Journal of Biomedical Informatics 67 (2017) 1–10

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Development and empirical user-centered evaluation of semantically-based query recommendation for an electronic health record search engine

David A. Hanauer [a,b], Danny T.Y. Wu [b,a], Lei Yang [b], Qiaozhu Mei [b,c], Katherine B. Murkowski-Steffy [d], V.G. Vinod Vydiswaran [e,b], Kai Zheng [d,b,*,1]

[a] Department of Pediatrics, University of Michigan Medical School, 5312 CC, SPC 5940, 1500 East Medical Center Drive, Ann Arbor, MI 48109, USA
[b] School of Information, University of Michigan, 105 South State Street, Ann Arbor, MI 48109, USA
[c] Department of Electrical Engineering and Computer Science, University of Michigan, 2260 Hayward Street, Ann Arbor, MI 48109, USA
[d] Department of Health Management and Policy, School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109, USA
[e] Department of Learning Health Sciences, University of Michigan Medical School, 1111 East Catherine Street, Ann Arbor, MI 48109, USA

ARTICLE INFO

Article history:
Received 26 August 2016
Revised 21 December 2016
Accepted 23 January 2017
Available online 25 January 2017

Keywords:
Information retrieval systems (L01.700.508.300)
Search engine (L01.470.875)
Electronic health records (E05.318.308.940.968.625.500)
Unified Medical Language System (L01.453.245.945.800)
Query recommendation
Query expansion

ABSTRACT

Objective: The utility of biomedical information retrieval environments can be severely limited when users lack expertise in constructing effective search queries. To address this issue, we developed a computer-based query recommendation algorithm that suggests semantically interchangeable terms based on an initial user-entered query. In this study, we assessed the value of this approach, which has broad applicability in biomedical information retrieval, by demonstrating its application as part of a search engine that facilitates retrieval of information from electronic health records (EHRs).
Materials and Methods: The query recommendation algorithm utilizes MetaMap to identify medical concepts from search queries and indexed EHR documents. Synonym variants from UMLS are used to expand the concepts along with a synonym set curated from historical EHR search logs. The empirical study involved 33 clinicians and staff who evaluated the system through a set of simulated EHR search tasks. User acceptance was assessed using the widely used technology acceptance model.
Results: The search engine's performance was rated consistently higher with the query recommendation feature turned *on* vs. *off*. The relevance of computer-recommended search terms was also rated high, and in most cases the participants had not thought of these terms on their own. The questions on perceived usefulness and perceived ease of use received overwhelmingly positive responses. A vast majority of the participants wanted the query recommendation feature to be available to assist in their day-to-day EHR search tasks.
Discussion and Conclusion: Challenges persist for users to construct effective search queries when retrieving information from biomedical documents including those from EHRs. This study demonstrates that semantically-based query recommendation is a viable solution to addressing this challenge.

1. Introduction

The widespread adoption of electronic health records (EHRs) in the U.S. and around the globe has led to the rapid growth of large repositories of unstructured, free-text clinical documents [1], resulting in a 'patient information explosion' [2]. Fortunately, extracting information locked in these documents can be aided with technologies such as medical information retrieval systems—or '*Google-like*' search engines—although few advanced search engines have thus far been developed specifically for patient records [3–9].

Retrieving information from such clinical documents is a difficult task due in part to the fact that clinicians may record the same medical concept in a variety of interchangeable forms (e.g., "Tylenol" vs. "acetaminophen"), in addition to the popular use of acronyms and abbreviations [10,11]. Further, healthcare professionals often lack proper training and skills to formulate effective

* Corresponding author at: M3531 SPH II, 1415 Washington Heights, Ann Arbor, MI 48109, USA.

E-mail addresses: hanauer@med.umich.edu (D.A. Hanauer), tzuyu@umich.edu (D.T.Y. Wu), yangle@umich.edu (L. Yang), qmei@umich.edu (Q. Mei), kbmurko@umich.edu (K.B. Murkowski-Steffy), vgvinodv@med.umich.edu (V.G.V. Vydiswaran), kzheng@umich.edu, kai.zheng@uci.edu (K. Zheng).

[1] Present address: 6095 Donald Bren Hall, Irvine, CA 92697, USA.

http://dx.doi.org/10.1016/j.jbi.2017.01.013
1532-0464/Published by Elsevier Inc.

(i.e., pertinent and inclusive) search queries [12–14]. For example, when searching for "breast cancer," few healthcare professionals would be able to compile a reasonably inclusive list of related search terms such as "breast ca," "BCA," "breast tumor," and "breast carcinoma." All of these are legitimate variations for describing this concept in patient records.

Computer-based query recommendation, also known as automatic query expansion [15–17], has proven to be an effective solution to assisting non-expert users in achieving better queries to improve both quality and efficiency of information retrieval tasks. Indeed, query recommendation has been commonly used by general-purpose web search engines to enhance search performance. For example, when a user enters "MI pain," popular search engines (e.g., Google, Bing) are intelligent enough to expand the acronym "MI" to include terms such as "myocardial infarction," or "Michigan" depending on the context, to help users retrieve the most desirable web pages. Similarly, the term "pain" could be expanded to a number of other related concepts such as "tenderness" and "discomfort". In healthcare, research has also shown that query recommendation is effective in enhancing search experience not only for consumers (i.e., patients, families, and the general public) [18–20], but also for professionals such as clinicians and health science researchers [21–23]. However, to date, studies conducted in professional settings have mainly focused on information retrieval from biomedical literature databases such as PubMed, rather than patient records.

In 2005, the University of Michigan Health System (UMHS) implemented a homegrown EHR search engine available for authorized users, known as EMERSE (http://project-emerse.org) [5]. With a user base of more than 1600, the system has played an instrumental role in supporting a variety of information retrieval tasks in areas such as clinical care, quality assurance, billing, and clinical and translational research [5,24]. Through several user behavior studies, we recognized that the utility of the system might have been severely limited due to users' inability to construct effective search queries [25,26]. As query recommendation has been shown to be advantageous in other settings, in this study we sought to develop this feature for EMERSE and conduct a user experiment to empirically evaluate its potential benefits in the context of retrieving information from EHRs. The U.S. National Library of Medicine (NLM)'s *Computational Thinking* program supported this work.

2. Background

Biomedical information retrieval systems are designed to provide users the capability of retrieving information by entering combinations of keywords, Boolean operators, and search queries in more advanced forms such as regular expressions [3–9]. EHR search engines provide a useful means for supporting tasks related to direct patient care (e.g., to locate the mention of a particular health event in the earlier care episodes of a patient); operational tasks that require routine chart auditing, such as quality improvement and billing; and research tasks that require chart review, such as patient eligibility screening, cohort identification, and phenotype characterization. For example, at our institution, EMERSE has been routinely used to perform data abstraction for submission to the Commission on Cancer Certified Tumor Registry, and by the billing team as a computer-assisted coding tool to improve the efficiency and inclusiveness of billing code assignments. EMERSE has also been used by numerous research groups in over 1110 clinical and translational studies, resulting in at least 134 peer-reviewed publications to date (full list at http://project-emerse.org) [e.g., 27–31].

Through several prior studies of EMERSE [25,26], we discovered that many end users of the system did not necessarily possess comprehensive clinical knowledge of the medical concepts they frequently searched for, e.g., research coordinators and student research assistants who were not clinically trained. In addition, even healthcare professionals with extensive clinical experience might lack the ability, or time and patience, to create a set of search terms that is 'minimally necessary' to ensure reasonably inclusive search results. These observations motivated the present research.

3. Materials and methods

3.1. The query recommendation algorithm

Development of the query recommendation algorithm evaluated in this study was informed by previous work in biomedical literature retrieval and information extraction from clinical text [21,23,32–37]. Fig. 1 illustrates the main building blocks of the algorithm and the typical information flow. First, the algorithm utilizes MetaMap to identify Metathesaurus concepts from target EHR documents. Then, the algorithm uses *Lemur*, a popular open-source search engine (http://www.lemurproject.org), to index the resulting concepts along with the EHR documents.

Because it has been shown that not all semantic types are crucial for information retrieval tasks with clinical text [34], the algorithm only retains 61 Unified Medical Language System (UMLS) Semantic Types, such as symptoms and disorders, to better ensure that only medically relevant concepts would be analyzed and expanded. For example, if a user entered a query "patients with heart disease," the concept "patients" would be dropped. Appendix *A* lists the 61 semantic types that are included, as well as the 72 other semantic types that are excluded.

In addition to UMLS, the algorithm also utilizes an empiric synonym set (ESS) that, at the time of this study, contained about 35,000 terms representing 8500 medical concepts and their synonyms and spelling variations. ESS is a heuristic synonym collection that we have accumulated over time from multiple sources including the search logs of EMERSE and an active working list of acronym expansions maintained by the medical coding team at UMHS. Appendix *B* displays the overlap of text strings from UMLS and from ESS related to the concept "hearing impairment," demonstrating that ESS provides additional synonyms and interchangeable forms that are not found in UMLS, but are commonly used in clinicians' clinical documentation.

Next, the algorithm applies MetaMap to process user-entered search queries both to extract relevant search terms and to identify the underlying medical concepts. These search terms, medical concepts, and expanded terms based on UMLS and ESS are then reconciled (e.g., duplicates removed) to produce search term recommendations. The recommended search terms are then used to query the indexed EHR documents. To rank the documents retrieved, the algorithm uses the *Pivoted Normalization* retrieval function [38], a classical measure of relevance of documents based on the vector space model, defined as follows.

Given a query q, the relevance score of a document d is expressed as [38]:

$$score(d) = \sum_{t \in q \cap d} \frac{1 + ln(1 + ln(c(t,d)))}{(1-s) + s\frac{|d|}{avdl}} \cdot c(t,q) \cdot ln\frac{N+1}{df(t)},$$

where $c(t, d)$ and $c(t, q)$ are the number of times that a search term t appears in the document and in the query, respectively; in our case $c(t, q) = 1$ for all terms t, which was done to counteract cases where a search term might be expanded to many additional terms, but should not be considered any more clinically important than the rest of search terms in the original query that do not have additional expansion concepts; $df(t)$ is the number of documents in the index that contain the search term ('document frequency'); N is the size of