Title

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

Authors

Hanauer, David
Wu, Danny
Yang, Lei
et al.

Publication Date

2017-03-01

DOI

10.1016/j.jbi.2017.01.013

Peer reviewed

Published in final edited form as:

J Biomed Inform. 2017 March ; 67: 1–10. doi:10.1016/j.jbi.2017.01.013.

Development and Empirical User-Centered Evaluation of Semantically-based Query Recommendation for an Electronic Health Record Search Engine

David A. Hanauer, MD, MS[a,b], **Danny T.Y. Wu, MS**[b,a], **Lei Yang, MS**[b], **Qiaozhu Mei, PhD**[b,c], **Katherine B. Murkowski-Steffy, MPH**[d], **VGVinod Vydiswaran, PhD**[e,b], and **Kai Zheng, PhD**[d,b]

[a]Department of Pediatrics, University of Michigan Medical School, Ann Arbor, Michigan, USA; 5312 CC, SPC 5940, 1500 East Medical Center Drive, Ann Arbor, MI 48109, USA

[b]School of Information, University of Michigan, Ann Arbor, Michigan, USA; 105 South State Street, Ann Arbor, MI 48109, USA

[c]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan, USA; 2260 Hayward Street, Ann Arbor, MI 48109, USA

[d]Department of Health Management and Policy, School of Public Health, Ann Arbor, Michigan, USA; 1415 Washington Heights, Ann Arbor, MI 48109, USA

[e]Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, Michigan, USA; 1111 East Catherine Street, Ann Arbor, MI 48109, USA

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

Address correspondence to: Kai Zheng PhD[1], M3531 SPH II, 1415 Washington Heights, Ann Arbor, MI 48109, Phone: +1 (734) 936-6331, kzheng@umich.edu.

[1]Present address: 5228 Donald Bren Hall, Irvine, CA 92697, USA; Phone: (949) 824-6920, kai.zheng@uci.edu.

Competing Interests Statement: The authors have no competing interests to declare.

Contributorship Statement: DAH, QM, and KZ designed the study. LY and DTYW implemented the query expansion feature. KMS managed the project and organized all user experiments. VGVV assisted DAH and KZ in analyzing the empirical evaluation data. All authors reviewed and approved this manuscript.

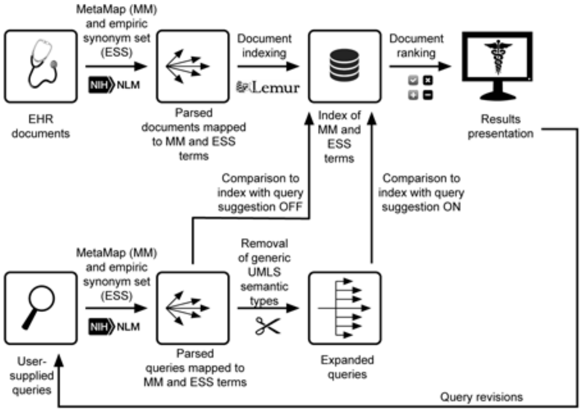
Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

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.

Graphical abstract


MetaMap (MM) and empiric synonym set (ESS)
NIH NLM
Document indexing
Lemur
Document ranking
EHR documents
Parsed documents mapped to MM and ESS terms
Index of MM and ESS terms
Results presentation
Comparison to index with query suggestion OFF
Comparison to index with query suggestion ON
MetaMap (MM) and empiric synonym set (ESS)
NIH NLM
Removal of generic UMLS semantic types
User-supplied queries
Parsed queries mapped to MM and ESS terms
Expanded queries
Query revisions

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

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 (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 1,600, 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 1,110 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] Figure 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 8,500 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]

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

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 the index (number of documents indexed); |*d*| is the length of a document; *avdl* is the average length of all documents contained in the index; and *s* is a smoothing parameter, empirically set as 0.1 as a commonly used value in information retrieval (see [39] table 7).

After query expansion has occurred, the search engine is potentially able to identify many additional documents that did not contain the terms in the original, user-entered search query. Further, when ranking documents, the synonyms of a search term are treated as a single term (i.e., a ‘concept term’). The term frequency of a ‘concept term’ is calculated as the sum of the term frequencies of all synonyms present in the document. The document frequency of a ‘concept term’ is calculated as the number of documents containing at least one of the synonyms. Retrieved documents are then re-ranked based on the new term frequencies and inverse document frequencies. Additional technical details about the implementation of the algorithm are reported elsewhere. [40]

3.2. Test Environment and Test Corpus

In this study, the semantically-based query recommendation feature was deployed in a test environment that we referred to as EHR search engine (EHR-SE). EHR-SE resembled a simplified version of EMERSE, and was customized to support the user experiment of this study (described in the next section).

Shown in Figures 2 and 3, EHR-SE is a web-based application with a look-and-feel similar to that of other popular search engines. A key function of EHR-SE is the provision of a toggle switch that turns the query recommendation feature off and on. When this switch is turned off, the search query that the user entered in the search box will be submitted to the search engine verbatim with no modifications (illustrated in Figure 2). When this switch is turned on, computer-recommended search terms will be presented to the user on the screen (Figure 3), and will be automatically added to the search query.

To ease patient privacy protection concerns, in this study, we obtained a data sharing agreement to use a test corpus that was originally made for the 2011 Text REtrieval

Conference (TREC) Medical Records track evaluation.[41]This TREC corpus contains 95,702 de-identified, free-text clinical documents from 17,198 patient visits, covering a variety of document types including 47,524 radiology reports, 13,168 emergency department notes, and 12,184 history and physical notes.

3.3. User Experiment Design

3.3.1. Search Tasks—To assess the potential benefits of semantically-based query recommendation in the context of EHR search, we designed a user experiment wherein a group of test users conducted simulated search tasks in EHR-SE. In the experiment, the test users were asked to use the toggle switch to control the behavior of the system, and to compare the performance of the search engine with the query recommendation feature first turned off vs. with it subsequently turned on.

We developed five one-sentence clinical vignettes describing the objective of each of the search tasks (Table 1). These search tasks were carefully designed and the vignettes carefully worded to represent a range of clinical contexts and levels of difficulty, as has been done with other information retrieval studies.[42] These tasks and vignettes were first drafted by a clinician on the research team (DAH), pilot-tested with several EMERSE users and laypersons, and then finalized through research group discussion. For each task, we verified that the 2011 TREC corpus contained at least a handful of clinical documents that would be matched and retrieved.

The search tasks were prepared at three difficulty levels. The vignette of the low difficulty task (#2) provided direct hints regarding the search term to use (“DCIS”), which was also how this concept would likely appear in the target documents. In contrast, the vignette of the high difficulty task (#4), to identify *“herbal supplements for the purposes of weight loss*,” was intentionally worded to be broad and vague. The vignettes of the other three tasks provided some hints for candidate search terms. However, turning them into effective search queries would likely require additional translation and expansion, e.g., “car accident” → “motor vehicle accident;” “smokers” → “tobacco use;” and “enlarged spleen” → “splenomegaly.”

Following these five ‘standard’ search tasks, participants of the user experiment were also given the option to use two additional search scenarios of their choice, which could be based on their area(s) of expertise or their prior EHR search experience.

3.3.2. Participants and Participant Recruitment—All participants were ‘active’ users of EMERSE, defined as those who logged into the system at least three times in the prior year. This criterion was used to ensure that participants of this study would already be familiar with retrieving information from EHRs, formulating search queries, and reviewing clinical documents returned. Most EMERSE users are clinicians and administrative or research staff at UMHS, a large, tertiary care health system comprised of three hospitals and over 120 ambulatory clinics in Southeastern Michigan. The participants were recruited by email messages; each received a $50 gift card after completing the user experiment. Our initial recruitment target was 25.

3.3.3. Experiment Protocol—The user experiments were conducted in a classroom equipped with desktop computers. Participants were scheduled in groups of up to four, at various days and times over a course of one month, to accommodate their busy work schedules. At the beginning of each experiment, a brief introduction to the study was provided followed by a live demonstration of EHR-SE. All participant questions were answered before proceeding.

Participants then went through the simulated EHR search tasks by following a printed protocol (Appendix *C*). They were asked to carefully consider each vignette and then enter the search terms of their choice, with the query recommendation feature first turned *off*. After reviewing the ranked documents retrieved, they could make as many rounds of modifications as they wished to refine the initial query. Once they were satisfied with the results, they were instructed to turn *on* the query recommendation feature and again review the updated, rank-ordered list of documents retrieved. The participants thus served as their own comparisons.

3.3.4. Evaluation Instrument—The printed protocol contained an evaluation instrument, shown in Table 2, for collecting user feedback after each search task. The first question (Q.A.1) asked the participants to compare the performance of the system with the query recommendation feature turned off vs. with it turned on. The second and third questions (Q.A.2 and Q.A.3) solicited the relevance of the search terms recommended by the computer, and whether the participants would be able to come up with these terms on their own without the computer's assistance. Following these three questions, a free-text area was provided to collect additional feedback regarding the query recommendation feature in an open-ended format.

After completing all search tasks, the participants were presented with three summative evaluation questions soliciting their perceived usefulness (Q.B.1) and perceived ease of use (Q.B.2) of the query recommendation feature, as well as whether or not they would like to see this feature adopted in EMERSE (Q.B.3). These three questions are based on the widely used technology acceptance model postulating that perceived usefulness and perceived ease of use are the two most important antecedents to people's technology acceptance behavior.[43-45]Another open-ended question was provided at the end of the evaluation instrument allowing the participants to describe their overall experience with the feature and to offer suggestions for improvement.

All quantitative questions included in the evaluation instrument were assessed on a five-point scale; actual wording varied according to the context of the question (see Table 2). No limitation was placed on how much time a participant could use to complete the experiment. The research protocol of this study was reviewed and approved by the University of Michigan Health Sciences and Behavioral Sciences Institutional Review Boards.

3.4. Data Analysis

The five-point quantitative responses were coded as an integer variable from -2 to +2, with 0 representing the middle or neutral point. All statistical analyses were performed in Stata 13 (StataCorp LP, College Station, Texas, USA). The open-ended feedback was qualitatively analyzed using inductive open coding and constant comparison to identify salient and recurring themes.[46,47]

4. Results

4.1. Participants

There were 214 EMERSE users who met the eligibility criteria; however, 17 of them were no longer with UMHS when this study was conducted. Our recruitment emails were thus sent to 197 prospective participants.

After the first round of recruitment emails, 33 individuals responded, which exceeded our initial recruitment target of 25. Because the objective of this study was to collect end user feedback about the query recommendation feature, a larger sample would be more desirable. We therefore decided to include all of them in the study (the IRB application was amended accordingly).

Characteristics of these 33 participants are summarized in Table 3. The majority of them (82%) were female; nearly half were coordinators or managers of clinical and translational studies at UMHS. These participants also represented a wide range of clinical and operational areas.

4.2 User-Entered Queries

Overall, users entered 1,526 queries across the five standard search scenarios and an additional 572 queries for the two optional scenarios. The average number of terms contained in a user-entered query, when the query suggestion feature was turned off, was 2.5; whereas the number of terms included in the final query with the expansion feature turned on was 30.5. Few user-entered queries included Boolean operators. For example, only 30 (1.4%) included the AND operator (e.g., “height and weight”), and none of them included the OR operator. An example of a user-entered query for scenario #5 was “mono, enlarged spleen”, which was expanded to a 19-term query with the query suggestion feature turned on: “enlarged spleen, large spleen, hsm, spleen enlargement, splenic enlargement, hepatosplenomegaly, enlargement spleen, splenomegaly, mono, monocyte, ebv, epstein, epstein-barr virus, monos, infectious mono, epstein-barr, mononucleosis, monocytic, ebstein.”

4.3. Quantitative Evaluation Results

Table 4 reports the effectiveness evaluation results. Across the five ‘standard’ tasks, the search engine's performance was rated consistently higher when the query recommendation feature was turned on, compared to when it was turned off (average score: 1.25, on a -2 to +2 scale). The performance gain was deemed higher for the medium-difficulty tasks (#1, #3, and #5; average score: 1.40); and more moderate for the easy (#2; score: 0.94) and the difficult task (#4; score: 1.09).

Similarly, the participants judged computer-recommended search terms to be relevant to highly relevant (average score: 1.52). They also reported that, on average, without the computer's assistance, they would only be able to come up with some, but not most, of the alternative search terms (average score: 0.012). The performance ratings (Q.A.1) are significantly correlated with the relevance scores (Q.A.2) (0.53, $p < 0.01$), and inversely with the participants' ability to think of the additional search terms (Q.A.3) (-0.46, $p < 0.01$). There is no statistically significant correlation between the relevance scores and participants' ability to think of additional terms (i.e. responses to Q.A.2 and Q.A.3).

The results based on the two user-initiated search tasks are similar except that in these scenarios, the participants were more confident in their ability to think of additional search terms on their own without the need for the computer-recommended terms. Appendix *D* provides a list of all user self-initiated search scenarios created during the evaluation experiment.

Table 5 shows the results from the three summative evaluation questions. Perceived usefulness and perceived ease of use of the query recommendation feature were rated overwhelmingly positive, suggesting that both the functionality and the usability of the feature were well received among the study participants. For the last question, "intention to adopt," 28 out of the 33 participants (84.8%) said that they would absolutely want to see the feature adopted in EMERSE, the EHR search engine that they had been routinely using.

4.4. Qualitative Analysis Results of Open-Ended Feedback

Seven salient and recurring themes emerged from the qualitative analysis of the open-ended feedback. They are reported in Table 6.

The most common theme was that the study participants liked the query recommendation feature and found the computer-suggested search terms to be very valuable (Theme *A*). This is consistent with the quantitative feedback received. They further described the time-saving benefits of the feature, particularly on how computer-based query recommendation could relieve their burden of manually compiling a reasonably inclusive list of relevant search terms (Theme *B*). Many study participants also appreciated the user interface design of EHR-SE that highlighted related or semantically interchangeable search terms using the same color (Theme *C*). This can be seen in Figure 3, where "hearing loss," "deaf," and "hard of hearing" are all highlighted in green.

Through open-ended feedback, the participants of the study also provided important improvement suggestions. First, many of them shared the worry that fully automated query recommendation might increase the odds of producing false positive results (Theme *D*). They preferred having more nuanced control over what terms to include, and what not to, so they could fine-tune a query according to the true intention of the search (Theme *E*). Further, some participants were able to think of additional search terms that were not on the computer-expanded list (Theme *F*). While they could manually add them each time, having a mechanism to permanently include these terms as part of the standard term recommendation was seen as a desirable feature to benefit future searches. For example, a participant specifically recommended adding standardized codes from ICD-9 and DSM-IV because some clinicians often include such codes in their clinical documentation. Finally, some participants would like to have the capability of assigning different weights to different search terms (Theme *G*); for example, when searching for "ductal carcinoma in situ," they preferred that the system would allow for "DCIS" to be weighted higher over "breast cancer" when retrieved results are ranked.

5. Discussion

The results of this study demonstrate that computer-based query recommendation has great potential to improve the information retrieval performance of EHR search engines and time efficiency for end users. Query recommendation for EHRs may thus help 'level the playing field,' considering that many users whose job involves routinely searching in EHRs may not have extensive clinical backgrounds. Even for search 'experts,' semantically-based query recommendation can still be beneficial to ensure research results are inclusive, and to relieve the burden of manually compiling comprehensive lists of search terms. As a participant of this study pointed out, without the feature, she or he would "*often spend much of my time googling similar words*."

It is worth noting that the effectiveness of semantically-based query recommendation could differ considerably according to the level of difficulty of a search task. For easy tasks wherein search terms are very specific and have few variant forms, the achievable improvement of semantically-based query recommendation may be marginal, and the likelihood of introducing irrelevant terms may increase, which can result in "query drift". [48] For difficult scenarios wherein search objectives are only vaguely defined, the usefulness of this feature may also be limited. This is not surprising because the efficacy of computer-based query recommendation, after all, builds upon the quality of the seed search terms that the user manually entered. For example, identifying "*herbal supplements for the purposes of weight loss*" (task #3) is challenging because coming up with the right set of initial search terms for "herbal supplements," as well as for "weight loss," is a nontrivial task. In addition, determining the causal relationship between these two concepts from the retrieved EHR documents can be very difficult. As a participant stated, "*I'm seeing a lot of people that are taking the supplement because they lost weight, not the other way around*." Another user stated, "*Most of what I was able to come up with didn't address the link between the herbal supplement and weight loss—just that the terms both appeared in the note*." Future research is therefore also needed to create computational tools that would aid users in the ideation of a search and in sense-making of the results retrieved.

The small number of queries with Boolean operators (1.4% with AND, 0% with OR) was lower than rates found in prior studies of users of general-purpose search engines (e.g., 6% overall use in a study involving Excite.com). [49] It was also lower than what was found among a large number of queries entered into the biomedical literature search tool, PubMed, which had 11% of queries with at least one Boolean. [50] The reason for this observation might be because the participants of this study were already familiar with the existing EHR search engine, EMERSE, that automatically included OR between terms when searching across a known set of patients.

This study has several limitations. First, because there is no true gold standard for determining how well the retrieved clinical documents meet a user's search objective(s), we relied on the subjective judgment of each study participant to assess the quality of the search results. That said, the participants of this study all had prior experience of routinely searching in EHRs. We therefore believe that their collective wisdom provided a reliable source of judgment on the performance of the search engine. Second, the quality of computer-based query recommendation is sensitive to the ability of MetaMap in identifying relevant medical concepts from user-submitted queries and from the EHR documents indexed. While prior research has found that MetaMap performs superiorly over baseline information retrieval systems,[51] there have also been critiques showing the potential limitations of MetaMap when applied to clinical documents.[52,53] Third, our study focused specifically on real end users' perceived value of automatic query expansion in the context of an EHR search engine. We used a basic, but standard, document ranking algorithm in our experiments, which we viewed as a strength as it reduced the variability in the before-after comparison. Future studies may consider using a larger sample of participants to assess the impact of various ranking algorithms on the perceived retrieval quality. Fourth, within this study we did not perform a thorough analysis of the search terms that the study participants entered during the user experiment. While we could confirm that expanded queries had more terms than the original queries, and that users were more satisfied with the performance of the system-expanded queries, we believe that further examination of these terms may provide additional insights into the struggles that the participants had when formulating search queries. We have recently completed this analysis and will report the findings in a follow-up publication.

6. Conclusions

In this study, we implemented a semantically-based query recommendation feature in an EHR search engine. We then empirically evaluated its performance by having 33 experienced users perform a set of simulated EHR search tasks with the feature turned off vs. with it turned on. The results show that semantically-based query recommendation has great potential to improve both the information retrieval performance of the search engine and time efficiency for end users. These findings may be generalized to search engines for other types of healthcare text, such as biomedical literature, forum messages, and insurance claims. We therefore encourage other EHR search engines, and other healthcare information retrieval systems in general, to consider incorporating this feature.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to the 33 clinicians and administrative or research staff from the University of Michigan Health System who participated in the study. We would also like to thank the University of Pittsburgh for making the 2011 TREC dataset available for use in this study.

Funding Statement: This work was supported by the U.S. National Library of Medicine Computation Thinking Award grant number HHSN276201000032C.

References

1. Adler-Milstein J, DesRoches CM, Furukawa MF, et al. More Than Half of US Hospitals Have At Least A Basic EHR, But Stage 2 Criteria Remain Challenging For Most. Health Aff (Millwood). 2014; 33(9):1664–1671. [PubMed: 25104826]
2. Berner ES, Moss J. Informatics challenges for the impending patient information explosion. J Am Med Inform Assoc. 2005; 12(6):614–617. [PubMed: 16049224]
3. Biron P, Metzger MH, Pezet C, Sebban C, Barthuet E, Durand T. An information retrieval system for computerized patient records in the context of a daily hospital practice: the example of the Leon Berard Cancer Center (France). Appl Clin Inform. 2014; 5(1):191–205. [PubMed: 24734133]
4. Gregg W, Jirjis J, Lorenzi NM, Giuse D. StarTracker: an integrated, web-based clinical search engine. AMIA Annu Symp Proc. 2003; 855
5. Hanauer DA, Mei Q, Law J, Khanna R, Zheng K. Supporting information retrieval from electronic health records: A report of University of Michigan's nine-year experience in developing and using the Electronic Medical Record Search Engine (EMERSE). J Biomed Inform. 2015; 55:290–300. [PubMed: 25979153]
6. Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE--An integrated standards-based translational research informatics platform. AMIA Annu Symp Proc. 2009; 2009:391–395. [PubMed: 20351886]
7. Natarajan K, Stein D, Jain S, Elhadad N. An analysis of clinical queries in an electronic health record search utility. Int J Med Inform. 2010; 79(7):515–522. [PubMed: 20418155]
8. Tawfik AA, Kochendorfer KM, Saparova D, Al Ghenaimi S, Moore JL. “I Don't Have Time to Dig Back Through This”: The Role of Semantic Search in Supporting Physician Information Seeking in an Electronic Health Record”. Performance Improvement Quarterly. 2014; 26(4):75–91.
9. Zalis M, Harris M. Advanced search of the electronic medical record: augmenting safety and efficiency in radiology. J Am Coll Radiol. 2010; 7(8):625–633. [PubMed: 20678732]
10. Barrows RC Jr, Busuioc M, Friedman C. Limited parsing of notational text visit notes: ad-hoc vs. NLP approaches. Proc AMIA Symp. 2000:51–55. [PubMed: 11079843]
11. Edinger T, Cohen AM, Bedrick S, Ambert K, Hersh W. Barriers to retrieving patient information from electronic health record data: failure analysis from the TREC Medical Records Track. AMIA Annu Symp Proc. 2012; 2012:180–188. [PubMed: 23304287]
12. Davies K, Harrison J. The information-seeking behaviour of doctors: a review of the evidence. Health Info Libr J. 2007; 24(2):78–94. [PubMed: 17584211]
13. McKibbon KA, Haynes RB, Dilks CJ, et al. How good are clinical MEDLINE searches? A comparative study of clinical end-user and librarian searches. Comput Biomed Res. 1990; 23(6): 583–593. [PubMed: 2276266]
14. Wachtel RE, Dexter F. Difficulties and challenges associated with literature searches in operating room management, complete with recommendations. Anesth Analg. 2013; 117(6):1460–1479. [PubMed: 24257396]
15. Balfe, E., Smyth, B. Proceedings of the 16th European Conference on Artificial Intelligence. Valencia, Spain: 2004. Improving Web search through collaborative query recommendation.
16. Baraglia R, Cacheda F, Carneiro V, et al. Search shortcuts. 2009:77.
17. Broccolo D, Marcon L, Nardini FM, Perego R, Silvestri F. Generating suggestions for queries in the long tail with an inverted index. Information Processing & Management. 2012; 48(2):326–339.
18. Mu X, Lu K, Ryu H. Explicitly integrating MeSH thesaurus help into health information retrieval systems: An empirical user study. Information Processing & Management. 2014; 50(1):24–40.
19. Suominen H, Salanterä S, Velupillai S, et al. Overview of the ShARe/CLEF eHealth Evaluation Lab 2013. 2013; 8138:212–231.
20. Zeng QT, Crowell J, Plovnick RM, Kim E, Ngo L, Dibble E. Assisting consumer health information retrieval with query recommendations. J Am Med Inform Assoc. 2006; 13(1):80–90. [PubMed: 16221944]

21. Griffon N, Chebil W, Rollin L, et al. Performance evaluation of Unified Medical Language System(R)'s synonyms expansion to query PubMed. BMC Med Inform Decis Mak. 2012; 12:12. [PubMed: 22376010]

22. Leroy G, Xu J, Chung W, Eggers S, Chen H. An end user evaluation of query formulation and results review tools in three medical meta-search engines. Int J Med Inform. 2007; 76(11-12):780–789. [PubMed: 16996298]

23. Yoo S, Choi J. On the query reformulation technique for effective MEDLINE document retrieval. J Biomed Inform. 2010; 43(5):686–693. [PubMed: 20394839]

24. Seyfried L, Hanauer DA, Nease D, Albeiruti R, Kavanagh J, Kales HC. Enhanced identification of eligibility for depression research using an electronic medical record search engine. Int J Med Inform. 2009; 78(12):e13–18. [PubMed: 19560962]

25. Yang L, Mei Q, Zheng K, Hanauer DA. Query log analysis of an electronic health record search engine. AMIA Annu Symp Proc. 2011; 2011:915–924. [PubMed: 22195150]

26. Zheng K, Mei Q, Hanauer DA. Collaborative search in electronic health records. J Am Med Inform Assoc. 2011; 18(3):282–291. [PubMed: 21486887]

27. Braley TJ, Segal BM, Chervin RD. Sleep-disordered breathing in multiple sclerosis. Neurology. 2012; 79(9):929–936. [PubMed: 22895593]

28. DiMagno MJ, Spaete JP, Ballard DD, Wamsteker EJ, Saini SD. Risk models for post-endoscopic retrograde cholangiopancreatography pancreatitis (PEP): smoking and chronic liver disease are predictors of protection against PEP. Pancreas. 2013; 42(6):996–1003. [PubMed: 23532001]

29. Jensen KM, Cooke CR, Davis MM. Fidelity of Administrative Data When Researching Down Syndrome. Med Care. 2013

30. Paczesny S, Braun TM, Levine JE, et al. Elafin is a biomarker of graft-versus-host disease of the skin. Sci Transl Med. 2010; 2(13):13ra12.

31. Walter JK, Benneyworth BD, Housey M, Davis MM. The factors associated with high-quality communication for critically ill children. Pediatrics. 2013; 131(1):S90–95. [PubMed: 23457155]

32. Aronson AR, Rindflesch TC. Query expansion using the UMLS Metathesaurus. Proc AMIA Annu Fall Symp. 1997:485–489. [PubMed: 9357673]

33. Hersh W, Price S, Donohoe L. Assessing thesaurus-based query expansion using the UMLS Metathesaurus. Proc AMIA Symp. 2000:344–348. [PubMed: 11079902]

34. Liu Z, Chu WW. Knowledge-based query expansion to support scenario-specific retrieval of medical free text. Information Retrieval. 2007; 10(2):173–202.

35. Martinez D, Otegi A, Soroa A, Agirre E. Improving search over Electronic Health Records using UMLS-based query expansion through random walks. J Biomed Inform. 2014

36. Zeng QT, Redd D, Rindflesch T, Nebeker J. Synonym, topic model and predicate-based query expansion for retrieving clinical documents. AMIA Annu Symp Proc. 2012; 2012:1050–1059. [PubMed: 23304381]

37. Zhu D, Carterette B. Improving health records search using multiple query expansion collections. 2012:1–7.

38. Singhal A. Modern Information Retrieval: A Brief Overview. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering. 2001; 24(4):35–42.

39. Fang H, Tao T, Zhai C. A formal study of information retrieval heuristics. 2004:49.

40. Wu DT, Hanauer DA, Yang L, Zheng K, Mei Q. Towards intelligent and socially oriented query recommendation for electronic health records retrieval.

41. Voorhees EM, Hersh WR. Overview of the TREC 2012 Medical Records Track. Paper presented at: TREC2012.

42. Zhang Y, Wang P, Heaton A, Winkler H. Health information searching behavior in MedlinePlus and the impact of tasks. 2012:641.

43. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly. 1989; 13(3):319.

44. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform. 2010; 43(1):159–172. [PubMed: 19615467]

45. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: Toward a unified view. MIS quarterly. 2003:425–478.

46. Charmaz, K. Constructing grounded theory. Thousand Oaks, Calif: Sage Publications; 2006. London

47. Glaser, BG., Strauss, AL. The discovery of grounded theory : strategies for qualitative research. New York: Aldine de Gruyter; 1999.

48. Carpineto C, Romano G. A Survey of Automatic Query Expansion in Information Retrieval. ACM Computing Surveys. 2012; 44(1):1–50.

49. Jansen BJ, Spink A, Saracevic T. Real life, real users, and real needs: a study and analysis of user queries on the web. Information Processing & Management. 2000; 36(2):207–227.

50. Herskovic JR, Tanaka LY, Hersh W, Bernstam EV. A day in the life of PubMed: analysis of a typical day's query log. J Am Med Inform Assoc. 2007; 14(2):212–220. [PubMed: 17213501]

51. Cohen, KB., Christiansen, T., Hunter, LE. MetaMap is a superior baseline to a standard document retrieval engine for the task of finding patient cohorts in clinical free text. The Twentieth Text REtrieval Conference (TREC) Proceedings, 2011; 2011.

52. Friedlin J, Overhage M. An evaluation of the UMLS in representing corpus derived clinical concepts. AMIA Annu Symp Proc. 2011; 2011:435–444. [PubMed: 22195097]

53. Zuccon, G., Holloway, A., Koopman, B., Nguyen, A. Identify disorders in health records using Conditional Random Fields and Metamap: AEHRC at ShARe/CLEF 2013 eHealth Evaluation Lab Task 1. Proceedings of CLEF Workshop on Cross-Language Evaluation of Methods, Applications, and Resources for eHealth Document Analysis; Valencia, Spain. 2013.

Highlights

- A user-centered evaluation is conducted to assess the value of query recommendation
- The feature is designed to facilitate retrieval of information from EHRs
- The algorithm utilizes MetaMap to identify medical concepts
- The performance is rated consistently higher with query recommendation turned on
- Perceived usefulness and perceived ease of use scores are overwhelmingly positive

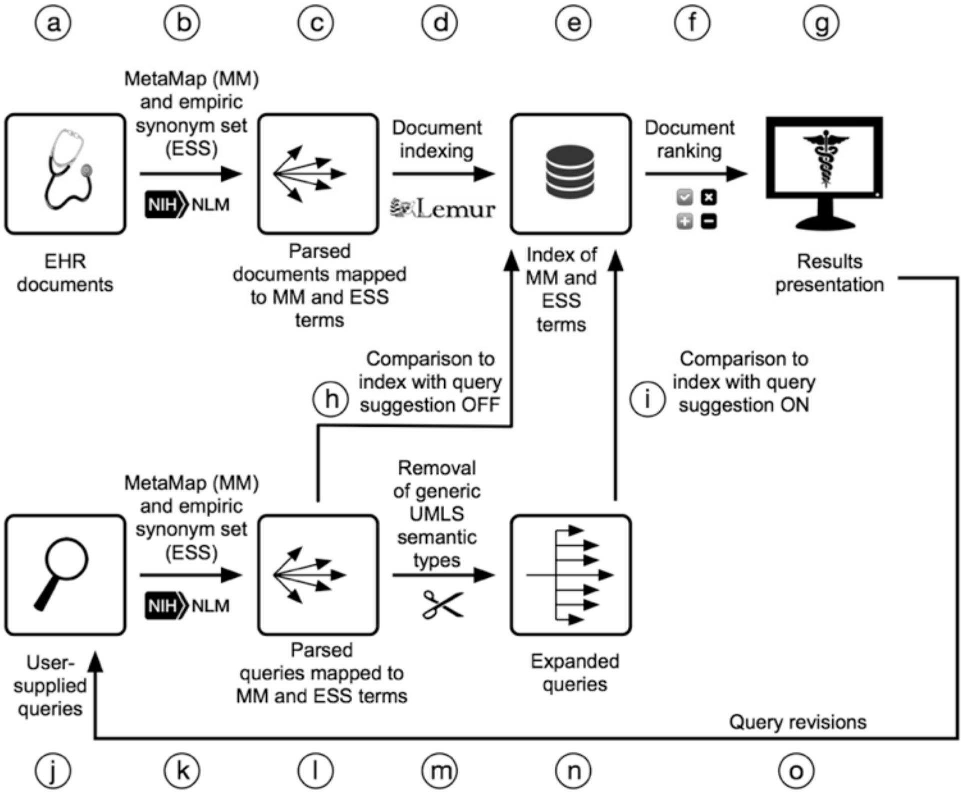
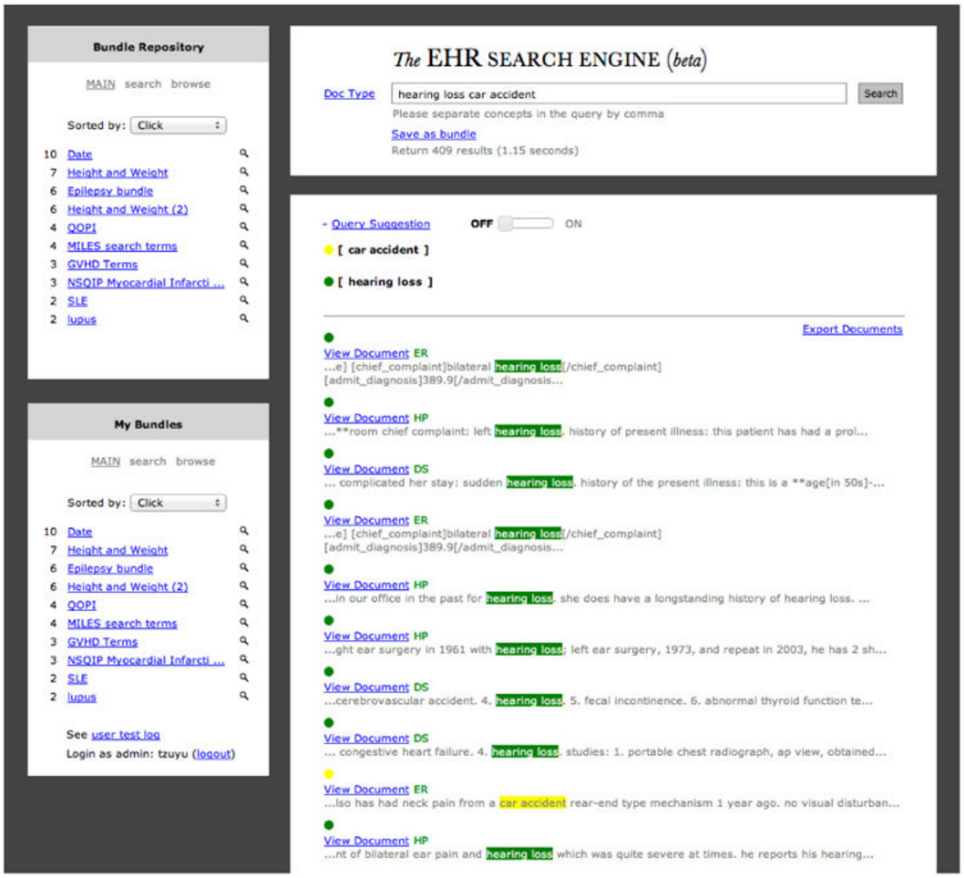

a
b
c
d
e
f
g
MetaMap (MM) and empiric synonym set (ESS)
NIH NLM
Document indexing
Lemur
Document ranking
EHR documents
Parsed documents mapped to MM and ESS terms
Index of MM and ESS terms
Results presentation
Comparison to index with query suggestion OFF
h
Comparison to index with query suggestion ON
i
MetaMap (MM) and empiric synonym set (ESS)
NIH NLM
Removal of generic UMLS semantic types
User-supplied queries
Parsed queries mapped to MM and ESS terms
Expanded queries
Query revisions
j
k
l
m
n
o

Figure 1. Components and Information Flow of the Query Recommendation Algorithm

EHR documents (a) are matched to UMLS terms using the output of the MetaMap natural language processing software as well as to terms in a locally developed empiric synonym set (b, c). These document terms are then indexed using Lemur (d, e). In a similar manner, with the query suggestion feature turned on, user-supplied queries (j) are also matched to UMLS and ESS terms (k, l), but generic UMLS semantic types are removed (m) to provide matching on the more relevant clinical concepts. This results in an expanded query (n) that is compared to the index (e) for subsequent document ranking (f) and presentation to the user (g). When the query suggestion feature was turned off (h), parsed queries (l) are compared directly with the parsed terms in the index (e) without any synonym expansion. When the query suggestion was turned on (i), the parsed terms were expanded using the concepts to which they mapped. In either mode (on or off) the user could revise their queries (o) and repeat their search as many times as desired.



Bundle Repository

MAIN search browse

Sorted by: Click

10	Date
7	Height and Weight
6	Epilepsy bundle
6	Height and Weight (2)
4	QOPI
4	MILES search terms
3	GVHD Terms
3	NSQIP Myocardial Infarcti ...
2	SLE
2	lupus

My Bundles

MAIN search browse

Sorted by: Click

10	Date
7	Height and Weight
6	Epilepsy bundle
6	Height and Weight (2)
4	QOPI
4	MILES search terms
3	GVHD Terms
3	NSQIP Myocardial Infarcti ...
2	SLE
2	lupus

See user test log

Login as admin: tzuyu (logout)

The EHR SEARCH ENGINE (*beta*)

Doc Type: hearing loss car accident

Search

Please separate concepts in the query by comma

Save as bundle

Return 409 results (1.15 seconds)

- Query Suggestion OFF ON

● [car accident]

● [hearing loss]

Export Documents

●
View Document ER
...e] [chief_complaint]bilateral hearing loss[/chief_complaint]
[admit_diagnosis]389.9[/admit_diagnosis...

●
View Document HP
...**room chief complaint: left hearing loss. history of present illness: this patient has had a prol...

●
View Document DS
... complicated her stay: sudden hearing loss. history of the present illness: this is a **age[in 50s]-...

●
View Document ER
...e] [chief_complaint]bilateral hearing loss[/chief_complaint]
[admit_diagnosis]389.9[/admit_diagnosis...

●
View Document HP
...in our office in the past for hearing loss. she does have a longstanding history of hearing loss. ...

●
View Document HP
...ght ear surgery in 1961 with hearing loss; left ear surgery, 1973, and repeat in 2003, he has 2 sh...

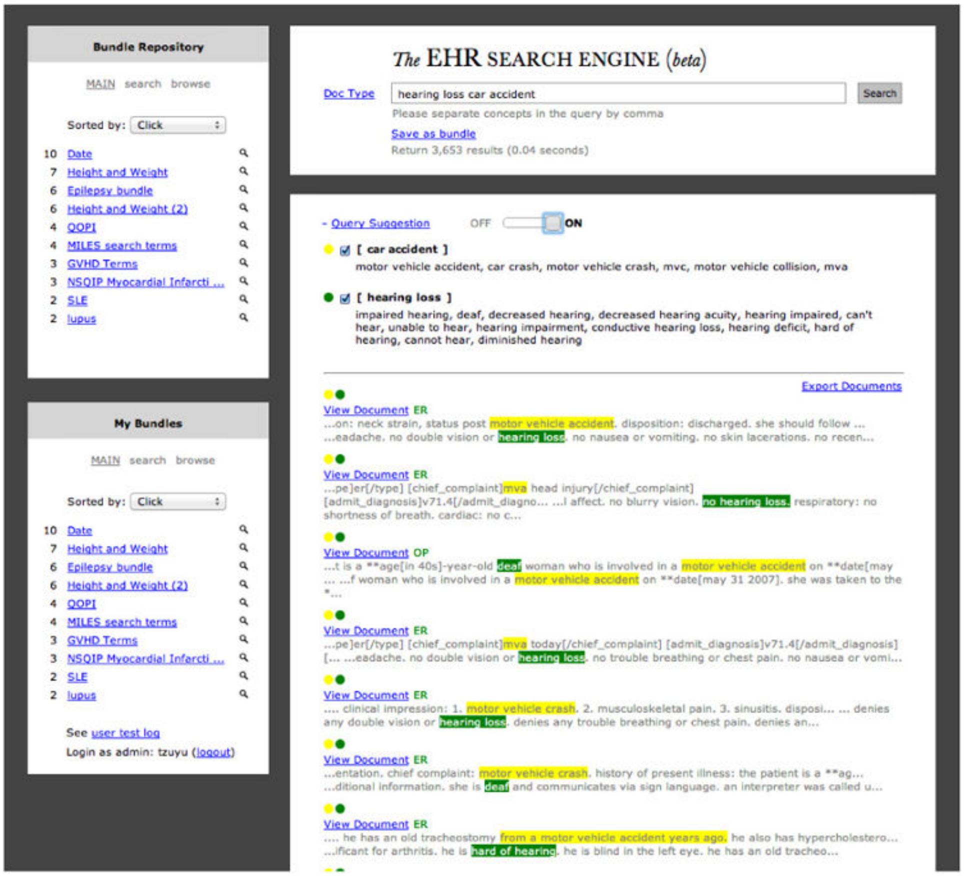
●
View Document DS
...cerebrovascular accident. 4. hearing loss. 5. fecal incontinence. 6. abnormal thyroid function te...

●
View Document DS
... congestive heart failure. 4. hearing loss. studies: 1. portable chest radiograph, ap view, obtained...

●
View Document ER
...lso has had neck pain from a car accident rear-end type mechanism 1 year ago. no visual disturban...

●
View Document HP
...nt of bilateral ear pain and hearing loss which was quite severe at times. he reports his hearing...

Figure 2. The Main User Interface of EHR-SE, with the Query Recommendation Feature Turned *Off*



hearing loss car accident

Please separate concepts in the query by comma

Return 3,653 results (0.04 seconds)

● ☑ [car accident]
motor vehicle accident, car crash, motor vehicle crash, mvc, motor vehicle collision, mva

● ☑ [hearing loss]
impaired hearing, deaf, decreased hearing, decreased hearing acuity, hearing impaired, can't hear, unable to hear, hearing impairment, conductive hearing loss, hearing deficit, hard of hearing, cannot hear, diminished hearing

●●
View Document ER
...on: neck strain, status post motor vehicle accident. disposition: discharged. she should follow ...
...eadache. no double vision or hearing loss. no nausea or vomiting. no skin lacerations. no recen...

●●
View Document ER
...pe]er[/type] [chief_complaint]mva head injury[/chief_complaint]
[admit_diagnosis]v71.4[/admit_diagno... ...l affect. no blurry vision. no hearing loss. respiratory: no shortness of breath. cardiac: no c...

●●
View Document OP
...t is a **age[in 40s]-year-old deaf woman who is involved in a motor vehicle accident on **date[may ... ...f woman who is involved in a motor vehicle accident on **date[may 31 2007]. she was taken to the *...

●●
View Document ER
...pe]er[/type] [chief_complaint]mva today[/chief_complaint] [admit_diagnosis]v71.4[/admit_diagnosis]
[... ...eadache. no double vision or hearing loss. no trouble breathing or chest pain. no nausea or vomi...

●●
View Document ER
.... clinical impression: 1. motor vehicle crash. 2. musculoskeletal pain. 3. sinusitis. disposi... ... denies any double vision or hearing loss. denies any trouble breathing or chest pain. denies an...

●●
View Document ER
...entation. chief complaint: motor vehicle crash. history of present illness: the patient is a **ag...
...ditional information. she is deaf and communicates via sign language. an interpreter was called u...

●●
View Document ER
.... he has an old tracheostomy from a motor vehicle accident years ago. he also has hypercholestero...
...ificant for arthritis. he is hard of hearing. he is blind in the left eye. he has an old tracheo...

Figure 3. The Main User Interface of EHR-SE, with the Query Recommendation Feature Turned *On*

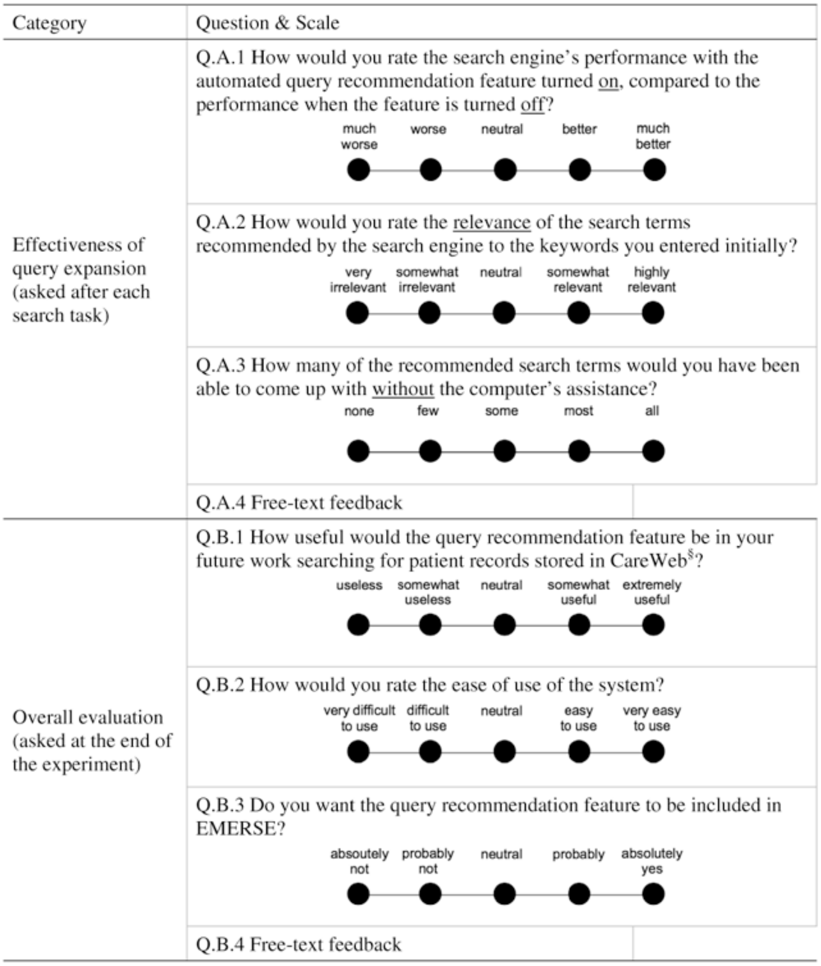
Table 1

The five simulated EHR search tasks and their vignettes

	Search task/vignette	Estimated level of difficulty
#1	"You are doing a research project in which you want to identify people who have had a concussive episode after being in a car accident."	*medium*
#2	"You are interested in identifying patients who have the non-invasive form of breast cancer known as DCIS."	*low*
#3	"Please try to identify patients who are smokers who have also been diagnosed with PTSD."	*medium*
#4	"You are interested in how many patients are taking herbal supplements for the purposes of weight loss."	*high*
#5	"Someone has asked you to determine if we have many patients diagnosed with mono who had an enlarged spleen."	*medium*
#6, 7	User self-initiated search scenarios	*variable*

Table 2

Evaluation instrument



Category	Question & Scale
Effectiveness of query expansion (asked after each search task)	Q.A.1 How would you rate the search engine's performance with the automated query recommendation feature turned on, compared to the performance when the feature is turned off? ● much worse ● worse ● neutral ● better ● much better
	Q.A.2 How would you rate the relevance of the search terms recommended by the search engine to the keywords you entered initially? ● very irrelevant ● somewhat irrelevant ● neutral ● somewhat relevant ● highly relevant
	Q.A.3 How many of the recommended search terms would you have been able to come up with without the computer's assistance? ● none ● few ● some ● most ● all
	Q.A.4 Free-text feedback
Overall evaluation (asked at the end of the experiment)	Q.B.1 How useful would the query recommendation feature be in your future work searching for patient records stored in CareWeb[§]? ● useless ● somewhat useless ● neutral ● somewhat useful ● extremely useful
	Q.B.2 How would you rate the ease of use of the system? ● very difficult to use ● difficult to use ● neutral ● easy to use ● very easy to use
	Q.B.3 Do you want the query recommendation feature to be included in EMERSE? ● absoutely not ● probably not ● neutral ● probably ● absolutely yes
	Q.B.4 Free-text feedback

[§] CareWeb was the EHR system used at UMHS when this study was conducted.

Table 3

Participant characteristics

Characteristic	***N* (%)**
Gender	
Female	27 (81.8)
Male	6 (18.2)
Clinical area[*]	
Internal medicine	9 (27.3)
Health system operations	7 (21.2)
Comprehensive Cancer Center	5 (15.2)
Pediatrics	4 (12.1)
Surgery	2 (6.1)
Psychiatry	2 (6.1)
Other (e.g., urology, pharmacy, ophthalmology)	4 (12.1)
Type of employment	
Staff	29 (87.9)
Faculty	4 (12.1)[§]
Main job title	
Research coordinator / manager	15 (45.5)
Coding, compliance, and administration personnel	4 (12.1)
Physician	3 (9.1)
Data manager / analyst	3 (9.1)
Research assistant	3 (9.1)
Lab/health technician	3 (9.1)
Educational nurse coordinator	1 (3.0)
Pharmacist	1 (3.0)

[§] 3 physicians (2 assistant professors, 1 full professor) and 1 pharmacist (assistant professor).

[*] This was the clinical area in which the participant was working but does not necessarily mean that the participant had clinical expertise in that area.

Table 4

Effectiveness evaluation results

Search task	Estimated level of difficulty	Q.A.1 "Performance"*	Q.A.2 "Relevance of recommended terms"**	Q.A.3 "Ability to come up with the terms w/o computer assistance"***
		mean [95% CI]	***mean [95% CI]***	***mean [95% CI]***
#1	*medium*	1.24 [0.96, 1.52]	1.73 [1.54, 1.91]	0.15 [-0.12, 0.42]
#2	*low*	0.94 [0.53, 1.35]	1.21 [0.80, 1.63]	0 [-0.34, 0.34]
#3	*medium*	1.42 [1.16, 1.69]	1.58 [1.30, 1.86]	0.09 [-0.18, 0.36]
#4	*high*	1.09 [0.81, 1.38]	1.18 [0.86, 1.51]	0.09 [-0.19, 0.38]
#5	*medium*	1.55 [1.31, 1.78]	1.88 [1.76, 2.00]	-0.27 [-0.57, 0.02]
#6, 7	variable	1.16 [0.95, 1.36]	1.53 [1.34, 1.73]	0.64 [0.37, 0.91]

[*] Response scale: much worse (-2); worse (-1); neural (0); better (1); much better (2).

[**] Response scale: very irrelevant (-2); somewhat irrelevant (-1); neural (0); somewhat relevant (1); highly relevant (2).

[***] Response scale: none (-2); few (-1); some (0); most (1); all (2).

Table 5

Summative evaluation results

Question	Response (count)				
	-2	***-1***	***0***	***1***	***2***
Q.B.1 “Perceived usefulness”*	0	0	0	11	22
Q.B.2 “Perceived ease of use”**	0	0	0	14	19
Q.B.3 “Intention to adopt”***	0	0	2	3	28

*Scale of response: useless (-2); somewhat useless (-1); neutral (0); somewhat useful (1); extremely useful (2).

**Scale of response: very difficult to use (-2); difficult to use (-1); neutral (0); easy to use (1); very easy to use (2).

***Scale of response: absolutely not (-2); probably not (-1); neutral (0); probably (1); absolutely yes (2).

Table 6

Qualitative themes identified from open-ended feedback, ordered from most (A) to least (G) frequent

	Theme	Examples
A	Users expressed appreciation for the value of the query recommendation feature	• "This search would have been impossible with the query suggestion off." • "I would definitely not have figured out the breast cancer search terms on my own." • "100% better with query search turned on."
B	Users believed that the query recommendation feature would improve time efficiency	• "I typically use EMERSE every day and often spend much of my time googling similar words to increase my results accuracy. Thi swould be a huge time-saver." • "Good terms search, I could have come up with some, but it would have taken extra time."
C	Users appreciated the use of consistent color groups for related concepts	• "I like the way all recommended search terms are the same color for each term - much easier to use this way." • "The color coding of search terms by reference is helpful."
D	False positives could undermine the utility of automated query recommendation	• "For this task, the computer has suggested too many additional terms, diluting the results. Particularly, adding 'PTS' as a suggestion for 'PTSD' is problematic because pts is a common abbreviation for 'patients'." • "I realized that different specialties use different words and abbreviations for things. So an allergist's 'PND' is not the same as a cardiologist's 'PND' etc."
E	Users desired more nuanced control over which recommended terms to include in the search	• "Most of the suggested additional terms were helpful, but not all were appropriate. It would be nice to be able to pick and choose which suggested additional terms to include, rather than all or none." • "Would be nice to be able to turn it on and off. Sometimes itincreases sensitivity and specificity, sometimes it worsens them."
F	Users desired the ability to add more terms to automated suggestions	• "Would have liked a few more terms to pop up on colonoscopy like sigmoidoscopy." • "Have you considered adding ICD-9 codes or DSM-IV§ codes to the suggestions. I work in mental health and many clinicians will use DSM-IV codes in their notes."
G	Users desired to 'weight' some of the recommended terms	• "Is there any way to weigh search topic i.e. DCIS drives the query?" • "Any way to say one term absolutely required then adding others i.e. DCIS, then breast may have been great."

[§] DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th Edition.