Suescún-Monsalve, Elizabeth; Sampaio-do-Prado-
Leite, Julio-Cesar; Pardo-Calvache, César-Jesús
Semi-Automatic Mapping Technique Using Snowballing to
Support Massive Literature Searches in Software Engineering

Semi-Automatic Mapping Technique Using Snowballing to Support Massive Literature Searches in Software Engineering

Elizabeth Suescún-Monsalve[1]
Julio-Cesar Sampaio-do-Prado-Leite[2]
César-Jesús Pardo-Calvache[3]

[1] Ph. D. Universidad EAFIT (Medellín-Antioquia, Colombia). esuescu1@eafit.edu.co. ORCID: 0000-0001-7872-7638
[2] Ph. D. Pontificia Universidad Catolica do Rio de Janeiro (Rio de Janeiro, Brasil). julio@inf.puc-rio.**br**. **ORCID**: 0000-0002-0355-0265
[3] Ph. D. Universidad del Cauca (Popayán-Cauca, Colombia). cpardo@unicauca.edu.co. ORCID: 0000-0002-6907-2905

Abstract

Systematic literature reviews represent an important methodology in Evidence-Based Software Engineering. To define the methodological route in these type of studies, in which a review of quantitative and qualitative aspects of primary studies is carried out to summarize the existing information regarding a particular topic, researchers use protocols that guide the construction of knowledge from research questions. This article presents a process that uses forward Snowballing, which identifies the articles cited in the paper under study and the number of citations as inclusion criteria to complement systematic literature reviews. A process that relies on software tools was designed to apply the Snowballing strategy and to identify the most cited works and those who cite them. To validate the process, a review identified in the literature was used. After comparing the results, new works that were not taken into account but made contributions to the subject of study emerged. The citation index represents the number of times a publication has been referenced in other documents and is used as a mechanism to analyze, measure, or quantitatively assess the impact of said publication on the scientific community. The present study showed how applying Snowballing along with other strategies enables the emergence of works that may be relevant for an investigation given the citations rate. That is, implementing this proposal will allow updating or expanding systematic literature studies through the new works evidenced.

Keywords: citation impact; evidence-based software engineering; massive literature searches; snowballing; software engineering; systematic mapping.

Proceso de mapeo semiautomático guiado por *snowballing* para apoyar búsquedas masivas de literatura en ingeniería de software

Resumen

Los estudios sistemáticos de literatura representan una metodología importante en Ingeniería de Software Basada en Evidencias, para definir la ruta metodológica en este tipo de estudios, en los cuales se realiza una revisión de aspectos cuantitativos y cualitativos de estudios primarios, con el fin de resumir la información existente

Elizabeth Suescún-Monsalve; Julio-Cesar Sampaio-do-Prado-Leite; César-Jesús Pardo-Calvache

sobre un tema en particular los investigadores utilizan protocolos que guían la construcción del conocimiento a partir de las preguntas de investigación. Este artículo presenta un proceso que utiliza la propuesta conocida como *Snowballing* hacia adelante, que identifica los artículos citados por el artículo en estudio y el número de citas como criterios de inclusión para complementar los estudios sistemáticos de literatura. Se diseñó un proceso que se apoya en herramientas de software para aplicar la estrategia *Snowballing* e identificar los trabajos más citados y quienes los citan. Se validó el proceso por comparación y surgieron nuevos trabajos que hicieron aportes al tema de estudio, que no habían sido considerados inicialmente. El índice de citas representa el número de veces que una publicación ha sido referenciada en otros documentos y se utiliza como mecanismo para analizar, medir o evaluar cuantitativamente el impacto de dicha publicación en la comunidad científica. La presente propuesta mostró cómo la aplicación del *Snowballing* con otras estrategias permite evidenciar nuevos de trabajos que pueden ser relevantes para una investigación dada la tasa de citas. Es decir, el uso de la presente propuesta permitirá actualizar o ampliar los estudios bibliográficos sistemáticos por los nuevos trabajos evidenciados.

Palabras clave: búsquedas masivas de la literatura; impacto de la citación; ingeniería de software; ingeniería de software basada en evidencia; mapeo sistemático; snowballing.

Processo de mapeamento semiautomático guiado por *Snowballing* para apoiar revisões massivas de literatura na engenharia de software

Resumo

Os estudos sistemáticos da literatura representam uma importante metodologia na engenharia de software baseada em evidências, para definir o caminho metodológico nestes tipos de estudos, nos quais é realizada uma revisão de aspectos quantitativos e qualitativos de estudos primários, a fim de resumir as informações existentes sobre um determinado tópico; pesquisadores utilizam

protocolos que orientam a construção do conhecimento a partir de questões de pesquisa. Este artigo apresenta um processo que utiliza a proposta conhecida como Snowballing para frente, que identifica os artigos citados pelo artigo em estudo e o número de citações como critérios de inclusão para complementar os estudos sistemáticos da literatura. Foi projectado um processo que conta com ferramentas de software para aplicar a estratégia Snowballing e identificar os trabalhos mais citados e aqueles que os citam. Para validar o processo, foi utilizada uma revisão identificada na literatura, os resultados obtidos foram comparados e surgiram novos trabalhos que trouxeram contribuições ao objeto de estudo, que não haviam sido levados em consideração. O índice de citação representa o número de vezes que uma publicação foi referenciada em outros documentos e é utilizado como mecanismo para analisar, medir ou avaliar quantitativamente o impacto dessa publicação na comunidade científica. A presente proposta mostrou como a aplicação do Snowballing com outras estratégias permite o surgimento de trabalhos que podem ser relevantes para uma investigação dada a taxa de citações. Ou seja, a utilização da proposta atual permitirá atualizar ou ampliar estudos sistemáticos da literatura pelos novos trabalhos evidenciados.

Palavras-chave: engenharia de software; engenharia de software baseada em evidências; impacto da citação; mapeamento sistemático; revisões massivas na literatura; snowballing.

Elizabeth Suescún-Monsalve; Julio-Cesar Sampaio-do-Prado-Leite; César-Jesús Pardo-Calvache

I. INTRODUCTION

Empirical research in Software Engineering (SE) is a form of experimentation or observation based on evidence. Between 2004 and 2005, Kitchenham, Dybå, and Jørgensen wrote three relevant works proposing the Evidence-Based Software Engineering (EBSE) concept [1,2,3]. This is based on previous works applied in the medicine, which were later taken up and adopted by other disciplines including economics, psychology, social sciences, and SE. The EBSE was adopted in the research processes to give methodological rigor to the results identified in the systematic literature studies, and to make these results impartial and more reliable [3]. The mechanism of systematic literature studies to identify and contribute evidence in medicine consists of several stages of SE [4]. It was structured in six steps organized into four methodological phases, i.e., (i) Set out a research question; (ii) Search for the evidence to answer the question; (iii) evaluate the evidence critically; and finally, (iv) use the evidence to address the question.

Systematic literature studies in EBSE are secondary studies classified into Systematic Mapping Studies (SMS), which are broad literature studies on a topic that seek to find the available evidence on the subject [4, 5]; and the Systematic Literature Reviews (SLR), which aim to identify, evaluate, and combine the evidence from primary studies and answer a research question in detail [6]. These systematic literature studies are supported by research guidelines or protocols that guide researchers to obtain results in less or greater depth. Examples of the application of these protocols and their results can be seen in [5, 7, 8, 9, 10, 11, 12, 13, 14, 15].

The protocols indicate how to proceed in a way that facilitates the analysis of results, the conclusions, and the possibility to replicate the study. K. Petersen et al. [6] describe the essential steps of the process to carry out an SMS from the definition of research questions to the result of the process. Petersen et al. [16] shows a protocol was updated and the concepts of PICO (Population, Intervention, Comparison, and Outcomes) were added to identify keywords and formulate search strings based on research questions and validity assessment such as descriptive validity, theoretical validity, generalization, and interpretive validity.

Additionally, strategies to include grey literature have emerged [17, 18, 19, 20]. Also, strategies to complement the protocols of systematic literature studies or to design a search strategy that appropriately balances result quality and review effort were suggested [21, 22, 23, 24]. For instance, the strategy presented in [25, 26], which aims to extend and detail the searches using the list of references or citations to identify additional works, better known as Snowballing. It could also be an alternative to update studies [27, 28, 29, 30, 31] or to maintain and manage traceability [32] to make the studies more reproducible [33, 34]. Henceforth, Snowballing will be the most widely used term in the scientific community.

It is evident that systematic literature studies benefit the researcher's work [35, 36, 37]; however, the execution of the protocols can be a repetitive, slow, laborious, and error-prone activity because there are many steps to follow, sources of information to consider, various jobs to manage. It is also a time-consuming activity [38], there can be problems caused by unreproducible research [39] or plagiarism [40]. Thus, systematic literature studies in SE require much more effort than traditional review techniques. Therefore, given the innumerable advantages of conducting systematic literature studies and the effort and additional work that they require, this work presents a semi-automatic process that supports the researcher in the execution of the research protocols in literature searches. In most cases, massive searches for information should be considered. The process is based on identifying the number of citations from lists of papers that enable finding new ones to be included. The proposed process will be used to complement traditional search processes and to cover more studies, especially those that may be disregarded by the manual treatment of information conducted by the researcher. In this vein, it will be possible to have a much more agile and precise protocol that can be replicated by implementing the proposal presented here, considering that the number of citations may change over time.

The article is organized as follows: Section 2 presents a theoretical framework and related works. Section 3 is an example of the application of the proposed process.

Section 4 presents the analysis of results. Finally, Section 5 presents the conclusions and recommendations for future work.

II. MATERIALS AND METHODS

A. Related Work

In general, the works mentioned below were identified. Wohlin and Rainer [41] argue that mistakes could be made, thus resulting in the production, consumption, and dissemination of invalid evidence. They propose a framework and a set of recommendations to help the community to produce and consume credible evidence. Pizard et al. [42] proposed to improve the practice to obtain better results. Therefore, provide a guidance for producers and consumers called "Bad smells" for software analytics papers [43]. Besides, present a process to select studies based on the use of statistics, in which the criteria are refined until reaching agreement [44]. After that, the researchers interpret the selection criteria and the bias is reduced as well as the time spent. In the same way, present an approach based on robust statistical methods for empirical SE that could be applied to systematic literature studies [45]. Finally, report several problems identified during their research that threaten any type of literature study and hinder the support of adequate tools [46]. They also recommend solutions to mitigate these threats.

Other works that propose ways to support researchers when they do systematic literature studies were identified. One of them is defined by Bezerra et al. [47], where an algorithm to perform forward and backward Snowballing is proposed, but the sources to replicate the study are not identified. Tsafnat et al. [48], a literature review is carried out to identify tools that support or automate the processes or tasks of systematic literature studies. Most of the findings focus on automating and/or simplifying specific tasks of the protocols. Likewise, some tasks are already fully automated while others remain largely manual. Moreover, some studies describe the effect that automation has on the entire process, summarize the support of the tool for each task, and highlight what subjects require more research to carry out

automation. As a research opportunity, they highlight the importance of integrating tools in literature reviews. In [49], a tool to support the implementation of a protocol for SMS-type studies is proposed. Finally, Marshall and Brereton [50], 14 tools that can help automate part or all of the process proposed in the protocols are identified and classified. The works found focus mainly on automating certain points of the protocol process that guide systematic literature studies or suggest the integration of several tools to have a more robust or unified proposal. Others focus on the analysis of the protocols. It is worth mentioning that none of them specifically focuses on the use of forward Snowball with citations as a strategy to include works. We highlight the statement of [51], who state that the quality of conclusions completely depends on the quality of the selected literature.
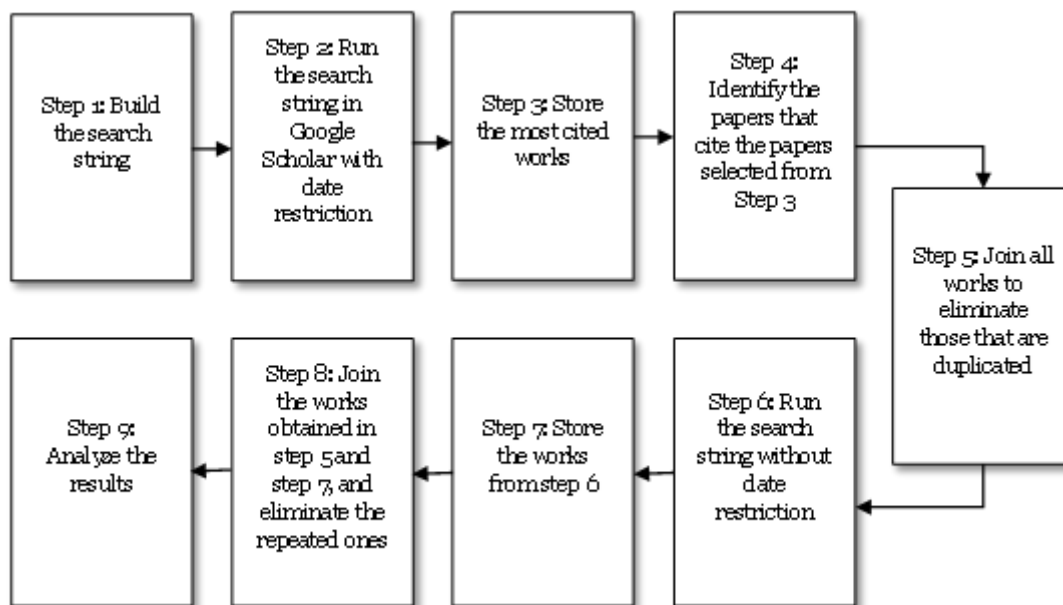

Fig. 1. Snowballing-based proposed process to support systematic literature studies.

B. Proposed Process Steps

The process proposes to automate some steps of the Snowballing technique presented by Wohlin [26]. Additionally, it is intended to extend and deepen the search for works to identify other documents. That is, to complement the search by

identifying the list of references and citations, being this number the inclusion or exclusion factor. This evaluation will be conducted based on the information provided by Google Scholar, a search engine for academic literature that has access to digital libraries and open scientific works. It is opensource and allows researchers to quickly obtain works that may be relevant to an investigation. In addition, Google Scholar provides information that may be useful for researchers, e.g., number of citations, related works, versions of the work, format for citations, among others. The number of citations will be the main element in this proposal.

The proposed process is summarized in Figure 1, as protocol with a series of steps to be executed sequentially. It is semi-automatic since in the first instance we will rely on a reference management tool. In this case we will use Zotero (https://www.zotero.org) as bibliographic manager and spreadsheets of Microsoft Excel to manage and store the results. In addition, the process is defined by the researcher's criteria, who finally decides which works to include or exclude from the research. As can be seen in Figure 1, some steps are proposed, and their details are shown in Table 1. It should be noted that the suggested steps can be adjusted according to the researcher's will, as well as the number of works selected in each. The example selection and the application example of the proposed process is described in detail in https://bit.ly/3qWUfXd.

Table 1. Steps of the process proposed in the present work.

Id.	Name	Description
1	Build the search string.	In this step, keywords are identified and strings are formulated to answer the research questions in the study.
2	Execute the search string in Google Scholar with date restriction.	When executing the string, it is possible to find several works, therefore, it is necessary to filter by the most cited ones and restrict the search by date.
3	Store the most cited works.	In this step, the selected documents are stored so that they can be processed in later stages. It should be noted that the selection of the most cited works is left to the discretion of the researchers.
4	Identify the papers that cite the papers selected from Step 3.	To the selected works, in Step 3, the Snowballing forward strategy is applied, that means to identify who is citing them and filter by the most cited ones.
5	Join all works in order to eliminate those that are duplicated.	To get a unified list, repeated works must be eliminated and from this point on managed the selected ones.

Id.	Name	Description
6	Run search string without date restriction.	This step aims to find works that have historically been highly cited or are benchmarks in the research area.
7	Store works from Step 6.	The data is stored from use in subsequent steps.
8	Join the works obtained in Step 5 and Step 7, and eliminate the repeated ones.	Both lists are unified and the repeated ones are eliminated to obtain a single list in which the data is processed.
9	Analyze the results.	Analysis of the results can be carried out to compare them or to identify whether they meet the research questions and quality criteria.

C. Analysis of the Results Obtained by Both Proposals

At this point it is important to consider the perspective of the research questions and the purpose of the work, and thus, make the necessary adjustments because the proposal does not seek to replace the protocols for systematic reviews of the existing literature. On the contrary, the objective is to complement the strategy the researcher is carrying out. Therefore, the first step was to compare the data obtained by executing the proposed process with the Coded Papers of Connolly et al. [52] (more information in: https://n9.cl/d3t70 and https://n9.cl/yx8p). It was possible to observe that: (i) the Coded Papers did not have the number of citations as a requirement or inclusion criteria, this is why the occurrences in both works are different; and (ii) some of the digital libraries used to elaborate the work of Connolly et al. [52] must be accessed through subscriptions and some like Google Scholar cannot track them. Some works in the Coded Papers do not have any citation, but their authors indicate that these works answer the research questions. Therefore, according to our proposal, works with a small number of citations or without citations would not be considered.
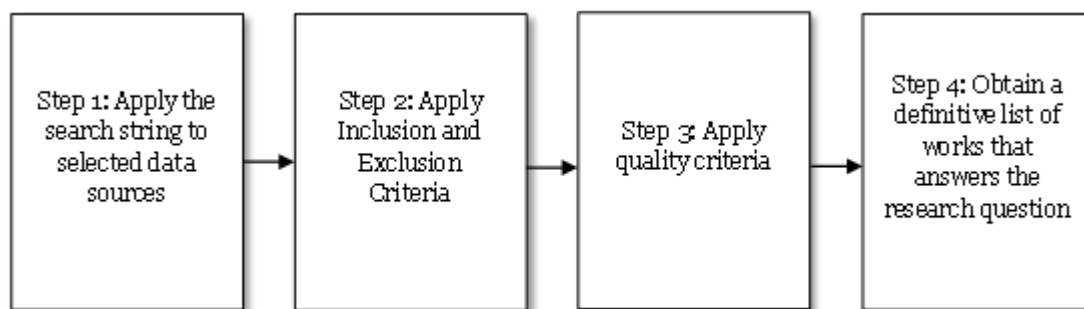
Fig. 2. Summary of the process applied by Connolly et al. [1].

On the other hand, in the process followed by Connolly et al. [52] represented in Figure 2, the inclusion and exclusion criteria are relevant to arrive at the definitive works. Therefore, the following steps are highlighted:

Step 1. The search string was applied to the selected data sources. According to the authors, 7,932 papers were found in ACM, ASSIA, BioMed Central, Cambridge Journals Online, ChildData, Index to Theses, Oxford University Press (journals), Science Direct, EBSCO, PsycINFO, SocINDEX, Library, Information Science and Technology Abstracts, CINAHL, ERIC, Ingenta Connect, InfoTrack, Emerald, IEEE Computer Society, and Digital Library.

Step 2. Inclusion and exclusion criteria were applied. This step made it possible to reduce the list of works to 129. According to the authors, the criteria were: (i) they include empirical evidence related to the impacts and results of the use of games; (ii) the works' data should be in a time window of only 5 years; (iii) they include an abstract, and (iv) they include participants older than 14 years of age in the studies.

Step 3. Quality criteria were applied to the 129 works obtained after applying the selection criteria: (i) game category; (ii) categorization of game effects; and (iii) coding methods. Then, the works were read and assigned a grade between 1 and 3 according to the following dimensions: research design, method and analysis, research results, relevance of the study focus, and whether the study could be extended.

Step 4: A final list of works was obtained with one more step after the Coded Papers (129). Then, the definitive list contained 70 works that answered the research

question, being considered by their authors as the best quality works for the investigation. Based on them, the final analysis was elaborated and presented.

III. RESULTS AND DISCUSSION

Results show that the number of citations can be used as a criterion to measure the importance of a work since those who are citing it find it useful. Hence, it indicates that the most cited works have an impact on a field. The proposal of this article helps to identify the most cited papers and iterate on those results. It does not pretend to replace the protocols already established, in turn, it is a support strategy. It is also necessary to use other criteria to identify whether the results of applying the proposal are answering the research questions, as described below.

Following the process proposed by Connolly et al. [52], the final quality criteria used as a new filter were: (i) empirical evidence on results and impacts related to the use of games; (ii) effects on games, mainly focused on positive ones; and (iii) the method used to evaluate the games, e.g., study with qualitative and/or quantitative results, examples, samples, sampling, data collection, data analysis, results, and conclusions.

In the application example based on our proposal, the quality criteria described by Connolly et al. [52], and subsequently, the number of citations criterion was no longer considered. By doing this, 58 works were discarded and a final list of 54 works was obtained (more information available at: https://n9.cl/3ek5i). They enabled drawing a final discussion (more information available at: https://n9.cl/jywvb and https://n9.cl/w8e3) and conducting the analysis of the subject. Regarding this last step, it should be noted that (i) all the selected works are in the context of serious games and computer games; (ii) the investigations whose scope was the evaluation of negative aspects of entertainment games were discarded; (iii) the works at design stages were discarded; (iv) the works including pedagogy were better valued; (v) many discarded papers were related to studies on the behavior of users when they played games, which was not within the scope of this study; (vi) the 54 works answered the research question and the number of citations is based on the

assumption that important works are usually cited (more information available at: https://n9.cl/3ek5i)

IV. CONCLUSIONS

To support the protocols used in systematic literature studies, a forward Snowball proposal with number of citations as inclusion criteria is an alternative that enables covering considerable volumes of information if used along with Google Scholar. The number of citations indicates the impact a work has had and also its quality by counting the number of times other authors cite it.

The work of Connolly et al. [52] was used for comparative analysis. New works that also answered the research questions and that may eventually be considered by the researcher to be included or to expand the study were found. By identifying these works and quantifying their citations, they can be considered important in the explored area of knowledge. The proposal does not necessarily reduce time or effort, but it does reveal works that could not have been considered due to previous strategies and that due to the number of citations may be impacting the area of interest of the research. The success of the application of the semi-automatic process based on Snowballing to support research protocols in massive literature searches lies fundamentally in the permanent validation of the procedure and its steps by the researcher.

As a future work, it is expected to fully automate the process and make more comparisons, as well, it would be useful to develop a tool that automates the process. However, we identified that there may be a limitation when trying to use a bulk search using Google Scholar, since it can detect and block the URL used. Other works showed that using massive queries can be detected as a security threat by the platform. Additionally, Google Scholar does not consider some digital libraries that require subscription or payment. Those ones should be added manually if the researcher considers it so. Moreover, some works may be left out of the search if Google Scholar is used without previous settings. The different standards of

scientific journals that belong to certain digital databases must be also considered when automating the process to back the proposal established in this work.

AUTHORS' CONTRIBUTION

Elizabeth Suescún-Monsalve: Supervision, investigation, writing -original draft, writing - revision and edition.

Julio-Cesar Sampaio-do-Prado-Leite: Supervision, investigation, writing - revision and edition.

César-Jesús Pardo-Calvache: Supervision, investigation, writing - revision and edition.

REFERENCES

[1] T. Dyba, B. A. Kitchenham, M. Jorgensen, "Evidence-based software engineering for practitioners," *IEEE Software*., vol. 22, no. 1, pp. 58–65, 2005. https://doi.org/10.1109/MS.2005.6

[2] M. Jorgensen, T. Dyba, B. Kitchenham, "Teaching evidence-based software engineering to university students," in *11th International Software Metrics Symposium*, 2005, p. 8. https://doi.org/10.1109/METRICS.2005.46

[3] B. A. Kitchenham, T. Dyba, M. Jorgensen, "Evidence-based software engineering," in *Proceedings. 26th International Conference on Software Engineering*, 2004, pp. 273–281. https://doi.org/10.1109/ICSE.2004.1317449

[4] T. Dybå, T. Dingsøyr, G. K. Hanssen, "Applying systematic reviews to diverse study types: An experience report," in *Proceedings 1st International Symposium on Empirical Software Engineering and Measurement*, 2007, no. 7465, pp. 225–234. https://doi.org/10.1109/ESEM.2007.59

[5] B. Barn, S. Barat, T. Clark, "Conducting systematic literature reviews and systematic mapping studies," in *10th Innovations in Software Engineering Conference*, 2017. https://doi.org/10.1145/3021460.3021489

[6] K. Petersen, R. Feldt, S. Mujtaba, M. Mattsson, "Systematic mapping studies in software engineering," in *12th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, 2008. https://doi.org/10.14236/ewic/EASE2008.8

[7] B. Martin, J. Irvine, "Assessing basic research," *Research policy*, vol. 12, no. 2, pp. 61–90, 1983. https://doi.org/10.1016/0048-7333(83)90005-7

[8] W. A. Chapetta, G. H. Travassos, "Towards an evidence-based theoretical framework on factors influencing the software development productivity," *Empirical Software Engineering*, vol. 25, no. 5, pp. 3501–3543, 2020. https://doi.org/10.1007/s10664-020-09844-5

[9] C. Wohlin, E. Papatheocharous, J. Carlson, K. Petersen, E. Alégroth, J. Axelsson, D. Badampudi, M. Borg, A. Cicchetti, F. Ciccozzi, T. Olsson, S. Sentilles, M. Svahnberg, K. Wnuk, T. Gorschek, "Towards evidence-

based decision-making for identification and usage of assets in composite software: A research roadmap," *Journal of Software: Evolution and Process*, vol. 33, no. 6, e2345, 2021. https://doi.org/10.1002/smr.2345

[10] L. Shanshan, H. Zhang, Z. Jia, C. Zhong, C. Zhang, J. Shen, M Babar, "Understanding and addressing quality attributes of microservices architecture: A Systematic literature review," *Information and software technology*, vol. 131, e106449, 2021. https://doi.org/10.1016/j.infsof.2020.106449

[11] V. Garousi, D. Pfahl, J.Fernandes, M. Felderer, M. Mäntylä, D. Shepherd, A. Arcuri, A. Coşkunçay, B. Tekinerdogan, "Characterizing industry-academia collaborations in software engineering: evidence from 101 projects," *Empirical Software Engineering*, vol. 24, no. 4, pp. 2540–2602, 2019. https://doi.org/10.1007/s10664-019-09711-y

[12] E. Souza, A. Moreira, M. Goulão, "Deriving architectural models from requirements specifications: A systematic mapping study," *Information and software technology*, vol. 109, pp. 26–39, 2019. https://doi.org/10.1016/j.infsof.2019.01.004

[13] J. Barros, F. Pinciroli, S. Matalonga, N. Martínez-Araujo, "What software reuse benefits have been transferred to the industry? A systematic mapping study," *Information and Software Technology*, vol. 103, pp. 1-21, 2018. https://doi.org/10.1016/j.infsof.2018.06.003

[14] T. Ribeiro, J. Massollar, G. H. Travassos, "Challenges and pitfalls on surveying evidence in the software engineering technical literature: an exploratory study with novices," *Empirical Software Engineering*, vol. 23, no. 3, pp. 1594–1663, 2018. https://doi.org/10.1007/s10664-017-9556-7

[15] M. Felderer, J. C. Carver, "Guidelines for systematic mapping studies in security engineering," in *Empirical Research for Software Security*, 2017, pp. 47–68. https://doi.org/10.48550/arXiv.1801.06810

[16] K. Petersen, S. Vakkalanka, L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Information and software technology*, vol. 64, pp. 1–18, 2015. https://doi.org/10.1016/j.infsof.2015.03.007

[17] V. Garousi, A. Rainer, "Gray literature versus academic literature in software engineering: A call for epistemological analysis," *IEEE Software*, vol. 38, no. 5, pp. 65–72, 2021. https://doi.org/10.1109/MS.2020.3022931

[18] X. Zhou, "How to treat the use of grey literature in software engineering," in *Proceedings of the International Conference on Software and System Processes*, 2020. https://doi.org/10.1145/3379177.3390305

[19] V. Garousi, M. Felderer, M. V. Mäntylä, "Guidelines for including grey literature and conducting multivocal literature reviews in software engineering," *Information and software technology*, vol. 106, pp. 101–121, 2019. https://doi.org/10.1016/j.infsof.2018.09.006

[20] A. Williams, "Using reasoning markers to select the more rigorous software practitioners' online content when searching for grey literature," in *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering*, 2018. https://doi.org/10.1145/3210459.3210464

[21] E. Mourão, J. Pimentel, L. Murta, M. Kalinowski, E. Mendes, C. Wohlin, "On the performance of hybrid search strategies for systematic literature reviews in software engineering," *Information and software technology*, vol. 123, no. 1, e106294, 2020. https://doi.org/10.1016/j.infsof.2020.106294

[22] Y. Shakeel, J. Krüger, I. von Nostitz-Wallwitz, O. von Guericke, C. Lausberger, G. Campero, G. Saake, T. Leich, "(Automated) literature analysis - threats and experiences," in *13th International Workshop on Software Engineering for Science*, 2018, pp. 20–27. https://doi.org/10.1145/3194747.3194748

[23] D. Carrizo, J. Manriquez, "Impact of assessment of empirical studies reliability: A revisited study," in *37th International Conference of the Chilean Computer Science Society*, 2018. https://doi.org/10.1109/SCCC.2018.8705250

[24] E. Hassler, D. Hale, J. Hale, "A comparison of automated training-by-example selection algorithms for Evidence Based Software Engineering," *Information and Software Technology*, vol. 98, pp. 59–73, 2018. https://doi.org/10.1016/j.infsof.2018.02.001

[25] C. Wohlin, R. Prikladnicki, "Systematic literature reviews in software engineering," *Information and software technology*, vol. 55, no. 6, pp. 919–920, 2013. https://doi.org/10.1016/j.infsof.2017.12.004

[26] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, 2014. https://doi.org/10.1145/2601248.2601268

[27] E. Mendes, K. Felizardo, C. Wohlin, M. Kalinowski, "Search strategy to update systematic literature reviews in software engineering," in *45th Euromicro Conference on Software Engineering and Advanced Applications*, 2019. https://doi.org/10.1109/SEAA.2019.00061

[28] E. Mendes, C. Wohlin, K. Felizardo, M. Kalinowski, "When to update systematic literature reviews in software engineering," *Journal of Systems and Software*, vol. 167, e110607, 2020. https://doi.org/10.1016/j.jss.2020.110607

[29] V. Nepomuceno, S. Soares, "On the need to update systematic literature reviews," *Information and software technology*, vol. 109, pp. 40–42, 2019. https://doi.org/10.1016/j.infsof.2019.01.005

[30] E. Mourao, M. Kalinowski, L. Murta, E. Mendes, C. Wohlin, "Investigating the use of a hybrid search strategy for systematic reviews," in *International Symposium on Empirical Software Engineering and Measurement*, 2017. https://doi.org/10.1109/ESEM.2017.30

[31] P. Singh, K. Singh, "Exploring automatic search in digital libraries: A caution guide for systematic reviewers," in *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering*, 2017. https://doi.org/10.1145/3084226.3084275

[32] V. Nepomuceno, S. Soares, "Maintaining systematic literature reviews: Benefits and drawbacks," in *Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, 2018. https://doi.org/10.1145/3239235.3267432

[33] B. Kitchenham, L. Madeyski, P. Brereton, "Meta-analysis for families of experiments in software engineering: a systematic review and reproducibility and validity assessment," *Empirical Software Engineering*, vol. 25, no. 1, pp. 353–401, 2020. https://doi.org/10.1007/s10664-019-09747-0

[34] Z. Li, "Stop building castles on a swamp! The crisis of reproducing automatic search in evidence-based software engineering," in *43rd International Conference on Software Engineering: New Ideas and Emerging Results*, 2021. https://doi.org/10.1109/ICSE-NIER52604.2021.00012

[35] Z. Yu, N. A. Kraft, T. Menzies, "Finding better active learners for faster literature reviews," *Empirical Software Engineering*, vol. 23, no. 6, pp. 3161–3186, 2018. https://doi.org/10.1007/s10664-017-9587-0

[36] N. Ali, M. Usman, "Reliability of search in systematic reviews: Towards a quality assessment framework for the automated-search strategy," *Information and Software Technology*, vol. 99, pp. 133–147, 2018. https://doi.org/10.1016/j.infsof.2018.02.002

[37] S. Barat, T. Clark, B. Barn, V. Kulkarni, "A model-based approach to systematic review of research literature," in *10th Innovations in Software Engineering Conference*, 2017. https://doi.org/10.1145/3021460.3021462

[38] J. C. Carver, E. Hassler, E. Hernandes, N. A. Kraft, "Identifying barriers to the systematic literature review process," in *International Symposium on Empirical Software Engineering and Measurement*, 2013. https://doi.org/10.1109/ESEM.2013.28

[39] L. Madeyski, B. Kitchenham, "Would wider adoption of reproducible research be beneficial for empirical software engineering research?," *Journal of Intelligent & Fuzzy Systems*, vol. 32, no. 2, pp. 1509–1521, 2017. https://doi.org/10.3233/JIFS-169146

[40] V. Nepomuceno, S. Soares, "Avoiding plagiarism in systematic literature reviews: An update concern," in *Proceedings of the 14th International Symposium on Empirical Software Engineering and Measurement*, 2020. https://doi.org/10.1145/3382494.3422170

[41] C. Wohlin, A. Rainer, "Challenges and recommendations to publishing and using credible evidence in software engineering," *Information and software technology*, vol. 134, e106555, 2021. https://doi.org/10.1016/j.infsof.2021.106555

[42] S. Pizard, F. Acerenza, X. Otegui, S. Moreno, D. Vallespir, B. Kitchenham, "Training students in evidence-based software engineering and systematic reviews: a systematic review and empirical study," *Empirical Software Engineering*, vol. 26, no. 3, pp. 1-53. 2021. https://doi.org/10.1007/s10664-021-09953-9

[43] T. Menzies. M. Shepperd, "'Bad smells' in software analytics papers," *Information and software technology*, vol. 112, pp. 35–47, 2019. https://doi.org/10.1016/j.infsof.2019.04.005

[44] J. Pérez, J. Díaz, J. Garcia-Martin, B. Tabuenca, "Systematic literature reviews in software engineering—enhancement of the study selection process using Cohen's Kappa statistic," *Journal of Systems and Software*, vol. 168, e110657, 2020. https://doi.org/10.1016/j.jss.2020.110657

[45] B. Kitchenham, L. Madeyski, D. Budgen, J. Keung, P. Brereton, S. Charters, S. Gibbs, A. Pohthong, "Robust statistical methods for empirical software engineering," *Empirical Software Engineering*, vol. 22, no. 2, pp. 579–630, 2017. https://doi.org/10.1007/s10664-016-9437-5

[46] V. Garousi, A. Rainer, M. Felderer, M. V. Mäntylä, "Introduction to the Special Issue on: Grey Literature and Multivocal Literature Reviews (MLRs) in software engineering," *Information and software technology*, vol. 141, no. 1, e106697, 2022. https://doi.org/10.1016/j.infsof.2021.106697

[47] F. Bezerra, C. H. Favacho, R. Souza, C. de Souza, *Towards supporting systematic mappings studies: An automatic snowballing approach*: https://bit.ly/3uIG890

[48] G. Tsafnat, P. Glasziou, M. K. Choong, A. Dunn, F. Galgani, E. Coiera, "Systematic review automation technologies," *Systematic reviews*, vol. 3, no. 1, p. 74, 2014. https://doi.org/10.1186/2046-4053-3-74

[49] R. Montebelo, A. Orlando, D. Porto, D. Zaniro, S. Fabbri, *Uma Ferramenta Computacional de Apoio à Revisão Sistemática*. https://bit.ly/3uRcBd8

[50] C. Marshall, P. Brereton, "Tools to support systematic literature reviews in software engineering: A mapping study," in *International Symposium on Empirical Software Engineering and Measurement*, 2013. https://doi.org/10.1109/ESEM.2013.32

[51] L. Yang, H. Zhang, H. Shen, X. Huang, X. Zhou, G. Rong, D. Shao, "Quality assessment in systematic literature reviews: A software engineering perspective," *Information and Software Technology*, vol. 130, e106397, 2021. https://doi.org/10.1016/j.infsof.2020.106397

[52] T. M. Connolly, E. A. Boyle, E. MacArthur, T. Hainey, J. M. Boyle, "A systematic literature review of empirical evidence on computer games and serious games" *Computers & education*, vol. 59, no. 2, pp. 661–686, 2012. https://doi.org/10.1016/j.compedu.2012.03.004