Using Forward Snowballing to update Systematic Reviews in Software Engineering

Katia Romero Felizardo
Federal Technological University of Paraná – UTFPR/CP
Departament of Computer
Cornélio Procópio, PR - Brazil
katiascannavino@utfpr.edu.br

Emilia Mendes
Blekinge Institute of Technology and University of Oulu
Depart. of Software Engineering
Karlskrona - Sweden
emilia.mendes@bth.se

Marcos Kalinowski
Fluminense Federal University
Graduate Programm in Computing
Niterói, RJ - Brazil
kalinowski@ic.uff.br

Érica Ferreira Souza
Federal Technological University of Paraná – UTFPR/CP
Departament of Computer
Cornélio Procópio, PR - Brazil
ericasouza@utfpr.edu.br

Nandamudi L. Vijaykumar
Department of Comp. and Applied Mathematics
National Institute for Space Research – INPE
São José dos Campos, SP – Brazil
vijay.nl@inpe.br

ABSTRACT

Background: A Systematic Literature Review (SLR) is a methodology used to aggregate relevant evidence related to one or more research questions. Whenever new evidence is published after the completion of a SLR, this SLR should be updated in order to preserve its value. However, updating SLRs involves significant effort. **Objective:** The goal of this paper is to investigate the application of forward snowballing to support the update of SLRs. **Method:** We compare outcomes of an update achieved using the forward snowballing versus a published update using the search-based approach, i.e., searching for studies in electronic databases using a search string. **Results:** Forward snowballing showed a higher precision and a slightly lower recall. It reduced in more than five times the number of primary studies to filter however missed one relevant study. **Conclusions:** Due to its high precision, we believe that the use of forward snowballing considerably reduces the effort in updating SLRs in Software Engineering; however the risk of missing relevant papers should not be underrated.

Keywords

Systematic literature reviews, forward snowballing.

1. INTRODUCTION

Evidence-Based Software Engineering (EBSE) provides knowledge about when, how and within which context technologies, processes, methods or tools are appropriate for software Engineering (SE) practices. Systematic Literature Reviews (SLR) have provided mechanisms to identify and aggregate evidence on several research topics in SE [1].

The SLR process was introduced in SE in 2004 [2]. Since its introduction, SLRs gained substantial importance [3][4][5][6] and have been applied to various topics of interest by both researchers and practitioners [7][8][9][10][11][12]. The reasons why more and more SLRs are conducted every year can be related to the advantages of SLRs, including reduced likelihood of bias in results and the potential ability to combine data from various quantitative studies by using meta-analysis. Despite its importance, SLRs that are not maintained (i.e., updated) might become outdated or misleading. Therefore, the update of SLRs is quite an important issue in SE. However, even when the same authors update their reviews, searching new evidence can take considerable time.

Thus, it is beneficial to have search approaches that support the update of SLRs, such as database searches and snowballing. In the case of database searches, looking for relevant studies, systematic searches are conducted in databases using well-defined search strings. The definition of keywords and their appropriate combinations for search purposes are a challenging aspect of the SLR process [13]. One reason for this difficulty is the lack of formalization of terminology [14]. The probability that two researchers use the same term to refer to the same concept is often lower than 20% [15]. Therefore, a direct comparison of terms may not be sufficient. Sjøberg et al. [16] agree that there is no common terminology and appropriate descriptors and keywords in the SE area. Additionally, a search strategy based on search strings usually results in a large set of studies to be read and analyzed by the researchers [17]. Moreover, electronic publication databases were not designed to support SLRs [18][19]. Consequently, the retrieval facilities offered by electronic resources are limited.

Although Webster and Watson [20] have recommended to use database search as the main approach for conducting the search of studies in SLRs, there are other search strategies that can be employed. One of which is called Snowballing, which is an alternative approach that does not require the use of predetermined search strings. The first step of the snowballing process involves the identification of a set of studies as a starting point (seed set) [21]. Next, the reviewer analyzes the reference lists of those studies looking for other relevant studies, and so on. The process ends when no more relevant studies are found. This search strategy is known as backward snowballing. Forward snowballing involves searching for studies that cite the studies contained in the seed set. It is a way to search forward in time counting from the publication of a relevant study to identify additional relevant studies published since then. One known difficulty relating to the use of snowballing is the selection of studies to be part of the seed set, as these should be relevant and seminal studies. However, in the context of updating SLRs, the key studies must have already been identified for inclusion in a previous SLR, thus the seed set already exists.

ESEM '16, September 08-09, 2016, Ciudad Real, Spain

DOI: http://dx.doi.org/10.1145/2961111.2962630

Overall, database searches using search strings have been in use more extensively than snowballing when carrying out SLRs in SE. However, within such context, a key issue is to evaluate the use of the snowballing as a search strategy to systematic reviews and to compare it against database searches [22]. Wholin [23] mentions that snowballing could be useful to update SLRs, since new studies are very likely to cite at least one study among those that were considered relevant in a previous SLR in the area. He hypothesizes that snowballing is a better approach than a database search to update SLR; however, there is little evidence in support of this [23].

The goal of the research detailed herein is to propose and evaluate the use of a forward snowballing approach to support finding new evidence to update SLRs. This approach is assessed through the replication of an updated and published SLR.

The remainder of this paper is organized as follows. In Section 2, an overview of related work is presented. In Section 3 we detail the forward snowballing approach, followed by a presentation and discussion of results and limitations in Section 4. Finally, Section 5 concludes our work and suggests directions for future work.

2. RELATED WORK

In terms of ensuring the ongoing relevance, three studies have proposed processes for updating SLRs in Software Engineering [24][25][26].

Felizardo et al. [25] proposed an approach named USR-VTM, to support the update of SLRs. This approach is based on Visual Text Mining (VTM) techniques, which are used to help in the selection of new evidence in the form of studies. They also presented a tool, named Revis, which implements the VTM techniques. Their approach was evaluated by comparing the outcomes achieved using USR-VTM with the traditional (database-search-based) selection. Their results showed that USR-VTM has the potential to increase the number of studies correctly included, when compared to the traditional approach.

The processes by Ferrari et al. [24] and Dieste et al. [26] are composed of three phases: planning, review execution, and analysis of results. During the planning phase, the review protocol is revisited in order to identify required changes to make it suitable for updating the SLR. For example, the exclusion criteria requires changes to restrict the search only to studies published after the previous review. The second phase includes discarding primary studies that overlap with the set of studies retrieved in the previous review. This phase also includes similar activities of the SLR process and, in particular, during data extraction, information extracted from studies selected in the new review must be merged with information from the previous review. Finally, the third phase also includes the same activities of the SLR process.

In spite of these initiatives, the search strategy of the processes above-mentioned is in general based on databases, and the use of search strings. In Kitchenham's and Brereton's guidelines [27], snowballing is recommended as part of a search strategy to retrieve studies for SLRs. The use of snowballing as the main search strategy in SLRs in SE has also been investigated by Skoglund and Runeson [28], who presented a search strategy based on references between papers. In their study, they used the reference list of a relevant paper, named as "*take-off paper*", to start the search. In addition, "c*ardinal papers*" (studies that are referred by many authors) are also used to guide the search for new evidence.

Other studies have also compared snowballing to database searches (e.g. [22][23]). Badampudi et al. [22] evaluated the efficiency and reliability of snowballing by comparing it with database search. They concluded that the efficiency of snowballing is comparable to that obtained via database search. Wohlin [23] carried out the comparison using the results from applying snowballing to the results from a previously conducted SLR that employed database searches. His conclusion was that snowballing is a good alternative to search for relevant studies in SLRs.

Similar to these works [22][23][28], we also investigate the use of snowballing as a search strategy to find evidence for SLRs, however, in our case, the goal is to employ forward snowballing to the specific problem of updating SLRs in SE, using the previously identified studies as the seed set.

3. CASE STUDY DESCRIPTION

The case study detailed herein hypothesizes that forward snowballing supports the search for relevant literature during the update of SLRs with higher recall and precision, when compared to employing database searches (traditional approach). The case study's description based on the case study guidelines from [29].

3.1 Case Study Objective

This case study investigates the recall and precision of using forward snowballing to support searching for new evidence to update SLRs in SE, and compares these results with the traditional approach. Our research questions (RQs) are as follows:

RQ1: *Does the use of forward snowballing improve the* ***recall*** *of the search for new evidence to update SLRs, when compared to the traditional approach?*

RQ2: *Does the use of forward snowballing improve the* ***precision*** *of the search for new evidence to update SLRs, when compared to the traditional approach?*

3.2 Instrumentation and Procedure

We chose the SLR by Kitchenham et al. [30] as the SLR to be updated using forward snowballing. This SLR was published in 2007 [30], and will hereafter called SLR1. We selected SLR1 for the following reasons:

1. SLR1 has been recently updated [31]. The updated version will be hereafter called SLR2. SLR2 used the traditional approach using most of the databases and the same search strings employed in SLR1; and
2. The research topic focus of SLR1 & SLR2 represents a very specific area within software engineering with a small number of studies; therefore, we believed this would be a more controlled context to test our hypothesis.

The goal of SLR1 and SLR2 was to gather evidence from studies that compared predictions between cross- and within-company models aiming to determine under which circumstances individual organizations would be able to rely on cross-company-based estimation models [30][31].

The forward snowballing approach was evaluated using the measures Recall and Precision [32]. Recall is the ability of a search approach to obtain all relevant studies. Since we cannot guarantee that the results from a SLR include ALL the relevant studies, true recall (also called sensitivity) cannot be calculated. An alternative is to calculate the relative recall. Therefore, we considered the overall existing set of relevant studies as the sum of unique relevant studies identified in our update efforts by both approaches: (i) searching using a search string and a set of databases; and (ii) searching using forward snowballing. This set of relevant studies retrieved in both approaches comprises 14 new studies, including 11 identified using the search string approach (SLR2) and 3 new studies identified with the snowballing approach (see Tables 1, 2, and 3). In the present study, we calculated relative recall (RC) as:

$$RC\_forward\_snowballing = \frac{a}{b} \quad (1)$$

where:
a = set of relevant studies retrieved by forward snowballing;
b= set of relevant studies retrieved by both approaches

$$RC\_search\_string = \frac{a}{b} \quad (2)$$

where:
a = set of relevant studies retrieved by the database searches;
b= set of relevant studies retrieved by both approaches.

The precision of a search approach is the amount of relevant studies amongst the studies retrieved by the search, i.e., the strategy's ability to detect no or few irrelevant studies. In the present study, we calculated precision (P) as:

$$P\_forward\_snowballing = \frac{a}{b} \quad (3)$$

where:
a = set of relevant studies retrieved by forward snowballing;
b= set of all studies retrieved by forward snowballing.

$$P\_search\_string = \frac{a}{b} \quad (4)$$

where:
a = set of relevant studies retrieved by the database searches;
b= set of studies retrieved by the database searches.

3.3 Results

This section details the steps and results of our case study, which are later on used to address research questions (*RQ1* and *RQ2*).

Table 1. SLR1: Included studies – start set of replication.

ID	Authors	Year
S1	K. Maxwell and L.V. Wassenhove and S. Dutta	1999
S2	L.C. Briand and K. El-Emam and K. Maxwell and D. Surmann and I. Wieczorek	1999
S3	L.C. Briand and T. Langley and I. Wieczorek	2000
S4	R. Jeffery and M. Ruhe and I. Wieczorek	2000
S5	R. Jeffery and M. Ruhe and I. Wieczorek	2001
S6	I. Wieczorek and M. Ruhe	2002
S7	M. Lefley and M. Shepperd	2003
S8	B. A. Kitchenham and E. Mendes	2004
S9	E. Mendes and B. A. Kitchenham	2004
S10	E. Mendes and C. Lokan and R. Harrison and C. Triggs	2005

Table 2. SLR2: Included studies – updated version.

ID	Authors	Year
S11	C. Lokan and E. Mendes	2006
S12	E. Mendes and S. Di Martino and F. Ferrucci and C. Gravino	2007
S13	C. Lokan and E. Mendes	2008
S14	E. Mendes and C. Lokan	2008
S15	E. Mendes and S. Di Martino and F. Ferrucci and C. Gravino	2008
S16	C. Lokan and E. Mendes	2009
S17	E. Mendes and C. Lokan	2009
S18	E. Kocaguneli and T. Menzies	2011
S19	O. Top and B. Ozkan and M. Nabi and O. Demirors	2011
S20	F. Ferrucci and F. Sarro and E. Mendes	2012
S21	L. L. Minku and X. Yao	2012

Our first step was to obtain the list of relevant studies selected for inclusion in SLR1, and to use it as our seed set (starting set). SLR1 identified a set of 10 relevant studies (see Table 1); we applied the forward snowballing approach to each of these 10 studies. A total of 11 new studies (see Table 2) have already been included in SLR2 – updated version, using the database search strategy.

Table 3. New studies revealed by forward snowballing.

ID	Authors	Year
N1	R. Premraj and T. Zimmermann	2007
N2	A. Bakır and B. Turhan and A. B. Bener	2010
N3	E. Kocaguneli and B. Cukic and T. Menzies and H. Lu	2013

We performed four iterations looking at the citations. The citations were extracted with the help of search engines, such as Google Scholar, ACM Digital Library and IEEE Xplore Digital Library, organized into an Excel file, and later filtered by one of the authors, who is an experienced researcher on the SLR1 topic. It is important to highlight that the filter was performed independently of the SLR2' results, i.e., the findings from applying forward snowballing were only compared to those from SLR2 at the end of the four iterations. These four iterations are detailed next.

3.3.1 Iteration 1

The results from iteration 1 of the forward snowballing approach are shown in Figure 1.

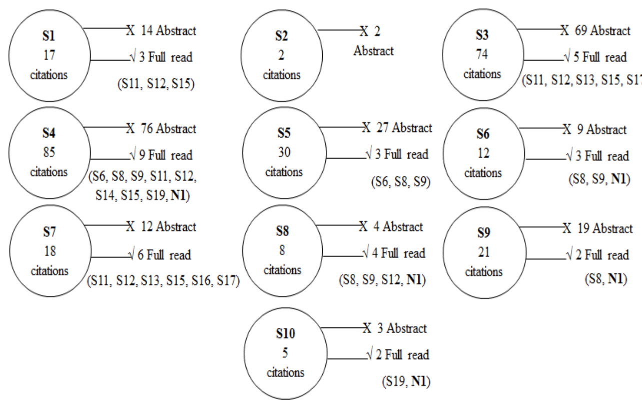

Figure 1: The first iteration results.

Figure 1 shows that study S1 was cited by 17 studies – candidates for inclusion. SLR2 was published in the beginning of 2014, and, due to this reason, we did not consider studies published in 2014 and 2015. Three (S11, S12 and S15) of these 17 citations were included, when applying the same inclusion criteria used in SLR1 and SLR2. Two studies have cited S2, however no relevant studies were identified from these candidates. S3 has already been cited by a large number of studies (74 studies). Out of these, five were included (S11, S12, S13, S15 and S17). Similar to S3, S4 has also already been cited by a large number of studies (85). Out of these, 9 were included (S6, S8, S9, S11, S12, S14, S15, S19, and N1). A new study, N1 (see Table 3), which was not identified in SLR2, was identified. S5 was cited by 30. Out of these, three studies were included (S6, S8 and S9). S6 was cited by 12. Out of these, three were included (S8, S9, N1) and one new study was identified. A total of 18 studies cited S7. Out of these, 6 were included (S11, S12, S13, S15, S16 and S17). Eight studies have cited S8 and four (S8, S9, S12 and N1) were identified as relevant. S9 was cited by 21. Two (S8 and N1) were identified as relevant. S10 has been cited by five studies. Out of these, two (S19 and N1) were included.

In summary, during the first iteration, 272 studies were found, representing 132 without duplications; of these, 9 studies were judged relevant to update SLR1 (S11, S12, S13, S14, S15, S16, S17, S19, and N1), of which one (N1) was a new study not included by SLR2. Thus, the first iteration of the forward snowballing approach identified 8 (72.72%) of the 11 studies included in SLR2 and one additional study. The overall precision of the first iteration was of 9/132 (6.82%) and the relative recall of 9/14 (64.29%).

3.3.2 Iteration 2

During iteration 2, the nine studies (S11, S12, S13, S14, S15, S16, S7, S19 and N1) identified in the previous iteration were analyzed. The results of this iteration are shown in Figure 2.

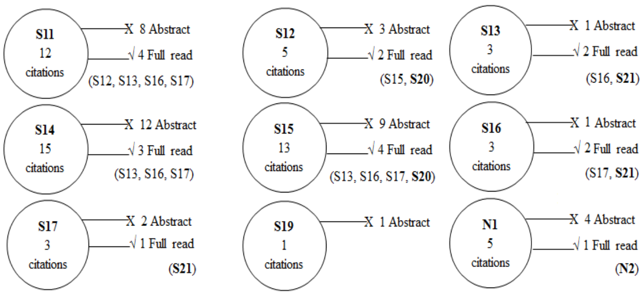

Figure 2: The second iteration results.

In summary, during the second iteration, 60 studies were found, representing 33 without duplications. Out of these, 8 studies (S12, S13, S15, S16, S17, S20, S21, and N2) were included, and one of these studies had not been included in SLR2. This iteration identified three additional new studies (S20, S21, N2), and its overall precision and relative recall were 8/33 (24.24%) and 8/14 (57.14%), respectively.

3.3.3 Iteration 3

This section presents the results of iteration 3, summarized in Figure 3.

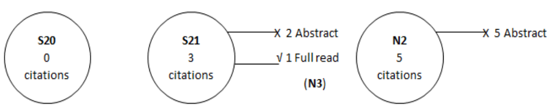

Figure 3: The third iteration results.

During the third iteration the three studies resulting from the previous iteration (S20, S21 and N2) were analyzed. S20 was not cited by others. S21 was cited by three studies, of which one was new - N3, which was included. N2 was cited by five studies, which were not included. The overall precision and relative recall of Iteration 3 were 1/8 (12.5%) and 1/14 (7.14%), respectively.

Finally, the in the last iteration (Iteration 4) study N3, selected in the previous iteration, was analyzed. N3 was not cited by others, thus finalizing the application of the forward snowballing approach.

4. DISCUSSION

In this section, we discuss issues related to the results we obtained in our case study and the limitations of our findings.

Using the database search, 1641 studies had been retrieved from the electronic databases, resulting in a set of 986 unique papers after removing the duplicates. Out of these 986 only 11 were included in SLR2, i.e., 975 (98.9%) were discarded. On the other hand, using the forward snowballing, 172 studies were retrieved, of which 13 were selected for inclusion. While it is noteworthy that the forward snowballing approach decreased five times the initial set of studies, there are also other important aspects, which we discuss next.

The forward snowballing approach identified all the studies included in SLR2, except for one – S18. Further investigation revealed that S18 only cited SLR1, rather than any of the previous primary studies in the same topic. Given that our snowballing approach focused solely upon primary studies published after SLR1, in order to use the same inclusion criteria previously employed by both SLR1 and SLR2, it would have been impossible to find S18. Of course, our assumption was that more recent primary studies in the topic originally investigated in SLR1 would cite other previously published relevant primary studies; however, such assumption did not hold true. Perhaps this is an isolated instance and other SLRs being updated using forward snowballing will not share the same issue; however, our recommendation, which is illustrated in Figure 4, is the following: the seed set to update an SLR should contain the studies included in the previous version of this SLR AND also the SLR itself. Had this strategy been used herein, all the relevant studies would have been identified by forward snowballing.

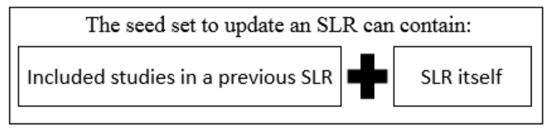

Figure 4: Lesson Learnt – construction of the seed set to update SRLs.

The forward snowballing identified three new studies (N1, N2 and N3) that were not identified by SLR2. Further investigation showed that two of these three studies (N1 and N2) had actually been retrieved using the search string strategy (were part of the 986 retrieved studies) but were not selected for inclusion by the SLR2 authors. Despite the use of team discussions and peer review, the authors of SLR2 agreed that this was an oversight in their filtering process. The large number of studies retrieved from the initial search (986 papers had to be analyzed to update SLR1 using the search string strategy) demanded significant manual and human intensive effort, and, as a consequence, was naturally error prone. Regarding the third study, N3, we noticed that, although it was not included amongst the 986 studies, it would have been retrieved now if we were to apply the same search string to the selected digital libraries. Given that the paper was published late in 2013 and SLR2 was conducted in the end of November 2013, we believe that N3 was not retrieved because it was not available in the digital libraries. Therefore, the search string strategy would have retrieved all the identified relevant papers. Such situation also prompted us to make another recommendation, which is for SLRs to be carried out in the year following the last year of coverage of that SLR. This way, the delay in having papers made available in the databases in which they are indexed would be mitigated.

Tables 4 and 5 present a summary of these results. To avoid any bias, as the focus of this paper is on the search strategy, the issue with the filtering process during SLR2 and not retrieving N3 due to timing of indexation in digital libraries are not considered. These results show that, for the specific case of updating SLR1, the search string strategy achieved a much lower precision but a slightly higher relative recall, since it was able to retrieve all the 14 identified relevant studies, while forward snowballing missed study S18. However, in SLRs in general, there is a trade-off between recall and precision and therefore it is necessary to strike a balance between them when choosing a search strategy in order to keep results manageable. Within the context of this work, forward snowballing had a much higher precision, retrieving a considerably lower amount of papers to be analyzed and reducing the overall SLR update effort. However, the difference in the precision between SLR2 and forward snowballing could be due to the specific search string employed in SLR1 and SLR2, which aimed to improve recall in order to avoid the risk of missing relevant papers. Skoglund and Runeson [28] point out that in a "good"' search strategy not too much of relevant material is missed and a manageable volume of irrelevant studies is retrieved. They observe that the existence of an optimal search strategy is unlikely.

Table 4. Case study: summary of results.

Type of search	Number of papers retrieved/identified (without duplications)	Number of papers included
Database searches	986	14
Forward snowballing (Iteration 1-4)	172 (132+33+7+0)	13 (8 +2 + 0 + 0) + (1 + 1 + 1 + 0) (10 in common to SLR2) + (3 new studies)

Table 5. Summary of results: precision and relative recall.

Type of search	Precision	Relative recall
Database searches	14/986 = 1.42%	14/14 = 100%
Forward snowballing (Iteration 1 – 4)	13/172 = 7.55%	13/14 = 92.86%
Forward snowballing (Iteration 1)	9/132 = 6.81%	9/14 = 7.44%
Forward snowballing (Iteration 2)	12/165 = 7.27%	12/14 = 85.71%
Forward snowballing (Iteration 3)	13/172 = 7.55%	13/14 = 92.86%
Forward snowballing approach (Iteration 4)	13/172 = 7.55%	13/14 = 92.86%

Based on our results, we believe that employing forward snowballing to update SLRs based on a seed set obtained from the previous SLR trial represents an interesting alternative to reduce the significant effort of updating SLRs. The main advantage of forward snowballing, when compared to the search string strategy for updating SLRs, is reducing the number of studies to be analyzed while still presenting a good recall of relevant papers.

A possible explanation for the forward snowballing results obtained herein may be the use of a good start set, comprising the primary studies included in SLR1. Badampudi et al. [22] argue that the reliability of the snowballing approach is highly dependent on the creation of a suitable start set

4.1 Reliability of Forward Snowballing

The forward snowballing approach missed study S18. This study is one of the studies that support the main findings of SLR2 [31]. The finding concerns the improvement in the accuracy of predictions from cross-company models, which started to provide results that were not significantly worse than predictions obtained using within-company models. Such improvement was associated with filtering mechanisms applied to the cross-company datasets in order to select samples for building the estimation models more similar to the within-company data.

While S18 was not the only primary study supporting the shift above-mentioned, in such a specific subject with few conducted studies a missing study could, in theory, make a large difference to the overall results. However, within the context of this work, this would not have been the case because out of the five new studies indicating that cross-company predictions are not significantly worse than within-company predictions, most of them used selection/filtering mechanisms; thus the shift in previous prediction accuracy patterns would have been observed even without including S18.

4.2 Threats to Validity

As in all studies, this research also has certain limitations. The main limitation is that our analysis was based on applying forward snowballing to update a single SLR, so further assessment is clearly needed. However, the results have provided some valuable insight and indication as to how much forward snowballing can support updating SLRs. Thus, as a first-cut assessment, we believe our case study met its goal. Other threats to validity are described below.

Internal validity. Concerning internal validity, it is important to highlight that we strictly followed the guidelines for conducting snowballing suggested by Wohlin [23].

Construct validity. In our study, the study inclusion criteria can be considered a potential confounding factors; however, it is documented [30][31] and its application was extensively discussed among the authors. The studies were selected by one of the authors who did not take part in SLR2 and discussion meetings were held where the judgments (inclusion or exclusion of a study) were reviewed by three authors. This step was carried out as part of each iteration.

Reliability. We minimized the risk of missing important papers by the extraction of citations with the help of search engines, such as *Google Scholar*, *ACM Digital Library* and *IEEE Xplore Digital Library*.

External validity. The study of applying forward snowballing was conducted on one specific SLR that had recently been updated. Both the original SLR and the updated one were peer reviewed and published. However, the study findings are not generalizable and replications on other SLR updates to reinforce our preliminary indications are required.

5. CONCLUSIONS

The goal of this paper was to compare database search with forward snowballing to support updating of SLRs in SE. Results were compared with respect to the precision and relative recall of the findings.

Searches of electronic databases using a search string presented an overall relative recall higher than that for forward snowballing (100% versus 92.86%). However, the overall precision of searches in electronic databases was low (1.42%). The overall precision of the forward snowballing approach – 7.55%, suggests it could be a competing choice to be used as a search tool for updating SLRs. Nevertheless, one relevant study that did not cite the other related primary studies was not retrieved. In light of these findings, the main contributions of this research are as follows:

1. we investigated the application of forward snowballing to support the update of SLRS; and
2. the results of our case study indicated that the forward snowballing approach reduces effort of updating SLRs, but exposes the risk of missing relevant studies whenever a primary study does not cite any related primary studies.

In addition replications of our case study using other SLRs in Software Engineering (SE), we believe that an interesting direction for future research is to use the results from all replications in order to identify typical values for the precision of SLRs in the SE field, which can later be used to guide other researchers carrying out SLRs evaluating their searches for evidence.

6. REFERENCES

[1] B. A. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering". Technical Report EBSE 2007–001, Keele University and Durham University Joint Report, 2007.

[2] B. A. Kitchenham, "Procedures for performing systematic reviews", Keele University and Empirical SE National ICT Australia Ltd, Joint Technical Report TR/SE–0401 (Keele) - 0400011T.1 (NICTA), 2004.

[3] J. Biolchini, P. G. Mian, A. C. C. Natali, G. H. Travassos, "Systematic review in software engineering". Technical Report ES 679/05, COPPE/UFRJ, 2005.

[4] H. Zhang and A. Muhammad, "An empirical investigation of systematic reviews in software engineering". In 5th Int. Symp. on Empirical Software Engineering and Measurement (ESEM' 11), pp. 1–10, 2011.

[5] K. Petersen and B. Nauman, "Identifying strategies for study selection in systematic reviews and maps". In 5th Int. Symp. on Empirical Software Engineering and Measurement (ESEM' 11), pp. 1–10, 2011.

[6] S. Imtiaz, M. Bano, N. Ikram, and M. Niazi, "A tertiary study: Experiences of conducting systematic literature reviews in software engineering". In 17th Int. Conf. on Evaluation and Assessment in Soft. Eng. (EASE' 13), pp. 177–182, 2013.

[7] E. Mendes, "A systematic review of web engineering research". In 4th Int. Symp. on Empirical Software Engineering (ISESE' 05), pp. 498–507, 2005.

[8] J. E. Hannay, D. I. K. Sjoberg, and T. Dyba, "A systematic review of theory use in software engineering experiments". Trans. on Soft. Eng., 33 (2), pp. 87–107, 2007.

[9] M. Sulayman and E. Mendes, "A systematic literature review of software process improvement in small and medium web companies". Advances in Software Engineering, 59(1), pp. 1–8, 2009.

[10] O. Dieste and N. Juristo, "Systematic review and aggregation of empirical studies on elicitation techniques". Transactions on Software Engineering, vol. 37, nº. 2, pp. 283–304, 2010.

[11] E. F. Souza, R. A. Falbo, and N. Vijaykumar, "Ontologies in software testing: A systematic literature review". In 6th Workshop on Ontology Brazil (ONTOBRAS' 13), Belo Horizonte/MG, pp. 71–82, 2013.

[12] F. J. Affonso, K. R. Felizardo, L. Oliveira, and E. Y. Nakagawa, "Reference architectures for self-managed software systems: a systematic literature review". In 8th Brazilian Symp. on Software Components, Architectures and Reuse (SBCARS' 14), pp. 1–10, 2014.

[13] J. Carver, E. Hassler, E. Hernandes, and N. Kraft, "Identifying barriers to the systematic literature review process". In 7th ACM-IEEE Int. Symp. on Empirical Software Engineering and Measurement (ESEM' 13), pp. 203–212, 2013.

[14] C. Wohlin, "Writing for synthesis of evidence in empirical software engineering". In 8th Int. Symp. on Empirical Software Engineering and Measurement (ESEM' 14), 2014.

[15] I. Spasic, S. Ananiadou, J. Mcnaught, and A. Kumar, "Text mining and ontologies in biomedicine: Making sense of raw text". Brief Bioinform, vol. 6, nº. 3, pp. 239–251, 2005.

[16] D. Sjøberg, T. Dybå, and M. Jørgensen, "The future of empirical methods in software engineering research". In 2007 Future of Software Engineering, ser. FOSE' 07. IEEE Computer Society, pp. 358–378, 2007.

[17] E. Engstrom, M. Skoglund, and P. Runeson, "Empirical evaluations of regression test selection techniques: A systematic review". In 2nd ACMIEEE Int. Symp. on Empirical Software Engineering and Measurement (ESEM' 08). ACM, pp. 22–31, 2008.

[18] P. Brereton, B. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain". Journal of Systems and Software, 80 (4), pp. 571–583, 2007.

[19] T. Dybå, T. Dingsøyr, and G. Hanssen, "Applying systematic reviews to diverse study types: An experience report". In 1st Int. Symp. on Empirical Software Engineering and Measurement (ESEM' 07), pp. 225–234, 2007.

[20] J. Webster and R. Watson, "Analyzing the past to prepare for the future: Writing a literature review". MIS Quarterly, vol. 26, nº. 2, pp. 13–23, 2002.

[21] S. Jalali and C. Wohlin, "Systematic literature studies: Database searches vs. backward snowballing". In 6th Int. Symp. on Empirical Software Engineering and Measurement (ESEM' 12), pp. 29–38, 2012.

[22] D. Badampudi, C. Wohlin, and K. Petersen, "Experiences from using snowballing and database searches in systematic literature studies". In 19th Int. Conf. on Evaluation and Assessment in Software Engineering (EASE' 15), pp. 17:1–17:10, 2015.

[23] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering". In 18th Int. Conf. on Evaluation and Assessment in Software Engineering (EASE' 14), pp. 38:1–38:10, 2014.

[24] F. Ferrari and J. Maldonado, "Experimenting with a multi-iteration systematic review in software engineering". In 5th Experimental Software Engineering Latin America Workshop (ESELAW' 08), pp. 1–10, 2008.

[25] K. R. Felizardo, E. Nakwgawa, S. MacDonell, and J. Maldonado, "A visual analysis approach to update systematic reviews". In 18th Int. Conf. on Evaluation and Assessment in Software Engineering (EASE' 14), pp. 1–10, 2014.

[26] O. Dieste, M. Lopez, and F. Ramos, "Formalizing a systematic review updating process". In 6th Int. Conf. on Software Engineering Research, Management and Applications (SERA' 08), pp. 143–150, 2008.

[27] B. A. Kitchenham and P. Brereton, "A systematic review of systematic review process research in software engineering". Information and Software Technology, vol. 55, nº. 12, pp. 2049–2075, 2013.

[28] M. Skoglund and P. Runeson, "Reference-based search strategies in systematic reviews". In 13rd Int. Conf. on Evaluation and Assessment in Software Engineering (EASE' 09), pp. 1–10, 2009.

[29] P. Runeson, M. Host, and A. Rainer, Case Study Research in Software Engineering: Guidelines and Examples, J. Wiley, Ed., 2012.

[30] B. A. Kitchenham, E. Mendes, and G. H. Travassos, "Cross versus within-company cost estimation studies: A systematic review". IEEE Trans. on Soft. Eng., 33 (5), pp. 316–329, 2007.

[31] E. Mendes, M. Kalinowski, D. Martins, F. Ferrucci, and F. Sarro, "Cross vs. within-company cost estimation studies revisited: An extended systematic review". In 18th Int. Conf. on Evaluation and Assessment in Software Engineering (EASE' 14), pp. 12:1–12:10, 2014.

[32] K. Dickersin, R. Scherer, and C. Lefebvre, "Systematic reviews: Identifying relevant studies for systematic reviews". British Medical Journal, 309 (6964), pp. 1286–1291, 1994.