SESRA – A Web-based Automated Tool to Support the Systematic Literature Review Process

Jefferson Seide Molléri

UNIVALI – Universidade do Vale do
Itajaí, Itajaí, SC – Brazil
jefferson.molleri@univali.br

BTH – Blekinge Institute of
Technology, Karlskrona – Sweden
jefferson.molleri@bth.se

Fabiane Barreto Vavassori Benitti

UFSC – Universidade Federal de Santa
Catarina, Florianópolis, SC – Brazil
fabiane.benitti@gmail.com

Abstract—Systematic Literature Review (SLR) is a key tool for evidence-based practice as it combines results from multiple studies of a specific topic of research. Due its characteristics, it is a time consuming, hard process that requires a properly documented protocol for scientific acknowledgment. In this context, this paper presents the SESRA – a web-based automated tool to support all phases of the SLR process, contributing to its productivity and reliability. We also present a use case on the software engineering field, applying specific knowledge to set the process in the discipline. Further, we discuss how to use the tool to establish a more formal and controlled process and to reduce effort on its repetitive activities. The outcomes and feedback obtained in early use have shown that SESRA could support the SLR process, automating some of its key activities.

Keywords—systematic literature review; SLR; automated tool

I. INTRODUCTION

Systematic Literature Review (SLR) is method to identify, evaluate and interpret the relevant research to a particular topic or phenomenon of interest. The conduction of an objective and rigorous systematic review present a fair evaluation of a research topic. Such reviews are a key tool for enabling evidence-based practice as they combine the results of multiple primary studies [1].

SLRs are intended to produce systematic comparisons between primary studies, allowing to create generalization from these. However, even when conducted according to their corresponding “best practices”, SLR process is not an easy task. Due its characteristics, it requires a more formal and controlled process than other unsystematic reviews. The lack of such scientific rigor in performing its several phases restrains the scientific acknowledgment of the review [2].

SLRs also require more effort than traditional literature reviews, providing additional information on identified primary studies in a wide variety of empirical methods [3]. Moreover,

SLR processes produces great volume of information and knowledge of complex management [4].

Given the above, the support of an automated tool is essential to provide higher quality and facilitate the systematic literature review process [2, 5]. Motivated by this need, we present the SESRA, an automated tool to support the SLR method, aiming productivity and reliability criteria of the process.

The tool was developed based on the systematic review process proposed by Kitchenham and Charters [1] and includes a series of expert’s recommendations to support specific phases and discrete activities of the process. We previously conducted a secondary study to identify the recommendations [5] and build a process model to integrate them to the SLR process [6].

This paper is organized as follows: section II presents the SESRA tool and how it supports the SLR process; section III discusses the outcomes and feedback obtained in early use; and section IV presents the final conclusions and further work.

II. SESRA TOOL

SESRA is a web-based software¹ that allows the researcher to manage and conduct SLR through three distinct phases: planning, conducting, and reporting the review; through a series of distinct stages and activities as proposed by [6].

Guidelines and support tips are provided in each stage to guide the researcher to avoid biases that compromise the review results. Developed to serve the SLR process in the software engineering (SE) discipline, SESRA includes specific activity settings proposed to the field, as describe in the phases below.

A. Phase 1: Planning the Review

SLR process starts with the **identification of the need for a review** (illustrated in Fig.1) that ensures the individuality of the proposed study. To achieve this, identification and assessment of any existing systematic reviews over the phenomenon of interest against appropriate evaluation criteria are required.

© 2015 Association for Computing Machinery. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

Fig. 1. Identification of the need for a review stage

The **commissioning of the review** stage allows the management of the researchers and other stakeholders involved in the process. The tool sends invitation and notification messages to stakeholders, informing their roles and assignments on the research process. Then, **the research question(s)** should be defined, addressing the five components of the PICOC acronym: population, intervention, comparison, outcomes and context. On software engineering context, SESRA provides automated expansion of the PICOC components through the SWEBOK [7] ontology, which contains the main concepts of the discipline.

To **develop the review protocol**, researchers should specify the methods to undertake the systematic review, fulfilling the following information:

- background and research question(s);
- search string, which is automatically generated by concatenating the PICOC components;
- strategy that will be used to search for primary studies;
- study selection criteria and procedures;
- study quality assessment instruments and procedures;
- data extraction and data synthesis strategy, including expected data format (i.e. text, number, boolean);
- dissemination strategy; and
- timetable, with stakeholders and their assignments.

To finish the planning phase, it is needed to **evaluate the review protocol**. Often, independent experts are invited to review and criticize the review protocol. SESRA also exports a document containing the information given in the protocol (as shown in Fig. 2).

Fig. 2. Evaluation of the review protocol stage

B. Phase 2: Conducting the Review

The second phase of SLR process starts with the **identification of research**. The search string obtained in the review protocol should be adjusted following the rules for each database, and then used to retrieve references for potentially useful studies. Contextual help supports the research in elaborate the specific search strings. SESRA also allows the document retrieval by manual inputting the references, by importing a BibTeX reference file, or by gathering studies from specific databases in SE field (i.e. IEEE Xplore, IET Digital Library and SpringerLink).

Follows the **study selection** stage (illustrated in Fig. 3), in which the studies should be assessed on inclusion and exclusion criteria. Researchers could undertake the selection process based on title and abstract, but is often necessary to obtain a copy of the paper to assess full text. When performed by multiple researchers is not uncommon to find divergences among the selection. SESRA tools helps in reaching an agreement by inviting an advisor to review the selection, or by applying an inter-rater reliability test to evaluate a random sample of the selected studies.

Fig. 3. Study selection stage

The **study quality assessment** stage (as shown in Fig. 4) involves applying a quality instrument or checklist to weight the importance of the included studies for the review results. SESRA provides such instruments suitable for the software engineering field, as Dyba and Dingsøyr [8], Kitchenham and Charters [1] and the CRD Database of Abstracts of Reviews of Effects [9]. Researchers can alternatively set specific quality instruments suitable for their SLR.

Fig. 4. Quality assessment stage

During **data extraction** stage, SESRA generate extraction forms based on the review protocol. Each field is validated according the data format set during the development of the review protocol. Some bibliographic data can be automatically extracted if a BibTeX reference document has been provided for the primary study.

Extracted data are summarized in the **data synthesis** stage. A comprehensible table (as shown in Fig. 5) is provided to demonstrate and compare qualitative and quantitative data of included studies. We also plan to offer, as further work, graphs and charts (such as forest plots and funnel plots) to present the results and illustrate correlations among specific data sets.

Title	Publication Year	Quality Score	Review Type	Review Focus	Review Topic	Cited EBSE paper?	Cited guidelines?	Number of Primary Studies
SE01	2008	4	MS	SERT	Human Aspects	N	Y	92
SE02	2008	4	SLR	RT	Knowledge Management	Y	Y	68
SE03	2008	1.5	MS	RT	Research Topics in Software	N	N	691

Fig. 5. Data extraction stage

C. Phase 3: Reporting the Review

The third phase of the process has three distinct stages. First, **specifying dissemination mechanisms** allows the researchers to select a dissemination format (such as journal paper or technical report). Follows the **formatting the main report** stage, in which specific information reporting the review should be filled, such as methods, results, discussion and conclusions, as shown in Fig. 6.

SESRA then exports a report document that should be evaluated by an external advisor at the stage of **evaluating the report**. After the approval, the review process is given as finished, and protocol and report documents remain available for further reference.

Abstract

After the introduction of evidence-based software engineering in 2004, systematic literature reviews (SLR) have become a common practice. Our goal is to analyze quality, coverage of software engineering topics, and potential impact of published SLRs. In July 2008 and December 2009, analyzed the relevant studies, compared and integrated our findings with previous studies. The results show that the overall quality of the studies is improving, suggesting that the systematic review method. However, the majority of the SLRs did not evaluate the quality of primary studies and fail to provide guidance for practice.

Research Questions

RQ1: Which automated approaches to support the secondary studies processes are used by researchers?
RQ2: Which the phases and stages are supported by the identified approaches?
RQ3: What limitations, phases and stages are not covered by the identified approaches?
RQ4: What are the impacts of using an automated approach into productivity and reliability criteria the secondary studies?

Methodology

We perform our search looking for peer-reviewed articles published between 1st July 2008 and 31st December 2009. The Automatic search was performed by R4 and R5 on 6 search engines and indexing systems: AC, Scopus, All searches were performed on the entire paper, including title and abstract, except for the ISI Web of Science. This is the string used in the automatic search:

('secondary study' OR 'systematic review' OR 'literature review' OR 'mapping review') AND (software OR 'conducting OR reporting OR 'research question' OR 'review protocol' OR 'study selection' OR 'quality extraction' OR 'data synthesis' OR 'meta analysis' OR meta-analysis) AND (productivity OR reliability OR 'data sources list above.

Fig. 6. Formatting the main report stage

III. DISCUSSION

We have shown how SESRA supports the SLR process through its very specific phases and stages, according with Kitchenham and Charters [1] guidelines and our own view of the process [6]. This section discuss how to use the tool to a more productive and formal process, generating reliable and verifiable results.

A. Productivity and Reliability

In the SLR context, the reliability criterion is often associated with the consistency and repeatability of the method, when the process can be redone in different conditions or by another group of researchers [10]. SESRA tool create a repository containing all the input data, as well the process documentation that researchers can still access to recreate or re-evaluate the process.

Reliability is also related to the study quality, ensured by the proper description of the methodology through the review protocol, the adoption of a quality assessment stage, and the conduction of study selection by multiple researchers [8, 11]. Such requirements are consistent with SESRA implementation, with comprehensive guidelines provided in each step.

To provide productivity and effectiveness on the SLR process, is essential to adopt an automated or semi-automated approach to support the process [12], as well to use project management techniques to time management and provide communication among stakeholders [13]. The tool maintains a timetable of ongoing processes, notifying researchers and advisors when their attention is required in a specific activity.

Productivity is also associated with the reduction of effort in performing the process activities, especially when related to the conduction phase [1, 2]. SESRA provides automation on search procedures and document retrieval of included studies.

B. Tool Evaluation

Since May 2013, SESRA tool is available to the research community. We first provide a version in Portuguese aiming the evaluation of the tool by a Brazilian group of researchers. Later, we have developed the English version for the international community.

An evaluation of SESRA using the GQM (goal, question, metric) model [14] have been conducted aiming the software engineering community [15]. The evaluation aims the productivity (G1) and reliability (G2) criteria of the tool on supporting SLRs. A mix of subjective and objective metrics were proposed, as shown in Table 1. Questions related to subjective metrics were investigate by a survey with researchers and practitioners, based on theirs experience on the tool. To obtain the objective metrics, we designed and conducted some experimental procedures on the automated features of SESRA tool.

Table 1. Questions for the tool evaluation

Question ²	Goal	Type
PQ1. The review protocol generated by the automated approach is as reliable as the document drafted by hand?	G2	Subj.
PQ2. Proposed automated approach is in accordance with the procedures for the planning the review phase of the systematic review process?	G2	Subj.
PQ3. Proposed automated approach reduces the effort in the planning phase of the review the systematic review process?	G1	Subj.
PQ4. Implemented automated expansion of the search terms using a body of knowledge in software engineering is a reliable technique?	G2	Obj.
CQ1. Proposed automated approach is in accordance with the procedures for the conducting the review phase of the systematic review process?	G2	Subj.
CQ2. Proposed automated approach reduces the effort in the conducting the review phase of the systematic review process?	G1	Subj.
CQ3. Results of the search for primary studies through automatically generated string are as reliable as the search performed manually?	G2	Obj.
CQ4. Supporting tools for automated selection of studies are as reliable as a manual selection of studies?	G2	Obj.
CQ5. Obtaining primary studies automatically via an API or importing a reference file reduces the effort in the identification of research stage of the review process?	G1	Obj.
CQ6. Obtaining primary studies automatically via an API or importing a reference file is as reliable as performed manually?	G2	Obj.
CQ7. Automated extraction of data from a BibTeX reference document provided reduces the effort in the data extraction stage of the systematic review process?	G1	Obj.
RQ1. The review report generated by the automated approach is as reliable as the document drafted by hand?	G2	Subj.
RQ2. Proposed automated approach is in accordance with the procedures for the reporting the review phase of the systematic review process?	G2	Subj.
RQ3. Proposed automated approach reduces the effort in the reporting the review phase of the systematic review process?	G1	Subj.

² PQ acronym stands for questions related to the planning review phase, CQ to the conducting the review phase, and RQ to reporting the review.

1) Subjective Evaluation

Survey answers indicate that users “broadly agree” with the tool support the SLR process in terms of productivity (average of 60.7%) and reliability (average of 72.1%), as shown in Figure 7. According to answers, the conduction phase presents better support than planning and reporting. Some “partially disagrees” were noted on the reliability of the review protocol and report.

Evaluators also supply us comments based on their expectations and usage experience, as such “the research question field should be stricter” and “generated documents might not be suited to expectations of the researcher”. We also receive positive feedback, asserting that “the main contribution of the tool is to organize the SLR process”.

Fig. 7. Distribution of survey answers according to each question

According to the results of subjective evaluation, the tool contributes to productivity and reliability criteria in the SLR process, providing a comprehensible software to support the SLR process and reducing effort by automation of its activities. However, the evaluation presents several limitations on generalizability, featuring only a small number of researchers in Brazil.

2) Objective Evaluation

To obtain the objective metrics, a series of experimental procedures were conducted to assess the main automated features on the SESRA tool: search string expansion (PQ4), generation of search strings for specific databases (CQ3), a similarity coefficient to support study selection (CQ4), reference retrieval through APIs or importing reference files (CQ5 and CQ6), and bibliographic data extraction (CQ7).

The procedures involves the selection of a sample of systematic review studies to provide benchmarks on the data we are investigating (e.g. number of collected studies by database, quantity of retrieved references per time period). Then, we run the automated activities on the SESRA tool using the same specification of the sample. Further, we collect the results and compare to the sample, using a percentage scale. We also synthesized the results on the same scale as the subjective metrics, as given in Table 2.

Table 2. Results on Objective Evaluation

Question	Results	Synthesis
PQ4	78.5%	broad contribution on reliability
CQ3	54.4%	partial contribution on reliability
CQ4	52%	partial contribution on reliability
CQ5	>100%	broad contribution on productivity
CQ6	98.5%	broad contribution on reliability
CQ7	93.8%	broad contribution on productivity

The interpretation of objective results suggests that SESRA tool can contribute to productivity and reliability on SLR process in software engineering. It is clear, however, that the procedures performed on automated features to reference retrieval and study selection contributed only partially (as demonstrated by results on CQ3 and CQ4), suggesting further enhancements and revaluation procedures. Full description of the procedures and supplementary data are presented in ours Master’s dissertation [15].

C. Further Work

Feedback from users during the subjective evaluation and analysis on experimental procedures were documented and analyzed face up to the process model [6]. Given that, the following opportunities for future work have been identified:

- Review the fields related to the research objective and research question, according to comments of users on the subjective evaluation;
- Enhance the automated generation of search strings for specific databases, and the similarity coefficient to support study selection, as identified on the objective evaluation;
- Integrate the document retrieval feature with additional databases, as such ACM Digital Library, CiteSeerX, and ScienceDirect;
- Apply snowballing sample techniques to obtain primary studies, using bibliographic references of already selected studies;
- Apply similarity techniques to propose agreements to solve selection conflicts, as proposed by Felizardo *et al.* [16];
- Generation of graphs and charts (such as forest plots and funnel plots) on the data synthesis stage;
- Support SLRs in other disciplines, including specific ontology and providing proper quality assessment instruments; and
- Conduct a more extensive evaluation of the tool, aiming the community of practitioners and researchers on software engineering and related fields.

IV. CONCLUSIONS

This paper presented the SESRA – a supporting tool to the systematic literature review process, providing a mean to perform all the phases and stages of the process [6]. The tool also implements automation techniques to reduce effort on the more time consuming activities, as identified in a previous study [5]. We outlined an evaluation undertaken, its overall results on productivity and reliability criteria, and the feedback on usage by practitioners on the software engineering field.

According to Biolchini *et al.* [2], to conduct a formal process and report reliable results is essential to obtain scientific rigor on systematic literature reviews. However, due its characteristics, it is a hard process that should benefit from computational tools to facilitate its several activities. In this context, SESRA aims at providing automated support to SLRs in software engineering context, contributing to productivity and reliability of the process.

ACKNOWLEDGEMENTS

The authors would like to thank Pedro Paulo de Mello Filho, Rafael Piemontez and Luiz Eduardo da Silva who contributed to the tool development. This work was supported by the Institutional Program of Initiation Scholarships for Technological Development and Innovation (PIBIT) of the Brazilian National Council for Scientific and Technological Development (CNPq).

REFERENCES

- [1] Kitchenham, B., Charters, S. Guidelines for performing systematic literature reviews in software engineering (version 2.3). Technical Report. University of Durham, 2007.
- [2] Biolchini, J., Gomes, P., Cruz, A., & Travassos, G. Systematic review in software engineering. Technical Report. Universidade Federal do Rio de Janeiro, 2005.
- [3] Shull, F.; Mendonça, M.; Basili, V.; Carver, J.; Maldonado, J. C.; Fabbri, S.; Travassos, G. H.; Ferreira, M. C. "Knowledge-Sharing Issues in Experimental Software Engineering". Empirical Software Engineering, 2004.
- [4] Ali, S., Briand, L. C., Hemmati, H., & Panesar-Walawege, R. K. A systematic review of the application and empirical investigation of search-based test case generation. Software Engineering, IEEE Transactions on, vol. 36(6), pp. 742-762, 2010.
- [5] Molléri, J. S., & Benitti, F. B. V. Automated Approaches to Support Secondary Study Processes: a Systematic Review. In Proc. 24th Conf. Software Engineering & Knowledge Engineering (SEKE), pp. 143-147, 2012.
- [6] Molléri, J. S., Silva, L. E. & Benitti, F. B. V. Proposal of an automated approach to support the systematic review of literature process. In Proc. 25th Conf. Software Engineering & Knowledge Engineering (SEKE), pp. 488-493, 2013.
- [7] Abran, A., Bourque, P., Dupuis, R., & Moore, J. W. Guide to the software engineering body of knowledge-SWEBOK. IEEE Press, 2011.
- [8] Dybå, T., & Dingsøyr, T. Strength of evidence in systematic reviews in software engineering. In Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement (pp. 178-187). ACM, October 2008.
- [9] Khan, K. S., Ter Riet, G., Glanville, J., Sowden, A. J., & Kleijnen, J. Undertaking systematic reviews of research on effectiveness: CRD's guidance for carrying out or commissioning reviews (No. 4 (2nd Edition)). NHS Centre for Reviews and Dissemination, 2001.
- [10] Biolchini, J. C. A., Mian, P. G., Natali, A. C. C., Conte, T. U., & Travassos, G. H. Scientific research ontology to support systematic review in software engineering. Advanced Engineering Informatics, vol. 21(2), pp. 133-151, 2007.
- [11] Alderson, P., Green, S., & Higgins, J. P. T. Cochrane reviewers' handbook 4.2. 2 [updated March 2004]. The Cochrane Library, vol. 1, pp. 1-59, 2004.
- [12] Zamboni, A. B., Di Thommazo, A., Hernandez, E. C. M., & Fabbri, S. C. P. F. StArt Uma Ferramenta Computacional de Apoio à Revisão Sistemática. In Proc.: Congresso Brasileiro de Software (CBSOFT'10), Salvador, Brazil, September 2010.
- [13] Baker, P. R., Francis, D. P., Hall, B. J., Doyle, J., & Armstrong, R. Managing the production of a Cochrane systematic review. Journal of public health, vol. 32(3), pp. 448-450, 2010.
- [14] Caldiera, V. R. B. G., & Rombach, H. D. The goal question metric approach. Encyclopedia of software engineering, vol. 2(1994), pp. 528-532, 1994.
- [15] Molléri, J. S. 2013. Automatização do Processo de Condução de Revisões Sistemáticas da Literatura em Engenharia de Software. Master's Dissertation. UNIVALI - Universidade do Vale do Itajaí.
- [16] Felizardo, K. R., Andery, G. F., Maldonado, J. C., & Minghim, R. Uma Abordagem Visual para Auxiliar a Revisão da Seleção de Estudos Primários na Revisão Sistemática. In Proceedings of 6th Experimental Software Engineering Latin American Workshop (ESELAW 2009), pp. 83, November 2009.