Analysis of the Tools to Support Systematic Literature Review in Software Engineering

Darko Stefanovic¹, Sara Havzi²*, Danilo Nikolic³, Dusanka Dakic⁴ and Teodora Lolic⁵

¹,²,³,⁴,⁵Department of Industrial Engineering and Engineering Management, Faculty of Technical Sciences, University of Novi Sad, 21 000, Serbia

E-mail: havzisara@uns.ac.rs

Abstract. A thorough and accurate review of the literature is a necessary condition for relevant research. However, the literature review is a laborious, error-prone process. For this reason, various software tools are being developed to support researchers in conducting literature reviews. In this paper, three tools to support a systematic literature review in software engineering were analysed and evaluated using the DESMET methodology. Tools were found by conducting mapping study and were selected since they were most promising to support an overall systematic literature review process. Although the tools do not yet support the whole systematic review process on a high level, they provide a reasonable basis for further development. It is needed to set features that need to be supported by the tools.

1. Introduction

A review of the literature is the basis for establishing comprehension of the state in any scientific area. Evidence-based research initially started in medicine, because research has shown that medical advice and conclusions based on expert opinion are not as reliable as advice based on the accumulation of the results of scientific experiments [1].

In order for the literature review to be correct and relevant, specific rules, as well as strategies and methods, need to be developed in the literature review process. For the reasons given, the literature review presents a complicated process, with a high possibility of error and subjectivity. All of this leads to a large amount of time required to develop a transparent literature review.

In software engineering, the most used approach for the review of the literature is a systematic literature review (SLR) developed by Barbara Kitchenham [2].

Approaches to reviewing the literature, especially the SLR, have reduced these problems to some extent. However, many authors point out that there are still problems with the amount of time spent reviewing the literature, the occurrence of errors and the subjectivity of researchers in this process [3]. In 2012, the research conducted by Hernandez and Craft showed that there are many activities of the SLR that need to be reduced in the difficulty level or time needed to perform an activity. The review of the literature is a manual process that requires a large amount of work [4]–[6] and is, therefore, the primary candidate to benefit aided by technology.

In recent years, software tools are being developed in order to solve the current problems that researchers are facing when conducting a literature review. In order to examine the usefulness of the developed tools in solving existing problems, it is necessary to evaluate these tools. Marshall [7]
conducted a feature analysis using DESMET methodology [8] on the set of candidate tools: SluRp [9], StArt [10], SLR-Tool [11] and SLRTOOL.

The study reported here aims to evaluate three other tools using the same method as Marshall. This way, there will be more tools evaluated by the same method, and researchers will have a broader scope of tools to consider.

The remainder of the paper is organised as follows. In Section 2, authors presented SLR as well as the literature review in the field software tools for literature reviews. In Section 3, the methodology for performing this study was presented. Results and discussion are presented in Section 4. Section 6 concludes the study, its limitations and suggests future research.

2. Tools to Support Systematic Literature Reviews in Software Engineering

SLR is one of the most used approaches in performing the review process of the literature [12]. Fink defined systematic literature review as “systematic, explicit, and reproducible method for identifying, evaluating, and synthesising the existing body of completed and recorded work produced by researchers, scholars, and practitioners” [13].

Therefore, researchers who conduct an SLR use transparent and a systematic process to define the research question, search the publication, assess their quality and qualitative or quantitative synthesis of research results.

SLR guidelines that are given by Kitchenham [2] consist of three main steps:

- planning,
- conducting and
- reporting the review.

Planning the review is the first phase, and it contains activities such as identification of the need for a review, development of protocol and validation of protocol. Conducting the review consists of the identification of research, study selection, quality assessment, data extraction, as well as data synthesis. Reporting means documenting the review, which is writing the review and validating the report. For the SLR to be correct, all of the steps above need to be well performed.

In order to support SLR process in:

- reducing the time required to complete the literature review process,
- increasing quality of assessment of the relevance of papers and review of literature,
- reducing the subjectivity of researchers in the process of reviewing the literature, as well as
- making it easier for researchers to review the literature,

software tools for supporting the SLR process where developed.

In order to map as many tools as possible that are developed to support the SLR, we conducted a mapping study and SLR study and found more than one hundred tools, that could, to some extent help in conducting the SLR. However, there are only a few tools that support the whole SLR process in a way, and some of them are not applicable in the field of software engineering.

3. Method

This section describes the method used to perform the evaluation of the three candidate tools, as well as the description of the chosen feature analysis approach. In addition, it describes how Marshall [14] applied this method evaluating three tools for SLR support, which is used in this paper.

3.1. DESMET method

DESMET is a methodology for evaluating tools or methods. It was developed by Barbara Kitchenham [8] and consisted of nine evaluation methods suitable for different tool analyses. There are quantitative and qualitative methods in DESMET methodology [15]. In the guidelines, the first step to perform DESMET methodology is selecting an appropriate evaluation method.
DESMET method is not flawless, so it has limits when trying to mix and match methods and tools as well as if there is no controllable development process in the organisation [15].

Kitchenham defined evaluation criteria on the basis of which the DESMET methods are chosen. Marshall [7] chose the feature analysis method for this type of evaluation.

3.2. Feature analysis

Feature analysis is a short type of evaluation method and is considered to have a low risk. The objective of the feature analysis provides an input for a decision whether to use or not use a tool in an organisation. Evaluation should consider the following areas [15]:

1. Suitability for purpose
2. Economic issues
3. Drawbacks
4. Other advantages

“A feature analysis type of evaluation is an attempt to put rationale and structure behind a “gut feeling” for the right product” [15]. This evaluation method is also considered as subjective.

Some of the criteria that should be considered when deciding whether to perform a feature analysis are [15]:

- a large number of methods/tools to assess and
- short timescales for evaluation exercise.

3.3. Candidates, Features and Scoring

This chapter presents candidate tools for evaluation as well as a brief overview of the features, feature analysis and scoring used in Marshalls evaluation. The detailed method can be found in [7]. The tool was evaluated by the authors based on a review of the tool as well as publicly available publications and tools documentation. Candidates were chosen from the unpublished SLR of software tools that support literature review in software engineering mentioned above. Candidate tools are chosen as their developers claim that they support each stage of the SLR process. Tools that were already evaluated by Marshall weren’t taken into consideration of this evaluation.

The three candidate tools are:

1. SESRA - a web application to support systematic literature review and is a completely free tool. The tool is available in Portuguese and English.
2. ReLiS – a web application to support the iterative systematic literature review process.
3. Parsifal - a web application that supports researchers in performing a systematic review of the literature.

A detailed explanation of the features and scoring can be seen in [7].
Subfeature levels of importance are [7]:
1. Mandatory feature – M – 4 points
2. Highly Desirable – HD – 3 points
3. Desirable – D – 2 points
4. Nice to have – N – 1 point

Interpretations of the Judgement Scales are [7]:
1. J11 – Is the feature present?
2. J12 – Is the tool simple to install and setup?
3. J13 – Is the activity supported?

### 4. Results

In this section, the authors give an overview of the evaluation results for every candidate tool. Feature sets for every tool are briefly described, and a table with grades and the overall score is presented for every tool. Summarisation of the results of all the papers is presented in Table 4.

#### 4.1. SESRA

Table 1 presents the scores for SESRA.

#### 4.1.1. Feature Set 1

In conducting this study, the authors noticed that there are parts of the tool that have not been translated from Portuguese into English, in the English version. The tool is not maintained often and the last changes, according to the author, happened two years ago. SESRA score 3 out of 6 for this feature set.

| id | Feature Set     | id         | Subfeature                          | Subfeature Level of Importance | Interpretation of Judgement Scale | Feature set Importance Weighting |
|----|-----------------|------------|-------------------------------------|--------------------------------|-----------------------------------|----------------------------------|
| F1 | Economic        | F1-SF01    | The tool does not require financial payment to use. | HD                             | III                              | 0.1                              |
|    |                 | F1-SF02    | Maintenance                         | HD                             | III                              |                                  |
| F2 | Ease of introduction and setup | F2-SF01    | The tool has reasonable system requirements. | M                              | III                              | 0.2                              |
|    |                 | F2-SF02    | Simple installation and setup.      | HD                             | III                              |                                  |
|    |                 | F2-SF03    | There is an installation guide.     | HD                             | III                              |                                  |
|    |                 | F2-SF04    | There is a tutorial.               | HD                             | III                              |                                  |
|    |                 | F2-SF05    | The tool is self-contained.         | HD                             | III                              |                                  |
| F3 | SR activity support | F3-SF01    | Protocol development                | D                              | JB                               | 0.4                              |
|    |                 | F3-SF02    | Protocol validation                 | D                              | JB                               |                                  |
|    |                 | F3-SF03    | Supports automated searches         | HD                             | JB                               |                                  |
|    |                 | F3-SF04    | Study selection and validation      | HD                             | JB                               |                                  |
|    |                 | F3-SF05    | Quality assessment and validation   | HD                             | JB                               |                                  |
|    |                 | F3-SF06    | Data extraction and validation      | HD                             | JB                               |                                  |
|    |                 | F3-SF07    | Automated analysis                  | HD                             | JB                               |                                  |
|    |                 | F3-SF08    | Text analysis                       | N                              | III                              |                                  |
|    |                 | F3-SF09    | Meta-analysis                       | N                              | III                              |                                  |
|    |                 | F3-SF10    | Report write up                     | N                              | JB                               |                                  |
|    |                 | F3-SF11    | Report validation                   | N                              | JB                               |                                  |
| F4 | Process Management | F4-SF01    | Support for multiple users          | M                              | III                              | 0.3                              |
|    |                 | F4-SF02    | Document management                 | M                              | III                              |                                  |
|    |                 | F4-SF03    | Security                             | D                              | III                              |                                  |
|    |                 | F4-SF04    | Management of roles                 | HD                             | III                              |                                  |
|    |                 | F4-SF05    | Support for multiple projects       | M                              | III                              |                                  |

*Figure 1. - Features used in the analysis [14]*
4.1.2. Feature Set 2
SESRA is based on guidelines for conducting a systematic literature review by Kitchenham. As for web applications, SESRA does not require any special installation. In order for the user to use the tool, it is necessary to register by entering data such as name, surname and email. The developers have made the tutorial available for SESRA users. However, the tutorial is only in Portuguese, and there is no English version. SESRA scored 11.5 out of 13 for this feature set.

4.1.3. Feature Set 3
SESRA makes it possible to identify the need for a systematic review of the literature. By filling out the form within the tool, the user has the opportunity to see if there are reviews of the literature in the area under study. A query made this way gets results from the SESRA public knowledge repository. After that, users are provided with a checklist based on which the user can see whether the currently existing systematic reviews of the literature in the area are sufficient and whether there is really a need for a review of the literature.

SESRA supports the development of research questions according to PICOC components. Adheres to the study selection stage, at which point studies should be examined and consequently included or excluded from further assessment. Although it is possible to undertake the selection process based on the title and abstract, researchers often opt to obtain a copy of the paper to assess the full text. If multiple researchers are involved, differences among their selection may occur. Hence, SESRA tools help in reaching an agreement through an advisor that reviews the selection, or by applying an inter-rater reliability test to evaluate a random sample of the selected studies. The tool does not have advanced techniques for automating SLR activities, except when it comes to graph extraction when writing reports. This tool scored 14.5 out of 23 for this feature set.

4.1.4. Feature Set 4
Within the tool, the user can create multiple projects. The disadvantage of SESRA is that it does not have the ability to store the entire text of the publication, and researchers can see the title, authors, abstract and keywords within the tool. In addition, there are different roles that can be assigned to users. For this feature set, SESRA scored 15 out of 17.

Table 1. Scores for SESRA

| ID | FEATURE SET       | ID | FEATURE SET SCORE | MAX SCORE | % FEATURE SET SCORE |
|----|-------------------|----|-------------------|-----------|---------------------|
| F1 | Economic          | F1SF01 | 3 | 3 | 6 | 50% |
|    |                   | F1SF02 | 0 | 0 | 0 | 0% |
|    |                   | F1SF03 | 0 | 0 | 0 | 0% |
| F2 | Ease of introduction and setup | F2SF01 | 4 | 10 | 13 | 77% |
|    |                   | F2SF02 | 3 | 3 | 3 | 100% |
| F3 | SRL activity support | F3SF01 | 2 | 14.5 | 23 | 63% |
|    |                   | F3SF02 | 2 | 2 | 2 | 100% |
|    |                   | F3SF03 | 1.5 | 1.5 | 1.5 | 100% |
|    |                   | F3SF04 | 3 | 3 | 3 | 100% |
|    |                   | F3SF05 | 3 | 3 | 3 | 100% |
|    |                   | F3SF06 | 1.5 | 1.5 | 1.5 | 100% |
|    |                   | F3SF07 | 0 | 0 | 0 | 0% |
|    |                   | F3SF08 | 0 | 0 | 0 | 0% |
|    |                   | F3SF09 | 0 | 0 | 0 | 0% |
|    |                   | F3SF10 | 0.5 | 0.5 | 0.5 | 50% |
| F4 | Process management | F4SF01 | 4 | 15 | 17 | 88% |
|    |                   | F4SF02 | 2 | 2 | 2 | 100% |
|    |                   | F4SF03 | 2 | 2 | 2 | 100% |
|    |                   | F4SF04 | 3 | 3 | 3 | 100% |
|    |                   | F4SF05 | 4 | 4 | 4 | 100% |
4.2. ReLis
Table 2 presents the scores for ReLis. Beside authors experience on using ReLis, the publication that presents [16] tool was also taken into consideration.

4.2.1. Feature Set 1
ReLis is entirely free and is maintained regularly. Users can register by filling an easy form. This tool scores maximum 6 points for this feature set.

4.2.2. Feature Set 2
It is easy to register users, as well as ReLis does not require any special system characteristics. There is a detailed tutorial on how to use ReLis, and the tool is self-contained. Regarding all this, ReLis scores 13 out of 13 for this feature set.

4.2.3. Feature Set 3
ReLis does not support automated searches, and it is one of its biggest disadvantages. Also, ReLis does not have report write up and report validation features, comparing to two other tools.

4.2.4. Feature Set 4
Within this tool, users can create multiple projects as well as to have multiple roles in multiple projects. Role management is possible in ReLis, and users can participate in projects in different ways. There is half-support when document management is considered.

| Table 2. Scores for ReLis |
|--------------------------|
| **ID** | **FEATURE SET** | **ID** | **WEIGHTED SCORE** | **FEATURE SET SCORE** | **MAX SCORE** | **% FEATURE SET SCORE** |
| F1 | Economic | F1SF01 | 3 | 6 | 6 | 100% |
| | | F1SF02 | 3 | | | |
| F2 | Ease of introduction and setup | F2SF01 | 4 | 13 | 13 | 100% |
| | | F2SF02 | 3 | | | |
| | | F2SF03 | 3 | | | |
| | | F2SF04 | 3 | | | |
| | | F3SF01 | 2 | 10 | 23 | 43% |
| | | F3SF02 | 2 | | | |
| | | F3SF03 | 0 | | | |
| | | F3SF04 | 3 | | | |
| | | F3SF05 | 0 | | | |
| | | F3SF06 | 1.5 | | | |
| | | F3SF07 | 0 | | | |
| | | F3SF08 | 0 | | | |
| | | F3SF09 | 1.5 | | | |
| | | F3SF10 | 0 | | | |
| | | F3SF11 | 0 | | | |
| F4 | Process management | F4SF01 | 4 | 15 | 17 | 88% |
| | | F4SF02 | 2 | | | |
| | | F4SF03 | 2 | | | |
| | | F4SF04 | 3 | | | |
| | | F4SF05 | 4 | 44 | 59 | 74.6% |

4.3. Parsifal
Table 3 presents the scores for Parsifal.

4.3.1. Feature Set 1
This tool is completely free and can be found in [17]. Parsifal provides the organisation of a systematic review of high-level literature. Parsifal is very similar to SESRA in the possibilities it offers to users.
It also works on the principle, filling out various forms and generating documents. Unlike the SESRA tool, Parsifal is in English and provides detailed monitoring of a systematic review of the literature. The tool is maintained regularly and the last changes, according to the evaluation, happened nine months ago.

4.3.2. Feature Set 2
Parsifal does not require any system requirements, and since it is a web application, installation, and setup consist of registering with email. There is a video tutorial on how to use this tool as well as Help and Blog pages on the website.

4.3.3. Feature Set 3
Parsifal is a tool that, through a series of steps, provides support to researchers in performing a systematic review of the literature. It enables researchers to easily plan, execute and document a systematic review of the literature through forms.

What is one of the biggest advantages of the Parsifal tool in relation to the SESRA tool, which enables almost the same functionalities, is that with the help of Parsifal it is possible to search the databases of Scopus and Science Direct publications directly from the tool. Finally, the tool allows reporting, but according to the evaluation performed for the purposes of this paper, the tool only allows the creation of a technical report and tables that would help researchers more easily classify publications in the review.

4.3.4. Feature Set 4
Parsifal offers the possibility to work on multiple projects. However, at the time that this study was taken, the invitation link for collaboration on the project led to “Error 404” page, and there was no way to add other people to project, even that tool offers to add collaborators. There is no possibility to manage roles.

| ID | FEATURE SET                     | ID   | WEIGHTED SCORE | FEATURE SET SCORE | MAX SCORE | % FEATURE SET SCORE |
|----|---------------------------------|------|----------------|-------------------|-----------|---------------------|
| F1 | Economic                        | F1SF01 | 3               | 4.5               | 6         | 100%                |
|    |                                 | F1SF02 | 1.5             |                   |           |                     |
| F2 | Ease of introduction and setup  | F2SF01 | 4               | 13                | 13        | 100%                |
|    |                                 | F2SF02 | 3               |                   |           |                     |
| F3 | SRL activity support            | F3SF01 | 2               | 15                | 23        | 65%                 |
|    |                                 | F3SF02 | 2               |                   |           |                     |
|    |                                 | F3SF03 | 3               |                   |           |                     |
|    |                                 | F3SF04 | 3               |                   |           |                     |
|    |                                 | F3SF05 | 3               |                   |           |                     |
|    |                                 | F3SF06 | 1.5             |                   |           |                     |
|    |                                 | F3SF07 | 0               |                   |           |                     |
|    |                                 | F3SF08 | 0               |                   |           |                     |
|    |                                 | F3SF09 | 0               |                   |           |                     |
|    |                                 | F3SF10 | 0.5             |                   |           |                     |
|    |                                 | F3SF11 | 0               |                   |           |                     |
| F4 | Process management              | F4SF01 | 4               | 8                 | 17        | 47%                 |
|    |                                 | F4SF02 | 2               |                   |           |                     |
|    |                                 | F4SF03 | 2               |                   |           |                     |
|    |                                 | F4SF04 | 3               |                   |           |                     |
|    |                                 | F4SF05 | 4               |                   |           |                     |

Table 3. Scores for Parsifal
Table 4. Feature Set Scores

| TOOL  | F1  | F2  | F3  | F4  | TOTAL |
|-------|-----|-----|-----|-----|-------|
| SESRA | 50% | 77% | 63% | 88% | 72%   |
| RELIS | 100%| 100%| 43% | 88% | 74.6% |
| PARSIFAL | 100%| 100%| 65% | 47% | 68%   |

4.4. Discussion
This section presents the discussion of the results, as well as the limitations of the study. ReLis achieved the best results in this analysis, i.e. 74.6% of the total score. As the differences in the percentages of other tools are not overly different, what ReLis pointed out as better than other tools is the regular maintenance of the tools and the management of roles within the tools. The tools show similar results, indicating that the authors chose tools with relatively similar capabilities and options for researchers. Comparing ReLis to two other tools that are evaluated in this study, its one of the biggest disadvantages is in feature set 3. ReLis does not support automated search, while Parsifal and SESRA do support it. Researchers that conduct SLR pointed out this feature as important [18].

The disadvantages of all tools are reflected when looking at the F3 feature set, which is the most important part of the evaluation. All tools support SLRs in organising reviews, but none of these tools supports researchers in increasing the quality of literature reviews or automating certain processes.

5. Conclusion
The paper presents an evaluation of three identified tools to support the entire SLR process. The methodology used within the paper is the DESMET, i.e. Feature analysis approach, based on the evaluation performed by Marshall in 2014.

There are a sufficient number of tools that support the whole SLR process. However, they do not support it by automating, but by managing the organisation of the process.

In the further development of the tool, developers should focus on automation based on machine learning, extracting knowledge from text and similar, to help reduce time and increase the quality of the SLR. To be able to fully conclude if these tools are useful to researchers, more researchers should use their tools for a longer period of time and conduct a qualitative analysis of their experience. The question that arises for further research is which tool combination gives the best results.

References
[1] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, “Systematic literature reviews in software engineering - A systematic literature review,” Inf. Softw. Technol., vol. 51, no. 1, pp. 7–15, 2009.
[2] B. Kitchenham and S. Charter, Guidelines for performing Systematic Literature Reviews in Software Engineering. 2007.
[3] M. J. Grant and A. Booth, “A typology of reviews: An analysis of 14 review types and associated methodologies,” Health Info. Libr. J., vol. 26, no. 2, pp. 91–108, 2009.
[4] 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,” J. Syst. Softw., vol. 80, no. 4, pp. 571–583, 2007.
[5] M. A. Babar and H. Zhang, “Systematic literature reviews in software engineering: Preliminary results from interviews with researchers,” 2009 3rd Int. Symp. Empir. Softw. Eng. Meas. ESEM 2009, pp. 346–355, 2009.
[6] M. Riaz, M. Sulayman, N. Salleh, and E. Mendes, “Experiences Conducting Systematic Reviews from Novices’ Perspective,” no. May 2014, 2010.
[7] C. Marshall, P. Brereton, and B. Kitchenham, “Tools to support systematic reviews in software
9

engineering: A feature analysis," ACM Int. Conf. Proceeding Ser., no. May, 2014.

[8] B. Kitchenham, S. Linkman, and D. Law, “A methodology for evaluating software engineering methods and tools,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 706 LNCS, no. June, pp. 121–124, 1993.

[9] D. Bowes, T. Hall, and S. Beecham, “SLuRp - A tool to help large complex systematic literature reviews deliver valid and rigorous results,” EAST’12 - Proc. 2nd Int. Work. Evidential Assess. Softw. Technol., pp. 33–36, 2012.

[10] E. Hernandes, A. Zamboni, S. Fabbri, and A. Di Thommazo, “Using GQM and TAM to evaluate StArt – a tool that supports Systematic Review,” CLEI Electron. J., vol. 15, no. 1, 2012.

[11] A. M. Fernández-Sáez, M. G. Bocco, and F. P. Romero, “SLR-Tool a tool for performing systematic literature reviews,” ICSOFT 2010 - Proc. 5th Int. Conf. Softw. Data Technol., vol. 2, no. January, pp. 157–166, 2010.

[12] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, “Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement,” Int. J. Surg., vol. 8, no. 5, pp. 336–341, 2010.

[13] F. Arlene, Conducting research literature reviews : from the internet to the paper. Thousand Oaks, CA, USA: Sage Publications, 2005.

[14] C. Marshall, P. Brereton, and B. Kitchenham, “Tools to support systematic reviews in software engineering: A feature analysis,” ACM Int. Conf. Proceeding Ser., no. September, 2014.

[15] B. A. Kitchenham and S. G. Linkman, DESMET : A method for evaluating Software Engineering methods and tools. 2000.

[16] B. Bigendako and E. Syriani, “Modeling a tool for conducting systematic reviews iteratively,” Model. 2018 - Proc. 6th Int. Conf. Model. Eng. Softw. Dev., vol. 2018-Janua, no. Modelsward, pp. 552–559, 2018.

[17] “Parsifal Tool.” [Online]. Available: https://parsif.al/.

[18] J. C. Carver, E. Hassler, E. Hernandes, and N. A. Kraft, “Identifying barriers to the systematic literature review process,” Int. Symp. Empir. Softw. Eng. Meas., pp. 203–213, 2013.