Do users benefit from controlled vocabularies in search interfaces?

Ying-Hsang Liu
School of Information Studies
Charles Sturt University
Wagga Wagga, NSW 2678, Australia
yingliu@csu.edu.au

Paul Thomas
CSIRO
GPO Box 664
Canberra, ACT 2601, Australia
paul.thomas@csiro.au

Jan-Felix Schmakeit
Research School of Computer Science
Australian National University
Canberra, ACT 2601, Australia
jan-felix.schmakeit@anu.edu.au

Tom Gedeon
Research School of Computer Science
Australian National University
Canberra, ACT 2601, Australia
tom@cs.anu.edu.au

ABSTRACT

Search providers in domains from medicine to news have long labelled documents with controlled vocabularies, to help users explore their collections. These vocabularies are expensive to build and use, however, and seem to be useful mostly for domain experts.

This paper describes an on-going gaze-tracking study which asks whether users notice controlled vocabularies when they are exposed in a search interface; whether they make use of them; and whether this improves search. We also hope to learn what effect several standard search interfaces have on the use of controlled vocabularies.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Search process; H.5.2 [**User Interfaces**]: User-centered design—*performance measures*

General Terms

Experimentation, Human Factors

Keywords

Search results presentation, individual differences, gaze behaviour, MeSH terms

1. INTRODUCTION

It has been recognised that people engage with different kinds of searching behaviours, but current information retrieval (IR) systems are primarily designed for specified search [1]. The simple search box is still the dominant interaction mode in modern search engines. However, a user-centred approach to interface design that takes into account individual differences, search goals and tasks, has the potential to support users interacting with IR systems more efficiently and effectively.

To this end researchers have advocated "natural" search user interfaces, arguing they are easier to use and require less user training [e.g. 9, 20]. It is however challenging to design natural interfaces because of the complexity of information problems and associated searching behaviours. For instance, user studies have demonstrated that user queries are typically very short representations of complex information needs [3, 11], and users have difficulty formulating queries to represent information problems. User interaction with IR systems is inherently interactive and exploratory [e.g. 2, 17], so usable interfaces for query formulation are important in support of natural search interactions. (See Wilson [24] for a recent comprehensive review of search interfaces, and Wacholder [22] for a review of interactive query formulation.)

One way to support query formulation is with a controlled indexing language, where each document is assigned terms from an predefined list or hierarchy of indexing terms. Examples include Medical Subject Headings (MeSH) terms and Library of Congress Subject Headings (LCSH). The usefulness of MeSH terms in biomedical searching is especially important because of the extreme popularity of the PubMed database[1], the publicly accessible version of MEDLINE on the web.

Controlled vocabularies are expensive to build, use, and maintain, and they may contribute to clutter in a search interface. There is some evidence that domain experts benefit from controlled vocabularies, but results have been mixed for ordinary users (e.g., [10, 15, 19]). Given these costs, and the unclear benefits for most searchers, we are interested in whether and how users make use of controlled vocabularies when they are available.

[1] `http://www.ncbi.nlm.nih.gov/pubmed`

This paper describes an on-going eye-tracking study of user gaze and search behaviours searching clinical search topics, with particular reference to the user's attention to and use of the document surrogates (i.e., MeSH terms, title and abstract). The specific research questions are:

1. What components of document surrogates do searchers look at when reformulating their queries? Do searchers even notice MeSH terms in standard search interfaces?

2. If they do notice them, how do searchers use the displayed MeSH terms in their search processes?

3. If they are used at all, do MeSH terms lead to better search performance and efficiency?

2. RELATED WORK

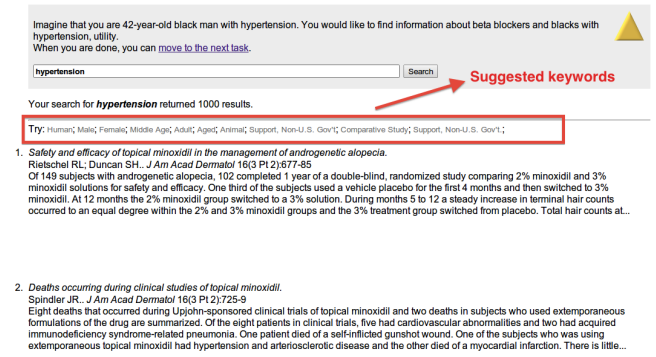
Past work has considered system designs to support query reformulation. From a system perspective, researchers have proposed visualizing document inter-relationships [21], explicit term distribution information [8] and search interfaces in support of search results navigation [18] to help users refine their queries. From a user perspective, research has revealed that searchers prefer to use such search interfaces for reformulating their queries and to have some degree of control over the search process [e.g. 12, 8, 13, 23]. In a recent study of search interfaces in support of interactive query expansion [7], it was found that displaying expanded terms and corresponding changes in summaries of search results was useful for the decision-making process in query reformulation; particularly for difficult search topics. However, it is still unclear whether users pay attention to these system features, and whether the use of these features contributes to better search performance and efficiency.

Recent HCI and IR research has focused on users' cognitive aspects in search interactions by measuring the gaze patterns, an indicator of searcher attention (see e.g. Dumais et al. [5] or Logio et al. [16]). The use of eye-tracking equipment for capturing searchers' fixation patterns provides a rich set of data to understand whether searchers read document surrogates (e.g. summary and metadata) and more importantly, how searchers attend to different components of search results or search interfaces [4, 14]. We are adopting similar techniques in our study.

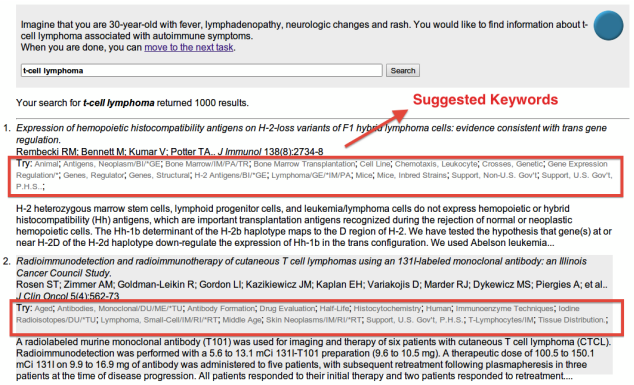
3. METHODS

We are conducting a user experiment to assess the effect of displayed MeSH terms on search behaviors and performance. The search task is to perform searches on clinical information for other patients, and find the best query to obtain as many relevant documents as possible. Our recruits are undergraduate and postgraduate students with search engine experience but without advanced academic background in the biomedical domain. Each user searches 8 topics in total, with a 7-minute limit for each topic, and the experiment takes about 90 minutes in total.

Participants are given brief instructions about the search task and system features, followed by a practice topic and then the searches proper. User interaction data is recorded: we are noting all queries, mouse clicks, retrieved documents, time spent, and eye movements. Electroencephalogram (EEG) readings are also captured.



(a) Screenshot of Interface "B", suggestions per-query and displayed at top.



(b) Screenshot of Interface "D", suggestions per-document and displayed with the document.

Figure 1: Two of the four search interfaces in our study.

3.1 Search interfaces

Participants search on four different search interfaces using a single search system. The four search interfaces are distinguished by whether the MeSH terms are presented and how the displayed MeSH terms are generated:

Interface "A" mimics web search and other search systems with no controlled vocabulary. This interface has a brief task description at top; a conventional search box and button; and each result is represented with its title, authors, publication details, and abstract where available.

Full text is not available, so the results are not clickable. Users must judge their success on the titles and abstracts alone.

Interface "B" (Figure 1(a)) adds MeSH terms to the interface. After the user's query is run, MeSH terms from all results are collated; the most frequent ten are displayed at the top of the screen. This mimics the per-query suggestions produced by systems like ProQuest[2].

MeSH terms are introduced with "Try:" and are clickable: if a user clicks a term, their query is refined to

[2]For example, see `http://www.proquest.co.uk/en-UK/products/brands/pl_pq.shtml`

Imagine that you are 63-year-old male with acute renal failure probably 2nd to aminoglycosides/contrast dye. You would like to find information about acute tubular necrosis due to aminoglycosides, contrast dye, outcome and treatment.

Figure 2: An example OHSUMED search topic, reworded for our participants.

include the MeSH term and then re-run. We hope that the label, and the fact they work as links, will encourage users to interact with them.

Interface "C" uses the same MeSH terms as "B" but displays them alongside each document, where they may be more (or less) visible. It is a hybrid of interfaces "B" and "D".

Interface "D" mimics EBSCOhost[3] and similar systems that provide indexing terms alongside each document. As well as the standard elements from interface "A", interface "D" displays the MeSH terms associated with each document, as part of that document's surrogate (Figure 1(b)).

Again, terms are introduced with "Try:" and are clickable.

Each interface is labelled with a simple figure—a square, circle, diamond, or triangle—which we refer to in our exit questionnaire.

3.2 Design

This experiment is a 4×4 factorial design with four search interfaces and four topic pairs. We are using a 4×4 Graeco-Latin square design [6] to arrange the experimental conditions. We expect to enroll 32 participants from the campus of a large university, which will give good statistical power (when $N = 32$, ANOVA $\beta < 0.01$ for "medium" effect of $\Delta = 0.75$).

Entry and exit questionnaires are collecting demographic information and information on participants' cognitive styles and their perception of the search process. We also ask participants' opinions of the tasks and the interfaces.

3.3 Topics

Search topics used here are a subset of the clinical topics from OHSUMED [10], originally created for batch-mode IR system evaluation. We have re-written the topics slightly so they read as instructions to our participants (see Figure 2 for an example).

We selected topics to cover a range of difficulties: we sorted the topics according to the number of judged relevant documents and selected two topics, at random, from each quartile. These eight topics were then randomly paired off to produce four pairs of topics. A final topic, the same for all participants, is used for training.

3.4 Software and hardware

The search system is built on Solr[4], with the search results ranked by default relevance score. The MeSH terms are not specifically weighted.

Gaze tracking uses FaceLab[5] software and hardware. We use Eyeworks software[6] for recording and basic analysis. EEG data is recorded with an Emotiv headset[7].

3.5 Analysis

With the design above, we expect to answer the three questions from Section 1.

Where do people look?. Recordings will be analysed to see how often there are fixations in different parts of document surrogates, and therefore how often people have looked at each part. In particular, for interfaces B, C and D we will consider how often participants look at the controlled vocabularies ("Try:..."). Any effect on gaze patterns due to interface would tell us which interfaces make the extra information easiest to discover.

Our exit questionnaire also asks whether users noticed the controlled vocabularies: we would not be surprised if there were differences between the self-reported data and the gaze data, for example if participants were trying to please us.

Do they use the controlled vocabulary?. Our software records all clicks on terms from the controlled vocabulary, so it will be easy to note how often it is used and whether there is any correlation with interface, task, sequence, or user. Again, an effect due to interface would suggest which style of interface makes features like the controlled terms most attractive.

Participants who merely read and re-type the controlled vocabulary may be picked up in query logs.

Again, we intend comparing these recordings with self-reports.

If so, does it help?. Assuming some participants do make use of the MeSH terms, we anticipate four ways to address this question. First, as before, we will consider self-reports of task difficulty to see whether these correlate with the use of controlled vocabulary features. Second, since participants' final queries on each topic should be the ones the like best, we can check how many of these use MeSH terms. Third, the judgements associated with OSHUMED topics will allow us to measure the actual effectiveness of queries with and without controlled terms. Finally, if participants do not use all their allocated time for each task, variations in completion time may be interesting.

4. FIRST RESULTS AND NEXT STEPS

We have conducted a small-scale pilot to test our design and instruments.

Our participants did glance at MeSH terms: 8% of fixations were on MeSH terms in interfaces B to D, which compares to 6% on document titles and 12% on abstracts. However, they were very seldom used – only one query, of 44 queries issued on these interfaces, used any MeSH terms at all. There are also some indications of a per-interface effect, with the MeSH terms at the top of interface D receiving little attention We will shortly be recruiting for the full-scale experiment. We hope this will offer some insight into the relationship

[3] http://www.ebscohost.com/
[4] http://lucene.apache.org/solr/
[5] http://www.seeingmachines.com/product/facelab/
[6] http://www.eyetracking.com/Software/EyeWorks
[7] http://www.emotiv.com/

between interface, reading patterns, search behaviour, and search effectiveness.

5. ACKNOWLEDGMENTS

Ying-Hsang Liu has been supported by the School of Information Studies Research Fellowship from Charles Sturt University and working as Visiting Fellow at Research School of Computer Science, The Australian National University.

6. REFERENCES

[1] N. J. Belkin. Some(what) grand challenges for information retrieval. *SIGIR Forum*, 42(1):47–54, 2008.

[2] N. J. Belkin, P. G. Marchetti, and C. Cool. Braque: Design of an interface to support user interaction in information retrieval. *Information Processing and Management*, 29(3):325–344, 1993.

[3] N. J. Belkin, R. N. Oddy, and H. M. Brooks. ASK for information retrieval: I. Background and theory. *Journal of Documentation*, 38(2):61–71, 1982.

[4] E. Cutrell and Z. Guan. What are you looking for?: An eye-tracking study of information usage in web search. *Proceedings of the SIGCHI Conference*, pages 407–416, 2007.

[5] S. T. Dumais, G. Buscher, and E. Cutrell. Individual differences in gaze patterns for web search. *Proceeding of the Symposium on Information Interaction in Context (IIiX '10)*, 3:185–194, 2010.

[6] R. A. Fisher. *The design of experiments*. Hafner Press, 9th edition, 1971.

[7] N. Gooda Sahib, A. Tombros, and I. Ruthven. Enabling interactive query expansion through eliciting the potential effect of expansion terms. *Lecture Notes in Computer Science*, 5993:532–543, 2010.

[8] M. A. Hearst. Tilebars: Visualization of term distribution information in full text information access. *Proceedings of the SIGCHI Conference*, pages 59–66, 1995.

[9] M. A. Hearst. 'Natural' search user interfaces. *Commun. ACM*, 54(11):60–67, 2011.

[10] W. Hersh, C. Buckley, T. J. Leone, and D. Hickam. Ohsumed: An interactive retrieval evaluation and new large test collection for research. *Proceedings of the ACM SIGIR Conference*, 17:192–201, 1994.

[11] B. J. Jansen, A. Spink, and T. Saracevic. Real life, real users, and real needs: A study and analysis of user queries on the web. *Information Processing and Management*, 36(2):207–227, 2000.

[12] D. Kelly and X. Fu. Elicitation of term relevance feedback: An investigation of term source and context. In *Proceedings of the ACM SIGIR Conference*, pages 453–460, New York, 2006. ACM.

[13] J. Koenemann and N. J. Belkin. A case for interaction: A study of interactive information retrieval behavior and effectiveness. *Proceedings of the SIGCHI Conference*, pages 205–212, 1996.

[14] B. Kules and R. Capra. Influence of training and stage of search on gaze behavior in a library catalog faceted search interface. *Journal of the American Society for Information Science and Technology*, 63(1):114–138, 2012.

[15] Y.-H. Liu and N. Wacholder. Do human-developed index terms help users? an experimental study of MeSH terms in biomedical searching. *Proceedings of the American Society for Information Science and Technology Annual Meeting*, 45(1):1–16, 2008.

[16] L. Lorigo, M. Haridasan, H. Brynjarsdóttir, L. Xia, T. Joachims, G. Gay, L. Granka, F. Pellacini, and B. Pan. Eye tracking and online search: Lessons learned and challenges ahead. *Journal of the American Society for Information Science and Technology*, 59(7):1041–1052, 2008.

[17] G. Marchionini and R. White. Find what you need, understand what you find. *International Journal of Human-Computer Interaction*, 23(3):205–237, 2007.

[18] X. Mu, H. Ryu, and K. Lu. Supporting effective health and biomedical information retrieval and navigation: A novel facet view interface evaluation. *Journal of Biomedical Informatics*, 44(4):576–586, 2011.

[19] M. L. Nielsen. Task-based evaluation of associative thesaurus in real-life environment. *Proceedings of the American Society for Information Science and Technology Annual Meeting*, 41:437–447, 2004.

[20] K. A. Olsen and A. Malizia. Interfaces for the ordinary user: can we hide too much? *Commun. ACM*, 55(1):38–40, 2012.

[21] R. C. Swan and J. Allan. Aspect windows, 3-D visualizations, and indirect comparisons of information retrieval systems. *Proceedings of the ACM SIGIR Conference*, 21:173–181, 1998.

[22] N. Wacholder. Interactive query formulation. *Annual Review of Information Science and Technology*, 45:157–196, 2011.

[23] R. W. White and G. Marchionini. Examining the effectiveness of real-time query expansion. *Information Processing and Management*, 43(3):685–704, 2007.

[24] M. L. Wilson. Search user interface design. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 3(3):1–143, 2011.