Assessing diversity in creating seed set for snowballing search for systematic literature review in software engineering

Katia Romero Felizardo
Francisco Carlos Souza
Alinne C. Corrêa Souza
Universidade Tecnol. Federal do Paraná/BRA
katiascannavino@utfpr.edu.br
franciscosouza, alinnesouza@utfpr.edu.br

Bianca Minetto Napoleão
Université du Québec à Chicoutimi/CAN
bmnapole@uqac.ca

Igor Steinmacher
Marco Gerosa
Northern Arizona University/USA
Igor.Steinmacher@nau.edu
Marco.Gerosa@nau.edu

***Abstract*—Background: Systematic literature reviews (SLRs) require robust search strategies to ensure comprehensive coverage. Although database searches have traditionally been the primary method, snowballing has emerged as an effective alternative strategy in software engineering research. However, the success of snowballing heavily depends on the initial seed set's composition, particularly regarding diversity across authors, publication years, and venues. Objective: This study investigates how different diversity characteristics in seed set creation influence snowballing performance and effectiveness in identifying relevant literature. Method: We conducted replication studies of two existing SLRs, comparing their conventional seed set creation approaches with our diversity-driven methodology, where we systematically incorporated diversity characteristics into constructing the seed sets. Results: Our diversity-based approach demonstrated substantial improvements, with a precision of 0.019 (compared to 0.006 in the original), a relative recall of 0.97 (versus 0.921), and an F-measure of 0.0372 (improving from 0.0119). Conclusions: The empirical evidence suggests that incorporating diversity criteria in seed set creation enhances snowballing efficacy while maintaining comprehensive coverage of relevant literature. This approach offers a systematic and effective method for conducting snowball-based literature reviews in software engineering research.**

***Index Terms*—Software Engineering, Snowballing, Systematic Literature Review, Systematic Mapping Review, Seed set**

I. Introduction

A systematic literature review (SLR) has a strong focus on comprehensiveness, as noted in its definition: "...a study that reviews all primary studies related to a specific research question..." [1]. To ensure comprehensive coverage in their evidence synthesis, the software engineering (SE) community has employed two principal search approaches [2]: database search and snowballing [3]. Each method offers distinct advantages and challenges in achieving comprehensive literature coverage.

The challenges in String-based database searches include the labor-intensive process of refining search strings, the effort required to sift through extensive irrelevant results, and the technical constraints of digital library platforms for performing reviews [2], [4]. These limitations significantly impact both the efficiency and comprehensiveness of the review process.

Snowballing has emerged as a promising alternative or complementary search strategy [3]. This methodology initiates with a carefully selected seed set of articles, followed by systematic forward and backward search strategies [3]. While backward snowballing examines reference lists to uncover additional evidence, forward snowballing identifies subsequent citations of the relevant studies.

Snowballing has been shown to contribute to 51% of included studies in systematic literature reviews [5]. Wohlin et al. [6] and Felizardo et al. [7], [8] compared snowballing with a database search strategy to update SLRs. They concluded that the approaches are comparable on the basis of the papers they found. However, snowballing is more efficient, reducing the number of primary studies that need to be analyzed more than five times. Recent research demonstrates that hybrid search strategies—defined as the systematic integration of multiple search approaches, such as combining database searches with snowball sampling from pre-identified relevant articles, yield superior results in identifying primary studies [2].

However, a fundamental challenge in snowballing is defining an optimal seed set. Few studies in the software engineering literature explicitly address strategies for defining a seed set in snowballing [1], [3]. Authors typically select seed articles based on either high citation counts within the research domain or keyword-based search results derived directly from research questions and their synonyms [7], [9]–[12]. While Wohlin [3] recommends heterogeneous seed sets that incorporate diversity in publishers, publication years, and authors, the optimal composition and necessary diversity levels have not yet been empirically investigated in the context of systematic literature reviews. To bridge this knowledge gap, we examine how specific diversity characteristics within seed sets impact the effectiveness of snowballing. This study is guided by the following research question:

How does the diversity of the seed set– considering publishers,

Brazilian National Council for Scientific and Technological Development (CNPq), Grant $302339/2022-1$

authors, and publication years –influence the effectiveness of snowballing in systematic literature reviews?

To address this research question, we replicate existing SLRs and evaluate snowballing performance using three metrics: recall, precision, and F-measure. We assess the diversity-based approach's ability to identify "all relevant studies" (recall) and determine the proportion of relevant studies among retrieved articles (precision). Since systematic literature reviews cannot guarantee complete identification of all relevant studies, we employ relative recall rather than true recall (sensitivity), calculating it based on the sum of relevant and unique studies identified in the original SLRs. The F-measure provides a balanced assessment of the relationship between recall and precision.

This research makes two primary contributions: (i) enhancing the understanding of snowballing methodology by providing empirical evidence on how seed set diversity characteristics (authors, years, and publishers) influence search effectiveness, thereby extending Wohlin's guidelines [3]; and (ii) developing an automated tool for diversity-based seed set recommendation that operationalizes Wohlin's guidelines [3], facilitating their application in review processes.

II. Related work

Snowballing begins with selecting a seed set of studies, typically identified through database searches. Next, backward snowballing is performed by examining the reference lists of the seed set to identify additional relevant studies. Backward is followed by forward snowballing, which involves using citation indexes, such as Google Scholar or Scopus, to find newer studies that cite the seed set papers. Each newly identified study is then assessed using predefined inclusion and exclusion criteria. The process is iterative, meaning backward and forward snowballing is repeated on newly included papers until no new relevant studies emerge [3].

Snowballing has emerged as a powerful approach in systematic literature reviews, demonstrating its value both as a complement to traditional database searches and as a standalone search strategy, with numerous studies validating its effectiveness [3], [4], [7], [8], [11], [13]. Badampudi et al. [11] evaluated the effectiveness and reliability of snowballing (backward and forward) compared to the database search. They found that the efficacy of both search strategies is comparable, contingent on creating a suitable seed set. Wohlin [3] established guidelines for snowballing as a search strategy, suggesting that it could serve as an alternative to searches in various databases, contributing to the retrieval of a comprehensive and relevant set of studies for analysis. Wohlin states that 5 to 10 initial studies are appropriate in a seed set, as this set would lead to convergence in some iterations. However, more research is needed to determine the optimal number of studies needed to compose a seed set for snowballing.

Felizardo et al. [7], [8] have investigated forward snowballing to update SLRs, concluding that it can effectively find newer, relevant studies. In the same line, Wohlin [6] recommended forward snowballing with Google Scholar and leveraging the original review's primary studies to compose the seed set. Mourao et al. [4], [13] explored hybrid search strategies that combine database searches and snowballing, suggesting that a hybrid strategy may be an appropriate alternative for searching for candidate studies in SLRs.

While several tools have been developed to support the snowballing process [14]–[17], they lack comprehensive support for the complete process and, notably, fail to address the critical aspect of diversity in seed set creation. Despite snowballing's growing recognition as a powerful strategy for systematic literature reviews, there remains a significant gap in research and tooling support for systematically incorporating diversity in the initial seed set generation —- a factor that can substantially influence the effectiveness of the overall review process.

III. Research design

We performed the snowballing replication in five steps:

Step 1. We defined a systematic approach for constructing a diverse seed set, as illustrated in Figure 1 and detailed in Section IV. A supporting tool was implemented to facilitate this step.

Step 2. We selected two systematic literature reviews (SLRs) for replication (see Section V).

Step 3. We applied the approach described in Step 1 to generate diversity-based seed sets for the selected SLRs (see Section VI-A).

Step 4. We executed snowballing using the seed sets generated in Step 3 (see Section VI-B).

Step 5. We evaluated snowballing effectiveness using relative recall (*RC*), precision (*P*), and the F-measure. Relative recall (*RC*) was computed as $\frac{Inc\_Rep}{Inc\_Rep \cup Inc\_Ori}$, where **Inc_Rep** is the set of studies included after snowballing iterations, and **Inc_Ori** is the set of studies originally included in the replicated SLR. Precision (*P*) was calculated as $\frac{Inc\_Rep}{Tot\_Rep}$, where **Tot_Rep** represents the total number of studies identified through snowballing (including both included and excluded studies). The F-measure, representing the harmonic mean between precision and relative recall, was calculated using the formula $2 \times \frac{P \times RC}{P + RC}$. Finally, we compared and discussed in detail the performance of different search strategies in Section VI-C.

The research team conducting this study has experience in this type of research. They have conducted several SLRs and researched the SLR method, including applying and experimenting with different snowballing approaches.

IV. Definition of an approach for creating a diversified seed set

Our approach to constructing a diverse seed set involves seven stages and an automated tool to support them. Initially, researchers extract keywords from the research question. Then, the researchers add appropriate synonyms. The finalized keywords and their synonyms are combined using logical operators: synonyms are linked with OR, while distinct keywords are joined using AND, forming a search string that the

researcher may further refine if needed (stage 1). Additional filters, such as publication year, may also be applied depending on the research objectives.

Next, a candidate seed search is performed considering titles, abstracts, and keywords (stage 2). The researchers manually excluded articles that do not meet the predefined selection criteria of the original SLR, including publication year. Specifically, papers published after the end date of the replicated SLR's search period were removed to ensure full comparability between the original and replicated sets. The remaining articles then undergo a two-stage labeling process. In the first labeling stage, articles are categorized based on three general criteria (GC1, GC2, GC3):

"GC1 – Keywords" – Articles whose titles contain at least one keyword from the search string.

"GC2 – Synonym" – Articles containing at least one synonym from the search string.

"GC3 – Most cited" – The five most-cited articles among the results.

An individual article may meet one or more of these general criteria simultaneously, resulting in multiple entries within the candidate seed list. Duplicate entries are thus removed before proceeding to the next labeling stage (stage 4, Figure 1), in which three diversity criteria — DC1, DC2, and DC3, are applied (Stage 5, Figure 1):

DC1 – Author: Author names are extracted from each article. When multiple articles share the same set of authors, regardless of the order in which they appear, only one of them is labeled as DC1. This criterion aims to promote authorial diversity by avoiding the repeated inclusion of articles from the same research team.

DC2 – Year: Articles published in a unique year, that is, a publication year not shared by any other article in the set are labeled as DC2. If multiple articles share the same publication year, only the most cited article among them is marked as DC2.

DC3 – Venue: The same logic applied to DC2 also applies here. Articles published in unique venues are labeled DC3. When multiple articles originate from the same venue, only the most cited article is labeled DC3.

Since individual articles may fulfill multiple diversity criteria simultaneously, duplicate entries are consolidated after this labeling stage (stage 6, Figure 1). Therefore, a single list is created, adding multiple tags to the articles.

Finally, in stage 7 (Figure 1), the seed set is generated through an incremental process. All articles labeled with the three diversity criteria (DC1, DC2, and DC3) are initially selected. If fewer than five such articles are available — the recommended minimum number for a seed set [3] — articles labeled with two diversity criteria are included next. If necessary, articles with only one diversity criterion are added until the set reaches the minimum required size.

To automate the diversity-based seed set recommendation process described in Section III, we developed an open source and online tool, available at https://seed-set-recommendation.onrender.com/. This tool is designed to support systematic literature reviews by generating diverse seed sets and offers five main functionalities: (a) keyword extraction; (b) generation of search strings including synonyms; (c) visualization of the search string; (d) retrieval of scientific articles from Scopus; and (e) seed set recommendation based on diversity criteria.

V. Selected SLRs

This section presents the two systematic literature reviews (SLRs) selected for the snowballing replication conducted in this paper. The choice of SLRs aligns with the authors' research expertise. Specifically, given that two authors have significant experience in software testing, we selected an SLR from that domain (SLR1). For the second review, we chose an SLR focusing on industry-academia collaboration, reflecting the specialization of another author and continuing the focus of a previous replication study utilizing the same dataset. This second review (SLR2) [2] is itself a replication of the original SLR by Garousi et al. [18].

The following selection criteria (C) guided our identification of suitable SLRs for replication: **C1.** The SLR was conducted according to rigorous guidelines [1], [3].

C2. The primary search strategy employed was snowballing (both backward and forward) [3].

C3. The process used to define the seed set was described.

C4. Keywords and their synonyms were explicitly described in the review protocol. Even though the main method of the SLR was snowballing, our strategy starts with the definition of a search string, used to identify candidate studies to compose the seed set.

C5. Details of the snowballing process were transparently reported.

C6. The complete list of included studies was made publicly available.

A. SLR 1: Software testing

To find a replication study addressing software testing, we searched the Scopus digital library using the following string: *TITLE-ABS-KEY(("systematic literature review" OR "systematic review") AND (snowballing) AND ("software testing"))*. We present the six (6) papers returned by the query and how they satisfy the criteria in Table I.

As can be seen, only the paper number 4 satisfies all six (6) criteria (Table I) [22]. The objective of the selected SLR [22] was *" ...Investigate which techniques of software testing receive more attention when applying Knowledge Management (KM), and identify the challenges faced due to the lack of KM practices..."*.

To identify studies of interest, the authors [22] used the snowballing process [3], detailed as follows.

The snowballing process started with creating the seed set, using database searches. A multi-stage screening process was applied, starting with 2,774 papers, narrowing down to 63 after title and abstract reviews, and then to 32 after introduction/conclusion checks. Full-text analysis led to 16 candidates, which were further reviewed for relevance. Finally, 13 highly cited and well-referenced papers were selected as seed.

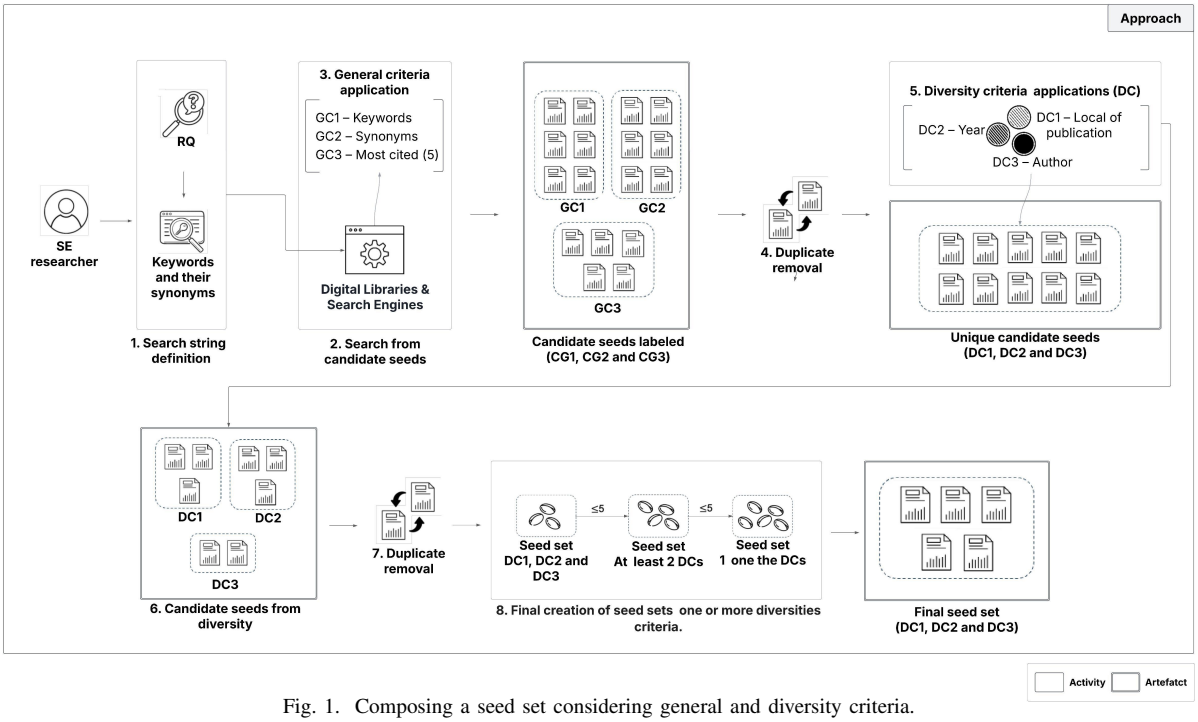

Fig. 1. Composing a seed set considering general and diversity criteria.

TABLE I
CANDIDATE STUDIES FOR REPLICATION AND HOW THEY MATCH THE SELECTION CRITERIA.

ID	Title	C1	C2	C3	C4	C5	C6	Ref
	Candidate studies on software testing							
1	A systematic review of cost reduction techniques for mutation testing: preliminary results	✓	×	✓	✓	✓	✓	[19]
2	A systematic review of the use of the definition of done on agile software development projects	✓	×	✓	✓	✓	✓	[20]
3	Industrial applications of software defect prediction using machine learning: a business-driven systematic literature review	✓	×	✓	✓	✓	✓	[21]
4	Knowledge management in software testing: a systematic snowballing literature review	✓	✓	✓	✓	✓	✓	[22]
5	On transforming model-based tests into code: a systematic literature review	✓	✓	✓	×	✓	✓	[23]
6	Testing and verification of neural-network-based safety-critical control software: a systematic literature review	✓	×	✓	✓	✓	✓	[24]
	Candidate studies on industry-academia collaboration							
7	Successful combination of database search and snowballing for identification of primary studies in systematic literature studies	✓	✓	✓	✓	✓	✓	[2]
8	Successful Practices in Industry-Academy Collaboration in the Context of Software Agility: A Systematic Literature Review	✓	×	✓	✓	✓	✓	[25]

The authors performed five (5) iterations of snowballing. To perform the forward citations, they employed Google Scholar. In the end, 1,457 candidate studies were analyzed (843 backward references and 614 forward citations), and 35 peer-reviewed studies (including the 13 seeds) passed the study criteria and were included. Twelve (12) studies were selected from backward references and 10 from forward citations.

B. SLR 2: Industry-academia collaboration

To identify a study suitable for replication that focuses on industry-academia collaboration in SE, we conducted a search in the Scopus digital library using the following query string: *TITLE-ABS-KEY ((“systematic literature review” OR “systematic review”) AND (snowballing) AND (industry) AND (academia) AND (collaboration) AND (“software engineering”))*. We present the two (2) articles retrieved from the query and explain how each meets the criteria outlined in Table I.

As shown in Table I, only the study 7 [2] fully satisfies all six (6) criteria.

The purpose of the study by Wohlin et al. [2] was to evaluate hybrid search strategies for SLRs. Therefore, the original SLR aims *“...Identify (a) the challenges to avoid risks to collaboration by being aware of the challenges, (b) the best practices to provide an inventory of practices (patterns) allowing an informed choice of practices to use when planning and conducting collaborative projects”*.

To identify relevant studies, Wohlin et al. [2] followed the snowballing process [3], as detailed below.

To construct the replicated seed set, the authors used the following search and applied it to the Scopus digital library:

"industry AND academia AND collaboration AND software AND engineering." The search targeted articles from 2010–2014, aligning with the original SLR time frame (conducted in early 2015 and published in 2016). In total, nine studies were selected as seed [26]. The seed set was determined by searching Scopus and evaluating the resulting papers. In total, 40 papers were found using Scopus; 15 papers met the inclusion criteria and went into full-text reading, resulting in nine (9) papers being selected. The authors identified in total 1942 papers of interest. The publications came from the start set, backward and forward snowballing, respectively: 40 papers from Scopus, 839 publications from BS (five rounds), and 1063 papers from FS (five rounds). In total, 78 articles were included in the full text assessment. It resulted in the inclusion of 43 articles (9 from seed set, 18 from BS and 16 from FS). The remaining 35 articles were excluded after reading their full text.

VI. Replication study

A. Seed set creation based on diversity

We applied the diversity-based process defined in Figure 1 to the selected SLRs to generate the diverse seed set.

Table II compares the original SLR seed set with the seed set generated in this replication. Based on the selection criteria of the original study, previously described in Section V-A, we excluded the following studies from the seed set: S16, S17, S18, S20, S21, S22, S23, and S24. S1, S3, S4, S6, S7, S9, and S11 appear in both the original SLR and our replication (labeled as "Both" in the last column). Studies S16 and S17 (both from 2009) were excluded because they were published in the same year as another study already selected (S9). Study S18 was excluded due to unavailability. Study S19 was excluded because it is not among the five most cited articles (GC3). Studies S20 and S21 were excluded because they did not present relevant keywords or synonyms in their titles (CG2). Studies S15 and S23 were discarded because they share the same authors (Souza, E. F. et al.) as studies S6. Moreover, S15 was published in the same year (2015) as S6.

Finally, study S24 was excluded because it was published in the same local journal (TSE) as study S11. Studies S2, S5, S8, S10, S12 and S13 are exclusive to the original SLR (marked as "Original"), while S14 is unique to the replication (labeled as "Replication").

For seeds used only in the original SLR, the diversity process did not suggest S5, S8, S10, S12, or S13 because they are not indexed in Scopus. Although S2 is indexed in Scopus, our diversity-based strategy did not recommend it.

Therefore, the final seed set consists of eight articles: S1, S3, S4, S6, S7, S9, S11, and S14.

Table III shows the seed set of the original SLR2 from Wohlin et al. [2] and the one created in this replication. The original SLR2 seed set includes nine studies: S1 through S9. Among these, only S1 is shared with the replication (denoted "Both" in the last column). Four studies (S11, S13, S15, and S18) were excluded from our replication because they did not meet the inclusion criterion IC1. In contrast, the diversity-based approach identified six unique studies not present in the original (S10, S12, S14, S16, and S17). In total, the replication seed set consists of six studies. In particular, S12, which is part of the replication seed set, was included in the first round of backward snowballing in the original study.

B. Snowballing iterations

In this section, we summarize the results of backward and forward snowballing replication using the diversified seed sets for SLR1 and SLR2.

*1) **SLR1**:* We conducted five iterations of backward and forward snowballing, reviewing both references and citations. Citation data was collected using Google Scholar, compiled in a spreadsheet, and filtered by two authors with expertise in the SLR1 topic. We applied the same inclusion criteria defined in the original SLR. Article selection was done independently of the original results and only compared to them after completing all iterations. Table IV summarizes the outcomes of the five iterations.

Iteration 1. We found 421 articles during the first iteration. Eleven studies (out of 35 in the original study) met the inclusion criteria and were included. Nine came from backward snowballing (BW1-–BW9) and two from forward snowballing (FW1 and FW2).

Iteration 2. In the second iteration, we analyzed the 11 studies identified previously (BW10–BW19 and FW3). We retrieved 548 articles and included 11 of them. The 11 matched studies already included in the original SLR.

Iteration 3. We identified 470 studies in the third iteration and included four studies (BW20–BW21 and FW4–FW5). One of the backward snowballing studies, BW–N1 [27], was a "new" inclusion that was not found in the original SLR.

Iteration 4. We retrieved 489 studies in the fourth iteration and included only one study (BW22), which had already been covered in the original SLR.

Iteration 5. We examined seven references and nine citations during the fifth iteration. None met the inclusion criteria, concluding the snowballing process.

In total, we included 37 studies in the replication: 28 identified from snowballing, nine (9) from the seed set (Table II). Of these, 34 were also found in the original SLR. Three studies (S14 [28] (seed), S19 [29] and BW-N1 [27]) were not part of the original. One study (S13) included in the original was not recovered in our replication.

*2) **SLR2**:* We performed six iterations of backward and forward snowballing.

We extracted citations and references using the tool described in [30]. We manually verified each reference by checking the content of the cited papers and used Google Scholar to double-check the citations and reduce bias. The results of the six snowballing iterations for SLR2 are summarized in Table IV. Detailed results are also available in the supplementary materials.

Iteration 1. We analyzed 499 studies and included 8 that matched the original SLR2: three (3) backward and five

TABLE II

SEED SET SLR1 – COMMON SEEDS OF THE ORIGINAL SLR1 AND THE DIVERSITY-FOCUSED STRATEGY SUGGESTIONS (S1, S3, S4, S6, S7, S9, S11). EXCLUSIVE SEEDS OF ORIGINAL SLR1 (S2, S5, S8, S10, S12, S13). S14 IS A SEED ONLY FOR REPLICATION. UNIQUE REPLICATION STUDIES WERE EXCLUDED FROM THE SEED SET BECAUSE THEY DID NOT MEET THE IC (S15, S16, S17, S18, S20, S21. S22, S23, S24) OR DURING THE DIVERSITY STRATEGY (S15 AND S19).

ID	Title	Authors	Year	Venue	Source
S1	A model of knowledge management system in managing knowledge of software testing environment	R. Abdullah, and Z.D. Eri, and A. M. Talib	2011	MySEC	Both
S3	Investigation of knowledge management methods in software testing process	Y. Liu, J. Wu, X. Liu, and G. Gu.	2009	ICITCS	Both
S4	Knowledge management and software testing	A. Desai, and S. Shah	2011	ICETAI	Both
S6	Knowledge management initiatives in software testing: A mapping study	E.F. de Souza, R.A. Falbo, and N.L. Vijaykumar	2015	IST	Both
S7	Observing software testing practice from the viewpoint of organizations and knowledge management	O. Taipale, K. Karhu, and K. Smolander	2007	ESEM	Both
S9	Research and implementation of knowledge management methods in software testing process	L. Xue-Mei, G. Guochang, L. Yong-Po, and W. Ji	2009	CSCE	Both
S11	The role of the tester's knowledge in exploratory software testing	J. Itkonen, M.V. Mäntylä, and C. Lassenius	2013	TSE	Both
S2	An architectural model for software testing lesson learned systems	J. Andrade, J. Ares, M.A. Martínez, J. Pazos, S. Rodríguez, J. Romera, and S. Suárez	2013	IST	Original
S5	Knowledge management approach in mobile software system testing	O.K. Wei, and T.M. Ying	2007	IEEM	Original
S8	Ontology-based testing platform for reusing	X. Li, and W. Zhang	2012	ICICSE	Original
S10	The role of experience in software testing practice	A. Beer, and R. Ramler	2008	DSD/SEAA	Original
S12	Using knowledge management to revise software-testing processes	K. Nogeste, and D. H. Walker	2006	JWL	Original
S13	A knowledge management approach for industrial model-based testing	D. Koznov, V. Malinov, E. Sokhransky, and M. Novikova	2009	ICKMIS	Original
S14	Outsourcing and Knowledge Management in Software Testing	K. Karhu, O. Taipale, K. Smolander	2007	EASE	Replication
S15	Knowledge management applied to software testing: A systematic mapping	E. F. Souza, R. Falbo, N. L. Vijaykumar	2013	SEKE	Replication [Exc. Diversity]
S19	Technology for knowledge management in software testing and its application	Y. P. Liu, L. Zou, M. Z. Jun, X. M. Liu	2008	CIMS	Replication [Exc. GC3]
S16	How do scientists develop and use scientific software?	J. E. Hannay, C. MacLeod, J. Singer, H. P. Langtangen, D. Pfahl, G. Wilson	2009	ICSE	Replication [Exc. Diversity]
S17	Insight knowledge in search-based software testing	Arcuri A.	2009	GECCO	Replication [Exc. Diversity]
S18	Software testing survey 2011: Knowledge objectives implementation and results	M. Winter, K. Vosseberg, A. Spillner, P. Haberl	2012	LNI	Replication [Exc. Not available]
S20	Test-driven evaluation of Linked Data quality	D. Kontokostas, P. Westphal, S. Auer, S. Hellmann, J. Lehmann, R. Cornelissen	2014	WWW	Replication [Exc. GC2]
S21	The role of replications in Empirical Software Engineering	F. J. Shull, J. Carver, S. Vegas, N. Juristo	2008	ESE	Replication [Exc. GC2]
S22	Transfer learning for cross-company software defect prediction	Y. Ma, G. Luo, X. Zeng, A. Chen	2012	ISE	Replication [Exc. IC]
S23	Using the Findings of a Mapping Study to Conduct a Research Project: A Case in Knowledge Management in Software Testing	E. F. Souza, R. Falbo; N. L. Vijaykumar	2015	SEA	Replication [Exc. Diversity]
S24	Predicting the location and number of faults in large software systems	T. J. Ostrand, E. J. Weyuker, R. M. Bell	2005	TSE	Replication [Exc. Diversity]

TABLE III

SEED SET SLR2 – S1 IS A COMMON SEED OF THE ORIGINAL SLR2 AND SUGGESTED BY THE DIVERSITY-FOCUSED STRATEGY. S2–S9 ARE EXCLUSIVE SEEDS OF THE ORIGINAL SLR2. FOUR STUDIES (S11, S13, S15 AND S18) WERE EXCLUDED FROM THE SEED SET BECAUSE THEY DID NOT MEET THE INCLUSION CRITERIA OR DURING THE DIVERSITY STRATEGY.

ID	Title	Authors	Year	Venue	Source
S1	Action research as a model for industry-academia collaboration in the software engineering context	Petersen, K., Gencel, C., Asghari, N., Baca, D., Betz, S.	2014	IWLICSE	Both
S2	Lessons learned on applying design science for bridging the collaboration gap between industry and academia in empirical software engineering	Rodriguez, P., Kuvaja, P., Oivo, M.	2014	IWCESI	Original
S3	Practical experiences in designing and conducting empirical studies in industry-academia collaboration	Martinez-Fernandez, S., Marques, H.	2014	IWCESI	Original
S4	The 4+1 view model of industry-academia collaboration	Runeson, P., Minor, S.	2014	IWLICSE	Original
S5	Get the cogs in synch-time horizon aspects of industry-academia collaboration	Runeson, P., Minor, S., Svenér, J.	2014	IWLICSE	Original
S6	Enablers and impediments for collaborative research in software testing: An empirical exploration	Enoiu, E., Causevic, A.	2014	IWLICSE	Original
S7	Foundations for long-term collaborative research	Kanso, A., Monette, D.	2014	IWLICSE	Original
S8	Empirical software engineering research with industry: Top 10 challenges	Wohlin, C.	2013	IWCESI	Original
S9	Agile collaborative research: Action principles for industry-academia collaboration	Sandberg, A., Pareto, L., Arts, T.	2011	IEEE Software	Original
S10	Understanding the link between information technology capability and organizational agility: An empirical examination	Lu Y.	2011	MIS	Replication
S11	Trust in a specific technology: An investigation of its components and measures	Mcknight D.H., Carter, M., Thatcher, J. B., Clay, P.F.	2011	TMIS	Replication [Exc. IC]
S12	It takes two to tango - An experience report on industry - Academia collaboration	Runeson P.	2012	ICST	Replication
S13	3-D object retrieval and recognition with hypergraph analysis	Gao Y., Wang, M., Tao, D., ji, R., Dai, Q.	2012	TIP	Replication [Exc. IC]
S14	Opportunities and challenges for collaboration industry-Academia via sponsored design competitions	Rodriguez J., Choudhury, A.	2014	ICL	Replication
S15	FeynRules 2.0 - A complete toolbox for tree-level phenomenology	Alloul A., Christensen, N. D., Degrande, C., Duhr, C., Fuks, B.	2014	CPC	Replication [Exc. IC]
S16	Industry-academia collaboration in software testing: An overview of TAIC PART 2015	Alshahwan N., Felderer, M. Ramler, R.	2015	ICSTW	Replication
S17	Potential of community of practice in promoting academia-industry collaboration: A case study	Pohjola I., Puusa, A., Iskanius, P.	2015	ICICKM	Replication
S18	Experience based co-design and healthcare improvement: Realizing participatory design in the public sector	Donetto S., Pierri, P, Tsianakas, V., Robert, G.	2015	Design Journal	Replication [Exc. IC]

(5) forward snowballing. The eight (8) were identified from snowballing in a single study.

Iteration 2. We recovered 368 studies and included 13: six (6) backward and seven (7) forwards. Seven of the eight seed studies used in this iteration led to at least one included study.

Iteration 3. We analyzed 1,145 studies and included 12: 2 backward and 10 forward. One study was responsible for nine (9) of the 10 forward inclusions, while the remaining three (3) were sourced from two other studies.

Iteration 4. We recovered 433 studies and included 7: 1 backward and 6 forward. Three of the 10 seed studies used in this iteration accounted for all inclusions.

Iteration 5. We examined 261 studies and included only one through the forward snowball.

Iteration 6. We analyzed 23 studies derived from the only inclusion in iteration 5. None met the inclusion criteria, and the process was concluded.

The original SLR2 included a total of 43 studies. As a result of our approach, all 43 studies from the original SLR2 were accounted for (42 through snowballing process and one study was a common study from both seed sets), ensuring that they were no losses in the included studies during this replication. Furthermore, we recovered four new studies.

C. Main results

This section summarizes the results of our snowballing replications for SLR1 and SLR2, comparing them against the original reviews. We focus on three main metrics: precision, relative recall, and the F-measure. Table V presents a side-by-side comparison of the key characteristics and quantitative results.

TABLE IV
SLR1 — A TOTAL OF 1944 STUDIES WERE ANALYZED (900 REFERENCES AND 1044 CITATIONS), FROM WHICH 28 WERE INCLUDED DURING THE FIVE ITERATIONS OF SNOWBALLING REPLICATION. SLR2 — A TOTAL OF 2729 STUDIES (1139 REFERENCES AND 1590 CITATIONS) WERE ANALYZED, OF WHICH 42 WERE INCLUDED DURING THE SIX ITERATIONS OF SNOWBALLING REPLICATION.

	Backward Snowballing				**Forward Snowballing**				**Total**			
	Returned		**Included**		**Returned**		**Included**		**Returned**		**Included**	
Iteration	SLR1	SLR2	SLR1	SLR2	SLR1	SLR2	SLR1	SLR2	SLR1	SLR2	SLR1	SLR2
Iteration 1	226	232	9	3	195	267	2	5	421	499	11	9
Iteration 2	288	88	10	6	260	280	1	7	548	368	11	13
Iteration 3	226	322	3	2	244	823	2	10	470	1145	5	12
Iteration 4	153	352	1	1	336	81	0	6	450	433	1	7
Iteration 5	7	124	0	0	9	137	0	1	16	261	0	1
Iteration 6	0	21	0	0	0	2	0	0	0	23	0	0
Total	900	1139	23	12	1044	1590	5	29	1944	2729	28	42

For SLR1, our replication achieved a relative recall of 0.974, recovering nearly all studies included in the original SLR1. We included 37 studies in total, 28 through snowballing and nine (9) from the seed set (including one that was initially excluded but later recovered). This exceeded the original SLR1, which included 35 studies and had a relative recall of 0.921. The precision in our replication was 0.019, an improvement over the original 0.006, despite analyzing fewer total studies (1,962 vs. 6,289). Measure F increased significantly, from 0.0119 to 0.0372.

For SLR2, we achieved perfect recall. The 43 studies from the original SLR2 [2] were recovered, and we identified four additional relevant studies, bringing the total to 47. Although the broader coverage slightly reduced precision (0.016), the F-measure remained competitive at 0.0315, confirming the effectiveness of the diversified seed set.

The answer to the RQ is that the diversity of the seed set reduces the effort to apply snowballing with a low risk of missing relevant studies. This approach led to fewer retrieved articles to filter while maintaining or improving evidence coverage compared to original studies.

VII. DISCUSSION

We chose to evaluate the creation of the diversity seed set by performing two replications to have a point of reference. In addition to the lessons learned and validity threats, some aspects are essential to reflect on.

Defining keywords – Mendes et al. [31] affirm that Ph.D. students often spearhead a significant portion of the conduction of SLRs, exceeding 50%. One difficulty these novices face on their first SLR face is related to the complexity of the SLR process [32]. The conduction of SLR by novices without expert and experienced researcher supervision can lead to several potential pitfalls [33]. For example, elaborating search strings requires knowledge of relevant keywords, databases, and search processes, which can be overwhelming for beginners. Therefore, the diversity strategy could also be considered if the researcher is a novice. The novice can insert an RQ and the strategy will suggest keywords. In our replication of SLR1, the novice could primarily rely on the keywords knowledge management" and software testing" and supplement them with other keywords suggested by specialists in the review domain added through the "OR" Boolean.

Effort to create the seed set – To create the original seed set for SLR1, 4,832 articles were retrieved from the Engineering Village, resulting in 13 selected seeds. In contrast, using the diversity strategy proposed here, only 18 studies were recovered, of which 8 were selected. This reduced the screening effort by two orders of magnitude, from 4,832 to 18, representing more 99% fewer articles to examine. Reducing the number of retrieved articles reduces manual workload and reduces the risk of human error. However, effort alone is not a sufficient measure; recall and precision must also be considered when evaluating seed set strategies.

As illustrated in Figure 2 (left) for SLR1, six studies were exclusive to the original SLR (S2, S5, S8, S10, S12 and S13). Five of these (S5, S8, S10, S12, and S13) were not returned by Scopus in our replication. However, three of them (S2, S8, and S10) were recovered during the first snowballing interaction. S5 and S12 were recovered in the second iteration. Only S13 was not recovered.

For SLR2, shown in Figure 2 (right), only one study of the original seed set (S1) overlapped with the diverse seed set. Another original study (S12) was reached during the first iteration of backward snowballing. All other studies, including the rest of the original seeds, were recovered in the six snowballing iterations. This confirms that no evidence from the original SLR2 was lost, even with a smaller seed set built from only 10 retrieved studies. These results suggest that the diversity strategy significantly reduces effort while maintaining high coverage.

Using SLR as a seed – Wohlin et al. [6] assert that "...Using an SLR as a seed set and its primary studies is the most cost-effective way to search for new evidence when updating SLRs". This approach is commonly justified by the typically high relevance and frequency of the citation studies included in SLRs. However, this strategy did not always yield the expected results in our replication. For example, the study BW3, "Knowledge management applied to software testing: A systematic mapping" did not contribute to the identification of any relevant new studies.

In replicating SLR1, we successfully recovered 34 of the 35 studies from the original review, resulting in a relative recall of 0.974, and identified three additional relevant studies: S14 [28] (used as seed), S19 [29] (included as relevant),

TABLE V
SUMMARY OF REPLICATION RESULTS COMPARED TO THE ORIGINAL SLRS.

Description	SLR1		SLR2	
Description	**Original**	**Replication**	**Original**	**Replication**
seed set creation	String (Engineering Village)	Diversity-process	String (Scopus)	Diversity-process
# articles analyzed to create the seed set	4832	18	40	10
# of seeds	13	8	9	6
Limited searches	2003—2015	2003—2015	until 2015	until 2015
Forward citations	Google Scholar	Google Scholar	Google Scholar	Semantic Scholar, ResearchGate and Google Scholar
Snowballing iterations	5	5	5	6
References analyzed	843	900	899	1139
References included	12	23	18	12
Citations analyzed	614	1044	1352	1590
Citations included	13	5	16	29
# articles analyzed (snowballing)	6289	1962	2251	2729
# studies included	35	37	43	47
Precision	0.006	0.019	0.018	0.016
Relative Recall	0.921	0.974	0.953	1
(Relative) F-measure	0.0119	0.0372	0.0353	0.0315

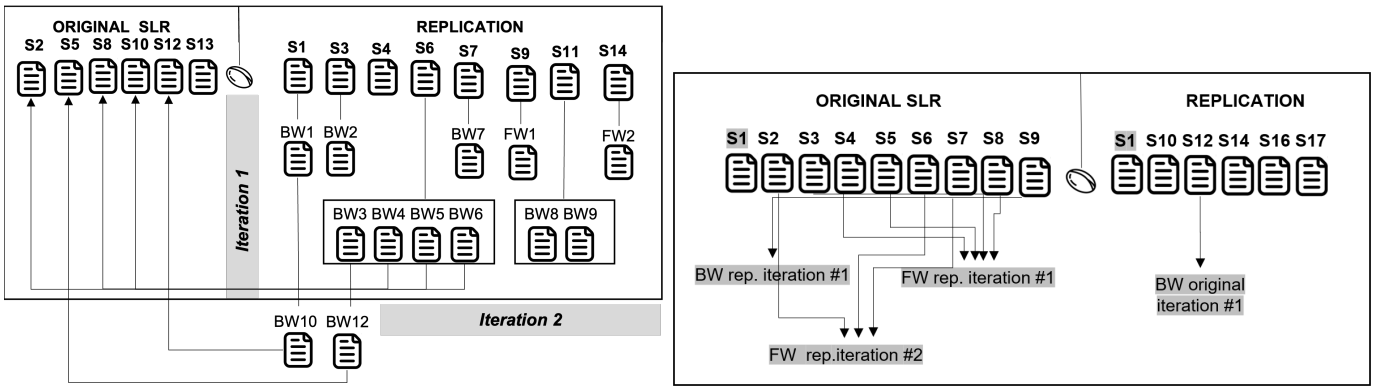

Fig. 2. Timeline from seed selection to return of unused seeds in SLR1 (left) and SLR2 (right).

and BW–N1 [27]. For SLR2, we achieved complete recall (100%), recovered all 43 studies from the original review, and identified eight new candidate studies, five of which were deemed relevant.

From an efficiency point of view, the replication of SLR1 required screening significantly fewer articles (1,944 compared to 6,289 in the original) while improving precision from 0.006 to 0.019. This shows that the diversity-based approach can reduce the workload of the reviewer while preserving, or even enhancing, the comprehensiveness of the evidence base. Although one study from the original SLR1 (S13) was not recovered, the replication revealed three relevant studies that were not present in the original, indicating that the original review may have overlooked important evidence.

The additional studies provided meaningful contributions to the field. S14 [28] investigated the intersection of software testing outsourcing and knowledge management, highlighting the need to integrate KM considerations into outsourcing decisions. BW–N1 [27] presented an automated method for generating test cases based on a requirements ontology supported by inference rules and genetic algorithms. Meanwhile, S19 [29] examined the application of knowledge representation, knowledge management models, and knowledge maps to support knowledge management in software testing. Collectively, these studies expand the scope of existing evidence and reinforce the practical advantages of adopting a diversity-oriented approach in systematic review methodologies.

Number of seeds – Further investigation is needed to determine the optimal size of the seed set. In our SLR1 replication, we used eight seeds; one of them (S4) did not lead to any relevant studies. For SLR2, we used seven seeds instead of the nine in the original review. In particular, only one of these seeds retrieved relevant studies in the first snowballing iteration. These findings suggest that not all seeds contribute equally and that effectiveness may depend more on seed quality than quantity. Additional replications are needed to better understand which factors, such as citation count, topical coverage, or diversity criteria, most influence seed effectiveness.

Author diversity – We start the snowballing with 22 authors for SLR1 and 14 for SLR2. Of the 58 authors in the original SLR1, 50 were not present in the initial seed set. For example, seed S6 (authors: E.F. de Souza, R.A. Falbo, and N.L. Vijaykumar; publication venue: Information and Software Technology Journal (IST)) led to the identification of articles BW3 and BW6. BW3 was co-authored by the same group,

while BW6 appeared in the same journal. Similarly, studies BW9, FW3 and FW5, identified through seeds S11, BW3, and BW14, share at least one author with the corresponding seed, indicating a pattern of author continuity throughout the snowballing process. For SLR2, among the 123 unique authors identified in the original review, only two, Petersen, K. and Runeson, P., were included in the seed sets used in both the original and our replication. This highlights the breadth of author diversity that emerged throughout the snowballing process and demonstrates the effectiveness of the strategy in expanding beyond the original seed set's author base.

Year diversity – For SLR1, the seed set initially covered the years 2007, 2009, 2011, 2013, and 2015. For SLR2, we included studies published up to 2015, following the approach of [2]. More recent studies tend to reflect ongoing research trends and are particularly valuable when the objective is to identify current or emerging evidence, such as when updating an SLR [6]–[8]. In contrast, using older references can help identify prior work and retrieve articles published in the same historical context. Ultimately, the included studies spanned from 2003 to 2015 for SLR1 and from 1999 to 2014 for SLR2.

Publication diversity – We started the snowballing covering eight (8) publication venues for SLR1 (CSCE, EASE, ESEM, ICETAI, ICITCS, IST, MySEC, TSE) and seven (7) for SLR2 (IWLICSE, MIS, ICST, TIP, ICL, ICSTW, ICICKM) and finalize it with a range of 29 for SLR1 and 25 unique for SLR2.

Stop decision – In addition to selecting appropriate seed studies, another critical challenge in snowballing is determining when to stop the iterative process. This decision depends on the researcher's assessment of whether additional relevant studies are still likely to be found. In our SLR1 replication, the number of newly included studies decreased consistently across iterations: 11 in the first, 11 in the second, 5 in the third, and 1 in the fourth. Although the fourth iteration returned only one new study, we conducted a fifth iteration to verify that no further studies would be identified; this served as our stopping criterion. The same rationale guided the fifth iteration in our SLR2 replication. This conservative approach ensures completeness and aligns with previous guidance that recommends continuing iterations until no new relevant studies are found [3].

Some lessons learned from our results are:

Number of seeds – Six (6) to eight (8) studies were appropriate for replication, since they allow the convergence of snowballing in five (5) or six (6) iterations.

Sample diversity – The diversity of seeds in terms of authors, year, and publication venue was observed in the final set of selected studies.

Effort to define seeds – For replication, using the diversity strategy streamlined the process of defining seeds six times.

Snowballing tool support – Google Scholar supports forward snowballing, although it does not allow automatic download of citations. The tool proposed by [30] facilitated the conduction of SLR2 replication by providing valuable support. However, a manual assessment was required because the tool utilized data from the ResearchGate and Semantic Scholar APIs instead of Google Scholar due to limitations imposed by Google.

Threats to validity – The selection of a particular SLR may be biased towards the seed set construction. However, the choice of the SLRs was guided by a set of criteria directed toward the SLR's content and did not favor the diversity strategy. The specific topic covered in the selected SLR was not crucial as the objective was to evaluate snowballing rather than synthesizing evidence from identified studies. However, replication researchers needed to be comfortable with the topic to ensure accurate inclusion or exclusion of articles. Furthermore, there is a risk that individual selection becomes biased. However, having two researchers perform independent evaluations on all articles helped mitigate individual assessor bias. Overall, the articles were judged according to the SLR's design and the predefined criteria used in the evaluation to help minimize the validity threats to the conclusion.

Concerning internal validity, we minimized the risk of missing essential studies by extracting citations with the help of Google Scholar, which offers a feature known as citation tracking. Researchers can see more recent articles that cite the original by finding a known article within the database. Moreover, we strictly followed the guidelines for the conduct of snowballs suggested by Wohlin [3]. The snowballing process is not without challenges. It is based on the assumption that other relevant studies are indexed in the same way as the seeds. This assumption was proven accurate in our replication. However, despite their high relevance to content, some studies may be indexed outside of known academic bases or as gray literature. We use Google Scholar to identify citations to mitigate this obstacle since "...The use of a generic database is sufficient to discover most of the studies..." [8].

We also recognize that one limitation of our study is that the diversity-based strategy is based solely on studies indexed in Scopus. As a result, potentially relevant studies not indexed in this database were not included in the candidate pool, which may have affected the completeness and coverage of our selection process. Future work could incorporate additional digital libraries, such as Web of Science, IEEE Xplore, and ACM Digital Library, to enhance the studies' diversity and representativeness.

VIII. Conclusions

In light of our findings, the primary contribution of this research is the exploration of a diversity-based strategy for constructing snowballing seed sets. Besides presenting our analysis, we provide a detailed replication to enable transparency and reproducibility. We acknowledge that our results are not definitive; rather, we encourage further studies exploring diversity-based seed set construction in SLRs within software engineering. Future research could validate or challenge our findings through additional empirical studies combining SLRs, diversity-driven seed sets, and snowballing methods, or even formal experiments. Ultimately, these results serve as a foundation for a deeper understanding and broader adoption of diversity strategies in seed set construction.

Supplementary data – Download all supplementary files included with this article: https://zenodo.org/records/15284025

IX. Acknowledgment

Professor Katia Romero Felizardo is funded by a research grant from the Brazilian National Council for Scientific and Technological Development (CNPq), Grant $302339/2022-1$.

References

[1] B. A. Kitchenham, D. Budgen, and Brereton, *Evidence-Based Software Engineering and Systematic Reviews*. United States: Chapman & Hall/CRC, 2015.

[2] C. Wohlin, M. Kalinowski, K. R. Felizardo, and E. Mendes, "Successful combination of database search and snowballing for identification of primary studies in systematic literature studies," *Information and Software Technology*, vol. 147, p. 106908, 2022.

[3] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *International Conference on Evaluation and Assessment in Software Engineering*, EASE' 14, (New York, NY, USA), pp. 321–330, ACM, 2014.

[4] E. Mourão, J. F. Pimentel, L. Murta, M. Kalinowski, E. Mendes, and C. Wohlin, "On the performance of hybrid search strategies for systematic literature reviews in software engineering," *Journal of Information and Software Technology*, vol. 123, no. 7, p. 106294, 2020.

[5] T. Greenhalgh and R. Peacock, "Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources," *Bmj*, vol. 331, pp. 1064—1065, 2005.

[6] C. Wohlin, E. Mendes, K. R. Felizardo, and M. Kalinowski, "Guidelines for the search strategy to update systematic literature reviews in software engineering," *Information and Software Technology*, vol. 127, p. 106366, 2020.

[7] K. R. Felizardo, E. Mendes, M. Kalinowski, E. F. Souza, and N. L. Vijaykumar, "Using forward snowballing to update systematic reviews in software engineering," in *International Symposium on Empirical Software Engineering and Measurement*, ESEM' 16, (New York, United States), ACM, 2016.

[8] K. R. Felizardo, A. Y. I. da Silva, E. F. de Souza, N. L. Vijaykumar, and E. Y. Nakagawa, "Evaluating strategies for forward snowballing application to support secondary studies updates: Emergent results," in *Brazilian Symposium on Software Engineering*, SBES' 18, (Brazil), pp. 184–189, Sociedade Brasileira de Computação, 2018.

[9] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *Journal of Systems and Software*, vol. 80, no. 4, pp. 571–583, 2007. Software Performance.

[10] S. Jalali and C. Wohlin, "Systematic literature studies: Database searches vs. backward snowballing," in *International Symposium on Empirical Software Engineering and Measurement*, ESEM' 12, (New York, United States), pp. 29–38, ACM, 2012.

[11] D. Badampudi, C. Wohlin, and K. Petersen, "Experiences from using snowballing and database searches in systematic literature studies," in *International Conference on Evaluation and Assessment in Software Engineering*, EASE' 15, (New York, NY, USA), ACM, 2015.

[12] C. Wohlin, "Second-generation systematic literature studies using snowballing," in *International Conference on Evaluation and Assessment in Software Engineering*, EASE' 16, (New York, NY, USA), pp. 1–6, ACM, 2016.

[13] E. Mourão, M. Kalinowski, L. Murta, E. Mendes, and C. Wohlin, "Investigating the use of a hybrid search strategy for systematic reviews," in *International Symposium on Empirical Software Engineering and Measurement*, ESEM' 17, (New York, United States), pp. 193–198, ACM, 2017.

[14] R. de Souza, C. Lopes, F. Bezerra, and C. R. B. de Souza, "Ramani: Uma ferramenta de apoio à colaboração durante a execução de estudos sistemáticos," in *Brazilian Symposium in Collaborative Systems*, SBSC' 13, (Brazil), pp. 144—147, Sociedade Brasileira de Computação, 2013.

[15] F. Bezerra, C. H. Favacho, R. Souza, and C. de Souza, "Towards supporting systematic mappings studies: Automatic snowballing approach," in *Simpósio Brasileiro de Banco de Dados*, SBBD' 14, (Brazil), pp. 1673–176, Sociedade Brasileira de Computação, 2014.

[16] S. C. P. F. Fabbri, C. Silva, E. Hernandes, F. Octaviano, A. Di Thommazo, and A. Belgamo, "Improvements in the start tool to better support the systematic review process," in *International Conference on Evaluation and Assessment in Software Engineering*, EASE' 16, (New York, NY, USA), ACM, 2016.

[17] E. S. Monsalve, E., J. C. S. P. Leite, and C. J. P. Calvache, "Semi-automatic mapping technique using snowballing to support massive literature searches in software engineering," *Revista Facultad de Ingeniería*, vol. 31, no. 60, p. e14189, 2022.

[18] V. Garousi, K. Petersen, and B. Ozkan, "Challenges and best practices in industry-academia collaborations in software engineering: A systematic literature review," *Information and Software Technology*, vol. 79, pp. 106–127, 2016.

[19] F. C. Ferrari, A. V. Pizzoleto, and J. Offutt, "A systematic review of cost reduction techniques for mutation testing: Preliminary results," in *2018 IEEE International Conference on Software Testing, Verification and Validation Workshops*, ICSTW' 18, (Vasteras, Sweden), pp. 1–10, IEEE Press, 2018.

[20] A. Silva, T. Araújo, J. Nunes, M. Perkusich, E. Dilorenzo, H. Almeida, and A. Perkusich, "A systematic review on the use of definition of done on agile software development projects," in *International Conference on Evaluation and Assessment in Software Engineering*, EASE' 17, (New York, NY, USA), p. 364–373, ACM, 2017.

[21] S. Stradowski and L. Madeyski, "Industrial applications of software defect prediction using machine learning: A business-driven systematic literature review," *Information and Software Technology*, vol. 159, p. 107192, 2023.

[22] K. Wnuk and T. Garrepalli, "Knowledge management in software testing: A systematic snowball literature review," *e-Informatica Software Engineering Journal*, vol. 12, no. 1, pp. 51–78, 2018.

[23] F. C. Ferrari, V. H. S. D. Durelli, S. F. Andler, J. Offutt, M. Saadatmand, and N. Mullner, "On transforming model-based tests into code: A systematic literature reviewg," *Frontiers in Research Metrics and Analytics*, vol. 33, no. 8, p. e1860, 2023.

[24] J. Zhang and J. Li, "Testing and verification of neural-network-based safety-critical control software: A systematic literature review," *Information and Software Technology*, vol. 123, p. 106296, 2020.

[25] D. Marques, T. Dallegrave, C. Oliveira, and W. Santos, *Successful Practices in Industry-Academy Collaboration in the Context of Software Agility: A Systematic Literature Review*, pp. 292–310. ACM, 07 2023.

[26] B. Minetto Napoleão, R. Sarkar, S. Hallé, F. Petrillo, and M. Kalinowski, "Supplementary material: Articles in the slr replication," 2024. Supplementary files of the paper "Emerging Results on Automated Support for Searching and Selecting Evidence for Systematic Literature Review Updates" published at WSESE' 24.

[27] V. Tarasov, H. Tan, A. Adlemo, A. Andersson, I. Muhammad, M. Johansson, and D. Olsson, "Ontology-based software test case generation (ostag)," in *European Projects in Knowledge Applications and Intelligent Systems*, (Lisbon, Portugal), pp. 135–159, INSTICC, SciTePress, 2015.

[28] K. Karhu, O. Taipale, and K. Smolander, "Outsourcing and knowledge management in software testing," in *International Conference on Evaluation and Assessment in Software Engineering*, EASE' 07, (Swindon, GBR), p. 53–63, BCS Learning & Development Ltd., 2007.

[29] Y.-P. Liu, L. Zou, M.-Z. Jin, and X.-M. Liu, "Technology for knowledge management in software testing and its application," *Computer Integrated Manufacturing Systems, CIMS*, vol. 14, no. 9, pp. 1805–1809+1844, 2008.

[30] B. Minetto Napoleão, R. Sarkar, S. Hallé, F. Petrillo, and M. Kalinowski, "Emerging results on automated support for searching and selecting evidence for systematic literature review updates," in *Proceedings of the 1st IEEE/ACM International Workshop on Methodological Issues with Empirical Studies in Software Engineering*, WSESE' 24, (New York, NY, USA), p. 34–41, ACM, 2024.

[31] E. Mendes, C. Wohlin, K. Felizardo, and M. Kalinowski, "When to update systematic literature reviews in software engineering," *Journal of Systems and Software*, vol. 167, p. 110607, 2020.

[32] A. Iwazaki, V. Santos, K. Felizardo, E. de Souza, N. Valentim, and E. Nakagawa, "Benefits and challenges of a graduate course: An experience teaching systematic literature review," in *Frontiers in Education*, FIE' 22, (Uppsala, Sweden), pp. 1–8, IEEE Press, 2022.

[33] M. Riaz, M. Sulayman, N. Salleh, and E. Mendes, "Experiences conducting systematic reviews from novices' perspective," in *International Conference on Evaluation and Assessment in Software Engineering*, EASE' 10, (Newcastle, UK), pp. 1–10, ACM, 2010.