Exploring Unsupervised Query Paraphrasing to Identify Relevant Search Phrases for a Literature Review

Laura-Maria PELTONEN [a],[1] Miretta TOMMILA [b] and Hans MOEN [c]

[a] *Department of Nursing Science, University of Turku, Turku, Finland*
[b] *Department of Perioperative Services, Intensive Care Medicine and Pain Management, Turku University Hospital and University of Turku, Turku, Finland*
[c] *Department of Future Technologies, University of Turku, Turku, Finland*

Abstract. Literature databases have multifaceted search options, but emerging research areas do not have an established terminology and therefore it is difficult to find relevant literature when conducting a review. This study aimed to explore if an unsupervised paraphrasing approach is useful in identifying relevant search phrases for a literature review on an emerging research topic – situational leadership in critical care. Using an initial set of 12 search phrases, the system was used to propose additional phrases, which were manually classified and further used in an expanded PubMed database search. Finally, we assessed the papers found with the expanded search and compared this to the initial search results. As a result, the expanded search more than tripled the search results, from 182 to 673 papers. The expanded search also more than tripled the number of relevant papers, from 12 in the original search to 39 in the expanded search.

Keywords. Literature review, search strategy, health care leadership, unsupervised query rewriting, unsupervised paraphrasing, PubMed

1. Introduction

Literature reviews present existing knowledge on a phenomenon of interest, by critically analysing and summarising previously reported research through a rigorous and systematic approach. There are a many types of literature reviews and the choice of review design is based on the purpose of the review [1]. Although reviews differ by type, the search process to find relevant literature is similar. The identification of relevant search terms, phrases, synonyms and language variations is important in an efficient search [2]. This includes identifying database specific terms and phrases, such as the controlled vocabulary of Medical Subject Headings (MeSH), which is maintained by the National Library of Medicine in the United States.

To date, literature databases have multifaceted search options developed for researchers for enabling robust search strategies [3]. The exploitation of various approaches and resources for bridging the gap between user queries and terminology used has been widely studied, such as query expansion using MeSH terms (see e.g. [4-

---

[1] Corresponding Author, Laura-Maria Peltonen, Department of Nursing Science, 20014 University of Turku, Finland; E-mail: laura-maria.peltonen@utu.fi.

8]). Previous work focusing on the use of database filters to improve precision of papers obtained have shown inconsistencies between different filters and raised the question of missing relevant literature if used as only search approach [9]. A more recent approach to advance search results includes the "Best Match" sort order for PubMed [10]. It is based on a two-stage ranking architecture where it first retrieves documents using a classic term-weighting algorithm and thereafter re-ranks the top 500 documents using an algorithm trained on past search statistics. However, the terminology in the body of evidence matures in time, and hence emerging research areas do not necessarily have an established terminology to be used. Therefore, finding appropriate search terms may still be difficult.

The aim of the present study is to explore the use of a recently proposed query rewriting and paraphrasing approach [11] – which relies primarily on distributional and sequential statistics about words – for the purpose of identifying relevant (multi-word) search phrases that can be used further when conducting a literature review on an emerging research area. In this paper, our phenomenon of interest is *situational leadership in critical care*.

2. Methods

In the first step, we identified a list of 12 initial phrases deemed to be central to the phenomenon of interest, i.e. situational leadership in intensive care, based on domain expertise (see Table 1 for original phrases). We searched the PubMed database (https://pubmed.ncbi.nlm.nih.gov/) with each of these 12 original phrases together with "intensive care" (e.g. "situational leadership" AND "intensive care"). The search was limited to full texts, papers published within 5 years and papers written in English.

As a second step, we used a recent query rewriting and paraphrasing approach (see [11]) that works in an unsupervised manner by combining 1) a semantic word n-gram (embedding) model, 2) a statistical language model and 3) a document search engine, where all these components have been indexed or trained on a PubMed abstracts data set. This system was used to extract 100 related (para-) phrase candidates for each of the 12 initial phrases. Two domain experts then classified these into four classes: 1 = synonym, 2 = relevant phrase, 3 = covers topic but is irrelevant for the search at hand, and 4 = irrelevant phrase. Consensus was sought if the domain experts disagreed.

In the third step we compared the papers obtained from searching with the initial phrases to a similar search using the automatically suggested phrases in addition to the initial phrases. All phrases that were classified as belonging to classes 1 and 2 in the second step were considered relevant and included in this expanded PubMed search on the topic. Two domain experts then manually classified the papers from these two approaches into three classes: A = relevant paper, B = on the topic but not important to the current review at hand, and C = irrelevant paper. We explored possible differences between the distribution of the classifications done to the papers found with the initial search phrases and the papers found with the expanded search with the Chi-square test.

3. Results

A total of 104 out of 1200 automatically generated phrases were classified as synonyms (n= 18) or relevant phrases (n= 86) by domain experts in the second step of the study.

In the third step, the expanded search more than tripled the results from 182 to 673 papers. The number of papers found with the initial phrases and the expanded search are presented in Table 1 together with the number of phrases added to each initial search phrase. Examples of automatically generated phrases used in the expanded search and classifications of these phrases by domain experts are found in Table 2.

Table 1. PubMed search findings with initial phrases and the expanded search

Initial search phrase	Initial search results	Number of phrases added to initial search phrase	Expanded search results
"situational leadership"	0	8	8
"situational medical leadership"	10	4	58
"medical team leadership"	68	13	168
"operational leadership"	0	4	134
"operational medical leadership"	70	15	100
"emergency leadership"	89	21	196
"medical emergency leadership"	63	13	331
"medical emergency team leadership"	19	0	19
"rapid response team leadership"	7	0	7
"resuscitation leadership"	0	25	355
"advanced cardiac life support leadership"	1	0	5
"advanced life support leadership"	7	1	7
Total	334	Total	1388
Duplicates	152	Duplicates	715
Final total	**182**	**Final total**	**673**

Table 2. Automatically generated new phrases for the initial phrase example "*medical team leadership*" with the classification done by the domain experts

Rewritten suggestions for the original search phrase: "medical team leadership"	Classification done by domain expert
Medical leadership, team	synonym
Leadership and medical teams	synonym
Physician leadership, interdisciplinary team	relevant phrase
Clinical team leadership	relevant phrase

The expanded search tripled the number of relevant papers retrieved (from 12 to 39). We found 182 papers with the initial search phrases, among these, 12 were classified as relevant, 32 were classified to be on the topic but not important to the review at hand, and 138 were classified as irrelevant. We found an additional 491 papers with the expanded search using the phrases suggested by the system that were classified relevant for the topic. Out of these 491 papers, 27 were classified as relevant papers, 89 were classified to be about the topic but not important to the review at hand and 375 were classified as irrelevant papers. The distribution of the papers that were classified to be relevant, on the topic but not important to the review at hand, and irrelevant in the search done with the initial phrases was similar to the distribution of the classifications done of the papers found with the expanded search X^2(2, N=673)= 0.10, p=0.86).

4. Discussion

The expanded search more than tripled the search results from 182 to 673 papers. Correspondingly, the expanded search also more than tripled the number of relevant papers from 12 to 39 papers. These findings indicate that the use of unsupervised query rewriting and paraphrasing approaches can be beneficial in identifying additional relevant search phrases in the context of searching for relevant literature in large databases such as PubMed. As future work we plan to explore how the approach used here may work in tandem with methods that are more directly tailored to PubMed searching, e.g. methods exploiting MeSH terms for query expansion [see e.g. 4-8]. Further, given that the present approach is trained in an unsupervised fashion, we also aim to explore how it can support search and information retrieval in text databases that lack relevant and/or tailored lexical resources, like MeSH and search history logs.

5. Conclusions

This study showed that the use of an unsupervised query paraphrasing system helped to identify additional relevant search phrases for the purpose of conducting a literature review on an emerging research area. Hence, the use of this or similar paraphrasing approaches may be used as one part of a rigorous search strategy.

References

[1] Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies, Health Information & Libraries Journal 2009;26: 91–108.
[2] Wichor BM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: an efficient and complete method to develop literature searches. Journal of the Medical Library Association 2018;106: 531-541.
[3] McKeever L, Nguyen V, Peterson SJ, Gomez-Perez S, Braunschweig C. Demystifying the search button: a comprehensive PubMed search strategy for performing an exhaustive literature review. Journal of parenteral and enteral nutrition 2015;39: 622–635.
[4] Lu Z, Kim W, Wilbur WJ. Evaluation of query expansion using MeSH in PubMed. Information retrieval 2009;12: 69-80.
[5] Dai S, You R, Lu Z, Huang X, Mamitsuka H, Zhu S. FullMeSH: improving large-scale MeSH indexing with full text. Bioinformatics 2020;36: 1533-1541.
[6] Crespo Azcárate C, Mata Vázquez J, Maña López M. Improving image retrieval effectiveness via query expansion using MeSH hierarchical structure. Journal of the American Medical Informatics Association 2013;20: 1014-1020.
[7] Abdeddaïm S, Vimard S, Soualmia LF. The MeSH-Gram Neural Network Model: Extending Word Embedding Vectors with MeSH Concepts for Semantic Similarity. Studies in health technology and informatics 2019;264: 5-9.
[8] Massonnaud C, et al.. Performance evaluation of three semantic expansions to query PubMed, Health Information & Libraries Journal; 2019. doi: 10.1111/hir.12291.
[9] Beynon R, et al. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. Cochrane Database of Systematic Reviews 2013;9. DOI: 10.1002/14651858.MR000022.pub3.
[10] Fiorini N, et al. Best Match: new relevance search for PubMed. PLoS biology 2018;16: e2005343. doi: 10.1371/journal.pbio.2005343.
[11] Moen H, Peltonen L-M, Suhonen H, et al. An Unsupervised Query Rewriting Approach Using N-gram Co-occurrence Statistics to Find Similar Phrases in Large Text Corpora, In NEAL Proceedings of the 22nd Nordic Conference on Computational Linguistics (NoDaLiDa), Linköping University Electronic Press, 2019.