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.

© 2010 Elsevier Ireland Ltd. All rights reserved.

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
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 CISeSearch, 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’ behaviors 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...
the most accessed. Our study followed the steps of TLA to understand clinician searches because of its unobtrusiveness in collecting data and its ability to examine the behavior of all search users.

2.2. CISeach

CISeach 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, CISeach displays results in reverse chronological order. In its current version, CISeach 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., CHEM-7 test), because accessing such information within our EHR is relatively user-friendly and efficient. CISeach was integrated into WebCIS in July 2008. The search box was placed at the top of the main, left navigation area within WebCIS so that it was easy to access. In order to reduce the barrier of implementation, we customized a widely used, open-source search engine, Lucene, to index and search clinical notes within a particular patient [40]. Lucene is based on the vector-space model and has several built-in features. Features utilized in CISeach were in-memory indexing, advanced query grammar, stop-word removal, text snippets, and results highlighting. At the time this study was performed, CISeach indexed and searched discharge summaries, radiology reports, and pathology reports only. It was decided that an in-memory search was acceptable for the initial version of CISeach because it did not require the creation and maintenance of a database of indexed documents and because the initial search document space was small. Fig. 1 shows the results for “diabetes” on a fictitious test patient.

3. Methods

3.1. Data collection

WebCIS log files were collected for 6 months (from July 14th to December 31, 2008). The files contained all CISeach transactions within WebCIS. There were two types of CISeach log entries: query and clickthrough. 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 clickthrough 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.

3.2. 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 clickthrough log entries were extracted and inserted into respective database tables.

3.3. 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 our 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 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. For example, the UMLS semantic type “Chemical Viewed Structurally” and its children were removed since, in a clinical context, it is more likely that clinicians refer to chemicals functionally (e.g., how a patient is reacting to an antibiotic) as opposed to structurally (e.g., the molecular structure of the chemical compound); likewise, the semantic types “Body Part” and “Body Location” were merged. The reviewers practiced categorizing on a standard set of 20 queries with an investigator before proceeding with their individual sets of 119 queries. The inter-annotator agreement between the three physicians was analyzed as well.
4. Results

4.1. General usage results

There were a total of 436 unique users of CiSearch in the first 6 months of its deployment within WebCIS. This represents roughly 5.3% of WebCIS users (the total number of WebCIS users was estimated to be the average number of active users within a month for the past year, which is approximately 8200). Fig. 2 shows the breakdown based on user types of the 436 unique users.

A total of 6117 search log lines were analyzed. We removed highly repetitive queries from our data set. In particular, three users conducted approximately 2200 searches with variations of the same query containing a specific drug and medical condition over several hundred patient records. All three individuals were researchers. These queries were considered outliers because of their high frequency and were excluded from the analysis. The users that only submitted these outlier queries were also removed from the analysis. Table 1 lists the general usage statistics of CiSearch. A unique query was defined as a distinct string, ignoring case as well as leading and ending white spaces. A view occurred when a user clicked on one of the documents returned by a query. Fig. 3 describes the monthly usage of the search utility. Of the 980 unique queries, only four utilized the built-in features in the query language of Lucene, such as quotes around queries (i.e., “chest tube”). Additionally, out of the 980 queries, 148 were abbreviations (e.g., “chf” for congestive heart failure) and 78 were part of a word (e.g., “tach” for either tachycardia or tachyarrhythmia).

### Table 1 – General usage statistics.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of queries after the removal of outlier queries</td>
<td>2207</td>
</tr>
<tr>
<td>Number of unique queries after the removal of outlier queries</td>
<td>980</td>
</tr>
<tr>
<td>Average number of terms per query</td>
<td>1.2</td>
</tr>
<tr>
<td>Number of clickthroughs (total number of queries = 4427)</td>
<td>618 (13.9%)</td>
</tr>
<tr>
<td>Number of unique users (approx # of active users = 8200)</td>
<td>436 (5.3%)</td>
</tr>
</tbody>
</table>

Table 2 shows the top 25 unique searches. The most frequent query, “class”, was used predominately in conjunction with “nyha”, “III”, and “IV.” In this case, we suspect that users were searching for mentions of New York Heart Association Classifications within the medical records.

### Table 2 – Top 25 queries.

<table>
<thead>
<tr>
<th>Query</th>
<th>Frequency</th>
<th>Percentage (N = 2207)</th>
</tr>
</thead>
<tbody>
<tr>
<td>class</td>
<td>217</td>
<td>9.8%</td>
</tr>
<tr>
<td>nyha</td>
<td>99</td>
<td>4.5%</td>
</tr>
<tr>
<td>hodgkins</td>
<td>64</td>
<td>2.9%</td>
</tr>
<tr>
<td>iii</td>
<td>52</td>
<td>2.4%</td>
</tr>
<tr>
<td>iv</td>
<td>50</td>
<td>2.3%</td>
</tr>
<tr>
<td>nephrogenic</td>
<td>39</td>
<td>1.8%</td>
</tr>
<tr>
<td>hysterectomy</td>
<td>33</td>
<td>1.5%</td>
</tr>
<tr>
<td>cva</td>
<td>24</td>
<td>1.1%</td>
</tr>
<tr>
<td>ef</td>
<td>23</td>
<td>1.0%</td>
</tr>
<tr>
<td>hf</td>
<td>19</td>
<td>0.9%</td>
</tr>
<tr>
<td>fibronectin</td>
<td>19</td>
<td>0.9%</td>
</tr>
<tr>
<td>cmv</td>
<td>17</td>
<td>0.8%</td>
</tr>
<tr>
<td>chf</td>
<td>16</td>
<td>0.7%</td>
</tr>
<tr>
<td>embol</td>
<td>15</td>
<td>0.7%</td>
</tr>
<tr>
<td>sirolimus</td>
<td>13</td>
<td>0.6%</td>
</tr>
<tr>
<td>hiv</td>
<td>13</td>
<td>0.6%</td>
</tr>
<tr>
<td>pericardiocentesis</td>
<td>13</td>
<td>0.6%</td>
</tr>
<tr>
<td>renal</td>
<td>13</td>
<td>0.6%</td>
</tr>
<tr>
<td>subdural</td>
<td>12</td>
<td>0.5%</td>
</tr>
<tr>
<td>lvad</td>
<td>12</td>
<td>0.5%</td>
</tr>
<tr>
<td>dsum</td>
<td>12</td>
<td>0.5%</td>
</tr>
<tr>
<td>placenta</td>
<td>12</td>
<td>0.5%</td>
</tr>
<tr>
<td>accreta</td>
<td>12</td>
<td>0.5%</td>
</tr>
<tr>
<td>mri</td>
<td>11</td>
<td>0.5%</td>
</tr>
<tr>
<td>inferior epigastric</td>
<td>11</td>
<td>0.5%</td>
</tr>
</tbody>
</table>

4.2. Analysis results

4.2.1. High-level classification of search queries

980 unique queries were categorized as informational (e.g., “chf”), navigational (e.g., medical record number), or transactional (e.g., “add drug”). The inter-annotator agreement was a Kappa of 0.93 [41]. A large majority of the queries were cate-
ties are being identified. Furthermore, we have not performed the beginning stages of development and new functionality in our institution because the search utility is at It is premature to perform a formal study of EHR-based search.

5.1. Adoption

Wish to integrate search into their EHR.

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

4.2.2. Semantic classification of informational search queries

357 unique queries were categorized, each by two clinicians. There were three reviewers, one of which was an investigator in this study (DS). Each reviewer categorized 238 queries of which 119 queries overlapped with one other reviewer. The inter-annotator agreement was a Kappa of 0.56 (the reviewers agreed on 161 queries and disagreed on 74). Among the disagreements, 52 were due to ambiguities with the semantic type “Laboratory or Test Results” (e.g., immunoglobulin can be classified as a laboratory test or a biologically active substance), and 16 were due to ambiguities with “Disease or Syndrome” (e.g., atria fibrillation can be classified as a finding or a disease). 122 of the total 357 queries were left uncategorized by either one or both reviewers due to uncertainty in the query. Some of these uncategorized queries were ambiguous abbreviations, part of a word, first names, or simply not medical terms.

The informational searches that the reviewers agreed upon showed that a majority of searches were about “Laboratory or Test Results” and “Disease or Syndromes.” Table 5 lists the top five semantic types of searches and their percentage from the total number of queries where the reviewers agreed with one another.

5. Discussion

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

5.1. Adoption

It is premature to perform a formal study of EHR-based search adoption in our institution because the search utility is at the beginning stages of development and new functionalities are being identified. Furthermore, we have not performed any marketing or training to WebCIS users. Yet, the consistent search usage from month-to-month of the system suggests the potential usefulness of such a tool. Once the planned enhancements of the search utility are implemented into production, a multi-year evaluation on the use and adoption of the system can be carried out, similarly to another system evaluation within our institution [42].

5.2. User type and 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 starkly different behavior, with hardly any navigational searches (95.3% informational and 4.7% navigational). Contrary 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. There have been systems and studies designed to examine the use of cross-patient searches for cohort eligibility through the use of EHRs [9,11,43,44]. Though the cross-patient searches imply the need for such a system, our objective is to understand how clinicians search within an individual patient’s record in order to extract pertinent information at the point of care.

5.3. EHR implications

Overall, the CISearch queries adhered to the broad Web search categories (transactional, navigational, and informational). While most of the searches were deemed informational, we did note occurrences of transactional and navigational queries. The finding of transactional searches was unexpected considering WebCIS is predominately a read-only system and is not an order-entry system; however, there were a few transactional searches such as “add note” or “add drug”. This may be a consequence of having multiple clinical information systems with overlapping functionality. The majority of the navigational searches were patient lookups (i.e., MRN or patient names). This might indicate the need for a more efficient way of switching patients within the EHR.

<table>
<thead>
<tr>
<th>Table 4 – Percentage of query type based on user type.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informational</td>
</tr>
<tr>
<td>Research (N=684)</td>
</tr>
<tr>
<td>Doctor (N=649)</td>
</tr>
<tr>
<td>Student (N=331)</td>
</tr>
<tr>
<td>Other Clinical (N=65)</td>
</tr>
<tr>
<td>Nurse (N=102)</td>
</tr>
<tr>
<td>Admin (N=63)</td>
</tr>
<tr>
<td>Other (N=313)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5 – Top 5 semantic types of searches.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic type</td>
</tr>
<tr>
<td>Laboratory or test result</td>
</tr>
<tr>
<td>Disease or syndrome</td>
</tr>
<tr>
<td>Body part, organ, or organ component</td>
</tr>
<tr>
<td>Pharmacologic substance</td>
</tr>
<tr>
<td>Diagnostic procedure</td>
</tr>
</tbody>
</table>
The semantic type “Laboratory or Test Result” was the most frequent informational search. This supports Chen’s findings that the most frequented section within WebCIS was laboratory results [29]. WebCIS is efficient at displaying laboratory results in a structured format, so it would be interesting to understand why users are looking for them with free-text queries. We could hypothesize that viewing laboratory results mentioned within clinical narratives conveys more relevant and contextualized information than merely viewing the raw data in the structured section of the EHR. This could be assessed through user interviews or surveys, and may inform better display of laboratory results. As more usage data is collected, patient displays and navigation within the EHR can be further tailored for individual users, potentially improving users’ ability to access patient information.

5.4. Concept-based searching

Our original plan 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 because of its multi-hierarchical structure. Table 3 suggests that semantic types could be leveraged to inform preference rules for disambiguating UMLS concepts during the retrieval process. For example, a preference rule that favors laboratory test/procedure types over biological substance types would classify the query “fibrinogen” as a laboratory procedure and then search for that concept within the machine processed notes.

On the other hand, 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 both free-text and concept-based indexing was needed for concept-based searching of clinical notes [45].

5.5. Semantic categorization

The low agreement between reviewers was predominately due to the ambiguous nature of the classification discussed in Section 5.6. For example, two reviewers classified the query “amylase” as a laboratory test and a biologically active substance. The reviewers commented that a query could easily be placed in either category.

5.6. Limitations

Log analysis is an efficient, unobtrusive 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 [27]. While it provides an abundant and rich source of data, TLA cannot be solely used to model a user’s information seeking behavior [27]. 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 clickthrough data [46]. 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 [47]. Though clickthrough 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 (e.g., the query was performed while in the laboratory section of WebCIS), resulting in the low kappa score. This manual process is also not scalable, a limitation which other search log studies face [24–26]. 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.

Finally, there were two limitations related to CISeach itself. First, no marketing or formal training on CISeach was provided to the users within the institution, which may explain the relatively low adoption rate of 5.3%, the infrequent use of the advanced search features, and the low number of click-throughs. Second, the search functionality was narrow in scope. For its initial release, CISeach only searched discharge summaries, radiology reports and pathology reports (in our current implementation, all note types are supported). As such, it is possible that the users were biased in their queries, based on the behavior of the search engine. However, when examining the search logs, it became evident that users were unaware of the limited search space and the internal workings of the search engine (this phenomenon also relates to the absence of user training), and instead queried the search functionality in a genuine fashion (as evidenced by the high frequency of navigational and transactional queries, typically not supported by domain-specific search engines). Overall, our approach to understanding information needs of clinicians in the context of EHR search follows an iterative process traditionally employed in software engineering, where a prototype with limited functionality is deployed in order to capture user needs in a real-world setting rather than in a laboratory setting. Once user needs are understood better, the prototype can be refined.

6. Conclusion

EHRs hold an abundant amount of information, especially in the narrative part of the record; however, this information is often cumbersome to access. Searching for information has become commonplace on the Internet, but little is known on its needs and use within the medical record. Our study showed that a variety of user types queried using an EHR-based search tool and that clinician searches in the EHR are largely informational, focusing on laboratory results and spe-
References


