

LETTER TO THE EDITOR

Use of text mining tools in the development of search strategies – Comparison of different approaches



We would like to comment on the article by Paynter et al. [1], who presented a prospective comparison of evidence synthesis search strategies with and without text mining tools (TMTs). The authors cite our research [2–4] and state that their study “expands on Hausner et al.’s previous work”, implying that the approaches applied were similar. However, we would like to note some key differences and refer to both the article [1] and their underlying full report [5].

We would like to first note that we welcome research on the use of TMTs in the development of search strategies. The Institute for Quality and Efficiency in Health Care (IQWiG) has been using TMTs for more than 10 years (first TM [6] in R, then WordStat [7]) and the results of our research and our experience in daily practice shows that TMTs can support the development of high-quality search strategies. IQWiG’s development and testing processes are based on the methodological standards for the development of objectively derived study filters [4,8].

Two IQWiG studies published in 2015 and 2016 (retrospective review based on 13 topics [2]; prospective review based on five topics [3]) showed considerably better results for TMTs than the study by Paynter et al. In our view, this finding can be explained by various factors differences in: 1) the selection and application of the TMTs, 2) the experience of the information specialists involved, and 3) the test set used for the text analysis.

1. Selection and application of TMTs

Our main point is that in the above study [1], the selection of the TMT tool and its application varied. Information specialists could choose one or more TMTs from the following predefined list: AntConc, PubReMiner, MeSH on Demand, and Yale MeSH Analyzer. The underlying full report also lists Carrot2 and VOSviewer, which are not mentioned in the article [1].

It was up to the information specialists to determine which TMTs were used, at which point in the search strategy development TMTs were applied, and how the search terms were chosen. The authors explain this by stating that no “best practice” methods were available. In contrast, IQWiG’s guidance

on TMTs (see supplementary data in [2]) supports the information specialists in their selection of search terms. For example, the guidance specifies that, as a rule, all citations from the test set should be found using both the free-text terms and the keywords. This is a very conservative approach, but leads to search strategies with high sensitivities.

2. Experience with TMTs

Information specialists with long-term experience in information retrieval (6 to 15 years) were involved in the study by Paynter et al. However, the following statements in Appendix C of the full report [5] indicate that some had only limited experience with TMTs:

“I found PubReMiner to be extremely helpful ... (once I figured out my method) in identifying MeSH terms and keywords.”

“PubReMiner—Still had to generate a PubMed query. I tried to build a query using my seed set of articles with their PMIDs, but this didn’t work (or I couldn’t get this to work)”.

“Once I realized that I could export the PMID list from Ovid in spreadsheet format, and then copy and paste the PMID column from the spreadsheet into PubReMiner, it became even easier and faster.”

“Time was added to search process to tweak and troubleshoot issues related to constraints of the TMTs ...”

The better results shown for TMTs in IQWiG’s studies might at least partially be explained by the greater experience of IQWiG’s information specialists with TMTs. For example, before the start of a project they are provided with detailed guidance on how to use these tools. In addition, two IQWiG information specialists are responsible for the maintenance, implementation and further development of TMTs as well as the training of other staff members.

3. Test set for the text analysis

A text analysis can only be as good as the underlying data, the test set (called “seed citations” or “seed set” in [1]) some

details on the test sets are presented in the full report [5], but these raise questions: The number of citations for the text analysis appears to be sufficient (11 to over 100 citations per report). However, some statements by the information specialists involved indicate that some of the citations were not sufficiently representative and may not have completely covered the research question. For example:

“It was difficult to determine how representative the known citations were of the topic area”

“The citations may have been a little broad, but generally seemed good”

“I had to add a number of terms that were not identified through TM, probably because there were so few seed citations”.

Particularly in the case of complex research questions, it is thus doubtful whether citations on all interventions were found in the test set. In this case, there is a risk that the text analysis did not reflect the entire research question.

IQWiG makes a substantial effort to systematically generate a representative test set at an early stage in a project. A search is conducted for systematic reviews on the topic under investigation. The quality of information retrieval in these reviews is assessed, and the citations included are then used as the basis for the text analysis. In addition, for each search strategy, it is checked whether the available citations from the test set are found. Citations not found are followed-up and if necessary, the strategy is adjusted or it is documented why they were not found.

4. Publication of further information

We would like to ask the authors to publish further information. This concerns in particular the search strategies (including numbers of hits). This would help to reproduce the study results.

In addition, the publication of the citations of the test sets and the reference standards would be useful. This information could be made available in the form of PMIDs. The following questions, which were not further investigated in the study by Paynter et al., could then be answered:

- Are the citations from the test set found with the TMT search strategy?
- If citations from the reference standard were not found, which search block was responsible?

5. Summary

In the study by Paynter et al., the expectations of TMTs were perhaps too high. The hope that, without (much) prior

experience and guidance, TMTs would produce high-quality search strategies with a few clicks was not fulfilled. This is in line with our experience. The use of TMTs alone is not in itself a sign of quality. The selection of a suitable test set forms the basis for a high-quality analysis using TMTs. In addition, the routine use of TMTs requires some practice, this should be taken into account when specifying information retrieval processes.

Thus, we disagree with the authors' conclusion that TMTs are “not ready to be used as the sole process for developing SR searches”. In our opinion, if the preconditions outlined above are met, TMTs can produce high-quality search strategies.

6. Outlook

For TMTs to become a valuable support for information specialists in their daily work, further steps are needed. First, the TMTs shown to be useful in previous studies should be documented. Second, a consensus should be reached on how search terms are selected (e.g., using cut-offs). In our opinion, it only makes sense then to conduct further studies. Like Paynter et al., we see the need for further research on the evaluation of test sets with regard to representativeness. The goal is to develop an evaluation tool for this purpose.

Acknowledgments

The authors thank Natalie McGauran for editorial support. We would also like to thank Tatjana Hermanns, Annika Orland and Dorothea Sow for content support.

Elke Hausner*

Marco Knelangen

Siw Waffenschmidt

Information Management Department

Institute for Quality and Efficiency in Health Care (IQWiG)

Im Mediapark 8, D-50670 Köln, Germany

*Corresponding author. Information Management Department, Institute for Quality and Efficiency in Health Care (IQWiG), Im Mediapark 8, D-50670 Köln, Germany. Tel.: +49 221 35685 258; fax: +49 221 35685 1.

E-mail address: Elke.Hausner@iqwig.de (E. Hausner)

References

- [1] Paynter RA, Featherstone R, Stoeger E, Fiordalisi C, Voisin C, Adam GP. A prospective comparison of evidence synthesis search strategies developed with and without text-mining tools. *J Clin Epidemiol* 2021;139:350–60.
- [2] Hausner E, Guddat C, Hermanns T, Lampert U, Waffenschmidt S. Development of search strategies for systematic reviews: validation showed the noninferiority of the objective approach. *J Clin Epidemiol* 2015;68:191–9.
- [3] Hausner E, Guddat C, Hermanns T, Lampert U, Waffenschmidt S. Prospective comparison of search strategies for systematic reviews: an

objective approach yielded higher sensitivity than a conceptual one. *J Clin Epidemiol* 2016;77:118–24.

- [4] Hausner E, Waffenschmidt S, Kaiser T, Simon M. Routine development of objectively derived search strategies. *Syst Rev* 2012;1:19.
- [5] Paynter RA, Fiordalisi C, Stoeger E, Erinoff E, Featherstone R, Voisin C, et al. A prospective comparison of evidence synthesis search strategies developed with and without text-mining tools Rockville (MD). 2021. Available at <https://www.ncbi.nlm.nih.gov/pubmed/33755394>. Accessed June 17, 2022.
- [6] Feinerer I. Package “tm”: text mining package. Available at <https://cran.r-project.org/web/packages/tm>. Accessed June 20, 2022.
- [7] Provalis Research. WordStat: Version 8.0.35. Available at <https://provalisresearch.com/products/content-analysis-software/>. Accessed June 20, 2022.
- [8] Jenkins M. Evaluation of methodological search filters: a review. *Health Info Libr J* 2004;21:148–63.

<https://doi.org/10.1016/j.jclinepi.2022.05.011>

Reply to Hausner et al. Re: use of text mining tools in the development of search strategies—comparison of different approaches



We thank Hausner et al. for their commentary and reply to each of their points below.

1. Hausner et al. Point 1: selection and application of text-mining tools

We are happy to agree it is entirely possible that our team would have achieved similar results if we had access to either of the fee-based software packages (i.e., SimStat/WordStat or the R Text Mining Package) used by Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Institute for Quality and Efficiency in Health Care) information specialists in their research along with their training and support [1–3].

Our interest lay in expanding the research base in this field by exploring the use of free text-mining tools (TMTs) available to evidence synthesis searchers and testing their performance on simple and complex topics (simple = one indication, one treatment clinical topics; complex = everything else) in a ‘real world’ environment, that is, in which searchers only have access to free tools for which no best-practice guidance currently exists [4,5]. In our study, for both processes, the recall was generally very high (i.e., we

are not saying the technology does not work) and perhaps with standardization, such as Hausner et al. describe, it could be even higher than our results. This initial foray into a different set of tools and using the creativity of information specialists to develop novel methods we hoped would lay some groundwork for future studies that may indeed establish best practices for these individual tools. Please observe our recently published short pros and cons article for more details on using TMTs [6].

Our current research project is creating a user-friendly, comparative guide to search tools listed on the SR Toolbox website. Using standardized criteria we are evaluating availability, usability, and performance, along with providing some usage tips and tricks with the results presented in an easy-to-understand graphic. With so many tools available but often too little time to try them out, we hope our fellow information specialists will find this guide useful.

2. Hausner et al. Point 2: experience with TMTs

We agree that the experience with TMTs and ongoing professional support are important ingredients in their successful use. We count ourselves among those evidence synthesis searchers who are not located in such an environment and thus wanted to investigate the performance of tools available to us and whether the time to learn a new tool paid off (even with no experience and no best-practices guidance available). We hope our brief comments have been useful to others regarding what works and what does not work so well.

3. Hausner et al. Point 3: test set for the text analysis

We agree that a representative seed set of citations is critical for generating a more accurate analysis. To our knowledge, there has not been any research on the effect of an unrepresentative (biased) seed set of citations on the development of searches. With the likely increase in the use of search tools, attention to this issue is critical in our opinion and the development of a risk of bias tool would be most welcomed.

4. Hausner et al. Point 4: publication of further information

The requested information would indeed make an interesting additional study; however, we did not systematically collect these, so are unable to provide them. This area of research would potentially benefit from the use of a standard dataset of search strategies and relevant citations, such as that created by Scells et al. [7], which could be used across evaluations to give more consistent results.

Funding: This project was funded under contract no. 290-2017-00003C from the Agency for Healthcare Research and Quality (AHRQ), United States Department of Health and Human Services. The authors are solely responsible for its content. Statements should not be construed as endorsement by the Agency for Healthcare Research and Quality or the United States Department of Health and Human Services.

Conflict of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.