Visualizing Search Sequences

Jacek Gwizdka
Department of Library and Information Science, Rutgers University
New Brunswick, NJ, USA
asist2011@gwizdka.com

ABSTRACT
This video presents a novel visualization technique of interactive search sessions. The objective of this work is to enable characterization and comparison of interactive search sessions with respect to strategies and tactics employed by different people and on different search tasks. The visual aspect of this approach aims to off-load cognition by shifting part of the required processing to perception enabling thus information science researchers to obtain quick overview of search sessions. The video explains the technique and its applications on examples created from data collected in a controlled Web search study.

Keywords

Search tasks, interactive search sessions, search sequences, sequence comparison, visualization

INTRODUCTION & MOTIVATION

Human behavior has been characterized by employing top-down perspective starting from goals/intentions, through plan formulation, tasks, subtasks, to actions. Similarly, models of human information seeking in information science characterize searcher's behavior by describing states or stages through which a searcher is progressing. Virtually all search models involve the notion of states (Wilson, 1999; Sarcevic, 1996; Ingwersen, 1996). The states are frequently further described by attributes and by actions that are typically conducted while in these states.

Search episodes were frequently described as series of states, sequences of search moves or searcher's actions. In particular, search tactics were described as individual search actions (Bates, 1979) and as a set of (temporally or semantically) related search moves. (Wildemuth, 2004). As Wildemuth pointed out "it is […] important to examine the sequences of moves made by searchers in order to understand the cognitive processes they use in formulating and reformulating their searches." Search tactics were described when information search was dominated by query-based search in specialized databases. New information technologies brought new interaction possibilities. One of the aims of our work is to broaden a set of search moves.

The importance of search state sequences motivates us to consider their new representations and, specifically, a visual representation that could support information scientists' work. The aims for creating this representation are to enable:

- characterization of users and search tasks by describing sequences of search moves and actions;
- comparative description of search sequences;
- relating task and user characteristics to sequences of moves.

User moves and actions can be considered at different levels. The lower level sequences (e.g., clicking on a link) can be aggregated into higher-level units. The simple task phase units that we use in our visualization can be used to create higher-level constructs that are closer to representing user information search strategies and their cognitive states.

METHOD

Experiment

Visualization presented in this paper and video use data collected in a controlled Web search experiment. In this short paper, we include only selected information that is useful for better understanding of our approach. For the full description of the study and the task phase data processing the reader is referred to Gwizdka (2010).

The experiment (N=48) involved performing 6 search tasks by each subject for a total of 288 search sessions. The tasks were designed to differ in the level of difficulty.

Task Phase Segmentation and Visualization

Task segment identification was based on the observable physical actions logged during search sessions (e.g., keyboard activity) and on the type of web page visited. Web page URLs were parsed and classified into four categories: 1) search engine home, 2) search results list page, 3) individual result page (content page), 4) bookmarking and tagging page. (Table 1). Bookmarking is an instance of a broader class of information keeping actions performed to make content findable or re-findable by the searcher and others. Other such actions might include saving a document, sharing, rating and commenting on a document.

ASIST 2011, October 9–13, 2011, New Orleans, LA, USA.

Table 1 Task Phases.

Phase code	Task phase name	Description
Q	Query formulation	Search query is formulated or re-formulated.
L	Examination of search result list (SERP)	In a response to entered query, a search engine results list is displayed and examined.
C	Examination of an individual result (content)	Individual result (content page) is visited and is visually scanned or read.
B	Bookmarking and tagging a relevant result	As information is extracted from individual results, decision about its relevance is made. Content documents judged as relevant are saved by means of bookmarking and tag description.

APPROACH

We treat the four task phases (QLCB) as fundamental units in our visualization. We mark these search task phases with the respective letters Q, L, C, B. We choose to represent them on a two-dimensional plane as vectors. Each vector's direction and color represents (uniquely) a different task phase (Figure 1). As the sequence continues, a new vector representing a current task phase is added to the end of the previous phase. Placing the C and B vectors at and angle ensures that the elements of the representation will not overlap. Next to vectors are numbers that uniquely identify pages or documents. They can be used to identify returns to the same search result lists and to the same documents.

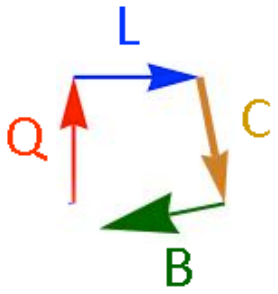

Figure 1. Basic elements of the visualization

Our ideas are best presented using examples; they follow in Figures 2-6. All examples are based on empirical data.

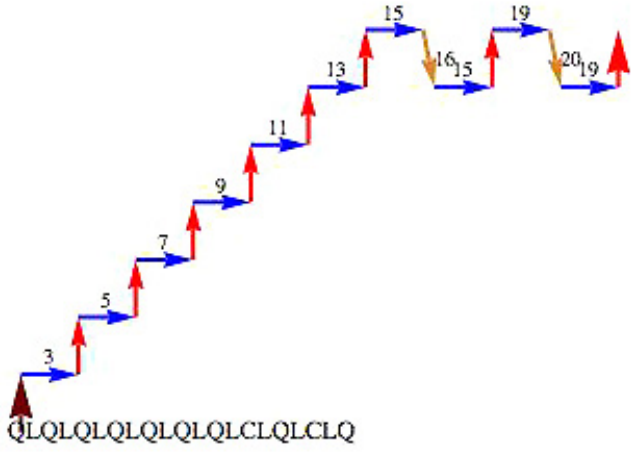

Figure 2. Upward slope pattern.

This pattern shows that a series of queries (original query and subsequent query reformulations) were entered and that for each a result list was examined. The brownish, downward pointing arrows appearing towards the end of the sequence represent reading individual documents. The overall shape of this kind of pattern has an upward slope. This pattern can be denoted as $(QL)^i$

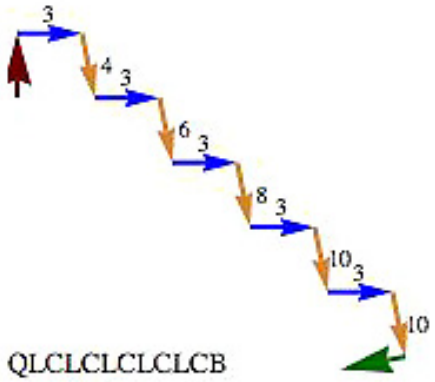

Figure 3. Downward slope pattern – hub-&-spoke.

In this pattern (Figure 3) one can see that user entered only one initial query, and then examined the search results list opening a series of content documents. This behavior is a hub-and-spoke movement, where a user returns to the same results list and opens next promising document. The typical downward slope of this pattern can be used to quickly indentify this type of user behavior. This pattern can be denoted as $(LC)^i$

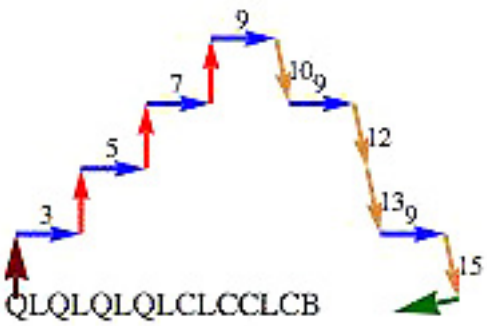

Figure 4. Upward and downward slope pattern.

This pattern (Figure 4) combines the two patterns shown in Figure 2 and 3. This pattern can be denoted as $(QL)^i$ $(LC)^i$

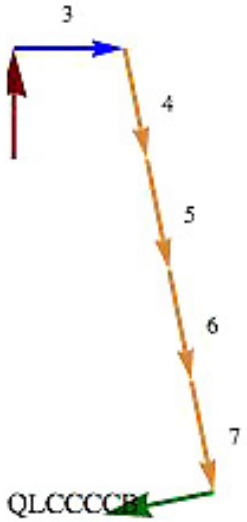

Figure 5. Downward slope – browsing documents.

Another typical user behavior is represented by the pattern of a vertical shape (Figure 5). In this case, a user entered an initial query, examined one search results page, and then started browsing documents by following links from one to another, eventually deciding to bookmark one of the documents. A good part of a search session of this kind involves browsing. As with previous examples, indentifying this pattern as a part of a search episode can help in understanding the searcher's behavior. This pattern can be denoted as $(C)^i$

The next example (Figure 6) presents visualizations of search processes of two users performing the same search task and using the same search interface. The differences in search strategies employed by these users are immediately apparent. The user on the left tends to issue few queries and visit more documents by following search result links, while the user on the right tends to issues more queries and visit fewer result documents.

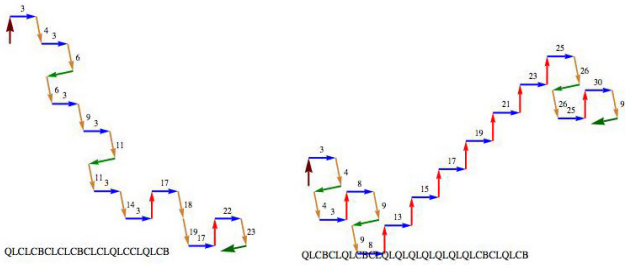

Figure 6. Search session patterns of two users on the same task and using the same search interface.

SUMMARY

We presented a novel visualization of search sequences. The technique is meant to complement, but not replace, quantitative methods of similarity calculation. This visualization allows for a quick overview of differences in user search task performance. The technique also enables possible identification of issues users might be encountering in the search process. More importantly, the technique can be helpful in identifying higher-level strategies and in detecting when people switch their search strategies.

ACKNOWLEDGEMENTS

This work was influenced not only by the work of the earlier mentioned information scientists, but also by researchers working on visualization of DNA sequences (Keogh et al, 2005) and on novel ways of calculating DNA similarity (Yu et al., 2011).

This research was supported, in part, by IMLS grant LG-06-07- 0105-07.

VIDEO

The video that accompanies this short paper has been uploaded to YouTube and is available at: http://bit.ly/ASIST2011video

REFERENCES

Bates, M.J. (1979). Information search tactics. *Journal of the American Society for Information Science*, 30(4), 205–214.

Bates, M.J. (1987). How to use information search tactics online. *Online*, 11(3), 47–54.

Gwizdka, J. (2010). Distribution of cognitive load in web search. Journal of the American Society for Information Science & Technology, 61(11), 2167 - 2187.

Ingwersen, P. Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. Journal of Documentation, 52(1), 1996, 3–50.

Keogh, E., et al. (2005). Visualizing the Similarity of Human and Chimp DNA. (Multimedia Video). http://www.cs.ucr.edu/~eamonn/DNA/

Saracevic, T. Modeling interaction in information retrieval (IR): a review and proposal. In: Hardin, S., ed. 59th ASIST. Silver Spring, MD: ASIST, 1996, 3–9.

Wildemuth, B. M. (2004). The effects of domain knowledge on search tactic formulation. Journal of the American Society for Information Science and Technology, 55(3), 246-258.

Wilson, T. D. (1999). Models in information behaviour research. Journal of Documentation, 55(3), 249-270.

Yu, C., Deng, M., & Yau, S. S.-T. (2011). DNA sequence comparison by a novel probabilistic method. Information Sciences, 181(8), 1484-1492. doi:10.1016/j.ins.2010.12.010