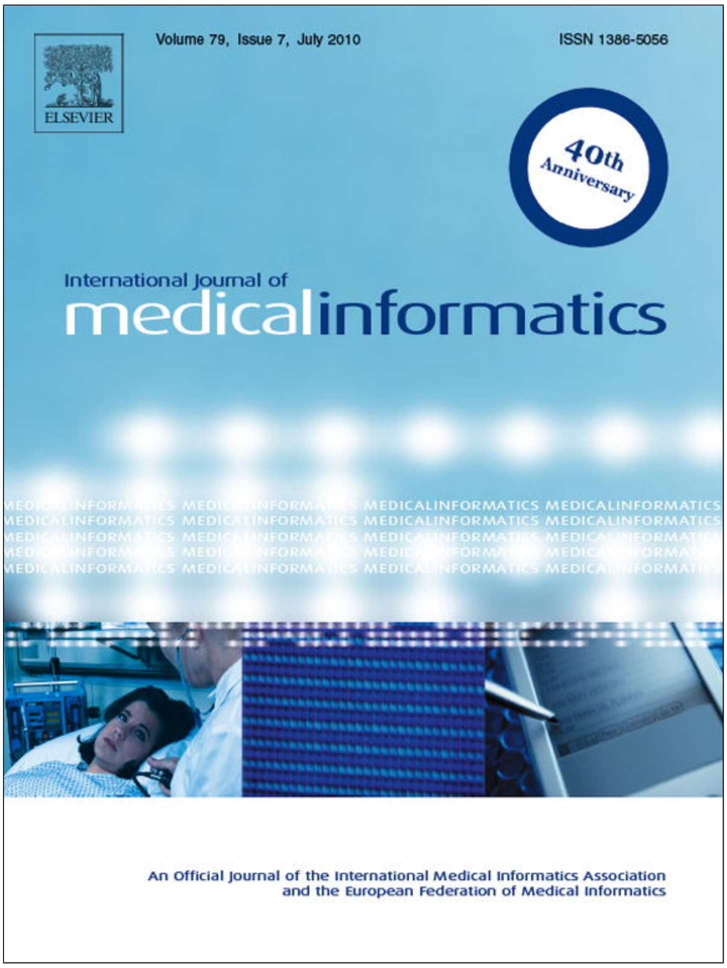
ELSEVIER
Volume 79, Issue 7, July 2010
ISSN 1386-5056
40th Anniversary
International Journal of
medicalinformatics
An Official Journal of the International Medical Informatics Association
and the European Federation of Medical Informatics

INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS 79 (2010) 515–522

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).

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

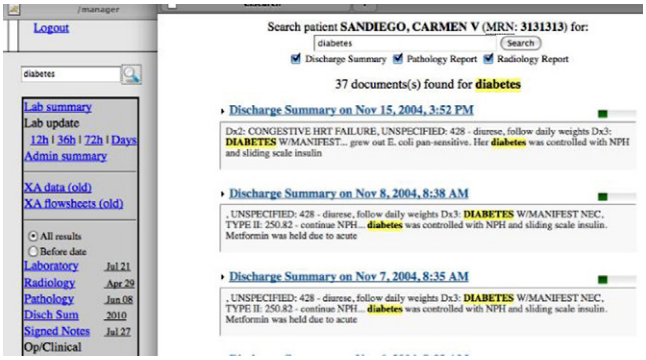

Fig. 1 – Results for query "diabetes" on a test patient.

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. *CISearch*

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. In its current version, 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., CHEM-7 test), because accessing such information within our EHR is relatively user-friendly and efficient. CISearch 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 CISearch were in-memory indexing, advanced query grammar, stop-word removal, text snippets, and results highlighting. At the time this study was performed, CISearch 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 CISearch 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 CISearch transactions within WebCIS. There were two types of CISearch 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.

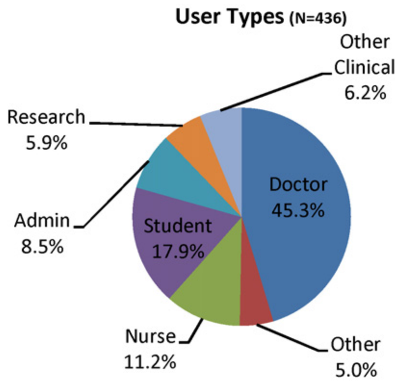

Fig. 2 – The various user types of the search utility.

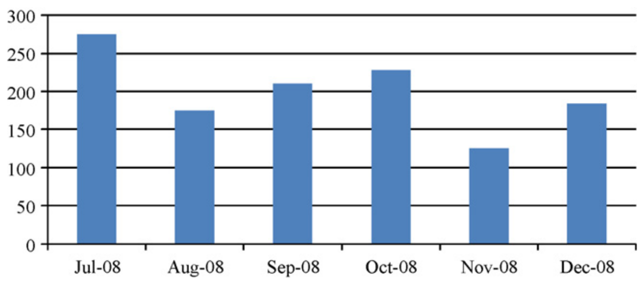

Fig. 3 – Number of unique queries per month.

4. Results

4.1. *General usage results*

There were a total of 436 unique users of CISearchin 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.

Number of queries after the removal of outlier queries	2207
Number of unique queries after the removal of outlier queries	980
Average number of terms per query	1.2
Number of clickthroughs (total number of queries = 4427)	618 (13.9%)
Number of unique users (approx # of active users = 8200)	436 (5.3%)

Table 2 – Top 25 queries.

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

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.

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-

Table 3 – Percentage of query types based on all unique queries.

Query type	Percentage (N = 980)
Informational	85.1%
Navigational	14.5%
Transactional	0.4%

Table 4 – Percentage of query type based on user type.

	Informational	Transactional	Navigational
Research (N = 684)	95.3%	0.0%	4.7%
Doctor (N = 649)	91.8%	0.0%	8.2%
Student (N = 331)	87.6%	0.0%	12.4%
Other Clinical (N = 65)	72.3%	0.0%	28.7%
Nurse (N = 102)	70.6%	0.0%	29.4%
Admin (N = 63)	52.4%	3.2%	44.4%
Other (N = 313)	94.9%	0.6%	4.5%

Table 5 – Top 5 semantic types of searches.

Semantic type	Percentage (N = 161)
Laboratory or test result	29.2%
Disease or syndrome	21.7%
Body part, organ, or organ component	8.1%
Pharmacologic substance	7.5%
Diagnostic procedure	6.2%

gorized as informational. Table 3 shows the breakdown of the queries and Table 4 shows what types of queries each user type performed.

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.

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 that 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 CISearch itself. First, no marketing or formal training on CISearch 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 clickthroughs. Second, the search functionality was narrow in scope. For its initial release, CISearch 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-

Summary points

What was already known before our study?

- Electronic health records are expanding as more and more information about patients, especially clinicians' notes, is collected.
- Accessing free-text information in the EHR can be difficult and time consuming.
- Little is known on what clinicians search for within the electronic medical record when given access to a search utility.

What did our study add to our body of knowledge?

- A variety of user types took advantage of a search utility within the EHR (e.g., doctors, nurses, students, researchers, and administrators).
- Most searches performed within the EHR search utility were informational (i.e., seeking information within the medical record) as opposed to navigational (i.e., searching for the purpose of navigating to a section of the medical record).
- When classified further, the majority of searches were related to specific laboratory results and diseases.

cific diseases. Understanding what clinicians search for within the EHR will inform new ways to present patient information and provide guidance on how to improve EHR-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.

Acknowledgements

This project was supported in part by National Library of Medicine Training Grant 5T15-LM007079-18. The authors would like to thank Herb Chase and Sandip Vaidya for reviewing the queries, Yiping Hu, Cynthia Chen, and George Hripcsak of the WebCIS team, and James Cimino for his guidance during the conception of CISearch.

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.R. Hersh, et al., CliniWeb: managing clinical information on the World Wide Web, J. Am. Med. Inform. Assoc. 3 (4) (1996) 273–280.
[6] W.R. Hersh, L.C. Donohoe, SAPHIRE, International: a tool for cross-language information retrieval, Proc. AMIA Symp. (1998) 673–677.
[7] N.C. Ide, R.F. Loane, D. Demner-Fushman, Essie: a concept-based search engine for structured biomedical text, J. Am. Med. Inform. Assoc. 14 (2007) 253–263.
[8] E. Meats, et al., Using the turning research into practice (TRIP) database: how do clinicians really search? J. Med. Libr. Assoc. 95 (2007) 156–163.
[9] W. Gregg, et al., StarTracker: an integrated, web-based clinical search engine, Proc. AMIA Symp. (2003) 855.
[10] D.A. Hanauer, EMERSE: the electronic medical record search engine, Proc. AMIA Symp. 331 (2006) 941.
[11] S. Schulz, et al., Evaluation of a document search engine in a clinical department system, Proc. AMIA Symp. (2008) 647–651.
[12] G. Hripcsak, J.J. Cimino, S. Sengupta, WebCIS: large scale deployment of a Web-based clinical information system, Proc. AMIA Symp. (1999) 804–808.
[13] Y. Senathirajah, S. Bakken, Development of user-configurable information source pages for medical information retrieval, AMIA Annu. Symp. Proc. (2007) 1109.
[14] P.C. Tang, et al., Semantic integration of information in a physician's workstation, Int. J. Biomed. Comput. 35 (1994) 47–60.
[15] H.J. Tange, et al., An experimental electronic medical-record system with multiple views on medical narratives, Comput. Methods Programs Biomed. 54 (1997) 157–172.
[16] Q. Zeng, J.J. Cimino, K.H. Zou, Providing concept-oriented views for clinical data using a knowledge-based system: an evaluation, J. Am. Med. Inform. Assoc. 9 (2002) 294–305.
[17] 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.
[18] E. Nygren, J.C. Wyatt, P. Wright, Helping clinicians to find data and avoid delays, Lancet 352 (9138) (1998) 1462–1466.
[19] W.R. Hersh, Information Retrieval: A Health and Biomedical Perspective, 3rd ed. Health Informatics Series, Springer, New York, NY, 2009, xvii, p. 486.
[20] J.A. Osheroff, et al., Physicians' information needs: analysis of questions posed during clinical teaching, Ann. Intern. Med. 114 (7) (1991) 576–581.
[21] L.M. Currie, et al., Clinical information needs in context: an observational study of clinicians while using a clinical information system, AMIA Annu. Symp. Proc. (2003) 190–194.
[22] R. Smith, What clinical information do doctors need? BMJ 313 (7064) (1996) 1062–1068.
[23] M. Allen, et al., The classification of clinicians' information needs while using a clinical information system, AMIA Annu. Symp. Proc. (2003) 26–30.
[24] A. Broder, A taxonomy of web search, SIGIR Forum 36 (2) (2002) 3–10.
[25] H. Li, et al., A new approach to intranet search based on information extraction, in: Proceedings of the 14th ACM International Conference on Information and Knowledge Management, ACM, Bremen, Germany, 2005.
[26] D.E. Rose, D. Levinson, Understanding user goals in web search, in: Proceedings of the 13th International Conference on World Wide Web, ACM, New York, NY, USA, 2004.
[27] 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.
[28] S.M. Beitzel, et al., Hourly analysis of a very large topically categorized web query log, in: Proceedings of the 27th Annual International ACM SIGIR Conference on Research

and Development in Information Retrieval, ACM, New York, NY, USA, 2004.

[29] E.S. Chen, J.J. Cimino, Automated discovery of patient-specific clinician information needs using clinical information system log files, AMIA Annu. Symp. Proc. (2003) 145–149.

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

[31] T. Joachims, Evaluating retrieval performance using clickthrough data, in: Proceedings of the SIGIR Workshop on Mathematical/Formal Methods in Information Retrieval, 2002.

[32] D. Kelly, J. Teevan, Implicit feedback for inferring user preference: a bibliography, ACM SIGIR Forum 37 (2003) 18–28.

[33] S. Klink, Improving document transformation techniques with collaborative learned term-based concepts, Lect. Notes Comput. Sci. (2004) 281–305.

[34] M. Mat-Hassan, M. Levene, Associating search and navigation behavior through log analysis, J. Am. Soc. Inform. Sci. Technol. 56 (2005) 913–934.

[35] T. Murata, K. Saito, Extracting users' interests from web log data, in: 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings) (WI'06), 2006, pp. 343–346.

[36] A. Scott-Wright, et al., Analysis of information needs of users of MEDLINEplus, 2002–2003, AMIA Annu. Symp. Proc. (2006) 699–703.

[37] C. Silverstein, et al., Analysis of a very large web search engine query log, ACM SIGIR Forum (1999).

[38] S. Fox, et al., Evaluating implicit measures to improve web search, ACM Trans. Inf. Syst. 23 (2) (2005) 147–168.

[39] J. Teevan, et al., Information re-retrieval: repeat queries in Yahoo's logs, in: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '07, 2007, p. 151.

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

[41] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Measure. 20 (1) (1960) 37–46.

[42] J.J. Cimino, Use, usability, usefulness, and impact of an infobutton manager, AMIA Annu. Symp. Proc. (2006) 151–155.

[43] B. De Bruijn, et al., Automated information extraction of key trial design elements from clinical trial publications, AMIA Annu. Symp. Proc. (2008) 141–145.

[44] S. Pakhomov, et al., Electronic medical records for clinical research: application to the identification of heart failure, Am. J. Manag. Care 13 (2007) 281–288.

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

[46] 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.

[47] 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.