Log Analysis to Understand Medical Professionals' Image Searching Behaviour[1]

Theodora TSIKRIKA[a,2], Henning MÜLLER[a] and Charles E. KAHN Jr[b]
[a] *University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland*
[b] *Medical College of Wisconsin, Milwaukee, WI, USA*

Abstract. This paper reports on the analysis of the query logs of a visual medical information retrieval system that provides access to radiology resources. Our analysis shows that, despite sharing similarities with general Web search and also with biomedical text search, query formulation and query modification when searching for visual biomedical information have unique characteristics that need to be taken into account in order to enhance the effectiveness of the search support offered by such systems. Typical information needs of medical professionals searching radiology resources are also identified with the goal to create realistic search tasks for a medical image retrieval evaluation benchmark.

Keywords. query log analysis, image retrieval evaluation, query modification

Introduction

Medical professionals increasingly interact with biomedical search engines in order to obtain information that will support their decision making during daily clinical routine, address their research questions, and aid them in preparing teaching files. Examination of their information searching behaviour aims at understanding the usage of such systems and aiding the development of user interfaces and mechanisms that assist users during search; it also benefits system design and medical information retrieval research.

The study of information searching behaviour is typically conducted with surveys and laboratory or field studies. One of the most effective methods, though, is the analysis of users' interactions logged by search engines [1]. Such logging mechanisms are able to unobtrusively record large amounts of the entire range of user-system interactions of a sizable number of users, in naturalistic settings, and over significant time periods. On the other hand, search logs do not record the underlying context and situation of users' searching process, or any other qualitative user aspects, and therefore, users' information needs, decision processes, and satisfaction with the system remain implicit [2]. Despite these inherent limitations, search logs have been widely examined in order to understand users' searching behaviour, mainly in the context of Web search engines [3], but also in the biomedical domain [4-7]. In both cases, though, the focus has been on logs of textual information retrieval systems,

[1] This work was partially supported by the EU in the context of the Promise and Khresmoi FP7 projects.
[2] Correspondence to: Theodora Tsikrika, HES-SO, 3960 Sierre, Switzerland. theodora.tsikrika@acm.org

while only few studies have examined users' behaviour when searching for visual information [7-9].

The study reported here examines the logs of a visual medical information retrieval system that provides access to radiology resources. Our work focusses on the analysis of the queries submitted by medical professionals and investigates the processes of query formulation and query modification, i.e., the processes involved when users engage in an iterative interactive searching process so as to satisfy their information needs. Our aim is to gain insights into the medical professionals' visual information searching behaviour with the goals: (1) to identify the typical information needs of medical professionals searching radiology resources so as to create realistic search tasks for ImageCLEFmed (http://imageclef.org/), a benchmark for evaluating medical image retrieval systems, and (2) to examine the query modification behaviour of users of medical search engines so as to inform the design and improve the effectiveness of the search support offered by such systems. The former has been previously investigated in the context of medical search logs not focussed on radiology resources [10], whereas the latter has not been examined in the context of biomedical search.

Methods

We obtained the query logs from ARRS GoldMiner (http://goldminer.arrs.org/), a medical search engine that provides access to a collection of 250,000 radiology images. Images are indexed and retrieved based on keywords and medical concepts. Results are displayed in a list of up to 10 results per page, each consisting of the retrieved image's thumbnail and caption, or in a grid of up to 40 thumbnail images per page. In particular, we acquired a consecutive sample of 25,000 search requests by medical professionals, each containing the submitted query and the number of results retrieved for that query. No further information, e.g., timestamps or click-through data, was supplied. Moreover, the logs were anonymous; no user information or identifiable IP address were provided.

The query logs were preprocessed as follows. First, the text of the submitted queries was normalised by converting it to lower case and removing punctuation, quotes, and special characters. Then, query terms corresponding to medical imaging modalities were normalised, e.g., `x-ray` and `x ray` were both converted to `xray`. Finally, consecutive identical queries that retrieved the same number of results were conflated into one, since such log entries most likely correspond to the server logging the same entry again when the next result page is viewed or the result display changes.

Log analysis is typically performed at three levels [2]: session, query, and term. Session level analysis provides insights into the complexity and the successfulness of each session, i.e., the series of search interactions performed by a single user in order to satisfy a single information need. The lack of user information in the logs does not allow us to perform session identification; user information is required so as to attribute related queries to the same searcher. Query level analysis focusses on the topics and frequency distribution of queries, on the number of their terms, and on query modifications, i.e., how often and in what way users modify their queries. Term level analysis examines the frequency of terms in queries and the topics of these query terms.

Query modification analysis is performed by identifying the relations between consecutive queries in a search session. We apply a term-based analysis that classifies query modifications based on the overlap between terms in consecutive queries and

examine whether terms have been added, eliminated, or substituted. When terms are added, the search is considered to become more specific (e.g., from query `epidermoid cyst` to query `testicular epidermoid cyst`), more general when terms are eliminated (e.g., from `osteochondrosis dissecans` to `osteochondrosis`), and when terms are substituted a parallel movement (reformulation) is made (e.g., from `ovarian cyst` to `ovarian fibroma`).

Since it is not possible to perform session identification and thus determine when two consecutive queries were submitted in a search session by a single user, we only analyse queries with at least one common term; these are considered to be submitted by a single user. Our analysis focusses on the frequency distribution of the different query modification types over all such pairs of queries, and also in relation to the number of results retrieved by the first query, a likely trigger for query modification. Finally, the frequency distribution of the number of terms that are modified is also examined.

Results

After preprocessing, there were 23,033 queries in the logs, 14,413 (63%) being unique. The average number of terms per query was 2.24 (2.46 for unique queries), while the median was 2 in both cases. This is on average one term less than PubMed queries [6], which target bibliographic information. It is closer to the number of terms in Web search queries [3] and journalist image search queries [9], but slightly less than Web image search queries [8]. Overall, queries are short; around 90% of them have 3 terms or less.

Table 1 lists the most frequent queries submitted and the most frequent terms extracted from them after excluding common stopwords (e.g., “of”, “in”, “the”, etc.). The medical imaging modalities that are searched for are (in brackets their frequency): mri (586), ct (425), ultrasound (199), xray (139), pet (34), pet/ct (13), angiography (13), echo (11), radiography (10), tomography (6), fmri (3), and pet/mri (1). This is of particular interest to search task development for the evaluation of medical visual information retrieval as it provides evidence on the modalities better corresponding to medical professionals' information needs. It is interesting to note the high occurrence of such modalities as part of the query text, despite GoldMiner's filtering functionalities.

Table 1: Most frequent queries and terms and the frequency of their occurrence.

	Query	Frequency		Term	Frequency
1	mega cisterna magna	118	1	cyst	801
2	baastrup disease	80	2	mri	545
3	limbus vertebra	74	3	disease	463
4	negative ulnar variance	67	4	ct	447
5	toxic	65	5	syndrome	438
6	cystitis cystica	50	6	fracture	404
7	throckmorton sign	46	7	sign	359
8	double duct sign	45	8	tumor	322
9	riedel lobe	40	9	bone	294
10	splenic hemangioma	40	10	pulmonary	293

Table 2 presents the results for the 23,032 pairs of consecutive queries that were analysed for determining their relations. Out of these, 17,238 (75%) share no terms, and therefore are not analysed any further since they are considered as unrelated, i.e., as belonging to different search sessions. A very small percentage of them (0.35%) are identical; although such queries were eliminated during the preprocessing of the log

files, these indicate queries that are identical with respect to their query texts, but differ in the number of results retrieved, e.g., due to different filtering parameters. These queries are also eliminated from any further consideration. What is analysed further is the 5,713 query pairs (25%) which share at least one common term and which users modified by adding, removing, or replacing search terms from one iteration to the next. Table 2 shows that term replacement (i.e., query reformulation) occurs roughly twice as often as term removal (i.e., query generalisation) and about three times as often as term addition (i.e., query specification). The large majority of the studies analysing Web search engines logs [9] also find that the most frequently used query modification type is reformulation. However, they find that this is followed by specification and generalisation, different to our case. Given the relatively small size of the sample used in our analysis, further investigation is required so as to determine whether this reflects a particularity of users searching in the biomedical domain.

Table 2: Frequency distribution of term-based query modification types.

Query modification	no common terms	term(s) replaced	term(s) removed	term(s) added	identical	
Number of	17,238	2,946	1,594	1,173	81	23,032
query pairs	74.84%	12.79%	6.92%	5.09%	0.35%	100.00%

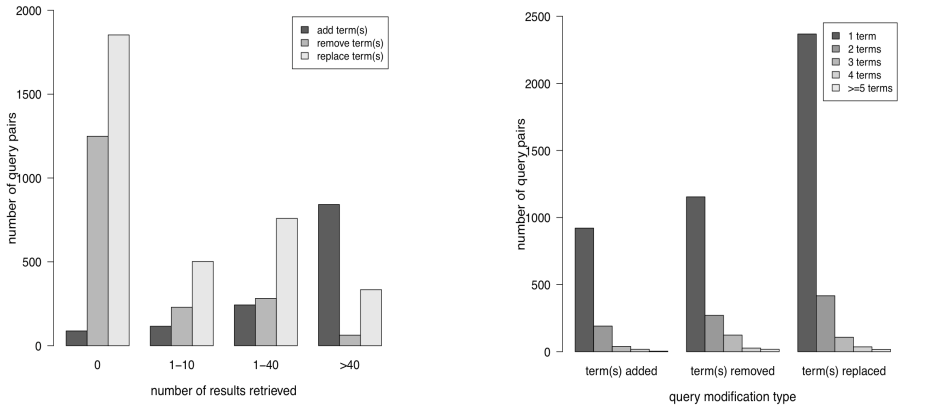

Figure 1: (a) Query modification types based on number of results. (b) Number of terms modified by type.

Next, we examine how users modify their queries based on how many results are retrieved. In particular, four different cases are considered: (i) no results retrieved, (ii) 1-10 results, i.e., when all results fit into the first page in the list display, (iii) 1-40 results, i.e., when all results fit into the first page in the grid display, and (iv) more than 40 results. Figure 1(a) shows that when no results are retrieved, users are more likely to reformulate their query, so as to clarify it (e.g., from `tb empyema` to `tuberculosis empyema`), or to change slightly the focus of their search (e.g., from `inflammatory tumor liver` to `inflammatory pseudotumor liver`), or to perform a spelling correction (e..g, from `fasciits` to `fasciitis`). They are also very likely to remove terms so as to broaden the scope of their search (e.g., from `paediatric brain tumors` to `brain tumors`), whereas they are not very likely to add terms. Similar behaviour is also observed when 1-10 or 1-40 results are returned, with the difference that term additions, although still the least observed, are becoming more frequent. When more than 40 results are returned, users' behaviour changes, as expected, with query specifications becoming more prominent

(e.g., from `fatigue fracture` (236 results) to `fatigue fracture foot` (2 results)), followed by reformulations (e.g., from `aneurysm carotid` (732 results) to `aneurysm brain` (192 results)), whereas generalisations do not occur often. Examination of the distribution of the number of terms added, removed or replaced shows that in the vast majority of cases, one or two terms are modified (Figure 1(b)).

Discussion

This study provides unique insights into medical professionals' visual information needs and searching behaviour. Our findings will guide the creation of realistic search tasks for ImageCLEFmed, a benchmark that strives to advance medical image retrieval. Furthermore, analysis of query modification behaviour can potentially enhance the effectiveness of search support through the query recommendations offered by medical information retrieval systems. Currently, support is mainly provided when no results are retrieved in the form of a correctly spelled variant of the submitted query. However, our analysis indicates that support is required also in other cases, e.g., when too many results are retrieved, and also in other forms, e.g., a query generalisation when no results are retrieved or a query specification when too many results are retrieved. Furthermore, the type of modification may be used by query recommendation approaches that combine several features obtained from co-occurring queries.

Our query modification analysis is limited by the relatively small size of available query logs, the lack of user information for identifying sessions, and the lack of click-through data for determining the modification type leading to successful queries (i.e., queries followed by at least one click). To validate the findings of this initial study, larger samples of GoldMiner's query logs (currently being collected) will be analysed, and a user study on radiologists' visual medical information searching behaviour will be carried out. Furthermore, the major limitation of this term-based method is that it can only classify pairs of queries that share at least one term; it cannot determine the relations between semantically related queries without common terms, such as `renal calculi` and `kidney stones`. We aim to complement it with a semantics-based method that maps queries into ontologies (e.g., MeSH or RadLex) and identifies their query modification type based on their semantic distance. Other types of query modification may be identified by replacing lexical tokens with corresponding part-of-speech categories, revealing the syntactic structures more likely to co-occur in query pairs. Application of such methods in tandem, and in conjunction with a morphological analyser to increase recall, may offer a new understanding of user searching behaviour.

References

[1] Jansen BJ, Spink A, Taksa I, editors. Handbook of research on Web log analysis. Hershey, PA: IGI Global; 2008. 628 p.
[2] Jansen BJ. The methodology of search log analysis. In [1]. p. 1-16.
[3] Jansen BJ, Spink A. How are we searching the World Wide Web? A comparison of nine search engine transaction logs. Inform Process Manag. 2006 Jan; 42(1): 248–263.
[4] Herskovic JR, Tanaka LY, Hersh W, Bernstam EV. A day in the life of PubMed: analysis of a typical day's query log. J Am Med Inform Assoc. 2007 Mar-Apr; 14(2): 212-220.
[5] Bernstam EV, Herskovic JR, Hersh WR, Query Log Analysis in Biomedicine. In [1]. p. 359-377.

[6] Dogan RI, Murray GC, Névéol A, Lu Z. Understanding PubMed user search behavior through log analysis. Database. 2009 Nov; 2009:bap018.
[7] Rubin DL, Flanders A, Kim W, Siddiqui KM, Kahn CE. Ontology-assisted analysis of web queries to determine the knowledge radiologists seek. J Digit Imaging. 2011 Feb; 24:160-164.
[8] Tjondronegoro D, Spink A, Jansen BJ. A study and comparison of multimedia web searching: 1997–2006. J Am Soc Inf Sci Technol. 2009 Sep; 60(9):1756-1768.
[9] Hollink V, Tsikrika T, de Vries AP. Semantic search log analysis: a method and a study on professional image search. J Am Soc Inf Sci Technol. 2011 Apr; 62(4):691-713.
[10] Müller H, Kalpathy-Cramer J, Hersh W, Geissbuhler A. Using MedLine queries to generate image retrieval tasks for benchmarking. In: Proceedings of the 21st MIE; 2008 May 25-28; p. 523-528.