SYSTEMATIC LITERATURE REVIEW (SLR) AUTOMATION: A SYSTEMATIC LITERATURE REVIEW

1 ZUHAL HAMAD, 2NAOMIE SAIIM

1PhD Student., SUDAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
2Assoc. Prof., UNIVERSITY TECHNOLOGY MALAYSIA

E-mail: zuhalahmed@yahoo.com, naomie@utm.my

ABSTRACT

Context: A systematic literature review (SLR) is a methodology used to find and aggregate all relevant studies about a specific research question or topic of interest. Most of the SLR processes are manually conducted. Automating these processes can reduce the workload and time consumed by human.

Method: we use SLR as a methodology to survey the literature about the technologies used to automate SLR processes.

Result: from the collected data we found many work done to automate the study selection process but there is no evidence about automation of the planning and reporting process. Most of the authors use machine learning classifiers to automate the study selection process. From our survey, there are processes that are similar to the SLR process for which there are automatic techniques to perform them.

Conclusion: Because of these results, we concluded that there should be more research done on the planning, reporting, data extraction and synthesizing processes of SLR.

Keywords: SLR, Automation, Planning, Reporting, Data Extraction, Synthesizing

1. INTRODUCTION

A systematic literature review or a systematic review is a means of identifying, evaluating and interpreting all available research relevant to a particular research question, or topic area, or phenomenon of interest.[2]

The systematic literature review methodology has a well-defined methodological steps or protocol. The methodological steps, search strategy and research question are explicitly defined so that other researchers can reproduce the same protocol.[2]

There are many reasons for undertaking a systematic review. The most common reasons are: to summarize the existing evidence concerning a treatment or technology, to identify any gaps in current research in order to suggest areas for further investigation and to provide a framework/background in order to appropriately position new research activities.[3].

As described in Figure1, a systematic literature review (SLR) consists of several activities. These activities can be grouped into three phases, as follows:

- Planning the review
- Conducting the review
- Reporting the review

Systematic reviews require considerably more effort than traditional reviews, and currently, most of its activities are done manually. Automating the SLR process will reduce most if not all of the human effort and time consumed to conduct it.

The aim of our SLR is to see if there are any techniques, or methods or approaches in the literature that are used or can be used for SLR process automation in any of its phases and how effective they are.
2. **METHOD**

**Research question**
An approach used to formulate research questions is to use PICOC criteria. Using this approach the research question structured in: 1. population. 2. Intervention. 3. Comparison 4. Outcomes. 5. Context

The attributes of our research question are shown in table1.

<table>
<thead>
<tr>
<th>Population</th>
<th>Studies about SLR automation or any of its processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>All possible techniques</td>
</tr>
<tr>
<td>Comparison</td>
<td>None</td>
</tr>
<tr>
<td>Outcome</td>
<td>Techniques that support to conduct SLR and to which SLR stage it is or can be applied.</td>
</tr>
<tr>
<td>Context</td>
<td>None</td>
</tr>
</tbody>
</table>

Figure1: Systematic literature review process[3].
The research questions addressed by this study are as follows:

RQ1: what are the techniques that support SLR processes and how good they are?
  - RQ1.1: what are the SLR processes that have been done automatically?
  - RQ1.2: what are the techniques that support each process?
  - RQ1.2: how effective are they?

Here we want to know what are the processes of SLR that have been supported by computer and what are the techniques that support the different processes of the SLR and how they are effective.

RQ2: Is there any similar process to SLR in the literature? How it is supported by computer?
  - RQ2.1: what are the processes that similar to each SLR process?
  - RQ2.2: how it is supported by computer?

Here we want to see what are the processes that are similar to each SLR stage and how it can be done automatically or what are the techniques used for these processes.

Search strategy
The strategy used for searching is automatic search

Search Strings
For the search string we take terms from research questions, alternative terms and synonyms and join the string using AND, OR connectors.

Strings for RQ1:
(strategy/technique/approach/supporting/searching/relevant categorization/classification/screening/Reduce workload/Data/knowledge/sentence/results/information extraction/collection/presentation/summarization) AND SLR OR
(systematic reviews OR systematic literature review OR meta-analysis OR scoping review OR evidence based OR Mapping studies OR systematic mapping OR scoping review).

Strings for RQ2:
1- (searching OR grouping OR clustering)
   AND (relevant articles OR papers OR similar articles OR papers)
2- (knowledge OR sentence OR information OR data) AND (extraction OR discovery OR mining)
3- Documents AND (classification OR categorization OR summarization OR clustering) AND (methods OR technique OR approach)

Data source
Databases to be searched for the primary studies are:
1- IEEE
2- ACM digital library
3- Science direct- Elsevier
4- Scopus – Elsevier
5- Wiley online library
6- Google scholar

Inclusion and exclusion criteria
Included studies
1- Journal and conference papers.
2- Publications written in English language.
3- That propose/implement/suggest methods/techniques to automate SLR complete process or automate any of SLR stages or similar process or sentence/knowledge/data extraction or documents classification/ categorization/ prioritization/ summarization
4- survey study about automatic SLR generation or any of it is stages.

Excluded studies
1- That describe theoretical aspects of SLR
2- Guidelines for doing SLR
3- SLR about other issues (not about SLR automation).
4- Studies that using manual techniques.

These criteria will be applied to the title, keywords, abstract and conclusion. This protocol will be reviewed by our supervisor.

Quality assessment
Table 2: Quality Assessment Criteria

<table>
<thead>
<tr>
<th>no</th>
<th>question</th>
<th>answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Is the technique or method used clearly stated?</td>
<td>Yes/no/partial</td>
</tr>
<tr>
<td>2</td>
<td>Does the article address one of the research questions?</td>
<td>Yes/no/partial</td>
</tr>
<tr>
<td>3</td>
<td>Does the article document the procedure used to validate its technique or method used?</td>
<td>Yes/no/partial</td>
</tr>
<tr>
<td>4</td>
<td>Is it not a duplicate paper?</td>
<td>Yes/no/partial</td>
</tr>
</tbody>
</table>

Search process

The following table contains the results of the search on the specified databases using the search strings for RQ1 and RQ2, our search start 3/5/2013 and end 11/5/2013.

Table 3: Search Process Preliminary Results

<table>
<thead>
<tr>
<th></th>
<th>Search results</th>
<th>Inclusion by title</th>
<th>Removing duplicates</th>
<th>Inclusion by abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1</td>
<td>2922</td>
<td>251</td>
<td>211</td>
<td>50</td>
</tr>
<tr>
<td>RQ2</td>
<td>1183</td>
<td>192</td>
<td>190</td>
<td>66</td>
</tr>
</tbody>
</table>

Studies selection process

During this process we apply the inclusion and exclusion criteria to the full study, starting with 116 studies plus 3 from one of the included studies reference.

The initial screening end with 40 relevant studies. And a review process by the author started with the same population along with the quality assessment criteria. A weight assigned to each study according to this rule (yes=1, partial= 0.5 and no=0), including papers with the weight(2 to 4) only.

Data extraction process

Data extraction process was carried out on 26 papers that passes the inclusion/exclusion check and the quality check, the data extracted after reading the full paper. Table 1 and Table 2 in appendix B summarize the data extraction process.

Data Synthesis process

For RQ1.1 synthesized data from all studies show that the study selection (initial screening and reviewing or validation of the selection process), data extraction and synthesizing have an automation support. The collected data show that the process which have more automatic support is the study selection process. It is very important to notes that there is no automatic support for the planning and the reporting phases of the SLR process.

For RQ1.2 the collected data show that For the study selection process the techniques used for the documents classification are the machine learning classifiers listed below:

1) Complement Naïve Bayes
2) Discriminative Multinomial Naïve Bayes
3) Alternating Decision Tree
4) AdaBoost (Logistic Regression)
5) AdaBoost (j48)
6) Support vector machine learning algorithm
7) A voting perceptron-based

In one paper graph representation is used as a technique to support the data extraction, for the search process a meta search is used in one paper and in another one the text mining is used to improve the search strategy by using an associative search and lastly a sentence extraction for multi documents summarization is used to support the data synthesis process.

For RQ1.3 the collected data show that the reduction of the human workload between 20%-50%, the papers about study selection reporting that no loss of relevant data and no inclusion of irrelevant one.

For RQ2.1(what are the processes that similar to each SLR process?), there are processes that are similar to the study selection process: filtering spam emails, news articles classification and data loss prevention. For Data Synthesis, similar process is Research paper recommender system and for reporting the review the similar processes are summarization of multiple news documents and summarization of dissertation abstracts. But no
similar process for planning the review process and data extraction process. For RQ2.2 (how it is supported by computer?), from the collected data the processes that are similar to the study selection process supported by using a machine learning classifiers, the data synthesizing is supported by using SCuBA algorithm, and for reporting the review process sparse predictive classification framework is used in addition to the hierarchical variable-based framework.

3-Discussion of the results
From the obtained results there are processes in SLR (planning and reporting the review) that have not been supported by computer and more research on these processes needs to be undertaken. From our survey there are similar processes to SLR processes and it is supported by computer, because of its effectiveness the technologies used can be applied to the SLR processes that have no computer support specially reporting the review process because it is the summary of the overall process.

4-Conclusion
In this paper we present a result of a systematic literature review aimed to investigate the use of computer to support systematic literature review processes, to identify the systematic literature processes that support by computer. The SLR study give us an identification of the current state of research and techniques to support research gaps and future work.

REFERENCES:


APPENDIX A

the following are references for the included studies that referenced by S

REFERENCES


Appendix B

Form 1: to answer research question 1 (RQ1): what are the techniques that support SLR processes and how good they are?

<table>
<thead>
<tr>
<th>Study Id</th>
<th>Author(s)</th>
<th>SLR process</th>
<th>Date</th>
<th>Method/technique</th>
<th>Performance measurements</th>
<th>Effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Cohen, A., Hersh, W., Peterson, K., &amp; Yen, P.</td>
<td>Study selection</td>
<td>2006</td>
<td>Machine learning based classifier: A voting perceptron-based automated citation classification</td>
<td>Recall, precision and F-measure</td>
<td>Reduction in the number of articles needing manual review (3 for each 15(20%))</td>
</tr>
<tr>
<td>S2</td>
<td>Tomassetti, F., Rizzo, G., Vetro, A., &amp; Ardito, L.</td>
<td>Study selection</td>
<td>2011</td>
<td>Extending technologies in the field of the linked data and text mining (Naive Bayes classifier)</td>
<td>Recall</td>
<td>Improving the second step in SLR by filtering the possible studies and automatically discarding non-relevant ones</td>
</tr>
<tr>
<td>S3</td>
<td>Felizardo, K. R., Andery, G. F., Paulovich, F. V., Minghim, R., &amp; Maldonado, J. C.</td>
<td>Study selection (review or validation)</td>
<td>2012</td>
<td>Visual text mining (VTM): 1- Content map 2- Citation map</td>
<td></td>
<td>The results have shown that employment of VTM techniques can successfully assist in the Selection Review task, speeding up the entire SLR process in comparison to the conventional approach.</td>
</tr>
<tr>
<td>S4</td>
<td>Bekhuis, T., Demner-Fushman, D</td>
<td>Study selection (the initial screening phase)</td>
<td>2010</td>
<td>Supervised machine learning Three types of classifiers: 1. decision trees. 2. EvoSVM 3. weighted one-dependence estimator (WAODE)</td>
<td>Mean recall, mean precision and harmonic mean of equally-weighted precision and recall (F1);</td>
<td>EvoSVM with a radial or Epanechnikov kernel may be an appropriate classifier when observational studies are eligible for inclusion in a systematic review.</td>
</tr>
<tr>
<td>S5</td>
<td>Wallace, B. C., Trikalinos, T. a, Lau, J., Brodley, C., Schmid, C. H.</td>
<td>Study selection (citation screening)</td>
<td>2010</td>
<td>Machine learning – support vector machine (SVM) Active learning strategy</td>
<td></td>
<td>The algorithm is able to reduce the number of citations that must be screened manually by nearly half in two of these, and by around 40% in the third, without excluding any of the citations eligible for the systematic review.</td>
</tr>
<tr>
<td>Page</td>
<td>Author(s)</td>
<td>Year</td>
<td>Methods</td>
<td>Results/Conclusions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>------</td>
<td>---------</td>
<td>---------------------</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| S6   | Ananiadou, S., Rea, B. | 2009 | -Text mining improves the search strategy by using an associative search which discovers the set of documents most similar to a given document.
-Document classification using support vector machine (SVM)
-Adaptable multi-document summarization | micro-average F1-measure and the multi-topic accuracy |
| S7   | Cohen, A. M. | 2008 | -machine learning techniques
-documents classifications (classification including feature systems unigram, n-gram, MeSH, and natural language processing (NLP) feature) | “AUC” using the area under the receiver operating curve as a measure of goodness. The best feature set used a combination of n-gram and MeSH features. NLP-based features were not found to improve performance. |
<p>| S8   | Cohen, A. M., Adams, C. E., Davis, J. M., Yu, C., Yu, P. S., Meng, W., Duggan, L., et al. | 2010 | -meta-search classifier (SVM based) clustering ranking | time and effort measurements (comparing the text mining-based pipeline for accelerating systematic reviews in evidence-based medicine will decrease the manual burden of systematic reviewers during the literature collection and review process, and increase the proportion of reviewer time spent synthesizing evidence, performing meta-analyses, and considering results. |
| S9   | Cohen, A. M., Ambert, K., &amp; McDonagh, M. (n.d.) | 2009 | -support vector machine learning algorithm was evaluated with cross-validation | “AUC” using the area under the receiver operating curve as a measure of goodness. On average, the method improves performance by about 20%, when the amount of topic-specific training data are scarce. |</p>
<table>
<thead>
<tr>
<th>Study</th>
<th>Authors</th>
<th>Date</th>
<th>Type of Study</th>
<th>Selection Method</th>
<th>Selection Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>S10</td>
<td>Felizardo, K. R., Nakagawa, E. Y., Feitosa, D., Minghim, R., Mapping, S., &amp; Mining, V. T.</td>
<td>2009</td>
<td>study selection</td>
<td>Visual text mining (VTM)</td>
<td>Effort reduction to conduct systematic mapping can be achieved, since the approach is automated using a supporting tool.</td>
</tr>
<tr>
<td>S12</td>
<td>Frunza, O., Inkpen, D., Matwin, S.</td>
<td>2010</td>
<td>Study selection</td>
<td>machine learning technique- CNB (Complement Naive Bayes) classifier</td>
<td>Recall, precision</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Our goal of improving the recall level from the first level of screening is achieved, since when both the classifier and the human judge are integrated in the workflow, the recall level jumps from 79.7% to 92.7%.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1-The experiments demonstrate that a committee of machine learning classifiers can rank biomedical research abstracts with a confidence level similar to human experts. 2-The ranking approach allows selecting abstracts that are classified as relevant or non-relevant with high level of prediction confidence 3-We tried our approach on data used in a real case systematic review. The papers selected with our ranking method are classified by the machine learning technique with a</td>
</tr>
<tr>
<td>Study</td>
<td>Authors</td>
<td>Year</td>
<td>Methodology</td>
<td>Task</td>
<td>Results</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>------</td>
<td>-------------</td>
<td>------</td>
<td>---------</td>
</tr>
<tr>
<td>S14</td>
<td>Malheiros, V., Hohn, E., Pinho, R., Mendonca, M., Maldonado, J.C.</td>
<td>2007</td>
<td>Visual text mining (VTM)</td>
<td>Study selection</td>
<td>Precision 84.3% for the class of interest. Recall 91.6%</td>
</tr>
<tr>
<td>S15</td>
<td>Matwin, S., Kouznetsov, A., Inkpen, D., Frunza, O., O’Blenis, P.</td>
<td>2010</td>
<td>Factorized version of the complement naive Bayes (FCNB) classifier</td>
<td>Study selection</td>
<td>The minimum workload reduction for a systematic review for one topic, achieved with a FCNB/WE classifier, was 8.5%; the maximum was 62.2% and the average over the 15 topics was 33.5%. This is 15.0% higher than the average workload reduction obtained using a voting perceptron-based automated citation classification system.</td>
</tr>
<tr>
<td>S16</td>
<td>Rizzo, G., Tomassetti, F., Ardito, L., Torchiano, M., &amp; Morisio, M.</td>
<td>2012</td>
<td>An automated pre-selection approach based on text mining and semantic enrichment techniques</td>
<td>Study selection</td>
<td>Results show a reduction of the manual workload of 18% that a human researcher has to spend. As baseline, we compared the enriched approach with one based on a normal Multinomial Naive Bayes classifier. The improvements range from 2.5% to 5% depending on the dimension of the trained model.</td>
</tr>
<tr>
<td>S11</td>
<td>Felizardo, K. R., Riaz,</td>
<td>2011</td>
<td>Graph representation</td>
<td>Data extraction</td>
<td>- Graphs were more efficiently...</td>
</tr>
<tr>
<td>M., Sulayman, M., Mendes, E., MacDonell, S. G., &amp; Maldonado, J. C.</td>
<td>understood - there is reduction in time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Form2: to answer research question2 (RQ2): Is there any similar process to SLR in the literature? How it is supported by computer?

<table>
<thead>
<tr>
<th>Study Id</th>
<th>Author(S)</th>
<th>date</th>
<th>SLR process</th>
<th>Similar process</th>
<th>Techniques/methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>S24</td>
<td>Ou, S., Khoo, C. S. G., &amp; Goh, D. H.</td>
<td>2005</td>
<td>Reporting the review</td>
<td>summarization of dissertation abstracts</td>
<td>hierarchical variable-based framework to integrate four kinds of information—research concepts, relationships between variables, contextual relations, and research methods extracted from different documents, and gives the user a map or overview of a specific topic which the user can explore and zoom in for more details.</td>
</tr>
<tr>
<td>S19</td>
<td>Androutsopoulos, I., Koutsias, J., Chandrinos, K. V., &amp; Spyropoulos, C. D.</td>
<td>2000</td>
<td>Study selection</td>
<td>Anti-Spam Filtering</td>
<td>Naive Bayesian classifier</td>
</tr>
<tr>
<td>S20</td>
<td>Hart, M., Manadhata, P. K., Johnson, R., &amp; Manadhata, P.</td>
<td>2011</td>
<td>Study selection</td>
<td>Data loss prevention</td>
<td>Support vector machine(SVM)</td>
</tr>
<tr>
<td>S22</td>
<td>Youn, S., &amp; Mcleod, D.</td>
<td>2007</td>
<td>Study selection</td>
<td>Spam email classification</td>
<td>Adaptive ontology-J48</td>
</tr>
<tr>
<td>S23</td>
<td>Ramdass, D., &amp; Seshasai, S.</td>
<td>2009</td>
<td>Study selection</td>
<td>Newspaper Articles Classification</td>
<td>Naive Bayesian classifier, Maximum Entropy Classification and Probabilistic Grammar Classification</td>
</tr>
<tr>
<td>S25</td>
<td>Diao, Y., Lu, H., &amp; Wu, D.</td>
<td>2000</td>
<td>Study selection</td>
<td>personal e-mail filtering</td>
<td>naive Bayesian classifier and decision tree based classifier was</td>
</tr>
<tr>
<td>S26</td>
<td>El-Halees, A.</td>
<td>2009</td>
<td>Study selection</td>
<td>Filtering Spam E-Mail</td>
<td>maximum entropy, decision trees, artificial neural nets, naive Bayesian , support vector machines and k-nearest neighbor.</td>
</tr>
</tbody>
</table>