journal homepage: www.intl.elsevierhealth.com/journals/ijmi

An analysis of clinical queries in an electronic health record search utility

Karthik Natarajan, Daniel Stein, Samat Jain, Noémie Elhadad*

Columbia University, New York, NY, United States

ARTICLE INFO

Article history:
Received 1 October 2009
Received in revised form
5 February 2010
Accepted 16 March 2010

Keywords:
Information storage and retrieval
Medical informatics
Medical informatics applications
Medical record systems
Computerized

ABSTRACT

Purpose: While search engines have become nearly ubiquitous on the Web, electronic health records (EHRs) generally lack search functionality; furthermore, there is no knowledge on how and what healthcare providers search while using an EHR-based search utility. In this study, we sought to understand user needs as captured by their search queries.

Methods: This post-implementation study analyzed user search log files for 6 months from an EHR-based, free-text search utility at our large academic institution. The search logs were de-identified and then analyzed in two steps. First, two investigators classified all the unique queries as navigational, transactional, or informational searches. Second, three physician reviewers categorized a random sample of 357 informational searches into high-level semantic types derived from the Unified Medical Language System (UMLS). The reviewers were given overlapping data sets, such that two physicians reviewed each query.

Results: We analyzed 2207 queries performed by 436 unique users over a 6-month period. Of the 2207 queries, 980 were unique queries. Users of the search utility included clinicians, researchers and administrative staff. Across the whole user population, approximately 14.5% of the user searches were navigational searches and 85.1% were informational. Within informational searches, we found that users predominantly searched for laboratory results and specific diseases.

Conclusions: A variety of user types, ranging from clinicians to administrative staff, took advantage of the EHR-based search utility. Though these users' search behavior differed, they predominantly performed informational searches related to laboratory results and specific diseases. Additionally, a number of queries were part of words, implying the need for a free-text module to be included in any future concept-based search algorithm.

1. Introduction

Electronic health records (EHRs) are used increasingly in the hospital and outpatient settings, and patients are amassing digitized clinical information. As patient records make the 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]. On the one hand, the traditional format is likely to lower adoption barriers and still maintain some of its useful aspects. On the other hand, preserving these older conventions results in missed

* *Corresponding author at*: Department of Biomedical Informatics, 622 W. 168th St, VC 5, New York, NY 10032, United States. Tel.: +1 212 305 5780.
E-mail address: karthik.natarajan@dbmi.columbia.edu (K. Natarajan).
1386-5056/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ijmedinf.2010.03.004

opportunities to create novel ways to organize the computerized patient record and improve the way its users seek and access information.

One glaring example of such a missed opportunity is that EHRs generally do not have a search utility. In a recent 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 especially 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 pertinent to the care of the patient. In such cases, an EHR-based search utility would alleviate 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 [4–8]. These studies have examined how literature searches are performed and have proposed novel approaches to improve search. There is sparse 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 [9–11]. However, no in-depth analysis has been performed to understand clinicians' specific information needs in the context of search.

Our institution has a Web-based clinical information system, WebCIS, that acts as a portal to all clinical narrative documents and laboratory test results within our clinical data repository [12]. 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 dearth of literature on the subject inspired us to build and study a search utility. We designed and implemented a simple keyword search utility called CISearch, which is integrated within WebCIS.

The topic of search within the EHR has many unexplored research questions. In this retrospective study, we attempt to answer one of the fundamental questions in order to guide future research: what are the characteristics of users' searches within the EHR? We hypothesize that general Web search classification schemas can be leveraged to categorize EHR-based queries and that these queries can be mapped further to medical semantic types derived from the Unified Medical Language System (UMLS).

2. Background

2.1. *Information overload and user intent*

The medical record is a source for clinical decision-making. It is thus essential to understand how and why clinicians use the information within it. Nygren and Henriksson conducted a study in 1992 to understand clinicians' use of medical records in order to inform computer interfaces [17]. 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 [18]." 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 alleviate the phenomenon of information overload.

In fact, researchers have been investigating for some time how to address the issue of information overload within the medical record by improving access to information. As the patient record moves to an electronic format, there have been novel solutions proposed, which range from system enhancements to improved user-interface designs [13–16]. 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 [19–23]. 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 [24]. 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). Other search taxonomies have had essentially the same three high-level categories [25,26]. 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 unobtrusive 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) [27]. TLA has been employed in studies across many domains in order to understand users' behavior when interacting with a system [4,28–39]. These studies range from examining general usage to examining implicit features such as clickthrough data to improve search. TLA has been utilized previously at our institution to study clinician information needs within the clinical information system [29]. The study found that laboratory and radiology reports were