SUPPLEMENTARY MATERIAL FOR:

AUTOMATING SYSTEMATIC LITERATURE REVIEWS WITH NATURAL LANGUAGE PROCESSING AND TEXT MINING: A SYSTEMATIC LITERATURE REVIEW

Girish Sundaram and Daniel Berleant, Department of Information Science, University of Arkansas at Little Rock

Introduction

Research in this area is growing continuously and as a result the number of papers being published in academic databases is growing exponentially. SLRs (both with or without meta-analysis) are being used to take informed decisions in many areas of healthcare such as treating a particular disease for a patient and broader levels such as taking a policy decision that is applicable to all sections of the society. Systematic reviews can be relevant to policy, clarifying the attendant problems, impacts and assumptions (Oliver and Dickson 2016 p. 235)[48].

Artificial Intelligence (AI) has been used in many domains to get a deeper understanding of the data in hand. Natural Language Processing (NLP) is a sub-area of AI that falls in the intersection of linguistics and computer science and is focused on human computer interaction and in particular the means to make sense of large volumes of unstructured natural language data. Most of the information that is being used by NLP applications today is unstructured text data. The primary goal of many applications is to analyze the data in a way that is close to humans and uses nuanced context based understanding of the information, using techniques that mimic human interaction. Other developing fields within the NLP arena are speech recognition, natural language understanding and natural language generation. NLP has its origins in the 1950’s when the famous scientist Turing published an article titled “Computing Machinery and Intelligence” where he proposes a variation on the “Imitation Game” where the participants are asked to evaluate if the other player is a computer or a human.[49]

Quality Criteria

Table A1 shows the quality criteria used in assessing the quality of the sources that were screened for review.

Table A1. Quality Assessment Criteria

No.	Criterion
	Problem Statement
Q1	Is the research objective sufficiently explained and well-motivated?
	Research Design
Q2	Is it clear which TM technique(s) can be used to support the SLR process?
Q3	Is it clear which SLR activities can be supported using the TM techniques or automation methodologies?
	Data Collection
Q4	Are the data collection and measures adequately described?

Q5	Are the measures and constructs used in the study the most relevant for answering the research question/issue?
	Data Analysis
Q6	Is the data analysis used in the study adequately described?
Q7 A	Qualitative study: Is the interpretation of evidence clearly described?
Q7 B	Quantitative study: Has the significance of the data been assessed?
Q8	Is it clear how the TM technique(s) or supporting tool(s) have been used?
	Conclusion
Q9	Are the findings of the study clearly stated and supported by the results?
Q10	Does the paper discuss the limitations or validity?
	Type of Study
Q11	Is this study a systematic literature review?

TM Methods

Table A2 shows the categories of text mining methods that were referred to.

Table A2. Categories of TM Methods (Adapted From Feng et al.[14])

TM Category	Description
Information Extraction (IE)	Finding a specific piece of information from a text document using a pattern-matching method to find key phrases and relationships in the text.
Information Retrieval (IR)	Investigation of appropriate mechanisms for searching relevant information from a collection of resources.
Information Visualization (IVi)	Put information in graphical form to support human understanding.
Classification (Categorization)	Finding interesting patterns/features that help define a grouping and assigning documents to known categories.
Clustering	Finding interesting traits associated with extracted data and grouping similar documents based on their content.
Summarization	Reducing the length and detail of the source text into a shorter version while preserving the gist of its Information.

Data Extraction Template

Table A3 shows the data extraction template used to collect the necessary information from the primary list of literature sources. The classification of TM methods is adapted from Feng et al.[14].

Table A3. Data Extraction Form Template

ID	Extraction Element	Possible Values	Notes

1	Title		
2	Passed inclusion criteria?	Y/N	
3	Year of publication		
4	Authors		
5	DOI		
6	Database for extraction (source repository)		
7	URL		
8	Document type	Journal article	
		Conference paper	
		Thesis	
		Working paper or in press	
		Article in periodical	
9	SLR steps automated	SLR1	Commissioning a review
		SLR2	Specifying the research questions
		SLR3	Developing a review protocol
		SLR4	Evaluating the review protocol
		SLR5	Identification of research
		SLR6	Selection of primary studies
		SLR7	Study quality assessment
		SLR8	Data extraction and monitoring
		SLR9	Data Synthesis
		SLR10	Specifying dissemination mechanisms
		SLR11	Formatting the main report
		SLR12	Evaluating the report
10	Level of automation	Complete/Partial	
11	Type of review	New/Update	
12	TM methods used (category)	Information Extraction (IE)	
		Information Retrieval (IR)	
		Information Visualization (IVi)	
		Classification	

		(Categorization)	
		Clustering	
		Summarization	
13	TM model/algorithm information		
14	TM model evaluation methodology used (if specified)	Cross-validation (specify type)	
15	Refer to additional details tab for more information	Hold-out sampling	
		Leave-One-Out	
		Bootstrap Sampling	
		Other	
		Unclear	
16	Evaluation metrics used	Recall	
		Precision	
		F-Measure (specify weighting)	
		ROC (AUC)	
		Accuracy	
		Coverage - indicates the ratio of positive instances in the data pool that are annotated during active learning.	
		Burden	
		Yield	
		Cost	
		Utility	
		Work saved (incl. WSS)	
		RMSE	
		Performance/efficiency	
		Time	
		True positives	
		False negatives	
		Specificity = TN/(TN+FP)	
		Baseline inclusion rate	
		Other	
		None?	
17	TM methods used as additional reviewer	Y/N	
18	Deep learning or AI used?	Y/N	
19	Sampling techniques used		
20	Overall results/conclusions (stated by authors)		

21	Performance gain over manual methods provided	Y/N	

List of Studies

Table A4 below lists the studies that were in the final list of works to analyze.

Table A4. Final List of Primary Studies

ID	Title (abbreviated)	SLR Step(s)	TM Methods	Algorithm(s)
[45]	Semi-automated screening of biomedical citations for systematic reviews	SLR6-Selection of primary studies	Classification (Categorization)	SVM
[46]	Text mining to support abstract screening for knowledge syntheses: a semi-automated workflow	SLR6-Selection of primary studies	Classification (Categorization)	LDA, Random forest
[50]	Supporting systematic reviews using lda-based document representations	SLR6-Selection of primary studies	Classification (Categorization)	SVM, LDA, BOW
[51]	Studying the potential impact of automated document classification on scheduling a systematic review update	SLR6-Selection of primary studies	Classification (Categorization)	SVM
[52]	Statistical stopping criteria for automated screening in systematic reviews	SLR6-Selection of primary studies	Classification (Categorization)	SVM
[53]	Reducing systematic review workload through certainty-based screening	SLR6-Selection of primary studies	Classification (Categorization)	SVM, Logistic regression, LDA, BOW
[54]	A novel framework to expedite systematic reviews by automatically building information extraction training corpora	SLR8-Data extraction and monitoring	Classification (Categorization)	SVM
[55]	Automatic text classification to support systematic reviews in medicine	SLR6-Selection of primary studies	Classification (Categorization)	Naïve Bayes, K-nearest neighbours (KNN), SVM, Rocchio
[56]	A machine learning approach for semi-automated search and selection in literature studies	SLR6-Selection of primary studies SLR5-Identification of research	Classification (Categorization)	SVM, Logistic regression, Decision trees
[57]	Building systematic reviews using automatic text classification techniques	SLR6-Selection of primary studies	Classification (Categorization)	Complement naïve Bayes (CNB), Multinomial naïve Bayes (MNB)
[58]	Advanced analytics for the automation of medical systematic reviews	SLR6-Selection of primary studies	Classification (Categorization)	Soft-margin based SVM

[59]		SLR6-Selection of primary studies	Classification (Categorization)	Soft-margin based SVM
[60]	Toxic effects of nanomaterials for health applications: How automation can support a systematic review of the literature	SLR6-Selection of primary studies SLR7-Study quality assessment SLR8-Data extraction and monitoring SLR9-Data Synthesis SLR11-Formatting the main report	Information Extraction (IE) Information Retrieval(IR) Classification (Categorization) Clustering	Various
[61]	The use of bibliography enriched features for automatic citation screening	SLR6-Selection of primary studies	Classification (Categorization)	SVM
[62]	Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error	SLR6-Selection of primary studies	Classification (Categorization)	SVMs, logistic regression, Random forests
[12]	Automating data extraction in systematic reviews: a systematic review	SLR8-Data extraction and monitoring	Information Extraction (IE)	SVM, Random forest, Naïve Bayes (NB), Multi-layer perceptron (MLP)
[63]	Extractive text summarization system to aid data extraction from full text in systematic review development	SLR8-Data extraction and monitoring	Information Extraction (IE) Summarization	SVM, Regression classifier, Sequential minimal optimization
[64]	Automated screening of research studies for systematic reviews using study characteristics	SLR6-Selection of primary studies	Classification (Categorization)	Unclear

[47]	Measuring the impact of screening automation on meta-analyses of diagnostic test accuracy	SLR6-Selection of primary studies	Classification (Categorization)	Logistic regression, Neural network
[65]	Systematic review automation methods	SLR6-Selection of primary studies SLR8-Data extraction and monitoring	Classification (Categorization)	Logistic regression, Others
[66]	Automating document discovery in the systematic review process: how to use chaff to extract wheat	SLR6-Selection of primary studies	Classification (Categorization)	Logistic regression
[67]	Evaluation of a rule-based method for epidemiological document classification towards the automation of systematic reviews	SLR6-Selection of primary studies	Clustering Classification (Categorization)	GATE
[68]	Extracting PICO sentences from clinical trial reports using supervised distant supervision	SLR8-Data extraction and monitoring	Information Extraction (IE)	Logistic regression
[69]	Automating risk of bias assessment for clinical trials	SLR7-Study quality assessment	Classification (Categorization)	SVM
[70]	Text classification on imbalanced data: application to systematic reviews automation	SLR6-Selection of primary studies	Classification (Categorization)	Naïve Bayes, Active decorate, SVM
[71]	Automating reviews using natural language processing-based extraction	SLR8-Data extraction and monitoring	Information Extraction (IE)	BioBERT-based NLP model
[40]	Automation of systematic literature reviews: a systematic literature review	Various	Various	Various
[72]	Automatic boolean query refinement for systematic review literature search	SLR6-Selection of primary studies SLR5-Identification of research	Information Extraction (IE) Classification (Categorization) Information Retrieval(IR)	K-nearest neighbour

[73]	Learning to identify relevant studies for systematic reviews using random forest and external information	SLR6-Selection of primary studies	Classification (Categorization)	Random forest

References

1-47. These references appear in the main article.
48. Dickson, K.: Systematic reviews to inform policy: institutional mechanisms and social interactions to support their production. Dissertation. University College London (2017). Http://discovery.ucl.ac.uk/id/eprint/10054092/1/KD_PhD_FinalAugust2018_Redacted.pdf.
49. Turing, A.: Computing machinery and intelligence. Mind LIX(236), 433–460 (1950). DOI:10.1093/mind/LIX.236.433.
50. Mo, Y., Kontonatsios, G., Ananiadou, S.: Supporting systematic reviews using LDA-based document representations. Syst Rev 4, 172 (2015). DOI: 10.1186/s13643-015-0117-0.
51. Cohen, A. M., Ambert, K., McDonagh, M.: Studying the potential impact of automated document classification on scheduling a systematic review update. BMC Med Inform Decis Mak 12, 33 (2012). DOI: 10.1186/1472-6947-12-33.
52. Callaghan, M. W., Müller-Hansen, F.: Statistical stopping criteria for automated screening in systematic reviews. Syst Rev 9, 273 (2020). DOI: 10.1186/s13643-020-01521-4.
53. Miwa, M., Thomas, J., O'Mara-Eves, A., et al.: Reducing systematic review workload through certainty-based screening. J Biomed Inf 51, 242–253 (2014). DOI: 10.1016/j.jbi.2014.06.005.
54. Basu, T., Kumar, S., Kalyan, A., et al.: A novel framework to expedite systematic reviews by automatically building information extraction training corpora. arXiv:1606.06424 [cs.IR] (2016). Https://arxiv.org/abs/1606.06424.
55. García Adeva, J. J., Pikatza Atxa, J. M., Ubeda Carrillo M., et al.: Automatic text classification to support systematic reviews in medicine. Expert Syst with Applications 41(4), 1498–1508 (2014). DOI: 10.1016/j.eswa.2013.08.047.
56. Ros, R., Bjarnason, E., Runeson, P.: A machine learning approach for semi-automated search and selection in literature studies. In: EASE '17: Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering, pp. 118–127. Association for Computing Machinery, New York (2017). DOI: 10.1145/3084226.3084243.
57. Frunza, O., Inkpen, D., Matwin, S.: Building systematic reviews using automatic text classification techniques. In: Proceedings of the 23rd International Conference on Computational Linguistics: Poster Vol. (COLING '10), pp. 303–311. Association for Computational Linguistics (2010).
58. Timsina, P., Liu, J., El-Gayar, O.: Advanced analytics for the automation of medical systematic reviews. Inf Syst Frontiers 18(2), 237–252 (2016).
59. El-Gayar, O. F., Liu, J., Timsina, P.: (2015). Active learning for the automation of medical systematic review creation. In: 21st Americas Conference on Information Systems (AMCIS). Puerto Rico (Aug. 13–15, 2015). Htttp://aisel.aisnet.org/amcis2015/BizAnalytics/GeneralPresentations/22.
60. Halamoda-Kenzaoui, B., Rolland, E., Piovesan, J., et al.: Toxic effects of nanomaterials for health applications: how automation can support a systematic review of the literature? J of Applied Tox 42(1), 41–51 (2021). DOI: 10.1002/jat.4204.
61. Olorisade, B. K., Brereton, P., Andras, P.: The use of bibliography enriched features for automatic citation screening. J of Biomed Inf 94, 103202 (2019). DOI: 10.1016/j.jbi.2019.103202.
62. Bannach-Brown, A., Przybyła, P., Thomas, J., et al.: Machine learning algorithms for systematic review: reducing workload in a preclinical review of animal studies and reducing human screening error. Syst Rev 8(1), 23 (2019). DOI: 10.1186/s13643-019-0942-7.

63. Bui, D., Del Fiol, G., Hurdle, J. F., et al.: Extractive text summarization system to aid data extraction from full text in systematic review development. J of Biomed Inf 64, 265–272 (2016). DOI: 10.1016/j.jbi.2016.10.014.
64. Tsafnat, G., Glasziou, P., Karystianis, G., et al.: Automated screening of research studies for systematic reviews using study characteristics. Syst Rev 7, 64 (2018). DOI: 10.1186/s13643-018-0724-7.
65. Norman, C.: Systematic review automation methods. Université Paris-Saclay; Universiteit van Amsterdam (2020). Https://tel.archives-ouvertes.fr/tel-03060620/document.
66. Norman, C., Leeflang, M., Zweigenbaum, P., et al.: Automating document discovery in the systematic review process: how to use chaff to extract wheat. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA), Miyazaki, Japan (2018). Https://aclanthology.org/L18-1582.
67. Karystianis, G., Thayer, K., Wolfe, M., et al.: Evaluation of a rule-based method for epidemiological document classification towards the automation of systematic reviews. J of Biomed Inf 70, 27–34 (2017). DOI: 10.1016/j.jbi.2017.04.004.
68. Wallace, B. C., Kuiper, J., Sharma, A., et al. Extracting PICO sentences from clinical trial reports using supervised distant supervision. J of Machine Learning Res 17, 132 (2016).
69. Marshall, I. J., Kuiper, J., Wallace, B. C.: Automating risk of bias assessment for clinical trials. J of Biomed and Health Inf 19(4), 1406-12 (2015). DOI: 10.1109/JBHI.2015.2431314.
70. Ma, Y.: Text classification on imbalanced data: application to systematic reviews automation. Dissertation. University of Ottawa (2007).
71. Begert, D., Granek, J., Irwin, B., et al.: Towards automating systematic reviews on immunization using an advanced natural language processing-based extraction system. Canadian Commun Dis Rep 46(6), 174–9 (2020). DOI: 10.14745/ccdr.v46i06a04.
72. Scells, H., Zuccon, G., Koopman, B.: Automatic boolean query refinement for systematic review literature search. In: The World Wide Web Conference (WWW '19), pp. 1646–1656. Association for Computing Machinery, New York (2019). DOI: 10.1145/3308558.3313544.
73. Khabsa, M., Elmagarmid, A., Ilyas, I., et al.: Learning to identify relevant studies for systematic reviews using random forest and external information. Mach Learning 102, 465–482 (2016). DOI: 10.1007/s10994-015-5535-7.