Investigation of text mining methodologies to aid the construction of search strategies in systematic reviews

Background

Systematic reviews are considered the highest level of evidence in the healthcare research hierarchy. They pose detailed questions and aim to assess all of the available evidence systematically and objectively. Information specialists are fundamental contributors to a systematic review. They provide the literature base which the review is founded on. This building block will inform every stage of the review process. Without a robust search strategy, it is difficult to capture all of the relevant literature. This introduces bias into the review and can negate the findings, providing erroneous information to healthcare providers and policy makers (1). The difficulties faced by information specialists are compounded by the continually expanding volume and formats of scientific and health literature. Review questions are becoming more complex to reflect this, yet information specialist methodologies largely remain the same (1). Designing a search strategy is a time consuming task. Information specialists must broach subjects which are unfamiliar and very rapidly accumulate enough knowledge to design a search strategy (2). Healthcare research encompasses a wide breadth of subjects and it is not realistic for an information specialist to have in depth knowledge of them all. This makes it incredibly difficult to design a search that will balance sensitivity and specificity (3). Thesauri differ between databases which means information specialists need to have a strong grasp on the subject to ensure all appropriate thesaurus headings are captured within the search strategy as well as building a body of free text terms that cover differences in terminology and include appropriate antonyms, synonyms and acronyms. Input from clinical experts can help with the initial design, however, information specialists can spend many hours reading around the subject to understand the terminology, acronyms and specificities of the topic (3). It is becoming increasingly apparent that methodologies for semi-automation are required to help information specialist grasp the extensive subject matter quickly and to facilitate search strategy design (4, 5).

Description of the Problem

Methodologies for semi-automation of search strategy design are becoming available yet they are only used in the minority of cases. Text mining applications make up the highest proportion of semi-automation methodologies. Some researchers have shown that the use of such applications does increase the precision and sensitivity of search strategies (3). Others argue that the inherent biases of using such applications in this context can skew the search strategy (6). Current reviews of text mining tools have explored their use in the development of the search strategy, however, they have widely been conducted by groups with prior knowledge of the applications and methodologies behind them (5, 7). This provides the benefit of deep investigation into the applications but lacks the perspective of the naïve user. If these applications are to come into mainstream use by information specialist, there is a need to explore not only their ability to contribute positively to the design of the search strategy but also how user friendly they are for individuals with no prior experience. This investigation proposes to assess 16 commonly referenced text mining applications and report the intuitiveness of use, potential biases and contributions to the design of a search strategy for a Diagnostic Test Accuracy review.

Objectives

This investigation will focus on the usability of the applications and the continuity between outputs, opinionising the overall benefits and detriments for novice users.

Applications being investigated

This investigation will focus on text mining applications which are regularly cited in information specialist review methods. A total of 16 applications, a mix of web-based and desktop, have been selected:

- Voyant
- VosViewer
- CitNetExplorer
- PubReMiner
- TerMine
- TextAlyser
- Text Analyzer
- Lingo3G (Carrot)
- Lingo4G
- Carrot2
- MeSHonDemand
- Yale MeSH Analyser
- EndNote
- Anne O'Tate
- BiblioShiny
- Tm for R

Why it is important to investigate these applications

The use of one or more of these applications may be able to help information specialists gain instant subject knowledge by providing a snapshot of the literature. There is the potential for saving a great deal of time and resources by using these applications, limiting the time spent reading around the subject. Text mining allows users to identify key concepts which are shaped by the words and phrases used by the authors. Including or removing these words and phrases in the search strategy can improve the performance (3). Text mining applications have the ability to find common terms or terms which are rare yet highly important for identifying relevant literature. This is particularly important when information specialists are not familiar with the subject as terms may easily be missed. Contrastingly, terms which are uninformative may be identified and can subsequently be removed from the search strategy. Being able to distinguish between informative and uninformative terms is, again, another issue with being unfamiliar with a subject. Using these applications to enhance the search strategy can limit the number of missed studies by balancing sensitivity and specificity. Ultimately this will increase the body of evidence, reduce bias and contribute to the applicability of the review findings. It is becoming increasingly important to advance the methodologies for information specialists as the volume of literature increases. Not only this, but new information sources are increasingly realised and sought after by stakeholders. A clear example of this is the need to access and summarise social media data such as Tweets. Text mining applications provide a quick and relatively simple methodology to analyse this information and draw out key concepts which can be used in a search strategy. It is important to ensure these applications are simple to use, provide clear outputs, maintain consistency and contribute positively to search strategy design.

Methods

An initial search strategy will be designed in Ovid-Medline by a trained information specialist (HO) and translated to Ovid-Embase. These will be used as the basis for the investigative analysis. Each application will be used by the information specialist with a selection of five reference publications relevant to the subject (see below 'Types of input data' for further details). Outputs from each

application will be saved for comparative analysis. Those with like-for-like inputs and outputs will be compared to ensure consistency between the results of the application. If inconsistencies are found, deeper exploration of the underlying algorithms will be conducted. The initial search strategy will be edited to reflect the results of the text mining, either including or excluding terms, and different version will be saved. Each version of the search strategy will be run on the same day and differences in the retrievals will be assessed (see below 'Search methods for identification of target studies' and 'Data collection and analysis' for further details).

Criteria for investigating applications

Types of target studies

The search strategy will be used to identify literature for a diagnostic test accuracy review. The review aims to determine if non-invasive or minimally invasive autopsy techniques are suitable alternatives to traditional autopsy for prenates, neonates and infants. Any study with a sequential design that assesses non-invasive or minimally invasive techniques against a traditional autopsy will be accepted. The initial search strategy will focus on the population, index test and reference standard concepts.

Types of applications

The applications considered for this investigation are based on text mining techniques. No limits have been applied to the type of text mining method or underlying algorithms deployed by these applications. Similarly, no restrictions have been placed on the input or output formats for the applications. However, the included applications have been limited to those widely cited as being specifically used by information specialists.

Types of input data

For the purposes of exploring the available applications, several approaches will be employed based on the applications requirements. The following publications will be used as a basis for all testing:

1. Brookes, J. et al. 1996. Non-invasive perinatal necropsy by magnetic resonance imaging (8).
2. Breeze, A. et al. 2007. Feasibility of percutaneous organ biopsy as part of a minimally invasive perinatal autopsy (9).
3. Sebire, N. et al. 2011. Minimally invasive perinatal autopsies using magnetic resonance imaging and endoscopic post-mortem examination ("Keyhole autopsy"): feasibility and initial experience (10).
4. Thayyil, S. et al. 2013. Post-mortem MRI versus conventional autopsy in fetuses and children: a prospective validation study (11).
5. Ruegger, C. et al. 2014. Minimally invasive, imaging guided virtual autopsy compared to conventional autopsy in foetal, newborn and infant cases: study protocol for the paediatric virtual autopsy trial (12).

The publications will be downloaded as PDFs and converted into an appropriate format for input to the applications where necessary. Citations will be exported and converted to .enw or .RIS files with EndNote. Web of Science will be used to directly import references into the applications where necessary or full reference and citation information will be downloaded as a plain text file. URLs will be supplied to the original publications and to the PubMed central (PMC) library entries where available. PubMed Identification codes (PMID) sourced for each publication will be used where required, PMID: 8888168; PMID: 18087719; PMID: 21740313; PMID: 23683720, PMID: 24438163, respectively. Finally, for those applications requiring a search string input, the following will be used: ((“minimally invasive” OR “minimally-invasive” OR “non invasive” OR “non-invasive”)[ti:ab] AND (autops*)[ti:ab] AND (prenatal OR prenate* OR perinatal OR perinate* OR neonatal OR neonate* OR infant)[ti:ab]).

Types of outcome measures

Outcome measures will be based on the outputs for each of the applications. These will include, but are not limited to, word frequency analysis, correlation and network mappings, phrase identification and term-frequency inverse document frequency analysis. No formal statistical measures will be applied. A summary opinion of the usability for each application will also be recorded. The terms identified for inclusion or removal from the search strategy will be listed with a record of the impact on the search retrieval.

Search methods for identification of target studies

Several version of the search strategy will be created to assess the impact of inclusion and removal of terms identified through the use of the applications against the initial design. All versions of the search, including the initial design, will be run on the same day to ensure the database content is consistent. Ovid-Medline and Ovid-Embase will be used to enable database specific comparison between different versions of the search strategies.

Data collection and analysis

Assessment of risk of bias within applications

Certain inherent biases are unavoidable given the nature of use. For example, information specialists are often given a small sample of papers from clinical experts which will not reflect the entire body of literature. Results based on this sample will be bias as the applications will only analyse the content of what is given to them (6). A lack of knowledge on the part of the information specialist can also subconsciously introduce bias, whether this is lack of knowledge regarding the subject or regarding the application of choice. However, this inherent bias does not negate the importance of this investigation. If anything, this investigation aims to reduce bias by providing a level of understanding to help information specialist choose an appropriate application for their task.
Nevertheless, other forms of bias may exist within the applications themselves, such as limits to the input formats. No formal assessment will be used but potential sources of bias within the applications will be investigated as well as potential bias introduced by the methodologies of this investigation.

Data management and synthesis

Data will be collated regarding the input formats used for each applications and all outputs from the applications for inspection and analysis where appropriate. Comparative analysis will be performed where applications accept like-for-like inputs and utilise the same type of text analysis. For example, those accepting URL inputs and performing word frequency analysis. The outputs will be compared through similarity measures (Percentage similarity) to ensure the applications are performing consistently. If inconsistencies are noticed, the underlying algorithms will be investigated and summarised with recommendations for improvement to the algorithms and/or suggestions for preference of an application. The following combinations accept the same input formats and it is anticipated that they will progress to the comparative analysis stage:

- VosViewer + CitNetExplorer + BiblioShiny
- Voyant + TextAlyser + Text Analyzer
- Lingo3G + Lingo4G + Carrot2 + Anne O'Tate

A record will be kept of the initial search strategy, what amendments have been made and the subsequent versions. This will include a list of the applications which provided the suggestions for amendment and the type of text mining approach that resulted in the suggestion. Data regarding the search retrievals will be collated, specifically the number of results retrieved, the difference between the amended version and the initial design terms and whether this difference lost/gained relevant results. Points of usability for each application will be recorded, such as ease of input, ease of

modifying inputs, accessibility of functions within the application, intuitiveness of the text mining methods and clarity of the outputs. A narrative synthesis will be used to describe the findings of this data synthesis. The narrative will be written from a novice users perspective and will include opinions on the use of each application in the context of this investigation. Narrative description will also be given to describe the robustness of the decisions made during this exploration, the impact this may have on the data synthesis and the conclusions drawn. Overall conclusions regarding the usefulness of each application will be drawn on a subjective 5 point scale – extremely useless, useless, no impact, useful, extremely useful. This will be based on the assessment of usability and the impact on the literature retrieval. A summary table will be produced to demonstrate the inputs, outputs, bias assessment and relative scale of usefulness for each application.

1. Metzendorf MI, Featherstone RM. Ensuring quality as the basis of evidence synthesis: leveraging information specialists' knowledge, skills, and expertise. Cochrane Database of Systematic Reviews. 2018(9).
2. Beverley CA, Booth A, Bath PA. The role of the information specialist in the systematic review process: a health information case study. Health Information & Libraries Journal. 2003;20(2):65-74.
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. Journal of Clinical Epidemiology. 2016;77:118-24.
4. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Systematic Reviews. 2015;4(1):5.
5. Glanville JW, H. . Text mining opportunities: White paper. 2018.
6. Thomas J, McNaught J, Ananiadou S. Applications of text mining within systematic reviews. Research Synthesis Methods. 2011;2(1):1-14.
7. Paynter RB, LL. Berliner, E. Erinoff, E. Lege-Matsuura, J. Potter, S., Uhl S. EPC Methods: An Exploration of the Use of Text-Mining Software in Systematic
Reviews. 2016.
8. Brookes JAS, Hall-Craggs MA, Sams VR, Lees WR. Non-invasive perinatal necropsy by magnetic resonance imaging. The Lancet. 1996;348(9035):1139-41.
9. Breeze ACG, Jessop FA, Whitehead AL, Set PAK, Berman L, Hackett GA, et al. Feasibility of percutaneous organ biopsy as part of a minimally invasive perinatal autopsy. Virchows Archiv. 2008;452(2):201-7.
10. Sebire NJ, Weber MA, Thayyil S, Mushtaq I, Taylor A, Chitty LS. Minimally invasive perinatal autopsies using magnetic resonance imaging and endoscopic postmortem examination ("keyhole autopsy"): feasibility and initial experience. The Journal of Maternal-Fetal & Neonatal Medicine. 2012;25(5):513-8.
11. Thayyil S, Sebire NJ, Chitty LS, Wade A, Chong WK, Olsen O, et al. Post-mortem MRI versus conventional autopsy in fetuses and children: a prospective validation study. The Lancet. 2013;382(9888):223-33.
12. Rüegger CM, Bartsch C, Martinez RM, Ross S, Bolliger SA, Koller B, et al. Minimally invasive, imaging guided virtual autopsy compared to conventional autopsy in foetal, newborn and infant cases: study protocol for the paediatric virtual autopsy trial. BMC Pediatrics. 2014;14(1):15.