in Section 3. In the first subsection, we will present an overview on the use of SLRs in SE and the second subsection we will discuss the visualization techniques and the SLR process.

2.1. Use of SLRs in software engineering

Three studies [23,26,9] have assessed the impact of SLRs as an EBSE method for aggregating evidence in SE. In 2009, Kitchenham et al. [23] 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. [26] 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. [26] found out that relatively few SLRs provided advice oriented to practitioners.

The results of an independent assessment study conducted by da Silva et al. [9] 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 et al. [23,26]. In addition, they support the findings from the Kitchenham et al. 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, da Silva et al. [9] found that they were poorly conducted in many SRLs. In this same context, Cruzes and Dybå [8] 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 is limited.

They reported that just half of the analyzed studies employ tables (i.e., the simplest type of graphic presentation) to show their findings, and that these tend to contain a lot of data from individual studies (e.g., title, authors, year, outline, strengths, amongst others). Other forms of visual representation were used in fewer than 20% of the studies. The adoption of approaches to aggregating research outcomes to provide a balanced, objective and more readily understood summary of research evidence appears to be an ongoing challenge on SLRs [4,8].

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. [34]. 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 search results are returned. Consequently, this leads to difficulties in reading and evaluating the state of the art of a current topic of interest [34]. Recent studies provide evidence that unstructured and poorly written abstracts can complicate the study selection process [4,12,24,11]. Riaz et al. [34] suggest that one of the causes of selection bias is that titles and unstructured abstracts may not be sufficient as a basis for the initial selection of primary studies in SLRs. It appears 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 unstructured abstracts often omit key information (e.g., background, aim, method, results and conclusions), but also they have been found to include irrelevant information [24]. One solution to minimize the difficulties encountered in SE SLR study selection is to promote the use of structured abstracts [24].

2.2. Visualization and the SLR process

The process used to extract high-level knowledge from low-level data is known as Knowledge Discovery in Databases (KDDs) [21]. Data Mining (DM) is a part of the KDD process responsible for extracting patterns or models from data. Visual data mining (VDM) is a combination of visualization and traditional DM techniques to explore large datasets [21,31]. The visual feedback allows users to have control of the mining task. A specific application of VDM, which is of interest for our work, is the joining of text processing algorithms with interactive visualizations in order to support users to make sense of text collections. In this case, a new area, called Visual Text Mining (VTM) emerges [21,27]. There are several different approaches to handle VTM. Here we employ the most common one, called document map [5]. A document map is a visual representation which aims at supporting the analysis of a number of documents either by similarity of content or through their relationships, such as citation and co-citation referencing, term co-occurrence and many others [5]. The use of document maps exploring text data has been the target of different techniques, as HiPP [33], and commercial systems, such as In-Spire [32] and Infosky [3], showing that it can really speed up the process of interpreting and extracting useful information from document collections of variable sizes.

Several studies have investigated the potential benefits of visualization in supporting the conduct of an SLR. El Emam et al. [15] 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. [1] 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 [24]. Garcia et al. [18] 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).

Three previous studies [16,17,29] have specifically investigated the use of VTM within the context of EBSE. Felizardo et al. [16] have conducted a systematic mapping on the use of VTM to support the conduct of SLRs. The authors reviewed 20 papers and their results indicated a scarcity of research on the use of VTM to help with conducting SLRs in the SE domain. However, most of the studies (16 of the 20 studies included in their mapping) have been conducted in the field of medicine and they revealed that the information extraction stage has more VTM support than other stages. In contrast, previous studies using VTM techniques with SLRs have not employed such techniques during the SLR's planning and analysis of results stages. Felizardo et al. [17] employed VTM to support categorization and classification of studies when carrying out systematic mapping studies; and Malheiros et al. [29] investigated the use of content-based VTM techniques to help with the selection of primary studies, using a feasibility study. The authors compared the reviewers' performances in carrying out the selection of studies reading abstracts or using a VTM tool based on similarity and concluded that VTM techniques can support a more precise selection of relevant studies, speeding up the selection process and also improving its quality. Similarly to the work proposed by Malheiros et al., the approach presented here also makes use of VTM techniques to support the process of SLR. Our proposal focuses on extending that approach to supporting the *review of the selection* task. Additionally, our expansion of the VTM approach for SLR supports both content-based analysis of documents, as in the work of Malheiros et al., and metadata, as citation maps. Therefore, our approach supports a more complete and detailed analysis of the data under investigation. No other paper which employs VTM techniques in the SLRs context has been identified in the literature.

3. Approach to use VTM in the *Selection Review* task

We propose an approach to supporting the *Selection Review* task of primary studies in the SLR process (see Fig. 2). Our approach involves three phases, and phases 1 and 3 are exactly the same previously defined by Kitchenham (2004) and explained in Section 1. We have modified phase 2 (gray boxes in the Fig. 2) incorporating two different VTM approaches to supporting the *Selection Review* task: (i) content map; and (ii) citation map visualizations. In the content map, we seek to give insights on how similar or dissimilar documents are to each other based on their contents. We propose two strategies to explore the content map: (i) exclusion history; and (ii) classification of study quality. In citation map, we aim at showing how documents are related to each other through direct citations or cross-citations. The maps (i.e. content and citation maps) can be analyzed together using the identity coordination mapping strategy. In this section, we described an overview of the approach that we have undertaken in our study and details follow in the next subsections.

The proposed approach is presented in the context of the guidelines suggested by Kitchenham [22], however, the VTM approaches

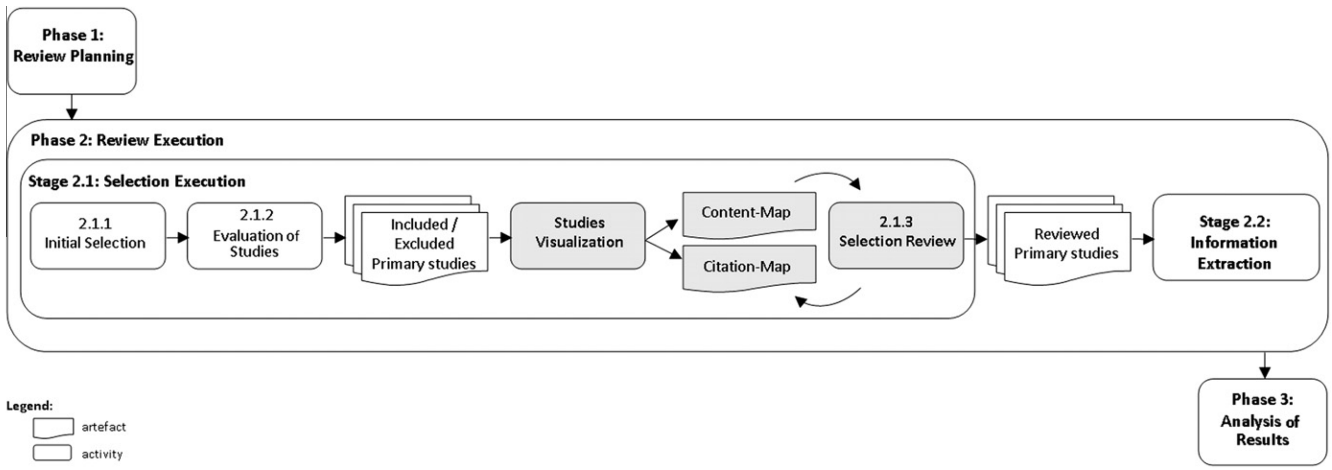


Fig. 2. SLR Process using VTM to support the *Selection Review* task. Adapted from [22].

suggested in our work can be applied isolated or together to SLRs in other computer science fields or even in other scientific disciplines as an activity to validate the selection of primary studies in SLRs. There is not any peculiarity of application of the proposed VTM approaches only to SE field.

In order to create the content maps we used the Revis,[1] which is a visualization and interaction tool that offers a framework of different projection techniques and methods to create data maps in general and document maps in particular based on the similarity between pairs of data points, supporting VTM exploration of document collections. In this work, we extended Revis to include features to generate and interact with citation maps, enabling tasks based on citation and co-citation relationships amongst documents. From now on, when we refer to Revis, in fact, we are referring to this extension.

The following sections present our strategies to apply document-maps (content and citation maps) to support the *Selection Review* task. The examples are given using 4 different real SLRs, numbered 1, 2, 3 and 4, containing papers (primary studies) of the software testing domain:

1. **SLR1:** 49 papers on Testing Distributed Software (38 excluded and 11 included);
2. **SLR2:** 34 papers on Agile Testing (16 excluded and 18 included);
3. **SLR3:** 37 papers on Software Component Testing (14 excluded and 23 included) and;
4. **SLR4:** 100 papers on Model-Based Testing (51 excluded and 49 included).

To present a general view of our approach, we decided to use different SLRs (i.e. several sizes and topics) to better illustrate the benefits and applicability of such approach, exploring the best aspects of VTM techniques in each example. A real and single example will be used in Section 4 from the begin to the end to exemplify the different strategies in a unique context.

3.1. Content map strategies

It is necessary to execute three different steps to create the similarity-based document maps: (1) text preprocessing; (2) similarity calculation; and (3) projection.

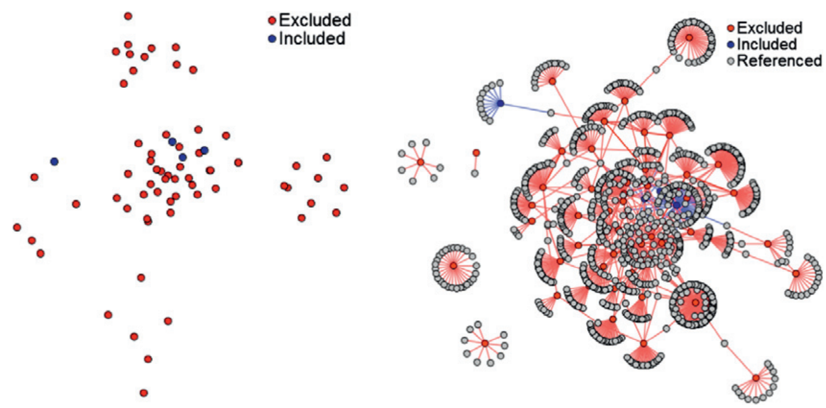
The text preprocessing step is responsible for structuring and cleaning data. It receives the set of primary studies selected in the previous stage as input. In our case, we employ only title, abstract and keywords of an article's content for two reasons: (i) there are many challenges to manipulate full-text articles, for example, the recognition and clean-up of embedded tags, non-ASCII characters, tables and figures, and even the need to convert from PDF into textual format. These issues are not found in title, abstract and keywords [7]; and (ii) Dieste and Padua [10] conducted an experiment to analyze if the strategy of searching titles and abstracts is appropriate for use in SLRs. Their results confirmed that searching titles and abstracts rather than the full text is a better strategy. The preprocessing step converts each document into a vector representation in which the dimensions represent the relevant terms present in the primary studies (known as *bag of words* [35]) and the coordinates are the frequencies of each term in each study weighted according to the *term frequency-inverse document frequency measurement* [19]. This measurement makes the importance (weight) of a word directly proportional to its frequency in each document, and inversely proportional to its frequency in the collection.

After the vector representation is built, the dissimilarity between documents is calculated as the distances between the vectors representing them. Although it is possible to employ any kind of vector-based distance function, it is known that the Minkowsky family of functions, which includes the well-known Euclidean formulation, fails to correctly capture the dissimilarities. In this scenario, a better choice is a function based on the cosine between these vectors [39]. Here we use the common formulation $d(\mathbf{x}_i, \mathbf{x}_j) = 1 - \cos(\mathbf{x}_i, \mathbf{x}_j)$, where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the vectors representing the ith and jth documents, respectively.

The final step to obtain the similarity-based document map is the projection, which maps each document as a point in a 2D or 3D space. This is achieved through point placement or multidimensional projection techniques [33]. Considering that the vector representation $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is embedded into an $\Re^m$ space, a multidimensional projection technique, or simply projection technique, can be viewed as a function that maps each m-dimensional instance into a p-dimensional instance, with $p = \{2, 3\}$, preserving as much as possible the reduced visual space similarity relationships defined in $\Re^m$. In other words, it is a function $f : \Re^m \rightarrow \Re^p$, which seeks to make $|d(\mathbf{x}_i, \mathbf{x}_j) - \hat{d}(f(\mathbf{x}_i), f(\mathbf{x}_j))| \approx 0, \forall \mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$, where $\hat{d}$ is the distance function in the p-dimensional visual space.

The outcome of a projection technique is a two-dimensional or three-dimensional visual representation, which we name content map, where each m-dimensional instance – in our case, a document (primary study) – is mapped on the screen as a graphic element, normally a circle (see Fig. 3a). Documents with similar

[1] Revis is freely available at http://ccsl.icmc.usp.br/redmine/projects/revis/files/



(a) Content map. (b) Citation map.

Fig. 3. Examples of maps, whose visual attributes (color) were changed to represent the status of the document (i.e., blue circles represent documents included in an SLR and red the excluded ones, gray circles are cited references). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

content are meant to be mapped close to each other and dissimilar ones are positioned far apart. Using the content map representation, a user can interactively browse the map, locating groups of similar documents checking the frontiers between different groups, and extracting information that is very difficult to attain by only reading the documents.

Additionally to content-based visualizations, metadata can be used to generate document maps that support other types of association. Examples of these relationship-based visualizations are citation (see Fig. 3b), co-citation and co-authoring [6]. A citation map is built by extracting papers and their references from a *bibtex* file, i.e., citation and co-citation relationships are extracted straight from the primary studies references. The reference extraction activity can be done manually but it is a very tedious and time-consuming task. Databases available online (e.g. ACM or SpringerLink) contain information on publications usually focusing on a certain field, such as the references. There are also reference managers (e.g. Mendeley), useful to read articles and export their references. The graph is displayed using a force-based strategy [13]. This means that studies attract or repel one another depending on how strong their connections (references to each other) are. Through this visualization, it is possible to see citations between papers and common references to other papers (references shared). Papers that do not share references are disconnected from the others in the map. Fig. 4a[2] presents a content map generated using Revis, composed of primary studies of the SLR1 dataset obtained using the traditional strategy to select papers reading abstracts or full texts. The colors identify the two possible classes of studies (included or excluded). Red points are studies excluded from the review and blue points represent the included ones. It takes Revis only a few seconds to create and present a map with a few hundred documents.

The strategy suggested here to review the studies selection from that map can be summarized into two main steps: (1) first, a clustering algorithm is applied to the content map, creating groups of highly related (similar) documents using the k-means algorithm [28] (the number of clusters is automatically detected based on the number of documents); (2) then, the resulting clusters are analyzed in terms of included and excluded documents in order to find inconsistencies. In this analysis, the possible situations a cluster can achieve and the possible consequences for the review process are:

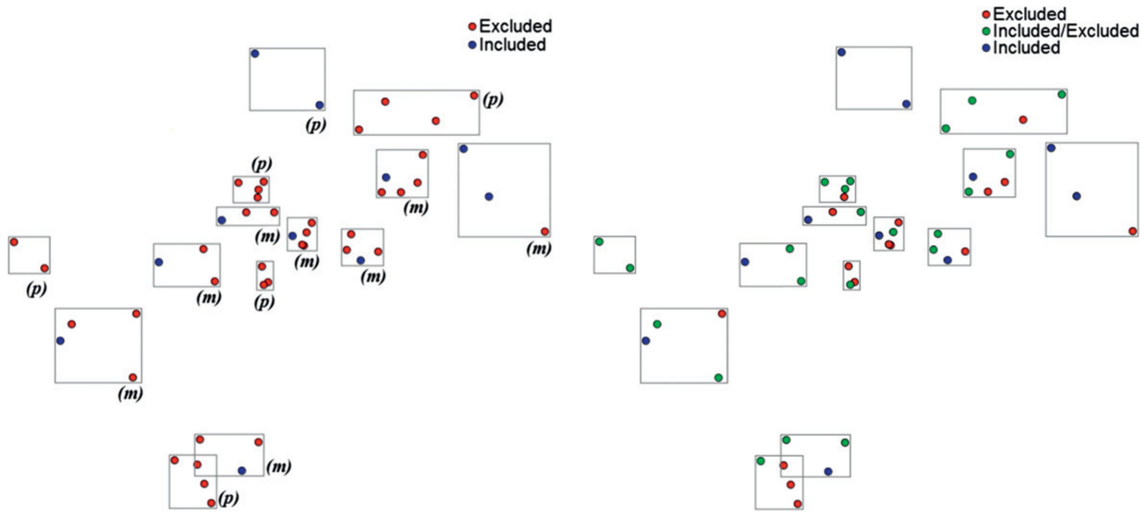
- Situation (a): **Pure Clusters** – all documents belonging to a cluster have the same classification (all included or excluded). Normally, such cases do not need to be reviewed.
- Situation (b): **Mixed Clusters** – there are documents with different classifications in the same cluster. These cases are hints to the reviewer that there are similar documents with different classifications. The studies grouped there should be reviewed following the traditional method.
- Situation (c): **Isolated Points** – there are documents that are not similar to others. These cases are also hints to the reviewer, and the isolated study, if classified as included, must be reviewed.

It is important to be clear that finding papers that are similar with respect to keywords, title and abstract with some included and excluded studies (i.e., mixed clusters) 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 and a journal paper) and decide to include only the most recent ones. Based on this justification, we can assume that similar papers with divergent classifications should be reviewed only if we guarantee that in the content map there is only the most recent versions of papers about each study. Examples of pure clusters are identified in Fig. 4a as ***p***. Mixed clusters are identified as ***m***. The evaluation of these clusters can be refined with the help of content-based strategies, which are described in the following section.

3.1.1. Exclusion history

The first strategy to review the selection activity considers the creation of content maps containing the studies collected and analyzed in an SLR and highlighting them using different colors in order to differentiate in which of the first two steps of the *Evaluation* task a study was removed from the review. In Fig. 4b the same content map of SLR1 shown in Fig. 4a is represented, but colored with the history of the inclusions and exclusions. The 16 red points represent studies excluded from the review in the first step (reading only the abstract); the 22 green points represent the documents included in the first step of selection, but excluded in the second (reading the full article); and the 11 blue points represent the included ones. Thus, green and red points, 38 in total, represent excluded documents and the blue points are the documents included in the final process. Considering the color as history, there are four

[2] In general, visualization techniques employ color in order to add extra information on a visual representation. Therefore we suggest the reading of a color printing version of this paper for fully understanding the pictures.



(a) Without history. (b) With history.

Fig. 4. Content maps related to SLR1. (a) Each point represents a primary study and the color indicates if a study has been included (blue points) or excluded (red points). The squares represent regions (clusters) where the content of the studies (i.e. title, abstracts and keywords) are similar; (b) the color indicates the history of the selection activity, red points are studies excluded in the first step (reading abstracts), green points are studies excluded in the second step (reading full text) and blue points are included studies. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

distinct possibilities for mixed clusters: (i) green and red; (ii) blue and green; (iii) blue and red; and (iv) blue, green and red. Clusters containing green and red studies, although mixed using the history representation, are equivalent to pure clusters (all excluded) and therefore do not need to be reviewed. Clusters containing blue (included) and green or red (excluded) studies represent mixed clusters in both representations, with or without the use of the history information. However, since blue and green points were evaluated in the second step by reading the whole study, only clusters that present blue and red points need to be reviewed. They represent studies excluded in the first step only by reading the abstract, but are similar to included studies entirely read. Finally, clusters containing the three colors should also be reviewed, since they configure the same situation previously mentioned, that is, red points similar to blue points.

It is worth mentioning that without using history information all mixed clusters should be reviewed. In the opinion of experts, the history information is a valuable mechanism to help analyze and support the initial decision on including or excluding a particular primary study. In the specific case of SLR1, the authors provide history information, which did not occur with the other studies used in this paper. These were obtained from publications in the literature.

3.1.2. Classification of study quality

When necessary, researchers perform a third step of selection, based on the quality criteria defined in the protocol. To support this task, we devised a strategy similar to the previous one. However, instead of coloring the studies based on their history, they are now colored according to the quality defined and assigned by the reviewers. Fig. 5 shows the content map of SLR2 colored using this strategy. In this review, the quality assigned by the reviewers seems to be very consistent and can be observed in the group composed of high-quality studies (Q1–Q3) at the top of the content map, which are all classified as included. In this example, an interesting situation is also presented in Fig. 5. An excluded paper (red point) is similar to an included paper of low quality *(Q7 – brown point)*, situation highlighted with a black box. In this case, two scenarios are possible and should be analyzed: (i) the study of low-quality *(Q7)* omitted details and therefore more information should be obtained from the authors of such a study to ensure its inclusion; (ii) the study has enough information and therefore its inclusion is questionable. This example shows that the strategy of coloring the content map based on quality criteria can support the review when combined to the inclusion and exclusion history.

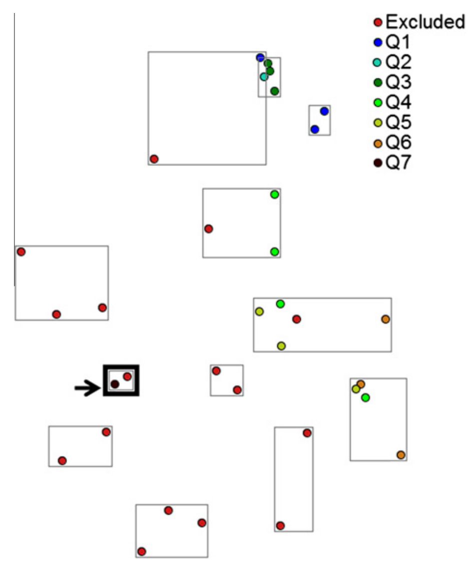


Fig. 5. Content map of SLR2. Red points are excluded studies and the other colors indicate the classification of the included studies according to the quality assigned by a reviewer. Q1 is the highest quality and Q7 the lowest. The highlighted brown point (included study) is a candidate to be reviewed since it is similar (content) to the excluded study and its quality is low. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Citation map strategy

The use of search strings to find relevant primary studies is not sufficient and it must be complemented by scanning reference lists [37]. Skoglund and Runeson [37] described and evaluated a search strategy for SLRs based on reference lists (i.e. references between papers that are used to find relevant papers). The authors believe that the semantic information in references between papers can be used to more efficiently identify the set of primary studies for the SLR.

The use of reference lists to identify relevant papers has been used in information science for several years [37] and it was suggested by Kitchenham's et al. guideline [22] for performing SLRs in SE. Practitioners of EBSE have used reference lists as a part of a search strategy, see e.g. [20].

In our approach, we use references to construct a different type of visualization, called citation map, providing additional support to the analysis of other properties for inclusion or exclusion for primary studies. The citation maps are important in our approach since only the content may not reflect the quality of a study precisely [25].

The most common way to represent citation maps visually is by means of graphs, which are composed of a set of vertices and edges representing objects and the relationships between them, respectively [2]. Here, vertices are studies and an edge indicates if a study cites another one. By employing this representation it is possible to identify, for instance, studies that are not connected to any other, that is, studies 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 requires attention occurs when a highly connected study, sharing citations with included studies, is not selected for inclusion. In this case, important studies may be missing since co-citation is also a valid criterion. In summary, 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. 6a and b shows the citation maps of SLRs 3 and 4, respectively. These citation maps were also generated using Revis. The red points represent the excluded studies, the blue points the included ones, and the gray points are referenced papers, which are not part of the initial set of documents considered. Edges reflect the citations between documents.

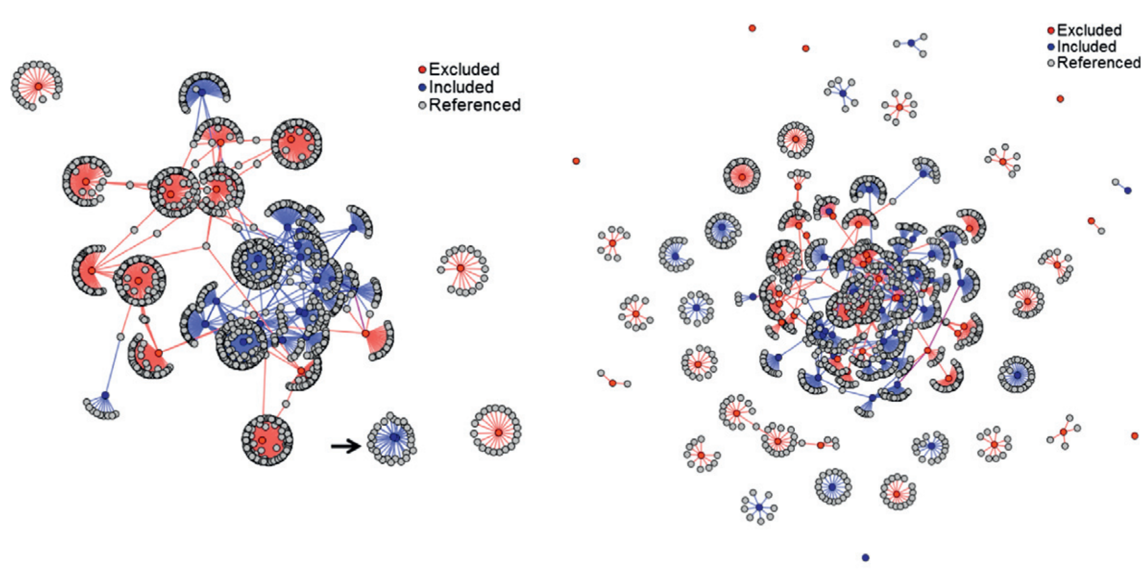
The citation map visualization presented in Fig. 6a shows that most included papers (located in the middle of the map) share the same references. The same is valid for the excluded ones (located on the top-left). The isolated studies (not sharing references) are all classified as excluded. The "blue point" located on the bottom right, if analyzed in detail, is not a single isolated document, but two documents sharing the same references not cited in any other study; it is therefore a candidate to be re-examined in the reviewing process.

A critical situation that should be reviewed is identified in the citation map of SLR4, presented in Fig. 6b. There are some included studies completely isolated from the others. These are the 11 blue points connected only with their own respective references, indicating that, for any reason, they are not really related to any other study under analysis. This scenario should be carefully verified.

3.3. Identity coordination mapping: coordinating content map and citation map

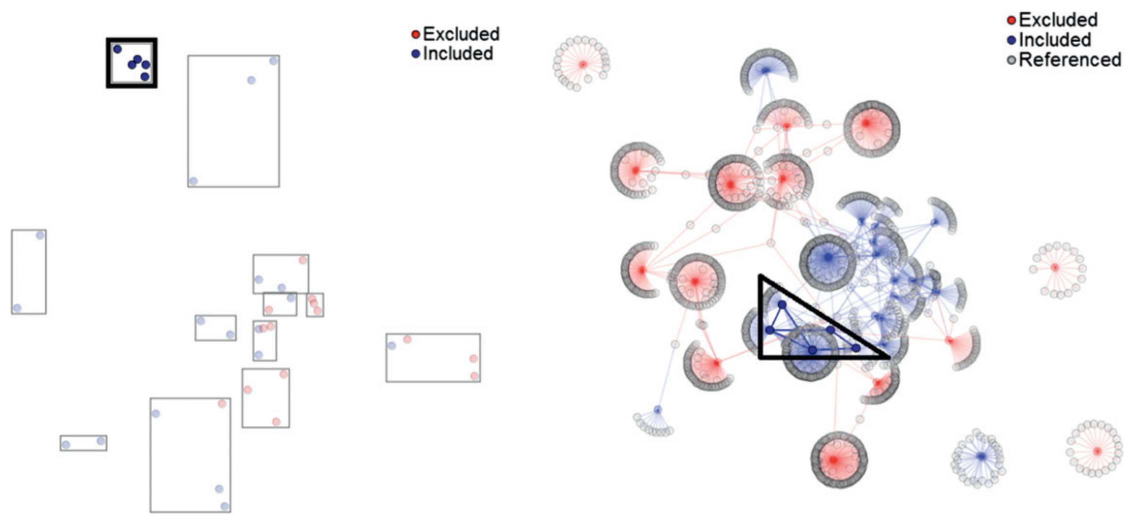
A Revis feature which can be used to support the analysis of content maps and citation maps together is the coordination between them. Links are created between the same documents in different visual representations. When a primary study, or a set of studies, is selected in a visual representation, the same studies are highlighted in the other representation.

Fig. 7 illustrates the coordination of the content map (Fig. 7a) and the citation map (Fig. 7b) of SLR3. Same documents selected in the content map (on the left, above) are highlighted in the citation map (center). Selected documents (points) maintain their original opacity and the other non-selected documents become semi-transparent. It is possible to see that the included studies located in the same cluster are both similar in content and have citations between themselves. This property supports their inclusion.



(a) Systematic review 3. (b) Systematic review 4.

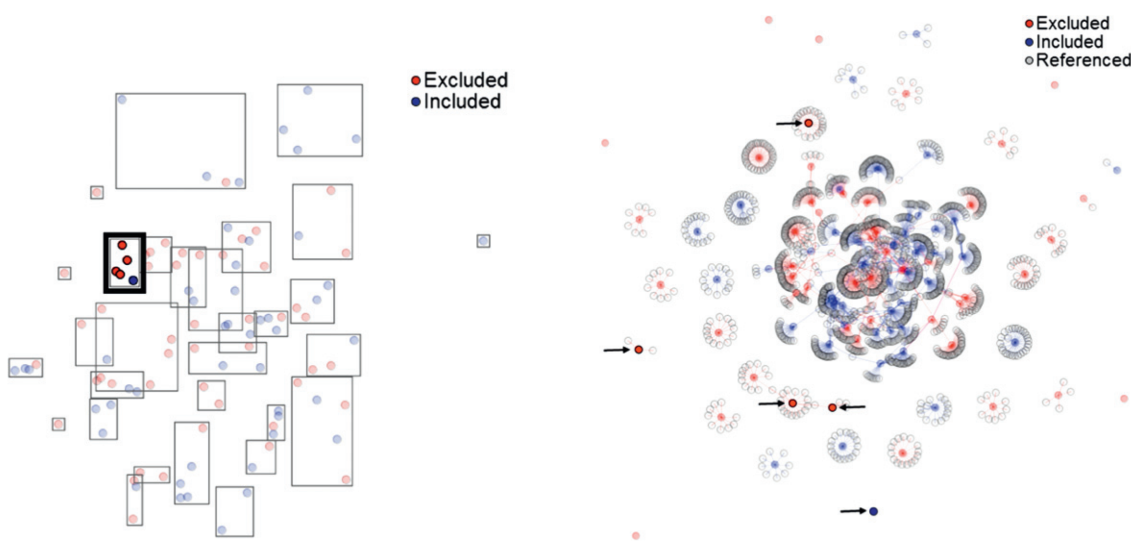
Fig. 6. Citation maps of reviews 3 and 4. The color indicates if a document has been included (blue points), excluded (red points) or cited by at least one of these documents, but not included in the review (gray points). Edges reflect the citations between the primary studies. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) Content map.

(b) Citation map.

Fig. 7. Coordination between the content map (a) and the citation map (b) of SLR3. The same documents are highlighted on both maps, showing documents which are similar in content and also share citations between themselves and strongly indicating they have been properly included.



(a) Content map.

(b) Citation map. Arrows indicate the same documents selected on the content map (a).

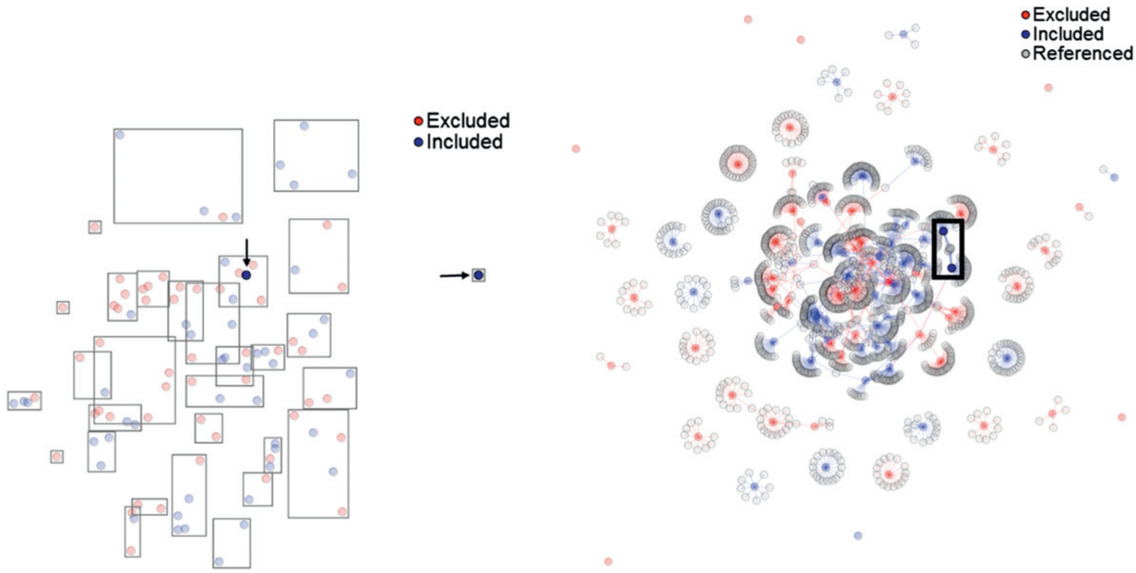
Fig. 8. Coordination between the content map (a) and the citation map (b) of SLR4. The included study (blue point) is similar in content to other excluded studies and isolated in terms of references, indicating that its inclusion should be reviewed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Using coordination it was possible to identify that the paper with no references included in SLR4 (blue point at the bottom of Fig. 6b) is similar to several other studies initially excluded from the review. This scenario is shown in Fig. 8. When selecting the group of studies highlighted in Fig. 8a, which displays the content map of that review, the same studies are highlighted in the citation map, as shown in Fig. 8b. This coordination suggests that the included study should be checked to verify why it has been included, or why similar studies have been excluded.

Fig. 9 presents another example of content-citation coordination in SLR4. This coordination reveals that an included study, which appears very isolated in the content map and is therefore different from any other study (see Fig. 9a, blue point on the right), actually shares many references with another included study (see Fig. 9a, blue point in the middle), suggesting that, although not really similar in content, both documents probably deal with similar issues. These two studies are highlighted on the citation map of Fig. 9b.

4. Case studies: investigating the use of VTM techniques to support the selection review activity

The examples used in Section 3 (i.e. SRLs 1–4) contain a rather small number of primary studies (dozens of articles), however, in real SRLs a large number of candidate studies are considered – hundreds and even thousands. The Revis and VTM approach suggested in our work can be used in SRLs with more articles. In this section, we present a case and a user study to confirm our hypotheses on real datasets.



(a) Content map. (b) Citation map.

Fig. 9. Another coordination between the content map (a) and the citation map (b) of SLR4. Although the included studies do not have similar contents, they share many references, supporting their inclusion in the review.

(a) Content map of SLR5 – Without history. (b) Content map of SLR5 – With history.

Fig. 10. Content maps of the pilot case study SLR. (a) The primary studies in blue are the included ones, and in red the excluded ones. (b) The color indicates the history of the selection activity. Only the red points (studies excluded in the first step – reading abstracts) placed in clusters containing blue points (included studies) need to be reviewed, speeding up the entire process by reducing the number of documents that need to be re-validated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.1. Case study 1: Demonstrating the use of the approach in a real systematic review

In order to demonstrate the use of our approach considering a larger quantity of primary studies, a complete case study, which is an additional SLR on *Software Testing of Concurrent Program* manually conducted by a collaborator (named SLR5) is used. SLR5 contains 261 primary studies, 14 included and 247 excluded. Among the excluded ones, 220 were excluded in the first step, that is, only by reading the title, abstract and conclusions, and 27 were excluded in the second step, by reading the full text. Fig. 10a shows the content map of SLR5. In Fig. 10b the same content map is colored with the history of the inclusions and exclusions. Clusters containing blue and red points are highlighted, helping the selection of the documents that should be reviewed (red points in Fig. 10b). In this case, only 31 primary studies are indicated for review, representing 11.8% of the total documents and potentially speeding up the entire process of selection review.

Fig. 11 shows the content map of SLR5 colored according to the quality of the documents assigned by the reviewers. Blue points represent high-quality studies (nine points), cyan points represent medium-quality studies (three points), green points represent low-quality studies (two points), and red points represent excluded studies. In this example, the quality assigned to the papers do

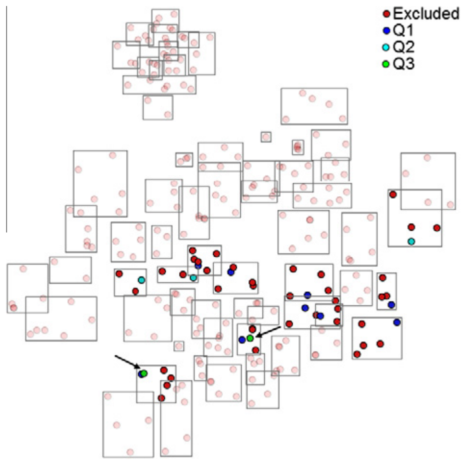


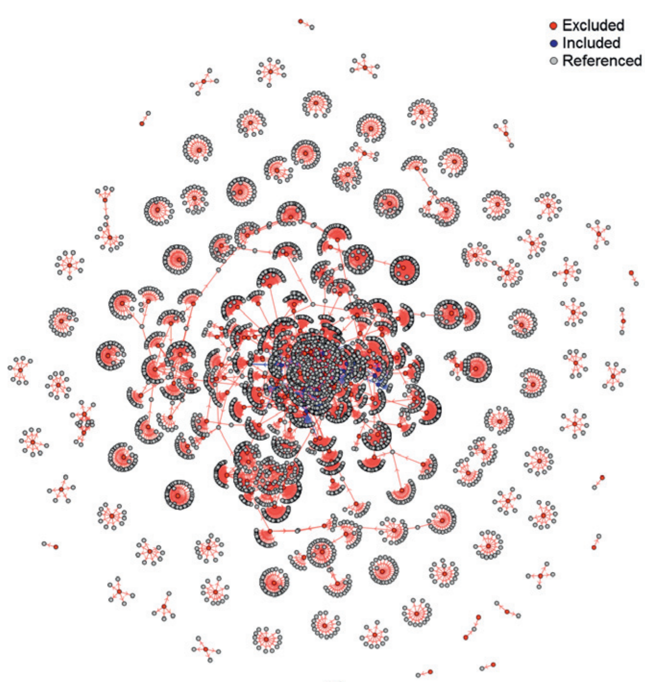
Fig. 11. Content map of the pilot case study SLR. The colors indicate the classification of the studies according to their quality. Two low-quality studies (green points) are similar in content to high-quality studies (blue points), indicating that the quality classification should be re-evaluated. This visual representation gives reviewers support to validate their quality classification, increasing the reliability of the entire process. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

not contribute to supporting the inclusion and exclusion decisions. For instance, there are two low-quality studies, indicated in the figure, which are very similar to two other high-quality studies, suggesting further analysis by the reviewer. This visual representation can help reviewers to make the subjective process of assigning more reliable quality rates.

Fig. 12a shows the citation map of SLR5. All the included papers are located in the middle of the map, therefore most of the included studies have references in common, or cite one another. In the presence of many documents, as in Fig. 12a, the visual representation can be cluttered. There exist several techniques to reduce this clutter [14]. Here we simply create a new citation map (see Fig. 12b) containing only the included studies and their references. It can be observed that there are several references to excluded studies, closer examination shows that some inclusions or exclusions are questionable and should be reviewed.

Fig. 13 shows the coordination of the content and citation maps of this SLR. In the citation map, an included study has two excluded references. In addition, these three documents are similar in content, suggesting that they should be reviewed for inclusion and exclusion since it is not possible to determine if the reviewer has properly included or excluded them from the review. Considering the clues given by the examples we can infer that the studies included and excluded in this SLR should be re-evaluated. The visual representations provide several indications that there are mismatches between the choice of including or excluding the studies and their relevance in the field under analysis. This conclusion is very difficult and much harder to draw using the usual approach to assist the selection review, that is, reading the documents twice.

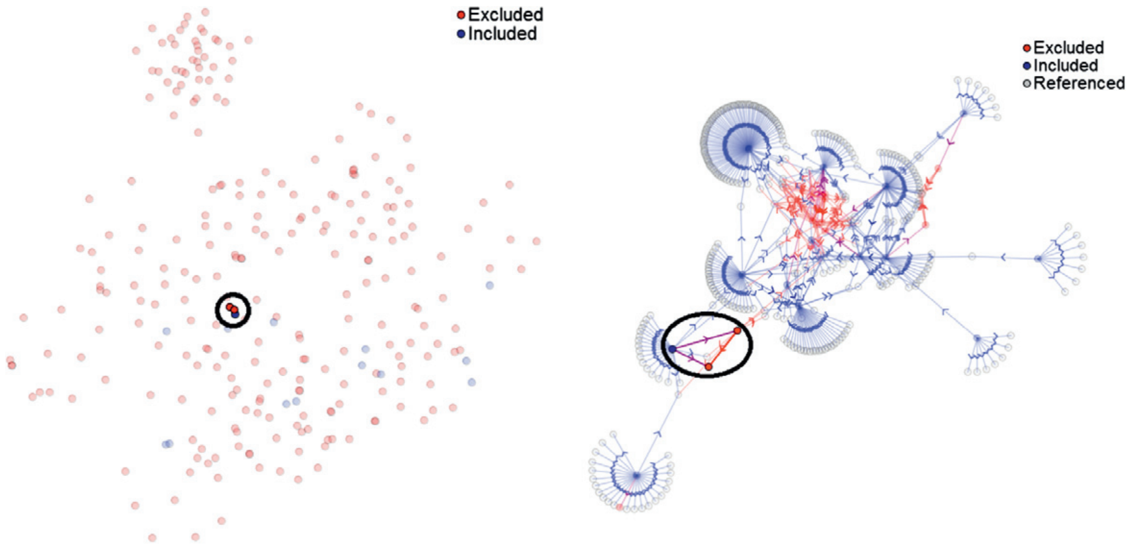
An SLR commonly involves a large set of data to be analyzed and interpreted. In the SLR tasks, the review of selection of primary studies is one of the most important activities that could impact the quality of the SLR's results. This work contributes with a new VTM approach and a tool (Revis) to support the review of study selection activity of the SLR process. Based on demonstration of our approach using the pilot case study, we found that the application of the VTM techniques in the *Selection Review* task has provided the following benefits; in particular, giving clues of relevant primary studies that should be included and inclusions done that should be reviewed. This is because the visualization supports user interaction with the mining algorithm and directs it towards a suitable solution to a given task. Moreover, VTM can be used to enhance user interpretation of mining tasks [31]. The



(a) Complete citation map.

(b) Citation map of included studies.

Fig. 12. Citation maps of the pilot case study SLR. Blue points are included studies, red are excluded studies and gray points are references, documents that are not part of the initial set of considered documents. Some included studies have excluded studies as references, then the inclusions and exclusions should be inspected. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) Content map. (b) Citation map.

Fig. 13. Coordination between the content map (a) and the citation map (b) of the pilot case study SLR. An included study (blue point) is similar in content to two other excluded studies and cites only excluded papers, a clue indicating that its inclusion should be reviewed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

results of our pilot case study lend support to this view since the VTM techniques facilitates exploration, interpretation, and decision-making in regard to the review of inclusion or exclusion of primary studies.

The next section presents a users' case study that assesses the utility of the Revis tool and our proposed VTM techniques.

4.2. Case study 2: Validating the use of the approach through a users' test

In order to validate our approach we conducted a case study involving four PhD students. In this paper, we argue that VTM techniques can support the selection review activity. Hence, our research questions (RQ) are:

1. **RQ1:** Do VTM techniques (content and citation maps) improve the performance (time taken) of the selection review activity in the SLR process?
2. **RQ2:** Do VTM techniques improve the accuracy (the agreement between systematic reviews as to which primary studies they include) of the selection review activity in the SLR process?

The subjects, datasets, users' task, metrics and case study conduction below and are followed by the results.

- Subjects and datasets

A call for volunteers was sent via email to PhD and Master's students in the Computing Systems Department at ICMC (Institute of Mathematics and Computer Sciences) at the USP (University of São Paulo), Brazil. We received responses from 5 PhD students. However, to validate the comparisons, we selected only four students who had prior experience in conducting SLRs.

The case study was organized in two sessions: (i) training; and (ii) execution. For training purposes, we used a small set of 20 primary studies (dataset 1). For execution purposes, the datasets used were the inclusion and exclusion criteria and studies from the SLR5 that had been previously conducted by one master student and her supervisor, the latter a collaborator from our research group and an expert in SLR (dataset 2). However, only 41 studies of a total of 261 were chosen to compose dataset 2. The studies chosen were only those selected by the reviewers (researchers that conducted the SLR5) reading the full text – 27 studies, and the included ones (14 studies). The studies excluded by reading the title, abstract and keywords (220) were discarded. We made this choice on the assumption that adding too many studies to our case study affect the motivation and performance of the participants in carrying out the assigned tasks.

- Definition of users' task and metrics

The users' task was to review the 41 studies from SLR5, to confirm their previous classification or to change it, that is, to ensure that the studies marked as included were in accordance to the inclusion criteria and that studies marked as excluded were in accordance to the exclusion criteria.

Subjects were required to record the time they took to execute the task. The *performance* of the subjects was calculated as $\frac{\text{chosen\_and\_relevant\_articles}}{\text{review\_time}}$. We considered *relevant* the articles marked as *included* by two or more subjects who participated in the case study. These articles were taken as the oracle. The *accuracy* was calculated as the number of included studies that belonged to the oracle.

- Case study conduction

The four subjects were split randomly into two groups of two students, one to conduct the selection review activity manually (group 1) and another to conduct it using the VTM approach (group 2). The case study comprised two sessions, i.e., training and execution (see Table 1).

Naturally only participants involved in the VTM based task (group 2) participated in the in training. During the training session one of the authors provided all subjects with an overview of the study and an explanation on the users' task. Subjects from group 2 were trained for approximately 30 min on how to use the Revis tool. During the training, subjects' doubts about the tool and the VTM techniques were clarified. For training purposes, dataset 1 was used.

Table 1
Groups and tasks.

Group	Subjects	Training (30 min)	Execution (no-limit time)
Reading papers (Group 1)	2	No training	Manual selection review (dataset 2)
VTM techniques (Group 2)	2	Revis tool (dataset 1)	VTM selection review (dataset 2)

Table 2
Summary of results.

Approach	Subject	Performance (time spend)	Accuracy (studies included as oracle)
Reading papers (Group 1)	#1	10.95 articles/h	9
Reading papers (Group 1)	#2	08.67 articles/h	8
VTM techniques (Group 2)	#3	14.83 articles/h	10
VTM techniques (Group 2)	#4	18.40 articles/h	9

In the execution session, the users' task was carried out using dataset 2. Group 1 was given the list of the papers to be reviewed and the previous classification of these papers (included or excluded). Subjects from group 2 received the visualizations (content and citation maps) containing the same papers used by group 1 (included papers were colored in blue and excluded papers in red). Both groups obtained the inclusion and exclusion criteria and a form to summarize their decisions.

No time limit was imposed to the subjects to execute the selection review and they were not allowed to communicate with each other. Table 1 shows a summary of the test setup.

4.2.1. Data analysis

This section presents the results of our users' case study addressing our specific research questions (RQ1 and RQ2). A summary of the results is presented in Table 2.

To answer our first research question (RQ1), we measured the subjects' performance (see third column on Table 2). Considering the performance, our results have shown that: (i) subject 1 reviewed 10.95 articles/h using manual review; (ii) subject 2 reviewed 8.67 articles/h using manual review; (iii) subjects 3 and 4 reviewed 14.83 articles/h and 18.40 articles/h respectively, applying VTM. The time taken by subjects 1 and 2 to perform the selection review activity by reading the papers was 4′20″ and 5′30″ min respectively, and the time taken by students 3 and 4 to perform the same activity using the VTM approach was 3′10″ and 2′50″ min respectively. The results allowed concluding that the answer to RQ1 is "Yes" – the results suggest that the performance of the subjects using the VTM is higher than that of the subjects using the manual method.

The 12 articles marked as "included" by at least two subjects were considered as the oracle. Considering accuracy (see fourth column of Table 2) our results have shown that: (i) subject 1 chose nine articles from the oracle, using manual review, that is, subject 1 chose 75% of the articles included in the oracle; (ii) subject 2 chose eight articles from the oracle, using manual review, that is, 66.66% of the articles included in the oracle; (iii) subject 3 chose nine articles from the oracle, applying VTM, that is, 83.3% of the included articles. In a similar manner, subject 4 chose 75% of the included articles (nine studies). The number of studies included correctly using the manual reading approach was similar to the VTM approach. Based on the results, it can be concluded that the answer to RQ2 is that there is no difference in accuracy using VTM or reading the papers.

The data of the examples and case studies are available for replications upon request.

4.2.2. Discussions

The results have suggested that there is a positive relationship between the use of VTM techniques and the time to conduct the selection review activity. The use of the VTM approach can help to improve the performance of the selection review activity in the SLR when compared to a manual reading method. Therefore the application of the VTM techniques is promising as it improved the performance of the selection review of primary studies. A broader case study is necessary to confirm that studies are most rapidly reviewed by the participants using VTM techniques, although our results are encouraging.

Kitchenham's guidelines for conducting SLRs [25] in SE emphasize the importance of properly conducting the primary study selection activity. It can be implicitly deduced that the quality of the primary study selection step impacts on the overall quality of the SLR. Therefore, in order to ensure better quality outcomes of the SLRs, it is important to conduct the primary study selection step as completely and reliably as possible, including the selection review activity. The results of this research suggest that the accuracy of the primary study selection activity carried out in SLRs by reading the papers is likely to be as good as that using VTM.

It is still to be determined what would happen with users performance if the number of studies were higher. Since VTM approach speeds up the process, it is likely that also supports keeping attention higher for longer. It is also likely that looking into the forms of visual representation will be useful in this context.

As in all research studies, this research also has certain limitations. The size of the subject set in the user study is very small as there were only two participants for each group. However we believe that the results have provided reasonable insight and indication as to how much the VTM techniques can support and speed up the selection review activity. As the case study was conducted with the same number of participants for each group to allow better comparisons, it was possible to obtain two voluntary participants for each group. In our view, a larger sample size of diverse SLR practitioners would help increase the reliability of the findings. However, as a first-cut assessment of the techniques, we believe our users' case study met its goal.

A second limitation is related to the fact that usually there is no agreement between reviewers on their selection of primary studies and it might be the case that the original reviewers would have made different choices for the basis SLR Today. For the purpose of this research, our assumption was not that the researchers who originally conducted the subject SLR made 100% correct decisions on the inclusion and exclusion of studies. On the other hand, we considered the articles marked as "included" by two or more participants relevant to be included in the SLR.

Another current limitation of the Tool is that the documents under analysis (papers) have to be in text format 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. Some research is currently being conducted to make this process as automatic as possible.

5. Conclusions and future work

We have proposed a novel interactive visual approach to help reviewers to comprehend primary studies analyzed in SLRs. Different VTM techniques are combined to support reviewers in the *Selection Review* task, aiming at defining a framework of tools that gives support to evaluate the decisions of including or excluding primary studies from an SLR and, mainly, to help reviewers to guarantee that important studies are not removed. The examples given and the users' case study conducted show that this approach speeds up the *Selection Review* task.

In the special case of an SLR executed by only one reviewer, the suggested approach eliminates the need for random choices of papers to be re-evaluated. Instead, such a selection is based on similarities and citations criteria revealed by the content and citation-based layouts. In addition, the manual approach implies additional effort to select studies for review. Using our approach, the visual representations can give solid clues about which studies should be checked, reducing the amount of documents that need to be re-evaluated and the time spent in the whole process.

In a scenario of a group of reviewers conducting an SLR, the results have shown that VTM techniques are valuable tools to reach to a conclusion on what should and should not be included. The employed visual representations can be used to compare and analyze the decisions made by the different reviewers, giving support to guide the group to a common sense about the inclusions and exclusions.

As a future work we will investigate measurements of citation maps, such as degree centrality, closeness centrality, or betweenness centrality [30] in order to verify if they can give additional support to reviewers in the process of selecting and reviewing primary studies. We also believe that interesting directions for future research are the exploration of other additional views, e.g. the textual difference between the referenced papers and shared references on a secondary level, i.e. references to referenced papers; and replication of our users' case study to compare the accuracy of the Revis tool against a manual selection with a larger number of users and a set of motivated users to handle a larger quantity of papers.

References

[1] S. Ananiadou, B. Rea, N. Okazaki, R. Procter, J. Thomas, Supporting systematic reviews using text mining, Social Science Computer Review 27 (4) (2009) 509–523.
[2] G.F. Andery, A.A. Lopes, R. Minghim, Visual exploration of multidimensional social networks, in: 2nd International Workshop on Web and Text Intelligence (WTI' 09), Spo Carlos, Brazil, 2009, pp. 1–9, (in Portuguese).
[3] K. Andrews, W. Kienreich, V. Sabol, J. Becker, G. Droschl, F. Kappe, M. Granitzer, P. Auer, K. Tochtermann, The infosky visual explorer: exploiting hierarchical structure and document similarities, Information Visualization 1 (3/4) (2002) 166–181.
[4] P.O. Brereton, B.A. Kitchenham, D. Budgen, M. Turner, M. Khalil, Lessons from applying the systematic literature review process within the software engineering domain, Journal of Systems and Software 80 (4) (2007) 571–583.
[5] K. Börner, C. Chen, K. Boyack, Visualizing knowledge domains, Annual Review of Information Science and Technology 37 (1) (2003) 179–255.
[6] C. Chen, F. Ibekwe-SanJuan, J. Hou, The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis, Journal of the American Society for Information Science and Technology 61 (7) (2010) 1386–1409.
[7] K.B. Cohen, H.L. Johnson, K. Verspoor, C. Roeder, L. Hunter, The structural and content aspects of abstracts versus bodies of full text journal articles are different, BMC Bioinformatics 11 (2010) 492.
[8] D.S. Cruzes, T. Dybs, Synthesizing evidence in software engineering research, in: 4th ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM' 10), ACM, New York, NY, USA, 2010, pp. 1–10.
[9] F.Q.B. da Silva, A.L.M. Santos, S. Soares, A.C.C. Franta, C.V.F. Monteiro, Six years of systematic literature reviews in software engineering: an extended tertiary study, in: 32th International Conference on Software (ICSE '10), IEEE Computer Society, Cape Town, South Africa, 2010, pp. 1–10.
[10] O. Dieste, A.G. Padua, Developing search strategies for detecting relevant experiments for systematic reviews, in: Proceedings of the 1st International Symposium on Empirical Software Engineering and Measurement. ESEM '07, IEEE Computer Society, Washington, DC, USA, 2007, pp. 215–224.
[11] T. Dybs, T. DingsØyr, Strength of evidence in systematic reviews in software engineering, in: 2nd ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '08), ACM, New York, NY, USA, 2008, pp. 178–187.
[12] T. Dybs, T. DingsØyr, G.K. Hanssen, Applying systematic reviews to diverse study types: an experience report, in: 1st International Symposium on Empirical Software Engineering and Measurement (ESEM '07), IEEE Computer Society, Washington, DC, USA, 2007, pp. 225–234.
[13] P. Eades, A heuristic for graph drawing, Congressus Numerantium 42 (1) (1984) 149–160.
[14] G. Ellis, A. Dix, A taxonomy of clutter reduction for information visualisation, IEEE Transactions on Visualization and Computer Graphics 13 (6) (2007) 1216–1223.
[15] K.E. Emam, E. Jonker, M. Sampson, K. Krleza-Jeric, A. Neisa, The use of electronic data capture tools in clinical trials: web-survey of 259 canadian trials, Journal of Medical Internet Research 11 (1) (2009) 1–8.
[16] K.R. Felizardo, S.G. MacDonell, E. Mendes, J.C. Maldonado, A systematic mapping on the use of visual data mining to support the conduct of systematic literature reviews, Accepted for Publication in Journal of Software (2011).
[17] K.R. Felizardo, E.Y. Nakwgawa, D. Feitosa, R. Minghim, J.C. Maldonado, An approach based on visual text mining to support categorization and classification in the systematic mapping, In: 14th International Conference on Evaluation and Assessment in Software Engineering (EASE '10), BCS-eWiC, Keele University, UK, April 2010.
[18] R.E. Garcia, M.C.F. de Oliveira, J.C. Maldonado, M.G.M. Neto, Visual analysis of data from empirical studies, in: 10th International Conference of Distributed Multimedia Systems, DMS' 04. San Francisco, 2004, pp. 225–230.
[19] J.H. Gennari, P. Langley, D.H. Fisher, Models of incremental concept formation, Artificial Intelligence 40 (1–3) (1989) 11–61.
[20] M. JØrgensen, M. Shepperd, A systematic review of software development cost estimation studies, IEEE Transactions on Software Engineering 33 (2007) 33–53.
[21] D.A. Keim, Information visualization and visual data mining, IEEE Transactions on Visualization and Computer Graphics 8 (1) (2002) 1–8.
[22] B.A. Kitchenham, Procedures for performing systematic reviews, in: Tech. Rep., Keele University and NICTA, 2004.
[23] B.A. Kitchenham, P.O. Brereton, D. Budgen, M. Turner, J. Bailey, S. Linkman, Systematic literature reviews in software engineering – a systematic literature review, Information and Software Technology 51 (1) (2009) 7–15.
[24] B.A. Kitchenham, P.O. Brereton, S. Owen, J. Butcher, C. Jefferies, Length and readability of structured software engineering abstracts, IET Software 2 (1) (2008) 37–45.
[25] B.A. Kitchenham, S. Charters, Guidelines for performing systematic literature reviews in software engineering, in: Tech. Rep. EBSE 2007-001, Keele University and Durham University Joint Report, 2007.
[26] B.A. Kitchenham, R. Pretorius, D. Budgen, P.O. Brereton, M. Turner, M. Niazi, S. Linkman, Systematic literature reviews in software engineering – a tertiary study, Information and Software Technology 52 (8) (2010) 792–805.
[27] A.A. Lopes, R. Pinho, F.V. Paulovich, R. Minghim, Visual text mining using association rules, Computers and Graphics 31 (3) (2007) 316–326.
[28] J.B. MacQueen, Some methods for classification and analysis of multivariate observations, in: 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Statistical Laboratory of the University of California, Berkeley, 1967, pp. 281–297.
[29] V. Malheiros, E.N. Höhn, R. Pinho, M. Mendonca, J.C. Maldonado, A visual text mining approach for systematic reviews, in: 1st International Symposium on Empirical Software Engineering and Measurement (ESEM' 07), IEEE Computer Society, Washington, DC, USA, 2007, pp. 245–254.
[30] M.E.J. Newman, The structure and function of complex networks, SIAM Review 45 (2) (2003) 167–256.
[31] M.C.F. Oliveira, H. Levkowitz, From visual data exploration to visual data mining: a survey, IEEE Transactions on Visualization and Computer Graphics 9 (3) (2003) 378–394.
[32] Pacific Northwest National Laboratory (PNL), IN-SPIRE™ Visual Document Analysis, 2008. <http://in-spire.pnl.gov/>.
[33] F.V. Paulovich, 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 14 (6) (2008) 1229–1236.
[34] M. Riaz, M. Sulayman, N. Salleh, E. Mendes, Experiences conducting systematic reviews from novices' perspective, in: 14th International Conference on Evaluation and Assessment in Software Engineering (EASE' 10), BCS-eWiC, Keele University, UK, 2010.
[35] G. Salton, A. Wong, C. Yang, A vector space model for automatic indexing, Communications of the ACM 18 (11) (1975) 613–620.
[36] D.I.K. SjØberg, T. Dybs, M. JØrgensen, The future of empirical methods in software engineering research, in: Future of Software Engineering (FOSE '07), IEEE-CS Press, 2007, pp. 358–378.
[37] M. Skoglund, P. Runeson, Reference-based search strategies in systematic reviews, in: 13th International Conference on Evaluation and Assessment in Software Engineering. EASE' 09, BCS, Durham University, England, UK, 2009, pp. 10.
[38] A. Tan, Text mining: The state of the art and the challenges, in: PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases (PAKDD '99), Springer, Beijing, China, 1999, pp. 65–70.
[39] P. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Addison-Wesley, 2006.