

What remains to be evaluated in Multilingual Medical Search?

Allan Hanbury

Institute of Software Technology and Interactive Systems, TU Wien
Favoritenstraße 9-11/188, 1040 Vienna, Austria
allan.hanbury@tuwien.ac.at

ABSTRACT

We present a summary of current text IR evaluation campaigns in the medical domain in terms of users, documents and tasks. We then identify areas of medical IR in which systematic large-scale evaluation remains to be done.

CCS Concepts

•Information systems → Information retrieval; •Applied computing → Health care information systems;

Keywords

evaluation; multilingual; medical search; position paper

1. INTRODUCTION

Medical search is an instance of domain-specific search. In [4], we presented the five dimensions that should be considered when modelling a search domain: subject areas; modality; users; tasks; and tools, techniques and algorithms. Evaluation campaigns evaluate the tools, techniques and algorithms, given fixed settings for the other four dimensions. In this position paper, we restrict the modality to text. We begin by considering a fine grained classification of subject areas, i.e types of documents in the medical domain; as well as users — people that carry out medical search (Section 2). We then briefly overview medical IR evaluation campaigns and identify which combinations of document types, users and tasks have been covered (Section 3). Finally, in Section 4, based on gaps identified in the two previous sections, we indicate some areas in which large-scale systematic IR evaluation has not yet taken place.

2. DOCUMENTS AND USERS

Table 1 shows document types in the medical domain, as well as the languages in which these documents tend to be written. Primary research is divided into theoretical and applied basic research, experimental and observational clinical

Document Type	Language
Medical records	Country-dependent
Doctor-patient communication e.g. discharge summaries	Country-dependent
Primary research	English, but important publications in Chinese, ...
Secondary research	English
Clinical guidelines	Country-dependent
Wikipedia medical pages	Multiple
Health web sites	Country-dependent
Patient information leaflets	Country-dependent
Regulatory documents	Country-dependent

Table 1: Types of medical documents and the language in which they are commonly written

research, and experimental and observational epidemiological research [9]. Secondary research is made up of meta-analyses and systematic reviews, and is less up-to-date but contains results with a higher level of certainty than primary research [5]. For example, given a specific question (such as whether a specific intervention is better than a placebo), a systematic review synthesises the relevant primary literature to create an outcome with higher certainty in the secondary research literature. Primary research and secondary research are almost always in the form of medical journal publications, although increasing amounts of pre-publication data is being released — an example is Clinical Study Reports, which are documents of several hundred pages documenting the actual details of carrying out a clinical trial. Clinical guidelines are documents closest to medical practice — they advise a physician on what should be done in specific cases. Clinical guidelines can be available at a national, regional (provincial) or even organisational level, and may be associated with a requirement for physicians to follow them. Wikipedia is a popular source of medical information, including for physicians [6]. Regulatory documents are associated with procedures for approval by medical regulatory organisations such as the US Food and Drug Administration (FDA) and European Medicines Agency (EMA).

People who use search in the medical domain include health care providers, researchers, information specialists, administrators and patients. Health care providers include physicians, nurses, dentists, mental health care workers, hospital social workers, at-home care providers, midwives and practitioners of alternative medicine [2]. Physicians can be further sub-divided by their specialities, such as primary care, neu-

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MedIR'16 July 21, 2016, Pisa, Italy

© 2016 Copyright held by the owner/author(s).

ACM ISBN 123-4567-24-567/08/06...\$15.00

DOI: 10.475/123_4

rology, oncology, pulmonology and cardiology, each having different information needs. Patients often search directly, but friends and relatives of patients also commonly perform searches.

3. EXISTING EVALUATION CAMPAIGNS

The TREC Medical Records Track (2011–2012) [11] proposed the task of searching through anonymised English medical records to identify cohorts for clinical trials, modelling a task likely carried out by a physician active in research. The TREC Clinical Decision Support Track (2014–2016) [8] models searching for medical articles (the Open Access Subset of PubMed Central) relevant to answering a question motivated by a description in an English patient medical record. Here, the task is likely carried out by a physician seeing a patient. TREC Genomics (2003–2007) modelled a task performed by medical researchers searching for genetics-related information in primary research articles. These tasks dealt only with English queries and English documents.

The Information Retrieval tasks of the CLEF eHealth Lab (running since 2013) [3] focus on medical information search by laypeople, modelling a task likely to be carried out by patients, with the documents searched being health web sites. In the two initial years, queries that individuals may realistically pose based on the content of their discharge summaries were created by medical professionals, but then approaches that elicit queries directly from laypeople were adopted [12]. In 2014 and 2015, multilingual queries (manual translations of all queries) were available. We omit the long-running ImageCLEF medical tasks due to our focus on text search.

In summary, the tasks that have been evaluated at large scale are: researchers searching medical records to identify cohorts, researchers searching the primary literature, physicians searching the literature, and patients searching the web. All searches of literature and medical records have only been evaluated in English, with the multilingual aspect only present in the patient search evaluation.

4. OPEN EVALUATION TOPICS

An important task for which the IR component has not been evaluated at large scale yet is systematic review creation. The creation of a single systematic review takes a significant amount of effort and time — for the Cochrane Collaboration, on average 23 months from protocol to publication [10]. Given the effort involved, it is essential that all relevant articles are found and included in the review, to ensure that the outcome contains all relevant input.

The multilingual aspects of medical literature search have not been evaluated yet. For systematic reviews, publications relevant for inclusion are written in languages other than English, with a significant amount in Chinese [7]. Given the amount of scientific work undertaken in China, this represents an immense loss of data for systematic reviews, reducing their value to practitioners of medicine. The insufficiency of current machine translation tools for this task is described in [1]. With the increasing use of Clinical Study Reports as information sources in Systematic Reviews, search for pertinent information within single, long documents is becoming crucial.

An aspect of medical search that has not been taken into account yet are different search approaches by people in dif-

ferent health care provision roles, or physicians with different specialities. In particular, general practitioners have very different information requirements to specialists [6], being more interested in secondary research literature, for example. Here there is also a multilingual aspect in search by professionals, as clinicians are often interested in obtaining relevant clinical guidelines (usually written in the national language) as well as international literature for more difficult cases. Search triggered in an automated way based on the contents of a patient record is also of interest.

Finally, there is also place for multilingual search within patient records. In Europe, there is increasing interest in establishing international cohorts for rare diseases in order to have large enough cohorts. For this, search in medical records in multiple languages is required.

5. ACKNOWLEDGMENTS

Thank you to members of the KConnect consortium for discussion on these ideas. This paper was funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 644753 (KConnect).

6. REFERENCES

- [1] E. Balk, M. Chung, N. Hadar, K. Patel, W. Yu, T. A. Trikalinos, and L. Chang. Accuracy of data extraction of non-English language trials with Google Translate. Technical Report 12EHC056EF, Agency for Healthcare Research & Quality, Rockville, MD, 2012.
- [2] D. O. Case. *Looking for Information A Survey of Research on Information Seeking, Needs and Behavior*. Emerald Group, 3rd edition, 2012.
- [3] L. Goeuriot, L. Kelly, H. Suominen, L. Hanlen, A. Névéol, C. Grouin, J. Palotti, and G. Zuccon. Overview of the CLEF eHealth Evaluation Lab 2015. volume 9283, pages 429–443. Springer LNCS, 2015.
- [4] A. Hanbury and M. Lupu. Toward a model of domain-specific search. In *Proc. 10th Conf. on Open Research Areas in IR*, pages 33–36, 2013.
- [5] W. R. Hersh. *Information Retrieval: A Health and Biomedical Perspective*. Springer, 2009.
- [6] M. Kritz, M. Gschwandtner, V. Stefanov, A. Hanbury, and M. Samwald. Utilization and perceived problems of online medical resources and search tools among different groups of european physicians. *J Med Internet Res*, 15(6):e122, 2013.
- [7] M. J. Prince. Dementia in China: east-west collaboration bears fruit. *The Lancet*, 381(9882), 2013.
- [8] K. Roberts, M. S. Simpson, E. M. Voorhees, and W. R. Hersh. Overview of the TREC 2015 Clinical Decision Support Track. In *Proc. of TREC*, 2015.
- [9] B. Röhrig, J.-B. du Prel, D. Wachtlin, and M. Blettner. Types of study in medical research. *Deutsches Arzteblatt Int.*, 106(15):262–268, 2009.
- [10] The Cochrane Library Oversight Committee. Measuring the performance of the Cochrane Library. *Cochrane Database of Systematic Reviews*, 11 2012.
- [11] E. Voorhees and W. Hersh. Overview of the TREC 2012 Medical Records Track. In *Proc. of TREC*, 2012.
- [12] G. Zuccon, B. Koopman, and J. Palotti. Diagnose this if you can. In *Proc. 37th European Conference on Information Retrieval (ECIR)*, volume 9022, pages 562–567. Springer LNCS, 2015.