Systematic Reviews: A Complex Search Episode for Evidence Based Policy and Practice

Sarvnaz Karimi
NICTA and the University of Melbourne
Department of CSSE
Parkville, VIC, Australia
skarimi@unimelb.edu.au

Falk Scholer
RMIT University
School of Computer Science and IT
Melbourne, VIC, Australia
falk.scholer@rmit.edu.au

ABSTRACT

Evidence based policy and practice – a paradigm that aims to ensure that decisions are based on consideration of research evidence that meets a high standard – began in the field of medicine, but is becoming widely used in other fields such as economic policy, education, and software engineering. Systematic reviews, the core tools of this evidence based approach, require stringent searching to identify sources of evidence that should inform a decision. We outline the systematic review process, an example of a *complex search episode*, and describe some of the challenges facing information retrieval in this domain.

Categories and Subject Descriptors

H.3.3 [**H.3.3 Information Search and Retrieval**]: Search Process

1. INTRODUCTION

Evidence based practice refers to the use of rigorous evidence, supported by systematic empirical research, to guide decisions. The paradigm was first developed in the field of medicine, aiming to ensure that medical decisions take the best available external evidence into account, rather than resting primarily on the basis of opinions and personal clinical experience [4]. The evidence is focused on rigorous, statistically significant results that are typically the outcomes of randomized controlled trials. Since being embraced in the medical field, the evidence based paradigm has been extended to many other areas of decision-making, from government policy, to software engineering, and product design.

The key tool used in evidence based policy and practice is the *systematic review*, a document which synthesizes available research on the topic of investigation. While most research work involves some sort of literature survey, a distinguishing feature of the systematic review is that it is carried out to agreed standards: using clear protocols in carrying out the process; focusing on specific questions; identifying as much of the relevant literature as possible; critically appraising the quality of the research included in the review; synthesizing research findings from included studies; being as objective as possible to remove bias; and, updating the review so that it remains relevant [1].

A key part of the systematic review, therefore, is the identification of the related literature. Indeed, achieving the objective of an unbiased synthesis of current evidence assumes that *all* relevant related work is identified and considered. In information retrieval terms, therefore, the systematic review process can be characterized as a *recall oriented* task, with the aim of finding all relevant documents that support the current review's underlying question.

SIGIR Workshop on "entertain me": Supporting Complex Search Tasks, July 28, 2011, Beijing.
.

2. SEARCH FOR SYSTEMATIC REVIEWS

We illustrate the challenges in conducing search for systematic reviews in the domain of evidence based medicine as an example of a complex search episode.

A focused research question is first specified by the researchers. An example is: *"Exercise in prevention and treatment of anxiety and depression among children and young people".*[1] Together with the research question, detailed inclusion and exclusion criteria are also specified. For the previous example, these are summarized as: *"Randomized trials of vigorous exercise interventions for children and young people up to the age of 20, with outcome measures for depression and anxiety".* However, we note that as part of the reported search strategy, the criteria are actually fully specified under four different headings: types of studies, participants, interventions and outcome measures.

The search process can then be viewed as consisting of three broad steps:

1. Search experts (e.g., health librarians) formulate complex Boolean queries – also known as *search strategies* – which are run over biomedical databases such as PubMed. The output is a large pool of document summaries consisting of titles, abstracts and authors.
2. The set of *summaries* is scanned by the investigators to identify a short-list of candidate documents that meet the systematic review inclusion criteria.
3. The investigators examine the *full text* of the articles in the short-list, and identify the final set of documents that will be included in the systematic review.

Each step of the process reduces the size of the candidate set drastically. For example, the MEDLINE bibliographic database of life sciences and biomedical information currently indexes around 20 million citations. The search strategy from Step 1 is typically formulated to retrieve a result set ranging from several hundred to a few thousand candidate documents. Triage based on summaries in Step 2 of the process reduced this candidate set to a few hundred items. The review of full text items in Step 3 the leads to

[1] `http://www2.cochrane.org/reviews/en/ab004691.html`

final included papers, typically from ten to a hundred documents [2].

3. COMPLEXITIES IN SYSTEMATIC REVIEWING

The primary *search* complexity in identifying papers that need to be included in a systematic review arises from the specific details of the information being sought. To effectively identify answer documents, it is for example necessary to understand the relationship between various entities in the query (in the example, this might include that patients are suffering from the specified condition, and that the condition could involve anxiety and depression, but only one is a necessary criterion for relevance), as well as the search context (for example, the fact that studies on older people should be excluded, and that only studies reporting specific outcome measures should be considered).

Currently, support for the multiple criteria which need to be considered in order to determine whether a document is likely to be relevant consists of the development of complex Boolean queries in Step 1 of the outlined process. These queries are often of the order of a hundred lines in length, and can take many weeks to develop. In the example systematic review on exercise, the search strategy involved Boolean queries over 7 biomedical databases, and the complex queries ranged from 37 to 79 lines in length. Moreover, these queries include the use of advanced operators for partial string matching, query expansion based on medical subject categories, and the complex manual combination of sub-sets of search results. Despite this intense human-driven effort, it is clear from current practice that specifying inclusion and exclusion criteria is insufficient: human intervention is needed at several steps of the process to remove many thousands of non-relevant items from the candidate set.

A further challenge is presented by the implicit assumption that the initial search strategy identifies all possibly relevant documents. This is fundamental to the evidence based paradigm, which posits that all high-quality evidence needs to be considered. The search task is therefore inherently recall focused: the cost of missing a relevant piece of evidence is high, potentially calling the findings of the final systematic review – a document that may take from 6 months to 2 years to produce – into question. Although the search strategies in Step 1 are typically developed by experts who are familiar with the domain in which the systematic review is being undertaken, it is still likely that some potentially relevant documents may be missed. The problem is further compounded by the fact that the reported search strategies sometimes contain errors, and on re-execution on the same document collection it often transpires that certain included documents in Step 3 are not in the candidate list from Step 1 [3].

Data complexities also exist; the key factor contributing to the difficulty of systematic review search episodes here is that the source collection to be searched over is often not in the form of full-text documents. For example, in PubMed – the most widely-used database for medical systematic reviews – only about 1 million of the 20 million indexed MEDLINE articles include full text, with the remainder consisting only of abstracts and metadata.

4. SUPPORT SYSTEMATIC REVIEWING

We contend that to effectively support search for complex scenarios such as evidence-based policy and practice, next generation information retrieval systems need to incorporate a range of features and technologies.

- To assist in query formulation for an initial search strategy, retrieval systems should aid the user in identifying relevant *entities*, this will relieve the need for searchers to construct long manual lists of synonyms. While attempts at synonym expansion using biomedical dictionaries or taxonomies (e.g., MeSH) are common, the naming conventions should be resolved with reference to the *current collection* that is being searched.
- Automated assistance in formulating the *relationships* between identified entities should be available, such that these accurately and directly map to the inclusion criteria. This is vital in reducing the complex re-combination of answer subsets that is currently required in the Boolean approach.
- While selecting individual documents for further consideration in each of the search steps, automated support for *consistency* is vital. If a reviewer selects one document, but later chooses to ignore a similar one, the system should flag this possible inconsistency.
- A dynamic *relevance feedback* approach that is active during the document selection process could rank the remaining documents based on estimated importance, assisting assessors in focusing their efforts. Moreover, such an approach might identify *additional* documents that exist in the collection but were missed by the initial search strategy.

While many of these items have been proposed and validated experimentally in isolation, we are unaware of a system that comprehensively includes all of these features.

5. CONCLUSION

Systematic reviews are a key tool for evidence based policy and practice, a decision making paradigm that is becoming increasingly widespread. The cost of producing such reviews is a direct function of the quality of the search used to identify relevant evidence. While there are a number of challenges that need to be resolved to allow the easy formulation of comparative search experiments in this paradigm, we believe that working to resolve these can offer significant benefits for information retrieval in evidence based policy and practice.

Acknowledgments NICTA is funded by the Australian government as represented by Department of Broadband, Communication and Digital Economy, and the Australian Research Council through the ICT centre of Excellence program. Thanks to our collaborators at the Global Evidence Mapping Initiative.

6. REFERENCES

[1] A. Boaz, D. Ashby, and K. Young. Systematic reviews: What have they got to offer evidence based policy and practice? *ESRC UK Centre for Evidence Based Policy and Practice*, 2002.

[2] A. Cohen, W. Hersh, K. Peterson, and P. Y. Yen. Reducing workload in systematic review preparation using automated citation classification. *JAMIA*, 13(2):206–219, 2006.

[3] S. Golder, Y. Loke, and H. M. McIntosh. Poor reporting and inadequate searches were apparent in systematic reviews of adverse effects. *J. Clin. Epidemiol.*, 61(5):440–448, 2008.

[4] D.L. Sackett, W. Rosenberg, JA Gray, R.B. Haynes, and W.S. Richardson. Evidence based medicine: what it is and what it isn't. *BMJ*, 312(7023):71, 1996.