Information and Software Technology 64 (2015) 1–18

Contents lists available at ScienceDirect

Information and Software Technology

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

Guidelines for conducting systematic mapping studies in software engineering: An update

Kai Petersen *, Sairam Vakkalanka, Ludwik Kuzniarz

Department of Software Engineering, Blekinge Institute of Technology, Sweden

ARTICLE INFO

Article history:
Received 1 September 2014
Received in revised form 23 February 2015
Accepted 14 March 2015
Available online 28 March 2015

Keywords:
Systematic mapping studies
Software engineering
Guidelines

ABSTRACT

Context: Systematic mapping studies are used to structure a research area, while systematic reviews are focused on gathering and synthesizing evidence. The most recent guidelines for systematic mapping are from 2008. Since that time, many suggestions have been made of how to improve systematic literature reviews (SLRs). There is a need to evaluate how researchers conduct the process of systematic mapping and identify how the guidelines should be updated based on the lessons learned from the existing systematic maps and SLR guidelines.
Objective: To identify how the systematic mapping process is conducted (including search, study selection, analysis and presentation of data, etc.); to identify improvement potentials in conducting the systematic mapping process and updating the guidelines accordingly.
Method: We conducted a systematic mapping study of systematic maps, considering some practices of systematic review guidelines as well (in particular in relation to defining the search and to conduct a quality assessment).
Results: In a large number of studies multiple guidelines are used and combined, which leads to different ways in conducting mapping studies. The reason for combining guidelines was that they differed in the recommendations given.
Conclusion: The most frequently followed guidelines are not sufficient alone. Hence, there was a need to provide an update of how to conduct systematic mapping studies. New guidelines have been proposed consolidating existing findings.

1. Introduction

Systematic mapping studies or scoping studies are designed to give an overview of a research area through classification and counting contributions in relation to the categories of that classification [1,2]. It involves searching the literature in order to know what topics have been covered in the literature, and where the literature has been published [2]. Though, a systematic mapping study and a systematic literature review share some commonalities (e.g. with respect to searching and study selection), they are different in terms of goals and thus approaches to data analysis. While systematic reviews aim at synthesizing evidence, also considering the strength of evidence, systematic maps are primarily concerned with structuring a research area.

Systematic mapping studies are used by many researchers on a number of areas using different guidelines or methods. A sample of mapping studies is mentioned below with their areas of research and the guidelines used.

- Condori-Fernandez et al. [3] provided a mapping of the research articles on software requirement specifications combining two guidelines (cf. [2,1]).
- Jalali and Wohlin [4] performed mapping of the literature available on Global software Engineering considering the guidelines by [2,1].
- Barreiros et al. [5] constructed systematic maps on the published research on software engineering test beds based on Kitchenham and Charters's [1] guidelines.
- Qadir and Usman [6] conducted a mapping on curriculum in software engineering using the guidelines by [2,1].

Recently, Wohlin et al. [7] compared systematic mapping studies that were conducted on the same topic by two groups of researchers working and publishing independently. That is, the review protocols of the two reviews were developed independently. Some questions on the reliability of systematic mapping

* Corresponding author.
E-mail addresses: kai.petersen@bth.se (K. Petersen), sava11@student.bth.se (S. Vakkalanka), ludwik.kuzniarz@bth.se (L. Kuzniarz).

http://dx.doi.org/10.1016/j.infsof.2015.03.007
0950-5849/

studies have been raised. For example, even though the same classification scheme has been used by the two studies the same articles have been classified in a different way. Both studies used the guidelines by Petersen et al. [2].

Overall, the number of mapping studies is continuously increasing, and there is a great interest in the methodology. To increase the confidence and reliability of mapping studies, there is a need to understand how they are conducted at large. Furthermore, given that many new insights have been gained through the conduct of systematic reviews, which have been synthesized by Kitchenham and Brereton [8], we may use this knowledge to determine the state of quality of mapping studies, taking the differences between SLRs and mapping studies into account.

This study makes the following contributions to improve systematic mapping guidelines:

- Assessing the current practice of conducting systematic mapping studies in software engineering.
- Comparing the identified guidelines for mapping studies with best practices as identified in Kitchenham and Brereton [8].
- Consolidating the findings to propose updates to systematic mapping guidelines.

The practical benefit is that the quality of systematic mapping studies could be improved further by consolidating the knowledge about good ides in the currently published guidelines.

The remainder of the paper is structured as follows: Section 2 presents the related work. Section 3 explains the research method used. Section 4 presents the results of the mapping of systematic maps in software engineering. Section 5 describes the updated guidelines. Section 6 concludes the paper.

2. Background and related work

2.1. The differences between systematic maps and reviews

Kitchenham et al. [9] contrasted the different characteristics of the process of systematic literature reviews and mapping studies. There are differences with respect to the research questions, search process, search strategy requirements, quality evaluation, and results.

The research questions in mapping studies are general as they aim to discover research trends (e.g. publication trends over time, topics covered in the literature). On the other hand, systematic reviews aim at aggregating evidence and hence a very specific goal has to be formulated (e.g. whether an intervention is practically useful by industry).

Given this main difference, the research process is impacted. As pointed out by Kitchenham et al. [9] the search for studies in systematic maps is based on a topic area, while systematic maps are driven by specific research questions. This also highlights why in one topic area, a variety of number of systematic reviews may be conducted (see Verner et al. [10] as an example).

Kitchenham et al. [9] argue that the search requirements are less stringent for mapping studies as we are only interested in research trends, while they argue that all studies must be found in systematic reviews. Though, as was discussed in Wohlin et al. [7] this is often not realistic, neither for systematic reviews nor mapping studies. Wohlin et al. argue for achieving a good sample with respect to the characteristics of the population the reviewers are aware of.

Quality assessment is more essential in systematic reviews to determine the rigor and relevance of the primary studies. In systematic maps no quality assessment needs to be performed. Using the classification of research types by Wieringa et al. [11] a category of solution proposals would contain articles with no empirical evidence. Articles in this category would generally not be included in a systematic review, though in systematic maps they are important to spot trends of topics being worked on.

The outcome of a mapping study is an inventory of papers on the topic area, mapped to a classification [11]. Hence, a mapping study provides an overview of the scope of the area, and allows to discover research gaps and trends [2].

2.2. Evaluations of systematic mapping studies

Wohlin et al. [7] compared the outcome of two independently conducted systematic reviews on the same topic and reflected on why differences have occurred. The two reviews were concerned with mapping the research of software product line testing (cf. [12,13]). Both studies started from quite similar research questions, though one studie's questions were more specific than those of the other study. Two important outcomes were observed. First, even though the studies had a similar aim, the articles included turned out to be different. Second, the classification of articles included in both studies was different. Both studies used the classification scheme proposed by Wieringa et al. [11], the use of that classification being recommended in Petersen et al. [2]. In particular, research types were classified differently. On the same set of articles included in both studies the research only agreed in 33% of the cases.

Multiple sources for differences have been discussed, e.g. research question phrasing (open versus specific), scoping (where to draw the boundaries for the research area), restrictions on research types and methods (e.g. only including empirical work), time span, and exclusion of specific publication types (e.g. gray literature). As classification has been an issue, it was suggested that refinements or clarifications (e.g. in terms of examples) may be needed. Standards or well known classification schemes (e.g. ACM Thesaurus) may also be of use. Also, classes may not be clearly disjoint and hence classification in more than one category may be possible.

An important point made was that one should aim for a good sample, rather than finding all articles. It is likely that this is not a realistic goal given the findings of Wohlin et al. [7] and other studies. In other words, more articles may not be better than fewer, if the fewer are a better representation of the population of articles for the targeted topic. As the population is not known before, and one cannot randomly sample from the population, the ability to reflect on the sample may be very limited. This point is further discussed in Section 5.

Similar observations were made by Kitchenham et al. [14] who conducted a mapping study on unit and regression testing to determine the completeness of their search and study selection. Several existing systematic literature reviews were used as a validation set to evaluate the completeness of the identified articles in the systematic map. Even though important groups of articles were identified sufficiently well, within each group represented by the different systematic reviews articles were missed. This also supports the sampling argument made by Wohlin et al. [7].

The actual value of mapping studies has been evaluated by Kitchenham et al. [15] from the perspective of education and the usefulness as input for further studies [9,16].

From an educational point of view Kitchenham et al. [15] conducted a study of student projects (three were graduate and three undergraduate students) where the task was to conduct a systematic mapping study. Thereafter, experiences and assessments were gathered from the supervisor as well as all students by means of a questionnaire with open-ended questions. Five propositions have been formulated and tested based on the feedback of the students. Multiple benefits have been experienced, such as learning research skills, and being beneficial for graduate students as a good baseline for Ph.D. studies. Graduate students did not find classification and search difficult, while this was a challenge for the undergraduate students. The same was observed for the documentation and reporting of the results. Other positive outcomes were that conducting systematic maps was perceived as challenging and at the same time enjoyable. It was highlighted that the maps provide a good overview of a topic.

Petersen et al. [2] proposed that a mapping study preceding a systematic review provides a valuable baseline. Kitchenham et al. [9,16] evaluated whether systematic maps are in fact useful as input for follow-up studies. The type of study conducted was a participant-observer case study. In the study, multiple benefits have been observed, namely: time-savings for follow-up studies (e.g. due to reuse of study protocols); good overview of an area and the ability to identify research gaps; visualization of research trends; related work identification; use as a validation set based on gathered references; educational source. Kitchenham et al. [9,16] also pointed out that it is important to have a well defined and reliable classification scheme. The scheme proposed by Wieringa et al. [11] and used by two independent studies on the same topic (cf. [12,13]) led to inconsistent classifications, which indicates that this is a major concern specific to systematic mapping studies.

3. Method

3.1. Research questions

The goal of this mapping study (following the guidelines in [2,1]) is to determine how systematic mapping processes have been executed in software engineering. This leads to the following research questions (RQs):

- RQ1: Which guidelines are followed to conduct the *systematic mapping* studies in *software engineering*?
- RQ2: Which software engineering topics are covered?
- RQ3: Where and when were mapping studies published?
- RQ4: How was the systematic mapping process performed? This includes, for example:
  - Identification of studies (search, inclusion and exclusion)
  - *Categorization* and *Classification* schemes and processes
  - Visualization of results

This information, in combination with best practices from existing literature, is then used to propose and develop an updated guideline.

In the following the search protocol is presented, which has been first developed by the second and third author, and later been reviewed by the first author.

3.2. Search

PICO (Population, Intervention, Comparison and Outcomes) suggested by Kitchenham and Charters [1] was developed to identify keywords and formulate search strings from research questions.

- *Population:* In software engineering, population may refer to specific software engineering role, category of software engineer, an application area or an industry group [1]. In our context, the population are systematic mapping/scoping studies.
- *Intervention:* In software engineering, intervention refers to a software methodology, tool, technology, or procedure. In the context of this study we do not have a clear intervention to be investigated.
- *Comparison:* In this study we compare the different processes of conducting the maps by means of identifying the different strategies that have been used by them. Though, no empirical comparison is made, at this stage alternative strategies are identified.
- *Outcomes:* No measurable outcome is considered, as we do not focus on empirical studies evaluating systematic mapping (these have been presented in Section 2).

The identified keywords are guidelines, systematic mapping and software engineering which were grouped into sets and their synonyms were considered to formulate the search string.

- Set 1: Scoping the search for software engineering, i.e. "software engineering".
- Set 2: Search terms directly related to the intervention, e.g. "systematic mapping", and "systematic maps".
- Set 3: Search terms related to the process of classification and categorization, e.g. methods, tools, classification, framework.

As the keywords identified from research questions and keywords identified from the PICO criteria are similar, we grouped them into sets. Each set of searches was performed on the databases of IEEE Xplore, ACM, Scopus, as well as Inspec/Compendex (Engineering village). The databases have been selected based on the experience reported by Dyba et al. [17]. The search strings used for each database can be found in Table 1, and have been used on all fields.

EndNote X6, a reference management tool, was used to remove duplicates and to manage the large number of references. This study has been conducted during 2013, the full year of 2012 and before have been considered during the search.

Table 2 shows the number of search results per database.

Table 1
Searches in databases.

Database	Search
IEEE	(("Systematic mapping" or "systematic map" or "systematic mapping study" or "systematic mapping studies" or "systematic maps") AND (("Methods" or "framework" or "model" or "practice") OR ("tools" OR "tool" OR "techniques") OR ("categorization" OR "classification" OR "grouping") OR ("guidelines" OR "rules")) AND "software engineering")
ACM	(("Systematic mapping" OR "systematic map" OR "systematic mapping study" OR "systematic mapping studies") AND software engineering)
Scopus	("Systematic mapping" OR "systematic map" OR "systematic mapping study" OR "systematic mapping studies" OR "systematic maps") AND ("software engineering") OR (model OR method OR approach OR tools OR tool OR techniques OR framework OR practice OR classification OR categorization OR process OR guidelines OR rules OR strategy OR way)
Inspec/ Compendex	((((Systematic mapping OR systematic map OR systematic mapping study OR systematic mapping studies OR systematic maps) WN All fields) AND (((Methods OR framework OR model OR practice) OR (tools OR tool OR techniques) OR (categorization OR classification OR grouping) OR (guidelines OR rules)) WN All fields)) AND ((software engineering) WN All fields))

Table 2
Number of studies per database.

Database	Search results
IEEE	5610
ACM	360
Scopus	1215
Inspec/Compendex	567

3.3. Study selection and quality assessment

We excluded articles based on titles and abstracts, as well as full-text reading and quality assessment. Studies also have been added through backward snowball sampling. The application of inclusion and exclusion criteria to titles and abstracts was conducted by the second author. Thus, each article was only reviewed by a single author, which poses a threat to the reliability of the mapping study (see also the discussion of validity threats Section 3.6). In order to reduce the threat, actions have been taken to evaluate the final set of articles included, which was done by the first author (see further below). If a very large number of studies is obtained and many of them are clearly identifiable noise the process may be conducted individually [18]. We were inclusive taking a paper to full-text reading when in doubt. The following inclusion criteria were applied to titles and abstracts:

- Studies present the research method and result of a systematic mapping study.
- Studies are in the field of software engineering.
- Studies were published online in the time frame 2004 (first systematic review and evidence-based software engineering guidelines were published) to 2012 (the study was conducted during 2013).

The following criteria state when a study was excluded:

- Studies presenting summaries of conferences/editorials or guidelines/templates for conducting mapping studies.
- Studies presenting non-peer reviewed material.
- Studies not presented in English.
- Studies not accessible in full-text.
- Books and gray literature.
- Studies that are duplicates of other studies.

The number of included and excluded articles is shown in Fig. 1 at each stage.

During full-text reading it became obvious that further articles should be removed as they were not in the scope based on the inclusion and exclusion criteria. The remaining articles (46) were used to conduct backward snowball sampling [19], which led to 11 studies being added.

On the set of 57 primary studies quality assessment was conducted. Three questions were answered to assess the quality of the primary studies:

- Is the motivation for conducting systematic mapping clearly stated?
- Is the process of conducting systematic mapping clearly defined (study identification, data extraction, classification)?
- Is there any empirical evidence for the defined mapping process? This question is concerned with whether the results of the mapping study are presented. That is, studies focusing on evaluating mapping studies without presenting their results are excluded.

The quality assessment was conducted by the second author, and hence assessed by an individual person.

To evaluate the study identification process the first author generated two validation sets after the second author completed the search and inclusion/exclusion process. The very first set consisted of eight papers that the first author was aware of, and knew that they ought to be included [20,21,13,12,22,23,4]. The second set consisted of papers based on the citations to the systematic mapping guidelines provided by Petersen et al. [2], and was created after defining the first set. A total of 321 citations are given to the article which were all retrieved. All articles that could not have been found due to their publishing date (i.e. 2013 and 2014) have been removed from the set. Thereafter, the first author applied the inclusion and exclusion criteria to the set of articles. Four further papers have been identified that were missed: [24–27], while Jalali and Wohlin [24] is a continuation of [4] (see Section 3.6 for further reflections).

There were also borderline papers that were deemed relevant during the inclusion and exclusion based on title and abstract, and were excluded later in the process. Hence, the first author reviewed all studies excluded during quality assessment and full-text reading, which led to eight studies being included, which have been excluded earlier.

3.4. Data extraction

To extract data from the identified primary studies, we developed the following template shown in Table 3. Each data extraction field has a data item and a value. The extraction was performed by the second author and reviewed by the first author by tracing back the information in the extraction form to the statements in each paper, and checking their correctness. Having another author check the extraction is a common practice in systematic reviews for social science [18].

3.5. Analysis and classification

The information for each item extracted was tabulated and visually illustrated (see Section 4). The extracted strategies were grouped and given a theme by the first author during analysis. Thereafter, the papers belonging to each theme were counted. As en example, for the study identification phase four main themes (choosing the search strategy, developing the search, evaluating the search, and inclusion and exclusion) and their sub-themes have been identified. The sub-themes for the inclusion and exclusion process were known a priori and are based on [28].

3.6. Validity evaluation

Petersen and Gencel [29] reviewed existing validity classification schemes and discussed their applicability to software engineering. The following types of validity should be taken into account: descriptive validity, theoretical validity, generalizability, interpretive validity.

Table 3
Data extraction form.

Data item	Value	RQ
General		
Study ID	Integer	
Article Title	Name of the article	
Author Name	Set of Names of the authors	
Year of Publication	Calendar year	RQ3
Area in SE	Knowledge areas in SWEBOK	RQ2
Venue	Name of publication venue	RQ3
Process		
Guidelines	Which guidelines were adopted	RQ1
Search strategy	What search strategy is followed, and how were studies selected	RQ4
Search type	Manual or automated or both	RQ4
Classification schemes	How were articles classified	RQ4
Visualization type	What visualization types were used in order to present the data in a pictorial manner	RQ4

3.6.1. Descriptive validity

Descriptive validity is the extent to which observations are described accurately and objectively. Threats to descriptive validity are generally greater in qualitative studies than they are in quantitative studies. To reduce this threat, a data collection form has been designed to support the recording of data. The form objectified the data extraction process and could always be revisited, as was the case for the primary studies. Hence, this threat is considered as being under control.

3.6.2. Theoretical validity

Theoretical validity is determined by our ability of being able to capture what we intend to capture. Furthermore, confounding factors such as biases and selection of subjects play an important role.

Study identification/sampling: During the search studies could have been missed, as pointed out by Wohlin et al. [7] two mapping studies of the same topic ended up with different sets of articles. To reduce this threat, we complemented the search with backward snowball sampling [19] of all studies after full-text reading (see Fig. 1).

Researcher biases may appear during the selection and extraction of data. The study selection was conducted by an individual author, which is the main threat to validity. Hence, important studies might have been missed. To reduce this threat and gain confidence in the results, we evaluated the study identification through the creation of a reference set of articles. This was done by the first author through forward snowball sampling on one of the guidelines specifically targeted towards mapping studies [2]. Only a small number of new studies have been obtained, indicating that the overall conclusions of this review would not change.

There is also the possibility that systematic mapping studies are referred to as systematic reviews. During the inclusion and exclusion phase there is a threat that these studies are dismissed.

We conducted the study during 2013 and wrote the report during 2014. Hence, only studies from 2012 and earlier are included in our analysis. We contemplated on whether to also include 2013 in the analysis and conduct the process for these studies as well. From a sampling perspective it was important to have a good representation of mapping studies overall (cf. [7]). We identified a total of 52 mapping studies, which were covering the different areas of software engineering well (see Fig. 3). Furthermore, different types of publication venues are well represented (see Fig. 4). Also, no new guidelines have been published that could have been well adopted during that time span. Hence, we aimed at developing the guidelines from the sample up till 2012 to assure timely availability.

There is a potential threat that activities may be overlooked, misunderstood, or incompletely reported as we based our analysis solely on the reporting. This may affect the completeness of the classification of activities. As a control action, the first author checked the extraction.

Data extraction and classification: During this phase researcher bias is also a threat. Kitchenham et al. [30] as well as Brereton et al. [31] found it useful to have one researcher extract the data and another reviewing the extraction. To reduce the bias, the first reviewer assessed all extractions made by the second reviewer. Though, given that this step involves human judgment, the threat cannot be eliminated.

3.6.3. Generalizability

Petersen and Gencel [29] present a distinction between external generalizability (generalizability between groups or organizations) and internal generalizability (generalization within a group). Given that a wide range of systematic maps on different topics and following similar guidelines have been identified, internal generalizability is not a major threat. Though, the results may not apply to systematic literature reviews as they are different in their goals (see Section 2.1). Individual practices and strategies to conduct search and study inclusion and exclusion are valuable for systematic reviews as well.

3.6.4. Interpretive validity

Interpretive validity is achieved when the conclusions drawn are reasonable given the data, and hence maps to conclusion validity. A threat in interpreting the data is researcher bias. The first author is a co-author on one of the guidelines [2], which may be a bias in interpretation. On the other hand, the experience with systematic review processes may also help in the interpretation of data.

3.6.5. Repeatability

The repeatability requires detailed reporting of the research process. We reported the systematic mapping process followed, and also elaborated on actions taken to reduce possible threats to validity. Repeatability was also aided by the use of existing guidelines.

4. Results of the mapping

4.1. Frequency of publication (RQ1)

Fig. 2 shows the number of mapping studies identified within the years 2007–2012. The first mapping study was published by Bailey et al. [32], and this was the only study in 2007. While the interest in mapping studies was moderately increasing 2008–2010, a significant increase can be observed in 2011 and 2012. Besides an increased interest, another potential reason may be the better distinction between systematic literature reviews and mapping studies. This increase in the number of mapping studies published indicates that they are considered highly relevant by the software engineering research community.

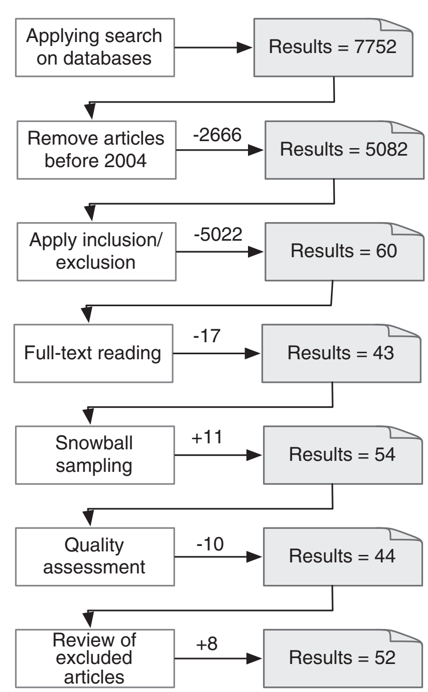

Fig. 1. Number of included articles during the study selection process.

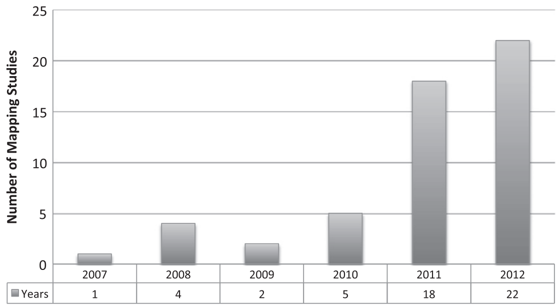

Fig. 2. Publications per year.

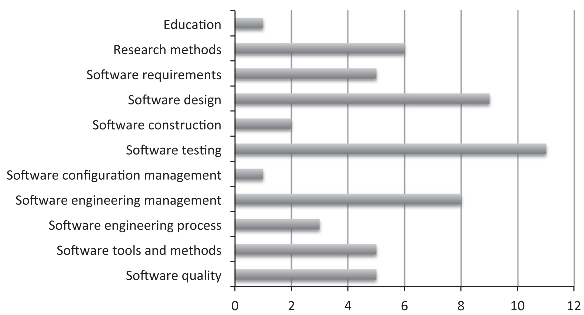

Fig. 3. Topics covered in systematic mapping studies.

4.2. Topics (RQ2)

The topics covered were classified based on the Software Engineering Body of Knowledge (SWEBOK) structure. Education and mapping studies on research methodologies had to be added as separate categories as they were not covered by SWEBOK. Fig. 3 shows the number of mapping articles per topic category. It is visible that all activities are well represented, while the emphasis is on requirements, design, testing, management, and tools and methods. The main life-cycle activities of software engineering are covered by the mapping studies, and they are not biased towards a specific topic. Though, there are gaps (e.g. in relation to configuration management and education).

4.3. Venues of publication (RQ3)

In this study, peer-reviewed venues (including journals, as well as peer-reviewed conferences and workshops) were considered. Fig. 4 provides an overview of the distribution of mapping articles between these venues. A similar number of articles is published in international conferences and scientific journals.

Looking at the specific venues, it was also clear which ones were mostly targeted by authors of mapping studies. The top three venues for mapping studies are stated in Table 4. 14 articles are published in the Information & Software Technology journal.

Overall, this indicates that systematic mapping studies are regarded as valuable scientific contributions, given that they are widely published in high quality forums.

4.4. Review process (RQ4)

We analyzed the review process with respect to the guidelines used, followed by an analysis of the mapping process itself (ranging from study identification to results visualization).

4.4.1. Guidelines

Ten different guidelines have been followed when designing the systematic mapping process (see Fig. 5).

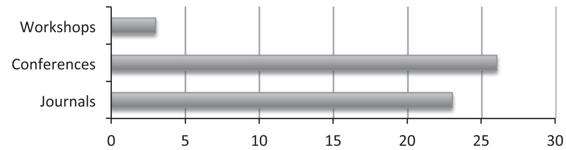

Fig. 4. Venue types.

Table 4
Targeted venues.

Rank	Venue	Studies
1	Information & Software Technology (IST)	[23,33,12,34–36,13,37,22,20,38,30]
2	Evaluation and Assessment in Software Engineering (EASE)	[39–46]
3	Empirical Software Engineering and Measurement (ESEM)	[32,5,3,47]

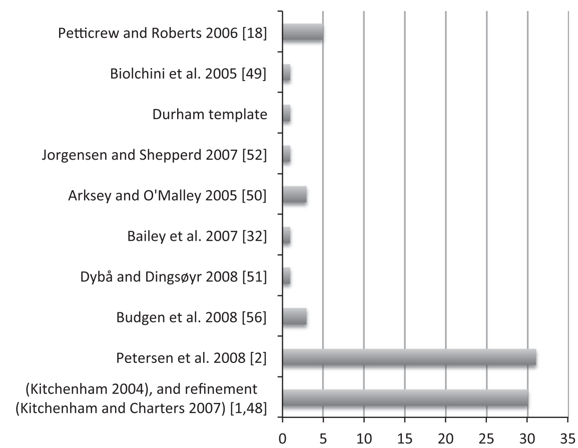

Fig. 5. Guidelines.

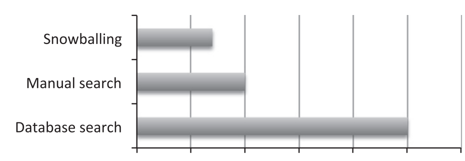

Fig. 6. Choosing search strategies.

The guidelines on systematic reviews (Kitchenham [48] and its refinement, Kitchenham and Charters [1]) and systematic mapping (Petersen et al. [2]) are most commonly followed.

The remaining guidelines only received a small number of citations. They comprise one guideline for conducting systematic reviews in software engineering (Biolchini et al. [49]) and a checklist for conducting mapping studies in software engineering (Durham template[1]).

Furthermore, guidelines from outside software engineering have been used. This includes one book on systematic reviews for social sciences [18], and a guide for mapping studies [50].

A set of studies also used existing literature studies [32,51,52] as templates for their systematic reviews.

It is noteworthy to observe that often multiple guidelines were combined as an individual guideline was not found to be sufficient.

[1] www.dur.ac.uk/ebse/resources/templates/MappingStudyTemplate.pdf.

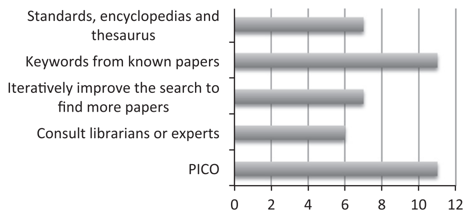

Fig. 7. Developing the search strategy.

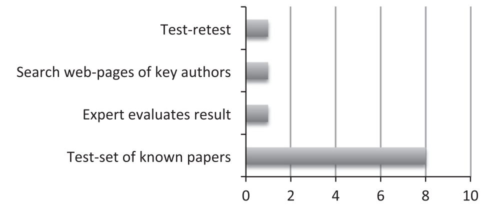

Fig. 8. Evaluating the search strategy.

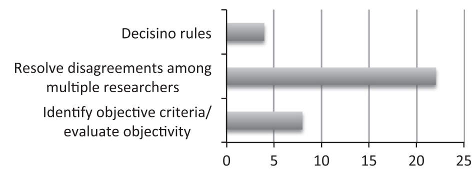

Fig. 9. Inclusion and exclusion.

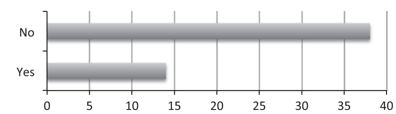

Fig. 10. Quality assessment.

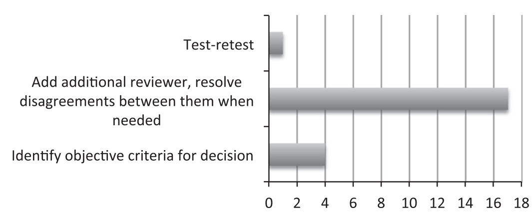

Fig. 11. Data extraction and classification process.

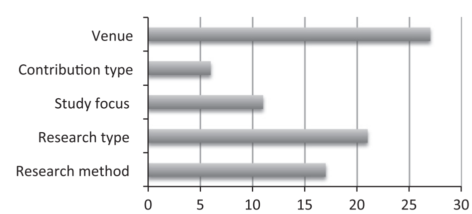

Fig. 12. Topic-independent classification.

Overall, 24 mapping studies used more than one guideline. Thus, individual guidelines do not appear to be complete enough to characterize the whole mapping process.

4.4.2. Study identification

The study identification includes the search for articles and their selection. Four main activities are performed, namely deciding on the search strategies to be followed, developing the search, evaluating the search, and conducting the inclusion and exclusion.

Three search strategies have been identified (see Fig. 6). The most common way of searching is the database search, followed by the manual search.

The strategies for developing the search are shown in Fig. 7. The search terms generally follow the research questions and topic studied for all included papers. However, there are further means of developing the search. These are: defining search terms in the categories of PICO (Population, intervention, comparison, and outcome); derive keywords from known papers; iteratively improve the search to find more relevant papers per iteration; consult librarians and experts for input. The most frequently applied strategies are PICO and extracting keywords from known papers.

The most common strategy for evaluating the search (see Fig. 8) by the definition of a test-set of known papers that ought to be found. Only few studies use the strategies of expert evaluation [39], checking websites of key authors [53], and utilizing test–retest [21].

During the inclusion and exclusion the researchers decide on whether the papers retrieved through searching are relevant, three strategies have been identified (see Fig. 9).

One strategy is to define direct decision rules on what to do with an article based on the outcome of evaluations of multiple researchers, e.g. *"To reduce bias, the first three steps were performed by two researchers independently, and the paper passed to the next step if at least one of them decided so"* [54]. In addition, further reviewers are added if there is a lack of clarity, or consensus meetings are held to resolve disagreements and uncertainty. Finally, the objectivity of criteria is assessed, either pre-review on a test set by measuring agreement, or post-review. As can be seen from Fig. 9 the most common approach is to resolve disagreements and uncertainty with additional reviewers and consensus meetings.

4.4.3. Quality evaluation

Multiple studies highlighted that quality assessment is more essential in systematic reviews as there the goal is evidence aggregation. As can be seen in Fig. 10, quality assessment is conducted in systematic mapping studies. Though, only 14 out of 52 studies assess the quality of primary studies, hence it is not very common to do so.

4.4.4. Data extraction and classification

To evaluate the data extraction and classification the same two strategies as for inclusion and exclusion of articles have been identified (see Fig. 11). The pattern is also similar with resolving disagreements through additional reviewers and consensus meetings being most common. However, the total number of studies evaluating the quality of data extraction is lower than the number for inclusion and exclusion. Even though, one may argue that the process for data extraction is equally prone to biases as the inclusion and exclusion of articles.

As one of the main goals of systematic mapping studies is to give an overview of an area the classification of articles is important. We distinguish between fixed classification schemes that are not highly dependent on the specific topic investigated (see Fig. 12), and topic specific classifications (see Fig. 13).

With respect to fixed classifications two dimensions used in the papers have been proposed earlier in the guidelines by Petersen et al. [2]. These are contribution type and research type (cf. Wieringa et al. [11]). Contribution type refers to determining the type of intervention being studied, this could be a process, method, model, tool, or metric. Study type consists of the categories evaluation research, validation research, solution proposal, philosophical paper, experience report, and opinion paper. Further reflections on

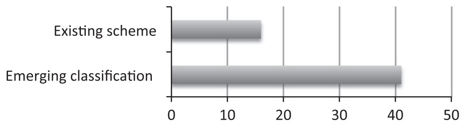

Fig. 13. Topic-related classification.

these will be provided when developing the updated guidelines [11].

Three new dimensions not highlighted by Petersen et al. [2] have been identified, namely venue, study focus, and research method. The study focus refers to the context being studied. Examples are distinctions between academic, industrial, government, project, and organization context. The venue is the type of publication venue (e.g. journal, conference, and workshop) as well as the concrete venues being targeted. The research method refers to the scientific method used, e.g. case study, experiment, or survey.

As shown in Fig. 12 the most frequently reported dimension is venue. The classification of the research intervention is mostly done using research type and research method. In comparison, contribution type is only considered by few studies, and hence appears to have a limited relevance or applicability to the majority of mapping studies.

The topic specific categorization has been done in two ways. Either a new classification was created, emerging from the papers being studied, or an existing classification has been used. Examples of existing classifications are SWEBOK or standards (e.g. IEEE or ISO/IEC standards). In most cases (see Fig. 13), a new classification was created following the strategy of identifying clusters based on keywords (cf. [2]), which bears similarities to open coding of grounded theory.

4.4.5. Visualization

Six approaches to visualize the classified data have been identified (see Fig. 14). The most common approaches are bar plots and bubble plots. Heatmaps are an interesting means of visualization as well as, similar to bubble plots, it directly visualizes the relative amount of publications in different categories. Though, only one individual study [55] used them.

4.4.6. Validity threats

A key to empirical studies is the discussion of validity threats. It is positive to see that the clear majority of studies reported the validity threats (see Fig. 15).

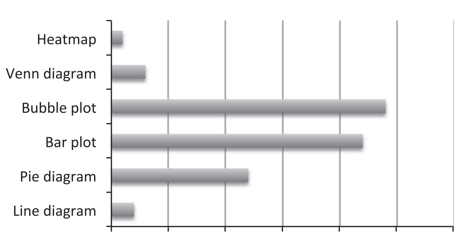

Fig. 14. Visualizations.

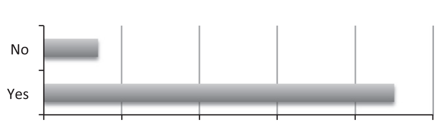

Fig. 15. Discussion of validity threats.

5. Guideline updates

In this section we propose updated guidelines for systematic mapping consolidating the practices identified through this mapping study and the practices suggested in the guidelines proposed by other researchers [48,1,2,56,57,49]. Furthermore, where applicable we also incorporate the learnings from a synthesis of research on evidence-based approaches in software engineering (cf. [8]). Thereby, important information on how mapping studies have been conducted in the past is concentrated in a comprehensive guideline.

The main activities of the systematic mapping process as identified in Section 4 are mapped to the guidelines in Table 5. As is evident from the table, the guidelines differ with respect to which activities are proposed. The mapping shows that existing guidelines do not represent the activities conducted in the mapping studies. Hence, to have a complete list (and thus being aware of the possible options to make informed decision to use an activity or now) is important, also motivating the need for an updated guideline. The information in Table 5 and obtained from the references is used to document the updated guidelines for systematic mapping studies.

The mapping process proposed consists of (1) planning, (2) conducting, and (3) reporting the mapping.

5.1. Planning the mapping

During the planning all the decisions relevant to the conduct of the mapping study are made.

5.1.1. Need identification and scoping

As highlighted in Section 2.1 systematic reviews and maps are different in their goals and hence there are partial differences in the process. This is also visible when comparing the proposals by Kitchenham and Charters [1] with those of Petersen et al. [2]. Activities related to thorough searching and finding all relevant evidence is much more intensively covered in Kitchenham and Charters [1], while the classification aspects and their representation are more emphasized in Petersen et al. [2].

A number of typical research goals for mapping studies were identified by Arksey and O'Malley, namely:

- *"To examine the extent, range and nature of research activity" [57]:* In software engineering, this may refer to the extent different practices are studied and reported in literature (see e.g. Jalali and Wohlin [19]).
- *"To determine the value of undertaking a full systematic review" [57]:* Systematic maps may lead the researcher to find existing systematic reviews. In addition, when classifying papers we often distinguish between evaluated or validated research and solution proposals (see Wieringa et al. [11]). Identifying evaluation and validation research studies provides the set of studies to continue further investigation on in the form of a systematic review. A similar process has been utilized by Barney et al. [23]. Also, the structuring of the area helps in refining research questions for the conduct of future systematic reviews [18]. They may also be used to determine the potential effort of a full systematic review [18].
- *"To summarize and disseminate research findings" [57]:* Systematic maps give a comprehensive overview over the area and can also be used as an inventory for papers. Specifically, graduate students may find them useful to orientate themselves in a new area early during the Ph.D. studies.
- *"To identify research gaps in the existing literature" [57]:* Based on the categorization areas, with very few studies or lack of evaluations, the need for future research becomes apparent. In case

Table 5
Guideline comparison.

Research process		References								
		Kitchenham [48]; Kitchenham and Charters [1]	Petersen et al. [2]	Budgen et al. [56]	Arksey and O'Malley [57]	Durham template	Biolchini et al. [49]	Petticrew and Roberts [18]	This study	Kitchenham and Brereton [8]
Need for the map	Motivate the need and relevance	√	•	•	•	√	√	•	√	•
	Define objectives and questions	√	•	•	•	√	√	•	√	•
	Consult with target audience to define questions	•	•	•	•	•	•	√	•	•
Study ident.	Choosing search strategy									
	Database search	√	√	√	√	√	√	√	√	√
	Snowballing	√	•	•	√	√	√	√	√	√
	Manual search	√	√	√	√	√	√	√	√	√
	Develop search									
	PICO(C)	√	•	•	•	•	•	√	√	•
	Consult experts	√	•	•	√	√	•	√	√	√
	Iteratively improve search	√	•	•	•	•	•	√	√	•
	Keywords from known papers	√	•	•	•	•	•	√	√	√
	Use standards, encyclopedias	•	•	•	•	•	•	•	√	•
	Search evaluation									
	Paper test-set	√	•	•	√	√	•	√	√	√
	Expert evaluation	•	•	•	•	•	•	√	√	•
	Authors' web pages	•	•	•	•	√	•	•	√	•
	Test–retest	•	•	•	•	√	•	•	√	•
	Inclusion/exclusion									
	Identify objective criteria for decision	√	•	•	•	√	•	√	√	√
	Resolve disagreements among multiple researchers	√	•	•	√	√	•	√	√	√
	Decision rules	•	•	•	•	•	•	•	√	√
Extr./Class.	Extraction process									
	Identify objective criteria for decision	√	•	√	•	√	•	•	√	•
	Obscuring information that could bias	•	•	•	•	•	•	√	•	•
	Resolve disagreements among multiple researchers	√	•	•	•	•	√	√	√	√
	Topic-independent									
	Research type	•	√	•	•	•	•	√	√	•
	Research method	•	•	•	√	•	•	√	√	•
	Study focus	•	•	•	•	•	•	•	√	•
	Contribution type	•	√	√	•	•	•	•	√	•
	Venuge type	•	√	•	•	•	•	•	√	•
	Topic specific									
	Emerging scheme	•	√	•	√	•	•	•	√	•
	Use of standards, etc.	•	•	√	•	√	•	•	√	•
Study validity	Discussion of threats	√	•	•	•	√	•	•	√	•
	Line diagram	•	•	•	•	•	•	√	√	•
	Pie diagram	•	•	•	•	•	•	•	√	•
	Bar plot	•	√	•	•	•	•	•	√	•
	Bubble plot	•	√	•	•	•	•	•	√	√
	Venn diagram	•	•	•	•	•	•	•	√	•
	Heatmap	•	•	•	•	•	•	•	√	•

there are only very few studies in a category, those can be investigated in more depth to see whether they are solving the problem in that category.

When defining the scope and research questions, Kitchenham et al. [15] points out that the questions in systematic mapping studies are less specific than in systematic reviews. Mapping questions are about what we are knowing with respect to a specified topic. A typical example of a research question for mapping studies is provided by Maglyas et al. [58]:

- *"What Do We Know about Software Product Management?"* Often those high level questions have to be further broken down to drive the data extraction, in the case of [58] the questions were:
  - *"What research questions in software product management are being addressed?"*
  - *"What original research exists in the intersection of software product management and cloud (service) environment?"*
  - *"What areas in software product management require more research?"*

While these questions were mainly focused on the coverage of the topic area, typically also questions regarding venues, research methods, and trends are included, for example with regard to combining static and dynamic quality assurance approaches, Elberzhager et al. [35], among others, is asking:

- *"What are existing approaches that combine static and dynamic quality assurance techniques and how can they be classified?"*
- *"In which sources and in which years were approaches regarding the combination of static and dynamic quality assurance techniques published?"*
- *"Is any kind of evidence presented with respect to the combination of quality assurance techniques and if so, which kind of evidence is given?"*

The authors of the research papers mentioned above [58,35] formulated their research questions and built the search with those questions in mind.

5.1.2. Study identification

The search strategies, as can be seen from Table 5, have been very intensively covered in the guidelines for systematic reviews.

We distinguished between developing the search, evaluating the search, and inclusion and exclusion.

Before discussing the individual approaches and when to apply them it is important to reflect on the aim of study identification. In systematic reviews the goal is an exhaustive search where we identify all the relevant evidence [18]. Based on previous evaluations for mapping studies there is an indication that this may not be a realistic goal [7,14]. Wohlin et al. [7] state that having more papers is not necessarily better than having fewer, it depends on whether the papers are a good representation of the population. One challenge is of course that the overall population is not well known beforehand, hence the evaluation can only be done with regard to our awareness. The following questions may help in gaining some initial insights about the population:

- Are different a priori known sub-areas of the field covered? Here, existing classifications of the field (e.g. software testing) may help, as well as experts who may point to classifications or could draw up a map of areas they believe are of relevance. This part is about understanding as much as possible about the population of relevant articles.
- Are the main publication forums specific to this area (e.g. conferences), or general software engineering forums (e.g. journals), represented when identifying relevant articles?
- Are there explanations for major changes in the number of studies published per year? For example, this may point to new areas that should be added to classifications established earlier.

To illustrate the above, we elaborate on an example from a study by Badampudi et al. [59]. In the study Badampudi et al. observed that using what we know about the population to reflect on the search and the results was useful, even though the knowledge of the population was limited. This is similar to the problem of testing, where we do not know the population of defects in the software when we are searching for them. As a solution, similar to testing we could partition the area of interest based on what we know. Badampudi et al. focused on finding studies about making trade-off decisions between different architectural asset types (e.g. open source versus in-house). Based on the asset types they could define partitions of the population that ought to be found (even though not knowing whether there are additional ones). The partitions identified are shown in Fig. 16 as circles. The snowballing result is shown in the figure as triangles, a triangle representing a study. It illustrates how studies were identified using snowball sampling based on citation relationships. With the results they knew that they at least missed out on the parts of the population that they were aware off. Hence, an independent researcher not involved in the snowballing activity conducted a database search to complement the findings. As a result, references in the empty partitions with the numbers four and five could be identified. With that they could not say how well they represent the population overall, but that they performed better with a combination of snowball sampling and database search.

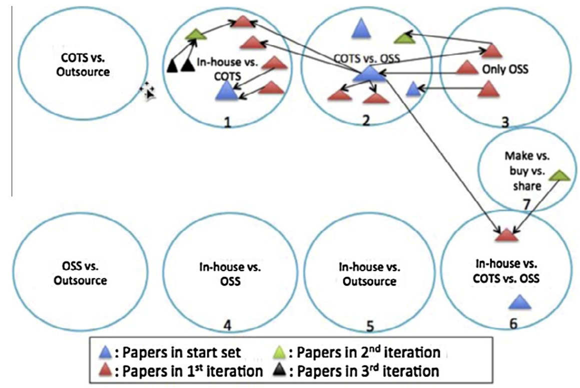

Fig. 16. Reflecting on the search results (from [59]).

Choosing the search strategy: First, one has to decide which strategies in Fig. 6 should be conducted. From this mapping study we found that database search is the most frequent approach. 23 studies did not take any additional help into account while developing the search, while several studies made use of several search strategies.

Though, conducting multiple search strategies may be quite time intensive. Thus, not all of them could be practically applied together in the systematic mapping study, as the trade-off between timely availability of the information and achieving a good overview of the research area needs to be considered. In particular, of the main importance is that the overall conclusions about trends and research gaps do not change. Therefore, we recommend to select one or a sub-set of the activities. Further empirical comparisons on a high number of cases is needed to recommend one approach over another.

Evidence has shown that manual searches are beneficial, and even be more effective in identifying relevant studies [8]. A few mapping studies included in our map have used manual search as their only study identification approach [60,30]. With respect to snowball sampling a recent study found that in the particular context studied, all relevant studies identified by the search could also be obtained by snowball sampling [61]. The key for success is to define a good start set. Wohlin [61] provides some advice to this end:

- Choose articles from different clusters (e.g. communities) that are not likely to cite each other, and hence cannot be found through citation relationships. They should of course be relevant for the research question.
- The number of articles in the start set should not be too small, the size depending on the focus and size of the area, which may not be known beforehand.
- Different authors, years of publication, and also publishers should be covered. For example, when choosing the same authors, these are likely to know about their own work, and hence may limit the breadth of the search.
- Keywords from the research question should be the base for formulating the start set, and are essential when searching for the initial set of papers to start the snowballing from.

However, given that too little evidence exist and the findings may depend on the topic studied, we do not know whether one particular approach should be preferred over another.

Developing the search: Multiple strategies have been proposed to help in developing the search. PICO(C) has only been used by 11 studies, though we believe it to be useful. As one has to determine what a good population is, it is important to reflect on the population and make a conscious decision on whether this should be taken into account in combination with the intervention when searching. The other dimensions (comparison, outcome, and context) may restrict the search too much and remove articles from the topic area. Hence, based on this reflection P and I may be most relevant for mapping studies and should be used.

For software engineering, standards from the IEEE (cf. [39]) as well as ISO/IEC standards are of relevance to identify keywords. Swebok [62] provides an overall structure of the field of software engineering that is widely known, and has been used in existing mapping studies (e.g. [63,39]).

The approaches identified (consult experts, iteratively improve the search, identify keywords from known papers, and use standards, encyclopedias, and thesaurus, see Table 5 and Fig. 7) are not highly time intensive and may greatly improve the quality of the search. Overall, such early quality assurance may save effort due to rework when mistakes were made in the early search activity. They may also help in focusing the search and hence reduce the noise, making the study selection process more time efficient.

When searching the databases, one has to select databases of interest. As recommended by Dyba et al. [17] and Kitchenham and Brereton [8] the use if IEEE and ACM as well as two indexing databases (e.g. Inspec/Compendex and Scopus) is sufficient.

When defining the search string based on the research questions and PICO(C) one also has to consider precision and noise. If a search returns a very large number of irrelevant hits, it may be useful to reflect on whether to further restrict the search (e.g. by making the population more precise). In the case of this study, we received a very high number of hits and had reasonable precision based on the test sets of articles created and from a sampling perspective, though a high number of irrelevant articles (noise) was also obtained. As an example, retrospectively we may have had an equally good result just searching for *("Systematic mapping" and "Systematic map") AND ("software engineering" OR "software development')*, but would have saved significant effort in the study selection phase.

Evaluate the search: To evaluate the search multiple proposals have been made. Utilizing a test-set of known papers was the most common one. This can, for example, be done by asking an expert in the area who should provide a set of ten papers that should be found. If no expert is available, key researchers in the field may have the relevant articles to be found on their web-pages. After the search, it may also be useful to have an expert evaluate the search result.

As the population is unknown when searching, Petticrew and Roberts [18] make the practical suggestion to define a stoppage criterion for the search. For example, if we are using database search and a complementary search approach (such as manual search, snowball sampling) does not add a specific number of new articles to the database search then no further searching is done. Another proposal is to set a time budget based on available funds and only include articles that have been identified, and list all other articles that were not considered [18]. Thereby, as mentioned earlier it is important to evaluate the set of obtained relevant articles against what we know about the population of the topic of interest.

Inclusion and exclusion: The inclusion and exclusion activity requires the definition of criteria. The inclusion and exclusion criteria may refer to (a) the relevance of the topic of the article, (b) the venue of publication, (c) the time period considered, (d) requirements on evaluation, and (e) restrictions with respect to language. Though, in the case of systematic maps (d) should be avoided to also see recent trends that have not reached the maturity for evaluation yet. An example, Laguna and Crespo [64] conducting a mapping of literature related to product line evolution defined the following to include articles:

- *"English peer-reviewed articles in conferences or journals published until Dec. 2011"*
- *"Articles that focus on software product lines"*
- *"Articles that provide some type of evolution of existing software artifacts, included among the terms selected"*
- *"Context, objectives, and research method are reasonably present"*

Also, exclusion criteria have been defined by Laguna and Crespo [64], these were:

- *"Articles that are not related to software product lines"*
- *"Articles that do not imply evolution of any software artifacts"*
- *"Non-peer reviewed publications"*
- *"Articles that are not written in English"*
- *"Context, objectives, or research method are manifestly missing"*

Three strategies have been identified to improve the reliability of the inclusion and exclusion process, namely identify objective criteria for decision (e.g. measure agreement), Identify objective criteria for decision (e.g. measure agreement) and decision rules (see Table 5). Ali and Petersen [65] conducted a study to determine an efficient way of combining these strategies. An overview of the process is shown in Fig. 17, which was derived from the practices identified by Petersen and Ali [28].

First, the selection criteria are specified from the review protocol, and are reviewed among the researchers. Thereafter, Ali and Petersen [65] propose to use a think-aloud protocol where one reviewer describes the thought process of inclusion/exclusion by applying them to one study, which aligns understanding. Thereafter, the reviewers pilot the criteria and determine the level of agreement. Each of these steps may lead to an update of the criteria. Thereafter, the selection is performed on titles and abstracts first. Each researcher evaluates the primary study with either "include", "exclude", or "uncertain".

For two reviewers (R1 and R2) the following combinations are obtained (see Table 6), and can be used for decision rules. Decision rules have only been utilized by very few studies, but have been found empirically useful to decide on whether to take a paper to the next stage in the review (e.g. reading introduction and conclusion due to an unclear abstract, or reading the full-text).

For each of the cases (A–F) decision rules are defined. The most inclusive strategy is to take all articles (A–E) to the next step and only exclude F right away. As has been shown in the study [65] following the most inclusive strategy (A + B + C + D + E) would find all relevant studies, but also has 25% more overhead to strategy (A + B + C + D), where 94% of all studies were identified. Overhead is defined as the *"percentage of irrelevant articles that had to be analyzed"* [65].

Quality assessment: In a systematic review the focus would now be on assessing the quality of included and excluded studies, following one or more of the approaches identified by Kitchenham and Brereton [8]. As we saw in this systematic map, only few studies conducted quality assessment (see Fig. 10). In some cases quality assessment may be useful for a systematic map as well (e.g. to assure that sufficient information is available to actually extract the information). As an example, in this study the mapping process needed to be documented to be able to answer the research

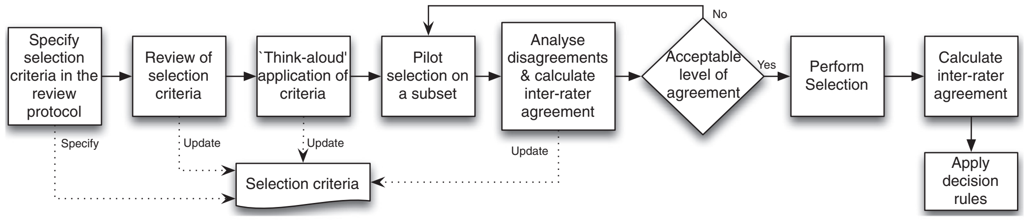

Fig. 17. Study selection process.

Table 6
Combinations for decision rules.

		R2		
		Include	Uncertain	Exclude
R1	Include	A	B	D
	Uncertain	B	C	E
	Exclude	D	E	F

question. In general, we concur with Kitchenham et al. [9] that quality assessment should not pose high requirements on the primary studies as the goal of mapping is to give a broad overview of the topic area.

5.1.3. Data extraction and classification

Extraction and classification process: For the data extraction process two alternatives have been identified. First, more than one researcher is involved and the additional researcher may either check the outcome or conduct all data extractions independently from the first reviewer; if needed a consensus meeting is held. Secondly, the objectivity of the criteria is assessed based on a pilot set of articles and/or post-extraction. As can be seen from Fig. 11 the most common strategy is the first one. Whether an agreement measure can be applied depends on the data extracted. In particular, mapping studies may support calculating agreement as papers are commonly classified into different categories. Though, given the trade-off between availability of the results and the reliability of the classification, Kitchenham and Brereton [31] found it useful to have another person check the data extraction.

For the classification we distinguished between general classification schemes and topic dependent classification schemes.

Topic-independent classification: The goal is that the topic-independent classifications are used by the majority of mapping studies conducted, hence they should be generally applicable. Only by using the same or similar classification schemes consistently enables comparisons. As an example, we could compare the types of research conducted in different fields (requirements elicitation and software product management) with each other, and gain some insights on the comparative maturity of the areas. Also, the consistent use of classifications may help to improve and clarify them. As an example, Wohlin et al. [7] was able to determine inconsistencies as two separately conducted studies utilized the same classification scheme.

Looking at Fig. 12 the most frequently applied facets for classification are venue, research type, and research method. Contribution type is only used by very few studies, and appears to not be of high relevance for the general population of mapping studies. Hence, we encourage the use of venue, research type, and research method.

For the classification of venues the ministry of education of Finland derived a classification[2] based on publication activities of

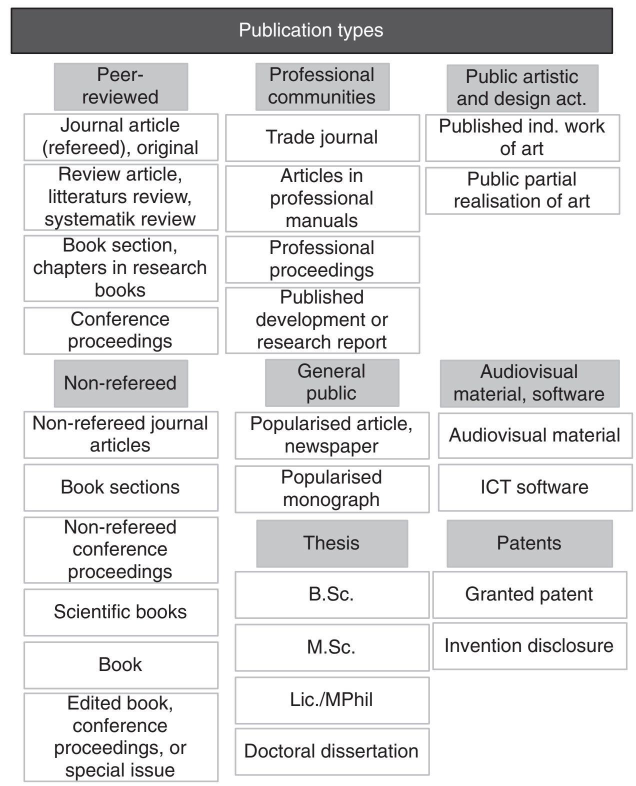

Fig. 18. Classification of publication venues.

universities since 1994. We propose to use their classification as it is a derivation of actual publication activity. The structure of the classification is shown in Fig. 18. The venue classification may also be of use when deciding which publication forums to include and exclude during study selection.

As highlighted by Wohlin et al. [7] researchers were not consistent in their classification when using the research types proposed by Wieringa et al. [11]. Suggestions were to provide more examples and the refinement of the classification. Which research type a study belongs to depends on a number of conditions. Table 7 presents a decision table to disambiguate the classification of studies.

The main confusion with respect to the classification originated from the distinction between validation and evaluation research (R1 and R4 in Table 7). Whether the solution validated or evaluated is novel is not a key criterion. Both have to be empirically evaluated, however, validation is not used in practice (i.e. it is done in the lab), while evaluation studies take place in a real-world industrial context. Furthermore, a new solution may be reported to be used in practice, though it is still a solution proposal if the empirical evaluation is missing.

The classification of research methods has also been frequently reported in mapping studies (see Fig. 12). Easterbrook [66],

[2] http://www.aka.fi/Tiedostot/Tiedostot/Liitetiedostot/OKM_julkaisutyyppiluettelo_2010_en.pdf.

Table 7
Research type classification (T = True, F = False, • = irrelevant or not applicable, R1–R6 refer to rules).

	R1	R2	R3	R4	R5	R6
Conditions						
Used in practice	T	•	T	F	F	F
Novel solution	•	T	F	•	F	F
Empirical evaluation	T	F	F	T	F	F
Conceptual framework	•	•	•	•	T	F
Opinion about something	F	F	F	F	F	T
Authors' experience	•	•	T	•	F	F
Decisions						
Evaluation research	√	•	•	•	•	•
Solution proposal	•	√	•	•	•	•
Validation research	•	•	•	√	•	•
Philosophical papers	•	•	•	•	√	•
Opinion papers	•	•	•	•	•	√
Experience papers	•	•	√	•	•	•

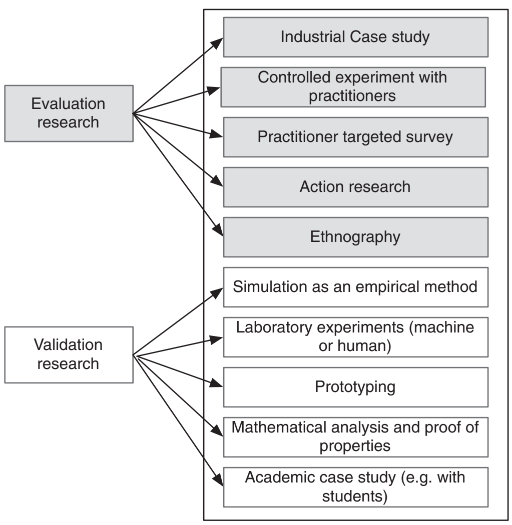

Fig. 19. Classification of research methods.

Wieringa et al. [11] and Wohlin et al. [67] describe the following set of research methodologies that are frequently applied in software engineering: Survey, case study, controlled experiment, action research, ethnography, simulation, prototyping, and mathematical analysis. The method selection has to be consistent with the classification of research type (evaluation research and validation research). As shown in Fig. 19, the methods to be classified map as follows to validation and evaluation research (cf. [11]). Please note that research methods may belong to both categories. For example, experiments with students are classified as validation research, while experiments with practitioners would be classified as evaluation research.

Topic-specific classification: The classification scheme related to the topics can either be emerging from the study, or is based on existing literature. In the mapping studies investigated in this paper the majority of studies built new classifications (see Fig. 13).

Based on availability, it is useful to take an existing classification as the baseline as this supports the comparability between mapping studies. As for the identification of keywords standards (IEEE and ISO/IEC) as well as Swebok [62] are of use. Before starting the classification we propose to consult experts in the area to identify existing classification schemes.

Petersen et al. [2] proposed to use keywording as a means to create a classification scheme and later count the number of articles per category. The process first identifies keywords and concepts from the papers' abstracts. These are then grouped and refined to create a classification scheme. As pointed out by Portillo-Rodríguez et al. [20] the keywording process is not clear. It was intended to be similar to open coding from grounded theory. During open we assign labels or keywords to concepts we find in the text. A number of open codes would be obtained, which thereafter have to be put into an overall structure. In the process the codes representing the categories may be merged or renamed [68]. After having identified the categories the papers are sorted into them, and the number of studies per category are represented. Depending on the quality of abstracts, the process may only be applied on the abstracts. If the abstracts are not clear, further parts of the paper may be considered (e.g. introduction and conclusion).

5.1.4. Visualization

The most commonly used means for visualization are bubble plots, bar plots, and pie diagrams (see Fig. 14). All visualizations are useful in the context of mapping studies. To illustrate the number of studies for a combination of categorizations (e.g. topic category with research type) bubble plots and heat-maps are particularly suited.

5.1.5. Validity threats

For any empirical study the discussion of validity threats is of importance and is a quality criterion for study selection. Hence, possible validity threats shall also be discussed in the context of mapping studies. A very high number of studies reported the validity threats (see Fig. 15). Possible threats for validity are, for example:

- Publication bias, as negative or new and controversial views may not be published (theoretical validity).
- Poorly designed data extraction forms and recording of data (descriptive validity).
- Potential researcher bias in the selection of studies and reporting of the data (theoretical validity).
- The quality of the sample of studies obtained with respect to the targeted population (theoretical validity).
- Generalizability of the results of the mapping. This includes within the population (internal generalizability) and between different populations (external generalizability). As an example, the guidelines identified here are suitable for mapping studies, but only partially so for systematic reviews.
- The reliability of the conclusions drawn in relation to the data collected, e.g. due to a possible bias of the researchers in the interpretation of that data.

5.2. Conducting the mapping

When conducting the mapping the process as defined during the planning phase has to be implemented. Kitchenham and Charters [1] and Petticrew and Roberts [18] recommend to record the information at all stages of the process. It is also noteworthy to highlight that the process is iterative and may require revisions. Furthermore, tools to record data are useful, such as spreadsheets and software for reference management.

5.3. Reporting the mapping

As far as possible, we should aim to have the same reporting structure and style in each systematic map. This makes them easier to evaluate and to compare. Based on the information gathered in this study, we propose the following structure.

- Introduction: Provide information on the background of the topic studied. Describe the need for the mapping, and highlight the usefulness.

Table 8
Activities conducted in this research.

Phase	Actions	Applied
Need for map	Motivate the need and relevance	√
	Define objectives and questions	√
	Consult with target audience to define questions	•
Study ident.	Choosing search strategy	
	Snowballing	√
	Manual	•
	Conduct database search	√
	Develop the search	
	PICO	√
	Consult librarians or experts	•
	Iteratively try finding more relevant papers	•
	Keywords from known papers	•
	Use standards, encyclopedias, and thesaurus	•
	Evaluate the search	
	Test-set of known papers	√
	Expert evaluates result	•
	Search web-pages of key authors	•
	Test–retest	•
	Inclusion and Exclusion	
	Identify objective criteria for decision	•
	Add additional reviewer, resolve disagreements between them when needed	•
	Decision rules	•
Data extr. and class.	Extraction process	
	Identify objective criteria for decision	•
	Obscuring information that could bias	•
	Add additional reviewer, resolve disagreements between them when needed	√
	Test–retest	•
	Classification scheme	
	Research type	•
	Research method	•
	Venue type	•
Validity discussion	Validity discussion/limitations provided	√

- Related work: Provide an overview of existing secondary and tertiary studies in the area.
- Research method: In the research method present the following in separate subsections: research question, search, study selection, data extraction (and if conducted quality assessment), analysis and classification, validity evaluation (discuss the different types [29]).
- Results: Present the outcomes of the study and structure the section with respect to the mapping questions.
- Discussion/Conclusions.
- Appendix with included as well as excluded borderline papers.

Having conducted and reported the process, an important step of evidence-based software engineering is to evaluate the evidence-based process.

5.4. Evaluate the mapping process

Kitchenham and Brereton [8] found that a common request to include in guidelines is a pocket guide for evaluation as support for researchers during the design, and reviewers during the assessment of mapping studies. Based on the findings of practices used, we devised an evaluation rubric. Table 8 presents the actions found relevant to conduct systematic mapping studies, actions that have been applied in this study are indicated by (√). A total of 26 actions have been identified that could be applied by a systematic mapping study.

Based on the quality checklist rubric criteria are defined for identifying the need, study identification, data extraction and classification, and validity discussion. A rubric defines levels of achievement. When reviewing an article, it has to be pointed out that the rating is conducted based on the reported information. Authors filling out the checklist and rubric themselves will notice whether they forgot to report an item of importance. Tables 9–13 present the scoring rubrics. The scores identified in this mapping study are highlighted as bold text.

We also suggest to calculate the ratio of the number of actions taken in comparison to the total number of actions. For this mapping study the ratio is 31%. The checklist and rubric allows to characterize the extent to which actions recommended/used in

Table 9
Rubric: need for review.

Evaluation	Description	Score
No description	The study is not motivated and the goal is not stated	0
Partial evaluation	Motivations and questions are provided	**1**
Full evaluation	Motivations and questions are provided, and have been defined in correspondence with target audience	2

Table 10
Rubric: choosing the search strategy.

Evaluation	Description	Score
No description	Only one type of search has been conducted	0
Minimal evaluation	Two search strategies have been used	**1**
Full evaluation	All three search strategies have been used	2

Table 11
Rubric: evaluation of the search.

Evaluation	Description	Score
No description	No actions have been reported to improve the reliability of the search and inclusion/exclusion	0
Minimal evaluation	At least one action has been taken to improve the reliability of the search xor the reliability of the inclusion/exclusion	**1**
Partial evaluation	At least one action has been taken to improve the reliability of the search and the inclusion/exclusion	2
Full evaluation	All actions identified have been taken	3

Table 12
Rubric: extraction and classification.

Evaluation	Description	Score
No description	No actions have been reported to improve on the extraction process or enable comparability between studies through the use of existing classifications	0
Minimal evaluation	At least one action has been taken to increase the reliability of the extraction process	**1**
Partial evaluation	At least one action has been taken to increase the reliability of the extraction process, and research type and method have been classified	2
Full evaluation	All actions identified have been taken	3

Table 13
Rubric: study validity.

Evaluation	Description	Score
No description	No threats or limitations are described	0
Full evaluation	Threats and limitations are described	**1**

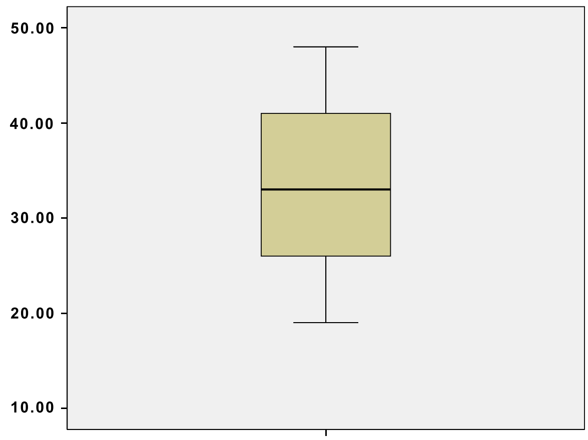

Fig. 20. Distribution of the ratio of the quality checklist items fulfilled.

Table 14
Rubric evaluation of existing studies (• = not applicable).

Activity	No Desc.	Min. E.	Part. E.	Full E.
Need for review	0	•	52	0
Search strategy	30	12	•	10
Evaluate search	23	22	7	0
Extract./Class.	5	32	15	0
Validity	7	•	•	45

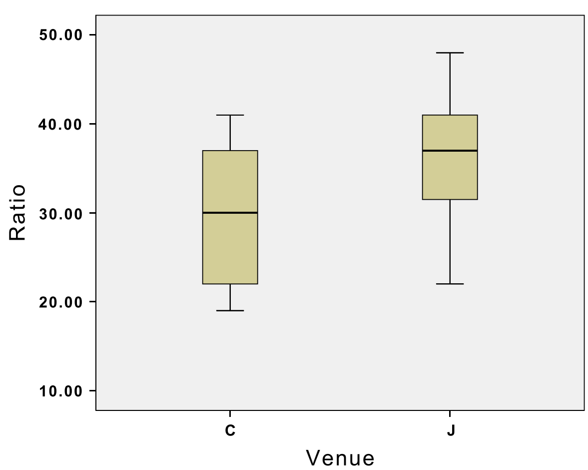

Fig. 21. Distribution of the ratio of the quality checklist items fulfilled between Journals (J) and Conferences/Workshops (C).

literature have been used, which may be an indicator of quality. That is, even though more actions taken increases the reliability of the mapping studies, there maybe a point where the return on investment is very low. Hence, to determine which are the right scores to achieve in the rubric further empirical studies are needed. An example of such a study is the comparison of snowballing and automated search [61].

Fig. 20 shows the distribution of the ratio of quality checklist items fulfilled by the included studies. It is visible that the median quality of studies is 33, with 25% of all studies having a quality score of above 40%.

With respect to the categorization of the rubrics the following scores have been obtained by the studies (see Table 14). As can be seen, the majority of studies falls into the "No description" or "Minimal evaluation" categories for the choice of search strategy and search evaluation. Furthermore, the result for validity and stating the need for the review is very good.

5.5. Dissemination

The classification of publication venues available for publication are provided in Fig. 18. As has been found in this review, a high number of mapping studies has been published in reputable journals. Fig. 21 shows the ratio of checklist items in comparison to the venue selected. It can be observed that the median value for journals is higher than for conference publications, which may be due to the space for reporting, which could be a motivation to favor journals over conferences when reporting extensive literature studies.

6. Conclusions

In this systematic mapping study we identified existing systematic maps and evaluated them with respect to topics investigated, frequency of publication over time, venues of publication, and the process of mapping studies.

Existing guidelines only partially represent the activities actually conducted in systematic mapping studies, hence to have a comprehensive list of activities (and thus being aware of the possible options to make informed decision to use an activity or not) motivated the proposal of an updated guideline. The guideline comprises of a description of activities related to planning, conducting, and reporting of mapping studies. We also propose a guide to evaluate mapping studies. The answers to the research question of our mapping study used to propose the updated guideline are presented below.

RQ1, Guidelines: The most commonly applied guidelines were Kitchenham and Charters [1] and Petersen et al. [2], these being from 2007 and 2008, respectively. It was evident that often multiple guidelines have been used in a single mapping study, which means that the existing individual guidelines are not sufficient to support the complete mapping process. This may also lead to variation in the execution of the process. Hence, the need for new guidelines were further supported by the data collected.

RQ2, Topics: With respect to topic areas covered based on the Swebok classification we found a good coverage of the majority of topics. Software testing is the subject for which the highest number of mapping studies have been performed. Only few studies have been identified in the education and configuration management context.

RQ3, Venues: Around half of the studies are published in conferences and workshops, and the other half in reputable journals. This indicates that mapping studies are considered a valuable scientific contribution. Furthermore, the number of studies has significantly increased in the years 2011 and 2012.

RQ4, Process: We identified the different strategies and application frequency for study identification, extraction and classification, and validity evaluation. Based on the findings the guidelines have been updated. Furthermore, a common request was to include a "pocket-book" for the evaluation of systematic reviews or mapping studies in the guidelines. We provided an evaluation rubric and exemplified it by application to this systematic map and thereafter applied it to the included studies.

An important discussion point in this review was the level of quality. An important reflection was that having a good sample and representation of studies is more important than having a higher number of studies. Furthermore, a high number of strategies has been identified. Conducting them will increase the reliability of the systematic mapping study, and with that the confidence. However, as the goal of a mapping study is to provide an overview and being an input for systematic reviews, a trade-off between effort and the reliability of the outcome has to be made. To make a more informed decision, further studies are needed that compare the different strategies (e.g. database search versus snowballing and manual searching).

Acknowledgments

This work was partly funded by ELLIIT (The Linköping-Lund Initiative on IT and Mobile Communication). We also would like to thank Claes Wohlin for valuable comments on the manuscript.

Appendix A. Included and excluded studies

The following systematic mapping studies are included: [69,70,32,39,40,71,23,5,41,72,3,33,12,34,73,74,60,35,36,13,63,75,76,37,4,77,42,64,58,78,79,22,20,6,43,54,38,80,30,44,81–84,21,45,85,55,86,53,47,46].

The following studies were excluded after full-text reading and quality assessment: [87–96,11,57,56,2,97,16,9,14,15].

Appendix B. Mapping of individual studies to categories

The following tables present the mappings of articles to the categorizations presented in Section 4.4 (see Tables B.15–B.27).

Table B.15
Studies per topic.

Topic	Studies
Software quality	[40,23,35,37,38]
Software tools and methods	[41,74,75,4,20,98]
Software engineering process	[39,74,98,38]
Software engineering management	[22,54,44,84,21,85,55,46]
Software configuration management	[63]
Software testing	[69,71,5,72,12,60,36,13,82,83,53]
Software construction	[64,43]
Software design	[99,70,32,40,42,79,81,45,86,47]
Software requirements	[71,3,77,58,78,100]
Research methods	[33,34,73,76,80,30]
Education	[6]

Table B.16
Venues.

Venues	Studies
Journal	[72,33,12,34,74,35,36,13,76,37,64,22,20,54,38,80,30,81,21,85,55,53,23]
Conference	[69,3,60,70,63,75,4,77,42,58,32,78,6,43,39,44,82–84,45,40,86,47,46,5,41]
Workshop	[73,79,71]

Table B.17
Guidelines followed.

Guide	Studies
(Kitchenham, 2004), and refinement (Kitchenham and Charters, 2007)	[69,39,23,5,41,72,3,33,12,34,73,74,35,36,76,37,4,58,22,6,54,38,80,30,21,45,85,53,47,46]
Petersen et al. (2008)	[70,39,40,71,23,41,72,3,12,74,60,35,36,13,63,75,37,4,77,64,78,79,22,20,6,54,38,81–83,86]
Budgen et al. (2008)	[37,77,81]
Dyb and Dingsyr (2008)	[83]
Bailey et al. (2007)	[69]
Arksey and O'Malley (2005)	[80,44,55]
Jorgensen and Shepperd (2007)	[43]
Durham template	[42]
Biolchini et al. (2005)	[84]
Petticrew and Roberts (2006)	[32,5,82,55,47]

Table B.18
Choosing the search strategy.

Action	Studies
Database search	[69,70,32,39,40,71,23,5,41,72,3,33,12,34,73,74,35,36,13,63,75,76,37,4,77,42,64,58,78,79,22,20,6,43,54,38,80,44,81–84,21,45,85,55,86,53,47,46]
Snowballing	[69,41,12,34,74,35,37,77,64,22,71,44,53,13]
Manual	[69,5,41,12,34,73,74,13,63,37,77,79,80,44,85,55,53,60,30,21]

Table B.19
Development of the search.

Action	Studies
PICO	[3,35,42,22,84,45,53,46,32,76,69]
Consult librarians or experts	[70,41,5,73,80,55]
Iteratively improve the search to find more papers	[71,23,13,4,58,38,80]
Keywords from known papers	[41,72,23,5,73,13,76,64,54,38,80]
Standards, encyclopedias and thesaurus	[69,23,35,36,79,22,38]

Table B.20
Evaluation of the search.

Action	Studies
Test-set of known papers	[71,32,34,13,76,37,22,55]
Expert evaluates result	[39]
Search web-pages of key authors	[53]
Test–retest	[21]

Table B.21
Inclusion and exclusion process.

Action	Studies
Identify objective criteria/evaluate objectivity	[46,36,76,37,22,54,85,35]
Add additional reviewer, resolve disagreements between them when needed	[39,53,69,23,4,30,73,77,80,46,34,71,3,60,76,85,86,54,12,5,35,36]
Decision rules	[40,71,3,54]

Table B.22
Quality assessment.

Value	Studies
Yes	[12,34,73,35,36,76,37,22,54,80,30,44,85,53]
No	[69,70,32,39,40,71,23,5,41,72,3,33,74,60,13,63,75,4,77,42,64,58,78,79,20,6,43,38,81–84,21,45,55,86,47,46]

Table B.23
Data extraction process.

Action	Studies
Identify objective criteria for decision	[80,20,45,37]
Add additional reviewer, resolve disagreements between them when needed	[73,64,76,41,6,30,80,20,72,33,60,77,45,85,86,13,37]
Test–retest	[21]

Table B.24
Topic-independent classification.

Facet	Studies
Research method	[32,5,72,3,34,73,36,63,37,58,78,6,54,38,80,85,86]
Research type	[40,71,23,41,12,74,60,13,63,4,77,42,64,22,38,82,83,21,85,86,53]
Study focus	[39,23,3,73,36,63,78,22,80,84,86]
Contribution type	[71,23,74,36,13,86]
Venue	[70,32,39,40,71,23,72,3,33,35,36,13,63,64,58,22,43,54,38,30,44,81,21,86,53,47,46]

Table B.25
Topic-related classification.

Classification	Studies
Emerging classification	[70,32,40,71,23,5,41,3,33,34,74,35,36,13,63,75,76,37,4,42,64,58,78,79,22,6,43,38,30,81–84,21,45,55,86,53,47,46,69]
Existing scheme	[72–74,60,77,42,20,54,38,80,44,45,85,39,69,63]

Table B.26
Visualizations.

Visualization	Studies
Line diagram	[39,74]
Pie diagram	[40,72,60,13,20,6,54,38,81,83,84,86]
Bar plot	[40,71,23,41,72,33,12,60,35,36,63,37,4,77,78,22,54,38,80,82,85,86]
Bubble plot	[69,70,39,71,23,41,74,60,35,36,13,75,37,4,77–79,22,38,80,44,81–83]
Venn diagram	[78,54,82]
Heatmap	[55]

Table B.27
Validity threats discussion.

Value	Studies
Yes	[69,32,23,5,41,72,3,33,12,34,73,74,60,35,36,13,75,76,37,4,77,42,64,58,78,79,22,20,43,54,38,80,30,44,81–84,21,45,85,86,53,47,46]
No	[70,39,40,71,63,6,55]

References

[1] B. Kitchenham, S. Charters, Guidelines for Performing Systematic Literature Reviews in Software Engineering, Tech. rep., Technical report, EBSE Technical Report EBSE-2007-01, 2007

[2] K. Petersen, R. Feldt, S. Mujtaba, M. Mattsson, Systematic mapping studies in software engineering, in: 12th International Conference on Evaluation and Assessment in Software Engineering, vol. 17, 2008, p. 1.

[3] N. Condori-Fernandez, M. Daneva, K. Sikkel, R. Wieringa, O. Dieste, O. Pastor, A systematic mapping study on empirical evaluation of software requirements specifications techniques, in: Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement, IEEE Computer Society, 2009, pp. 502–505.

[4] S. Jalali, C. Wohlin, Agile practices in global software engineering-a systematic map, in: 2010 5th IEEE International Conference on Global Software Engineering (ICGSE), IEEE, 2010, pp. 45–54.

[5] E. Barreiros, A. Almeida, J. Saraiva, S. Soares, A systematic mapping study on software engineering testbeds, in: 2011 International Symposium on Empirical Software Engineering and Measurement (ESEM), IEEE, 2011, pp. 107–116.

[6] M.M. Qadir, M. Usman, Software engineering curriculum: a systematic mapping study, in: 2011 5th Malaysian Conference in Software Engineering (MySEC), IEEE, 2011, pp. 269–274.

[7] C. Wohlin, P. Runeson, P.A. da Mota Silveira Neto, E. Engström, I. do Carmo Machado, E.S. de Almeida, On the reliability of mapping studies in software engineering, J. Syst. Softw. 86 (10) (2013) 2594–2610.

[8] B. Kitchenham, P. Brereton, A systematic review of systematic review process research in software engineering, Inf. Softw. Technol. 55 (12) (2013) 2049–2075.

[9] B.A. Kitchenham, D. Budgen, P. Brereton, The value of mapping studies-a participant-observer case study, in: Proceedings of the 14th International Conference on Evaluation and Assessment in Software Engineering, British Computer Society, 2010, pp. 25–33.

[10] J.M. Verner, O.P. Brereton, B.A. Kitchenham, M. Turner, M. Niazi, Risks and risk mitigation in global software development: a tertiary study, Inf. Softw. Technol. 56 (1) (2014) 54–78.

[11] R. Wieringa, N. Maiden, N. Mead, C. Rolland, Requirements engineering paper classification and evaluation criteria: a proposal and a discussion, Requirements Eng. 11 (1) (2006) 102–107.

[12] P.A. da Mota Silveira Neto, I.d. Carmo Machado, J.D. McGregor, E.S. De Almeida, S.R. de Lemos Meira, A systematic mapping study of software product lines testing, Inf. Softw. Technol. 53 (5) (2011) 407–423.

[13] E. Engström, P. Runeson, Software product line testing–a systematic mapping study, Inf. Softw. Technol. 53 (1) (2011) 2–13.

[14] B. Kitchenham, P. Brereton, D. Budgen, Mapping Study Completeness and Reliability-A Case Study.

[15] B. Kitchenham, P. Brereton, D. Budgen, The educational value of mapping studies of software engineering literature, 2010 ACM/IEEE 32nd International Conference on Software Engineering, vol. 1, IEEE, 2010, pp. 589–598.

[16] B.A. Kitchenham, D. Budgen, O. Pearl Brereton, Using mapping studies as the basis for further research–a participant-observer case study, Inf. Softw. Technol. 53 (6) (2011) 638–651.

[17] T. Dyba, T. Dingsoyr, G.K. Hanssen, Applying systematic reviews to diverse study types: an experience report, in: First International Symposium on Empirical Software Engineering and Measurement, 2007, ESEM 2007, IEEE, 2007, pp. 225–234.

[18] M. Petticrew, H. Roberts, Systematic Reviews in the Social Sciences: A Practical Guide, John Wiley & Sons, 2008.

[19] S. Jalali, C. Wohlin, Systematic literature studies: database searches vs. backward snowballing, in: Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, ACM, 2012, pp. 29–38.

[20] J. Portillo-Rodríguez, A. Vizcaíno, M. Piattini, S. Beecham, Tools used in global software engineering: a systematic mapping review, Inf. Softw. Technol. 54 (7) (2012) 663–685.

[21] B. Kitchenham, Whats up with software metrics?–A preliminary mapping study, J. Syst. Softw. 83 (1) (2010) 37–51.

[22] K. Petersen, Measuring and predicting software productivity: a systematic map and review, Inf. Softw. Technol. 53 (4) (2011) 317–343.

[23] S. Barney, K. Petersen, M. Svahnberg, A. Aurum, H. Barney, Software quality trade-offs: a systematic map, Inf. Softw. Technol. 54 (7) (2012) 651–662.

[24] S. Jalali, C. Wohlin, Global software engineering and agile practices: a systematic review, J. Softw.: Evol. Process 24 (6) (2012) 643–659.

[25] A. Tahir, S.G. MacDonell, A systematic mapping study on dynamic metrics and software quality, in: 2012 28th IEEE International Conference on Software Maintenance (ICSM), IEEE, 2012, pp. 326–335.

[26] J.W.C. Llanos, S.T.A. Castillo, Differences between traditional and open source development activities, in: Product-Focused Software Process Improvement, Springer, 2012, pp. 131–144.

[27] J. Sousa Gomes, P.A. da Mota Silveira Neto, D.S. Cruzes, E. Santana de Almeida, 25 years of software engineering in brazil: an analysis of sbes history, in: 2011 25th Brazilian Symposium on Software Engineering (SBES), IEEE, 2011, pp. 4–13.

[28] K. Petersen, N.B. Ali, Identifying strategies for study selection in systematic reviews and maps, in: Proceedings of the 5th International Symposium on Empirical Software Engineering and Measurement, ESEM 2011, 2011, pp. 351–354.

[29] K. Petersen, C. Gencel, Worldviews, research methods, and their relationship to validity in empirical software engineering research, in: 2013 Joint Conference of the 23rd International Workshop on Software Measurement and the 2013 Eighth International Conference on Software Process and Product Measurement (IWSM-MENSURA), IEEE, 2013, pp. 81–89.

[30] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, S. Linkman, Systematic literature reviews in software engineering–a systematic literature review, Inf. Softw. Technol. 51 (1) (2009) 7–15.

[31] P. Brereton, B.A. Kitchenham, D. Budgen, M. Turner, M. Khalil, Lessons from applying the systematic literature review process within the software engineering domain, J. Syst. Softw. 80 (4) (2007) 571–583.

[32] J. Bailey, D. Budgen, M. Turner, B. Kitchenham, P. Brereton, S.G. Linkman, Evidence relating to object-oriented software design: a survey, in: ESEM, vol. 7, Citeseer, 2007, pp. 482–484.

[33] D.S. Cruzes, T. Dybå, Research synthesis in software engineering: a tertiary study, Inf. Softw. Technol. 53 (5) (2011) 440–455.

[34] F.Q. Da Silva, A.L. Santos, S. Soares, A.C.C. França, C.V. Monteiro, F.F. Maciel, Six years of systematic literature reviews in software engineering: an updated tertiary study, Inf. Softw. Technol. 53 (9) (2011) 899–913.

[35] F. Elberzhager, J. Münch, V.T.N. Nha, A systematic mapping study on the combination of static and dynamic quality assurance techniques, Inf. Softw. Technol. 54 (1) (2012) 1–15.

[36] F. Elberzhager, A. Rosbach, J. Münch, R. Eschbach, Reducing test effort: a systematic mapping study on existing approaches, Inf. Softw. Technol. 54 (10) (2012) 1092–1106.

[37] A. Fernandez, E. Insfran, S. Abrahão, Usability evaluation methods for the web: a systematic mapping study, Inf. Softw. Technol. 53 (8) (2011) 789–817.

[38] R. Wendler, The maturity of maturity model research: a systematic mapping study, Inf. Softw. Technol. 54 (12) (2012) 1317–1339.

[39] S.T. Acuña, J.W. Castro, O. Dieste, N. Juristo, A systematic mapping study on the open source software development process, in: Evaluation & Assessment in Software Engineering (EASE 2012), 16th International Conference on, IET, 2012, pp. 42–46.

[40] A. Arshad, M. Usman, Security at software architecture level: a systematic mapping study, in: 15th Annual Conference on Evaluation & Assessment in Software Engineering (EASE 2011), IET, 2011, pp. 164–168.

[41] J.F. Bastos, P.A. da Mota Silveira Neto, E.S. de Almeida, S.R. de Lemos Meira, Adopting software product lines: a systematic mapping study, in: 15th Annual Conference on Evaluation & Assessment in Software Engineering (EASE 2011), IET, 2011, pp. 11–20.

[42] D.S. Kusumo, M. Staples, L. Zhu, H. Zhang, R. Jeffery, Risks of Off-the-Shelf-Based Software Acquisition and Development: A Systematic Mapping Study and a Survey.

[43] T. Shippey, D. Bowes, B. Chrisianson, T. Hall, A Mapping Study of Software Code Cloning.

[44] L.L. Lobato, I. do Carmo Machado, P. d. M.S. Neto, E.S. de Almeida, S.R. de Lemos Meira, Risk Management in Software Engineering: A Scoping Study.
[45] M. Anjum, D. Budgen, A mapping study of the definitions used for service oriented architecture, in: 16th International Conference on Evaluation & Assessment in Software Engineering (EASE 2012), IET, 2012, pp. 57–61.
[46] J. Saraiva, E. Barreiros, A. Almeida, F. Lima, A. Alencar, G. Lima, S. Soares, F. Castor, Aspect-oriented software maintenance metrics: A systematic mapping study, in: 16th International Conference on Evaluation & Assessment in Software Engineering (EASE 2012), IET, 2012, pp. 253–262.
[47] R. Pretorius, D. Budgen, A mapping study on empirical evidence related to the models and forms used in the uml, in: Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, ACM, 2008, pp. 342–344.
[48] B. Kitchenham, Procedures for performing systematic reviews, Keele, UK, Keele University, vol. 33, 2004, p. 2004.
[49] J. Biolchini, P.G. Mian, A.C.C. Natali, G.H. Travassos, Systematic review in software engineering, System Engineering and Computer Science Department COPPE/UFRJ, Technical Report ES vol. 679(05), 2005, p. 45.
[50] H. Arksey, L. O'Malley, Scoping studies: towards a methodological framework, Int. J. Soc. Res. Meth. 8 (1) (2005) 19–32.
[51] T. Dybå, T. Dingsøyr, Empirical studies of agile software development: a systematic review, Inf. Softw. Technol. 50 (9) (2008) 833–859.
[52] M. Jorgensen, M. Shepperd, A systematic review of software development cost estimation studies, IEEE Trans. Softw. Eng. 33 (1) (2007) 33–53.
[53] M. Palacios, J. García-Fanjul, J. Tuya, Testing in service oriented architectures with dynamic binding: a mapping study, Inf. Softw. Technol. 53 (3) (2011) 171–189.
[54] I. Steinmacher, A.P. Chaves, M.A. Gerosa, Awareness support in distributed software development: a systematic review and mapping of the literature, Computer Supported Cooperative Work (CSCW).
[55] F.Q. Silva, R. Prikladnicki, A.C.C. França, C.V. Monteiro, C. Costa, R. Rocha, An evidence-based model of distributed software development project management: results from a systematic mapping study, J. Softw.: Evol. Process 24 (6) (2012) 625–642.
[56] D. Budgen, M. Turner, P. Brereton, B. Kitchenham, Using mapping studies in software engineering, in: Proceedings of PPIG, vol. 8, 2008, pp. 195–204.
[57] H. Arksey, L. O'Malley, Scoping studies: towards a methodological framework, Int. J. Soc. Res. Meth. 8 (1) (2005) 19–32.
[58] A. Maglyas, U. Nikula, K. Smolander, What do we know about software product management?–A systematic mapping study, in: 2011 Fifth International Workshop on Software Product Management (IWSPM), IEEE, 2011, pp. 26–35.
[59] D. Badampudi, C. Wohlin, K. Petersen, Worldviews, research methods, and their relationship to validity in empirical software engineering research, in: S19th International Conference on Evaluation and Assessment in Software Engineering (EASE 2015), ACM, 2015.
[60] V.H. Durelli, R.F. Araujo, M.A. Silva, R.A. Oliveira, J.C. Maldonado, M.E. Delamaro, What a long, strange trip it's been: past, present, and future perspectives on software testing research, in: 2011 25th Brazilian Symposium on Software Engineering (SBES), IEEE, 2011, pp. 30–39.
[61] C. Wohlin, Guidelines for snowballing in systematic literature studies and a replication in software engineering, in: 18th International Conference on Evaluation and Assessment in Software Engineering, EASE '14, 2014, p. 38.
[62] A. Abran, P. Bourque, R. Dupuis, J.W. Moore, Guide to the Software Engineering Body of Knowledge-SWEBOK, IEEE Press, 2001.
[63] S.S.M. Fauzi, P.L. Bannerman, M. Staples, Software configuration management in global software development: A systematic map, in: 2010 17th Asia Pacific Software Engineering Conference (APSEC), IEEE, 2010, pp. 404–413.
[64] M.A. Laguna, Y. Crespo, A systematic mapping study on software product line evolution: From legacy system reengineering to product line refactoring, Science of Computer Programming.
[65] N.B. Ali, K. Petersen, Evaluating strategies for study selection in systematic literature studies, in: Proceedings of the 8th International Symposium on Empirical Software Engineering and Measurement, ESEM 2014, 2014.
[66] S. Easterbrook, J. Singer, M.-A. Storey, D. Damian, Selecting empirical methods for software engineering research, in: Guide to Advanced Empirical Software Engineering, Springer, 2008, pp. 285–311.
[67] C. Wohlin, P. Runeson, M. Höst, M.C. Ohlsson, B. Regnell, A. Wesslén, Experimentation in Software Engineering, Springer, 2012.
[68] J. Corbin, A. Strauss, Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory, Sage, 2008.
[69] W. Afzal, R. Torkar, R. Feldt, A systematic mapping study on non-functional search-based software testing, in: SEKE, vol. 8, 2008, pp. 488–493.
[70] E.A. Antonio, F.C. Ferrari, S. Ferraz Fabbri, A systematic mapping of architectures for embedded software, in: 2012 Second Brazilian Conference on Critical Embedded Systems (CBSEC), IEEE, 2012, pp. 18–23.
[71] Z.A. Barmi, A.H. Ebrahimi, R. Feldt, Alignment of requirements specification and testing: a systematic mapping study, in: 2011 IEEE Fourth International Conference on Software Testing, Verification and Validation Workshops (ICSTW), IEEE, 2011, pp. 476–485.
[72] C. Catal, D. Mishra, Test case prioritization: a systematic mapping study, Softw. Qual. J. 21 (3) (2012) 445–478.
[73] F.Q. da Silva, M. Suassuna, R.F. Lopes, T.B. Gouveia, A.C.A. França, J.P.N. de Oliveira, L.F. de Oliveira, A.L. Santos, Replication of empirical studies in software engineering: preliminary findings from a systematic mapping study, in: 2011 Second International Workshop on Replication in Empirical Software Engineering Research (RESER), IEEE, 2011, pp. 61–70.
[74] I.F. da Silva, P.A. da Mota Silveira Neto, P. O'Leary, E.S. de Almeida, S.R. de Lemos Meira, Agile software product lines: a systematic mapping study, Softw.: Pract. Exper. 41 (8) (2011) 899–920.
[75] D. Feitosa, K.R. Felizardo, L.B.R. de Oliveira, D. Wolf, E.Y. Nakagawa, Software engineering in the embedded software and mobile robot software development: a systematic mapping., in: SEKE, 2010, pp. 738–741.
[76] 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, J. Softw. 7 (2) (2012) 450–461.
[77] J. Lemos, C. Alves, L. Duboc, G.N. Rodrigues, A systematic mapping study on creativity in requirements engineering, in: Proceedings of the 27th Annual ACM Symposium on Applied Computing, ACM, 2012, pp. 1083–1088.
[78] J.G. Mohebzada, G. Ruhe, A. Eberlein, Systematic mapping of recommendation systems for requirements engineering, in: 2012 International Conference on Software and System Process (ICSSP), IEEE, 2012, pp. 200–209.
[79] E.Y. Nakagawa, D. Feitosa, K.R. Felizardo, Using systematic mapping to explore software architecture knowledge, in: Proceedings of the 2010 ICSE Workshop on Sharing and Reusing Architectural Knowledge, ACM, 2010, pp. 29–36.
[80] F.Q. Da Silva, M. Suassuna, A.C.C. França, A.M. Grubb, T.B. Gouveia, C.V. Monteiro, I.E. dos Santos, Replication of empirical studies in software engineering research: a systematic mapping study, Empirical Softw. Eng.
[81] M. Guessi, L.B.R. Oliveira, E.Y. Nakagawa, Extensions of uml to model aspect-oriented software systems, CLEI Electron. J. 14 (1) (2011). 3–3.
[82] A. Sharma, T.D. Hellmann, F. Maurer, Testing of web services-a systematic mapping, in: 2012 IEEE Eighth World Congress on Services (SERVICES), IEEE, 2012, pp. 346–352.
[83] T.D. Hellmann, A. Sharma, J. Ferreira, F. Maurer, Agile testing: past, present, and future–charting a systematic map of testing in agile software development, in: Agile Conference (AGILE), 2012, IEEE, 2012, pp. 55–63.
[84] J. Menezes Jr., C. Gusmão, H. Moura, Indicators and metrics for risk assessment in software projects: a mapping study, in: Proc. 5th Experimental Software Engineering Latin American Workshop (ESELAW 2008), 2008.
[85] F.O. Bjørnson, T. Dingsøyr, Knowledge management in software engineering: a systematic review of studied concepts, findings and research methods used, Inf. Softw. Technol. 50 (11) (2008) 1055–1068.
[86] E. Murugesupillai, B. Mohabbati, D. Gašević, A preliminary mapping study of approaches bridging software product lines and service-oriented architectures, Proceedings of the 15th International Software Product Line Conference, vol. 2, ACM, 2011, p. 11.
[87] T.R. Cavalcanti, F.Q. da Silva, Historical, conceptual, and methodological aspects of the publications of the brazilian symposium on software engineering: a systematic mapping study, in: 2011 25th Brazilian Symposium on Software Engineering (SBES), IEEE, 2011, pp. 14–23.
[88] V.H. Durelli, K.R. Felizardo, M.E. Delamaro, Systematic mapping study on high-level language virtual machines, in: Virtual Machines and Intermediate Languages, ACM, 2010, p. 4.
[89] K.R. Felizardo, E. Nakawaga, D. Feitosa, R. Minghim, J.C. Maldonado, An approach based on visual text mining to support categorization and classification in the systematic mapping, in: Proc. of EASE, vol. 10, 2010, pp. 1–10.
[90] L.L. Fortaleza, T. Conte, S. Marczak, R. Prikladnicki, Towards a gse international teaching network: mapping global software engineering courses, in: Collaborative Teaching of Globally Distributed Software Development Workshop (CTGDSD), 2012, IEEE, 2012, pp. 1–5.
[91] N. Kerzazi, M. Lavallée, P.N. Robillard, Mapping knowledge into software process, in: Computing in the Global Information Technology (ICCGI), 2010 Fifth International Multi-Conference on, IEEE, 2010, pp. 180–184.
[92] L. Mufioz, J.-N. Mazon, J. Trujillo, Etl process modeling conceptual for data warehouses: a systematic mapping study, Latin Am. Trans. IEEE (Revista IEEE America Latina) 9 (3) (2011) 358–363.
[93] P.A. da Mota Silveira Neto, I. do Carmo Machado, J.D. McGregor, E.S. de Almeida, S.R. de Lemos Meira, Corrigendum to: a systematic mapping study of software product lines testing [inf. softw. technol. 53(5) (2011) 407–423], Inf. Softw. Technol. 54 (7) (2012) 802.
[94] N.M.N. Daud, W.W. Kadir, Systematic mapping study of quality attributes measurement in service oriented architecture, 2012 8th International Conference onInformation Science and Digital Content Technology (ICIDT), vol. 3, IEEE, 2012, pp. 626–631.
[95] F.Q. Da Silva, A.L. Santos, S. Soares, A.C.C. França, C.V. Monteiro, F.F. Maciel, Six years of systematic literature reviews in software engineering: an updated tertiary study, Inf. Softw. Technol. 53 (9) (2011) 899–913.
[96] E. Takeo Ueda, W.V. Ruggiero, A systematic mapping on the role-permission relationship in role based access control models, Latin Am. Trans., IEEE (Revista IEEE America Latina) 10 (1) (2012) 1243–1250.
[97] K. Petersen, N.B. Ali, Identifying strategies for study selection in systematic reviews and maps, Empirical Softw. Eng.
[98] I.F. da Silva, P.A. da Mota Silveira Neto, P. O'Leary, E.S. de Almeida, S.R. de Lemos Meira, Agile software product lines: a systematic mapping study, Softw.: Practice Exper. 41 (8) (2011) 899–920.
[99] M. Guessi, L.B.R. Oliveira, E.Y. Nakagawa, Modeling aspect-oriented software systems using uml: a systematic mapping, in: Proceedings of the 36 Latin American Conference of Informatics (CLEI 2010), IEE, 2011, pp. 269–274.
[100] S. Mujtaba, K. Petersen, R. Feldt, M. Mattsson, Software product line variability: a systematic mapping study, School of Engineering, Blekinge Inst. of Technology.