ISSN No. 0976-5697

Volume 3, No. 3, May-June 2012

International Journal of Advanced Research in Computer Science

REVIEW ARTICLE

Available Online at www.ijarcs.info

A Review of Clinical Queries for Electronic Health Record

Dr. G. R. Bamnote
Professor & Head, Dept. of CSE.
PRMITR, Badnera, India
grbamnote@rediffmail.com

Prof. S. S. Agrawal
Asst. Prof. Dept. of CSE
COE & T, Akola, India
sachin.s.agrawal@gmail.com

Abstract: Database systems must be able to respond to requests for information from the use, i.e. process queries. Obtaining the desired information from a database system in a predictable and reliable fashion is the art of Query Processing. Searching clinical information in a large collection of medical data is a complex task. The use of query processing tools and clinical resources could simplify the retrieval of the information. Clinical information available in electronic format is increasing rapidly. There are large collections of clinical data that contain visual and textual information available for researchers, health care providers and all types of users interested in this kind of information. Still, it is not always easy to access this large volume of data and, thus, there is a need to avail themselves with tools to enhance data accessibility and management of retrieval systems. Search engines have become nearly everywhere on the Web, clinical records lack search functionality; also, there is no knowledge on how and what healthcare providers search while using an clinical records search utility.

I. INTRODUCTION

Query processing and optimization is a fundamental part of any DBMS. To be utilized effectively, the results of queries must be available in the timeframe needed by the submitting user, be it a person, robotic assembly machine, doctor or even another distinct and separate DBMS.

Electronic Health records (EHR) are used increasingly in the hospital and outpatient settings, and patients are gathering digitized clinical information. As patient records shift from paper to digital format, many of the traditional organizational conventions of the paper chart are preserved, such as chart "sections" and labeled "tabs" for easier data browsing. There has been much debate as to the relative benefits of old and new ways of organizing patient data [1, 2].

Instead, the traditional format is likely to lower adoption barriers and still maintain some of its useful aspects. In contrast, preserving these older conventions results in missed opportunities to create novel ways to organize the computerized patient record and improve the way its users seek and access information. Example of such a missed opportunity is that EHRs generally do not have a search utility. In one qualitative study in Norway, where EHR adoption reached 95% nationally, researchers observed general practitioners' use of EHRs and reported that many of them found it difficult to find information, thereby hindering access to the information within the EHR.

This was specially true in lengthy medical records, like those of chronically ill patients [3]. Ironically, it is these very patients who require the most care, and the information within these records is especially applicable to the care of the patient. In such cases, an EHR-based search utility would improve information overload. It would do so by helping clinicians search for specific information within the patient record, the same way Web-based search engines help Web surfers find relevant information on the Web. While there is research on the use of search engines for clinical purposes, it is generally focused on searching for medical literature [4]. These studies have examined how literature searches are performed and have proposed novel approaches to improve search.

There is literature on the design of search tools to help users find clinical information within the EHR. It has been shown that clinicians find search functionality useful for both searching within and across patient records [5]. However, no in-depth analysis has been performed to understand clinicians' specific information needs in the context of search.

A Web-based clinical information system, WebCIS, that acts as a portal to all clinical narrative documents and laboratory test results within clinical data repository [6]. It is used regularly during clinical workflow for accessing clinical information; however, it lacks search functionality. The absence of an EHR search feature and the relative shortage of literature on the subject and study a search utility. The topic of search within the EHR has many unknown research questions.

II. BACKGROUND

A database query is the tool for instructing a DBMS to update or retrieve specific data to/from the physically stored medium. The actual updating and retrieval of data is performed through various "low-level" operations. Examples of such operations for a relational DBMS can be relational algebra operations such as project, join, select, Cartesian product, etc.

A. Information overload and user intent:

The clinical record is a source for clinical decision-making. It is essential to understand how and why clinicians use the information within it. A study in 1992 was conducted to understand clinicians' use of medical records in order to inform computer interfaces [7]. They identified three primary uses of the medical record by clinicians: "to gain an overview of a familiar or new patient, to search for specific details, and to prompt or explore hypotheses [8]."

A search utility can be useful in achieving the latter two goals, especially as more and more information becomes available within the EHR. Search functionality could help improve the occurrence of information overload.

In fact, there is need to investigate how to address the issue of information overload within the medical record by improving access to information. As the patient record

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html

moves to an electronic format, there have been different solutions proposed, which range from system enhancements to improved user-interface designs [9]. Though these alternative approaches reduce information overload, they focus primarily on structured data, such as laboratory data, and ignore free-text notes.

In order to improve search utilities and the search experience of any system, understanding users' search intent is essential. Although the medical informatics field has studied search and clinician information needs, the research has focused on accessing medical reference information, which is different from EHR-based search [10]. From a different perspective, investigators in the computer science and information science fields have examined search on a broad scale.

Broder was the first to categorize and study why people searched the Web [11]. He determined three broad search categories: navigational, informational, and transactional. Navigational searches are searches that involve a user seeking a specific site (e.g., searching for the International Journal of Medical Informatics homepage). Informational searches are searches that involve a user seeking information on a topic (e.g., searching "what is biomedical informatics"). Transactional searches are searches that involve a user seeking a site to perform another transaction (e.g., searching for PubMed in order to search for this article).

Li et al. analyzed intranet queries in a more domain-specific setting than Broder. Their high-level classification followed Broder's scheme, and they expanded the analysis to include domain-specific sub-categories of search types. The categories were derived in an iterative process by manually examining the intranet queries. Li's intranet search study suggests that medical searches within EHRs, which are also domain specific, can be categorized into Broder's three search categories.

There are many ways to capture users' information needs in order to understand search intent. Research methods, such as surveys, interviews, and focus groups, provide a deep understanding of the subjects' behaviors and needs. Another method, the analysis of transaction logs, provides an modest way to capture user behavior. Transaction logs are files that contain records of the interactions between a system and its users. The methodology of analyzing these transaction logs in order to investigate research questions is called transaction log analysis (TLA) [12].

TLA has been used in studies across many domains in order to understand users' behavior when interacting with a system [4]. These studies range from examining general usage to examining implicit features such as click through data to improve search.

B. CISearch (Example):

CISearch is a general search utility, which searches free-text clinical reports within patient's electronic medical record. Unlike most search engines, which display search results based on relevancy,CISearch displays results in reverse chronological order. CISearch indexes all free-text, clinical documents (e.g., radiology reports, discharge summaries, and nursing notes). It does not search structured, coded data that can be represented numerically, such as laboratory results (e.g., CHEM7 test), because accessing such information within EHR is relatively user-friendly and efficient.

The search box was placed at the top of the main, left navigation area so that it was easy to access. In order to reduce the barrier of implementation, a customized open-source search engine, Lucene, to index and search clinical notes within a particular patient [13]. Lucene is based on the vector-space model and has several built-in features.

Features utilized in CISearch were in-memory indexing, advanced query grammar, stop-word removal, text snippets, and results highlighting.

III. METHODS

A. Data collection:

The user log files were collected. The files contained all search transactions. There were two types of CISearch log entries: query and click through. We define query as the entire string that a user enters and define query term as the individual strings separated by white space that comprise a query. The query entry contained timestamp, the user identifier and its IP address, the patient medical record number for the patient currently viewed, the document types that were selected to be searched, the search query, the number of documents retrieved from the search, the total number of documents in the patient record, and the document retrieval time.

The click through entry is similar to the query view. It contained the document selected, the document's relevancy score, and the document's rank in the result set.

B. Pre-processing:

Once the data was collected, the log files were cleaned before analysis. First, the log files were filtered to remove entries of hospital information-technology employees and system developers. Then the log files were de-identified by replacing Medical Record Numbers (MRN) and user ids with unique numbers. Finally, the query and click through log entries were extracted and inserted into respective database tables.

C. Analysis :

The analysis of the queries was carried out using Broder's categories (navigational, transactional, and informational). Two investigators (KN and NE) manually categorized all the unique queries and inter-annotator agreement was analyzed. For example, a query containing a patient MRN was labeled as a navigational search because it was most likely that the user was trying to switch patients rather than searching for the MRN within the current patient's medical record. Queries that represented an action were labeled as a transactional search.

For instance, the query "add note" most likely referred to the user's intent to create a new note as opposed to searching for those words within the medical record. All other queries were labeled as informational searches. During the analysis it became apparent that informational searches were most frequently performed. Considering the large proportion of informational searches and future goal of extracting pertinent information from the medical record, we further categorized informational searches. Three physicians categorized a random sample of informational searches with semantic information. The reviewers were given

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html

overlapping data sets so that two clinicians categorized each query.

In order to reduce the burden of categorizing the queries, an abbreviated list of UMLS semantic types was provided to the clinicians. The abbreviated list was created by iteratively filtering and clustering UMLS concepts with similar meaning from a clinical perspective.

IV. DISCUSSION

The analysis of search logs yielded several design implications, and possibly for others who wish to integrate search into their EHR.

A. User type & high-level query classification :

Overall, users show a strong bias toward informational searches. When stratified by user types, however, different user behaviors emerge. All clinical users (e.g., doctors, nurses, and students) who provide direct care to patients tend to perform more informational searches (with doctors at 91.8%). Administrative staff's queries are evenly balanced between navigational and informational searches, confirming that their information needs differ from clinical users. Finally, researchers exhibit simply different behavior, with hardly any navigational searches (95.3% informational and 4.7% navigational).

Opposing to clinical users, researchers approach the EHR as an interface tool for cohort selection, explaining the negligible number of navigational searches. The unanticipated use of the system to frequently search the same set of terms across multiple patients suggests that cross-patient search functionality would be useful for research purposes.

B. Concept-based searching:

To improve search was to map queries to UMLS concepts within machine processed notes. When entering a query, a user would be prompted to select the semantic type that best represents the query. However, we found that mapping query terms to the UMLS is inherently ambiguous as of its multi-hierarchical structure. Alternatively, the large presence of queries with abbreviations and incomplete words, which do not map to the UMLS, suggests that indexing and searching based on UMLS concepts cannot be the sole solution.

Rather, a combination of free-form text and concept-based search is needed. This finding is supported by Nadkarni et al.'s study that determined that both free-text and concept-based indexing was needed for concept-based searching of clinical notes [14].

C. Limitations:

Log analysis is an efficient, modest way to obtain information about a user's actions; however, it does not give insight into the user's underlying motivations or background for performing a search. While it provides an abundant and rich source of data, TLA cannot be solely used to model a user's information seeking behavior [12]. There have been many studies examining log files to determine features that represent a successful search. One such feature that has proven to be representative of whether a search result is relevant is click through data [15]. It is mainly used for ranking search results, and it is effective because the search results contain query-based snippets, allowing a user to determine whether a document is relevant or not before clicking on it [16]. Though click through analysis is useful in determining a document's relevancy, it is also limited because it does not account for documents that the user deems relevant based on the snippet. Thus, to truly understand what users are searching and the usefulness of a search utility, log analysis must be supplemented with observational and survey studies.

Another limitation in the study concerns the semantic categorization of queries. Besides the inherent ambiguity of labeling with the UMLS, it was difficult to disambiguate the queries because the reviewers were not provided the context of the queries, resulting in the low kappa score. This manual process is also not scalable, a limitation which other search log studies face. The only solution to this problem would be a semi-manual approach whereby a trained classifier program would categorize a random sample of queries, and then these categorized queries would be manually reviewed. This review process would occur rarely.

V. CONCLUSION

One of the most critical functional requirements of a DBMS is its ability to process queries in a timely manner. This is particularly true for very large, mission critical applications. Clinical data hold a rich amount of information, especially in the description part of the record; however, this information is often huge to access. Searching for information has become commonplace on the Internet, but little is known on its needs and use within the medical record. Study showed that a variety of user types queried using search tool and that clinician searches are largely informational, focusing on laboratory results and specific diseases.

Understanding what clinicians search for within the clinical data will inform new ways to present patient information and provide guidance on how to improve clinical data specific search engines. Ultimately, this will allow physicians, nurses, laboratory practitioners, and others in the health care field to access pertinent patient information with greater ease and possibly result in better health care delivery.

VI. REFERENCES

[1]. J. Hippisley-Cox, et al., The electronic patient record in primary care—regression or progression? A cross sectional study, BMJ 326 (7404) (2003) 1439–1443.

[2]. H.J. Tange, The paper-based patient record: is it really so bad? Comput. Methods Programs Biomed. 48 (1–2) (1995) 127–131.

[3]. T. Christensen, A. Grimsmo, Instant availability of patient records, but diminished availability of patient information: a multi-method study of GP's use of electronic patient records, BMC Med. Inform. Decis. Making 8 (2008) 12.

[4]. C. Chisnell, K. Dunn, D.F. Sittig, Determining educational needs for the biomedical library customer: an analysis of end-user searching in MEDLINE, Medinfo 8 (Pt 2) (1995) 1423–1427.

[5]. W. Gregg, et al., StarTracker: an integrated, web-based clinical search engine, Proc. AMIA Symp. (2003) 855.

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html

[6]. G. Hripcsak, J.J. Cimino, S. Sengupta, WebCIS: large scale deployment of a Web-based clinical information system, Proc. AMIA Symp. (1999) 804–808.

[7]. E. Nygren, P. Henriksson, Reading the medical record. I. Analysis of physicians' ways of reading the medical record, Comput. Methods Programs Biomed. 39 (1–2) (1992) 1–12.

[8]. E. Nygren, J.C. Wyatt, P. Wright, Helping clinicians to find data and avoid delays, Lancet 352 (9138) (1998) 1462–1466.

[9]. Y. Senathirajah, S. Bakken, Development of user-configurable information source pages for medical information retrieval, AMIA Annu. Symp. Proc. (2007) 1109.

[10]. W.R. Hersh, Information Retrieval: A Health and Biomedical Perspective, 3rd ed. Health Informatics Series, Springer, New York, NY, 2009, xvii, p. 486.

[11]. A. Broder, A taxonomy of web search, SIGIR Forum 36 (2) (2002) 3–10.

[12]. B.J. Jansen, Search log analysis: what it is, what's been done, how to do it, Libr. Inform. Sci. Res. 28 (3) (2006) 407–432.

[13]. Apache Lucene - Query Parser Syntax, 2007; available from: http://lucene.apache.org/java/docs/queryparsersyntax.html.

[14]. P. Nadkarni, R. Chen, C. Brandt, UMLS concept indexing for production databases: a feasibility study, J. Am. Med. Inform. Assoc. 8 (2001) 80–91.

[15]. J. Thorsten, Optimizing search engines using clickthrough data, in: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining %@ 1-58113-567-X, ACM, Edmonton, Alberta, Canada, 2002, pp. 133–142.

[16]. A. Tombros, M. Sanderson, Advantages of query biased summaries in information retrieval, in: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval – SIGIR '98, 1998, pp. 2–10.