Quantitative Analysis of Search Sessions Enhanced by Gaze Tracking with Dynamic Areas of Interest

Tuan Vu Tran and Norbert Fuhr

University of Duisburg-Essen
Information Engineering
Duisburg, Germany
vtran@is.inf.uni-due.de
norbert.fuhr@uni-due.de

Abstract. After presenting the ezDL experimental framework for the evaluation of user interfaces to digital library systems, we describe a new method for the quantitative analysis of user search sessions at a cognitive level. We combine system logs with gaze tracking, which is enhanced by a new framework for capturing dynamic areas of interest. This observation data is mapped onto a user action level. Then the user search process is modeled as a Markov-chain. The analysis not only allows for a better understanding of user behavior, but also points out possible system improvements.

1 Introduction

The analysis of user search behavior has a long tradition. The major focus of this research has been the development of cognitive models [5], which improved our understanding of the user's actions, but also has been considered in more recent designs for developing user-friendly systems.

On the other hand, quantitative analyses mostly have focused on system log analyses, but it is difficult to apply user-oriented criteria when regarding this data. More valuable information can be derived from observation data if we are able to relate this data to cognitive actions of the user, e.g. via intellectual mapping in the form of a post-hoc analysis (see e.g. [4]).

Log file studies have also been described by Jansen [6], Jones [8], Zhang and Kamps [11] who concluded that transaction log analysis can provide valuable insights into user-system interactions. Hoare et al. [4] developed a model of states and transitions that users traverse during a search session to understand user search processes. This model allows predictions of how changes to the interface might change user's behavior and thus enables comparisons between designs. Kumpulainen and Järvelin [10] modeled interaction between different information channels to optimize the process of task-based information access.

In recent years, gaze tracking has become a valuable tool for capturing implicit feedbacks from the users. Studies like [3], [7] and [1] analyze gaze-tracking data to understand the user search process.

In the remainder of this paper, we first present the ezDL experimental framework for the evaluation of user interfaces to digital library (DL) systems. Then we describe a new quantitative analysis method based on logging and gaze tracking; by combining these two sources of evidence, we are able to perform an automatic mapping onto cognitive actions. As a major improvement over previous approaches using gaze tracking, we are able to consider dynamic areas of interest, i.e. display objects that appear only for a short period of time or change their position during the search session.

2 An experimental framework for DL user interfaces

ezDL[1], the successor of Daffodil [9] is a configurable open source software for the design of advanced user interfaces for DLs and their evaluation. For the user, ezDL provides a desktop interface offering a configurable set of tools. Current tools comprise functions such as database selection, query formulation, result lists with sorting, grouping and clustering options, detail view, co-author graphs, search history, basket (clipboard) and a personal library for storing retrieved items in folders.

ezDL supports user-oriented experimentation at all levels of a DL: in the *user interface* existing tools can be replaced by variants with different designs or visualizations. At the *functional* level, new tools can be easily integrated into the system. ezDL can be connected with arbitrary *back-ends* via adapters and wrappers for retrieval systems (e.g. Solr), digital library systems or web-accessible digital libraries (currently, we have wrappers for about 10 libraries from computer science and medicine). Finally, at the *collection* level, experiments with new collections can be performed by running them with one of the back-ends currently supported.

Besides being a flexible, extensible system, ezDL comes with specific components for the user-oriented evaluation of digital library systems. The core ezDL system consists of the front-end (running on the user's computer), the back-end (running on a server) and the IR engine(s) managing the content. For running lab experiments, there is an experiment database that contains data for controlling the experiments and for collecting observation data. The former is used by an additional web application which schedules the user tasks (e.g. based on a latin square design) and starts the ezDL front-end, but also presents questionnaires to the users (e.g. before and after each task, as well as at the beginning and the end of an experimental session). As observation data, the experiment database collects the answers to the questionnaires as well as the logging data from the ezDL system, which can be configured according to the desired granularity. In addition, we also gather data from a gaze tracking tool.

The analysis of gaze tracking data is usually based on the definition of 'areas of interest' (AOI), which are rectangular areas on the screen like e.g. query interface, result items or details of a document. However, standard gaze-tracking

[1] www.ezdl.de

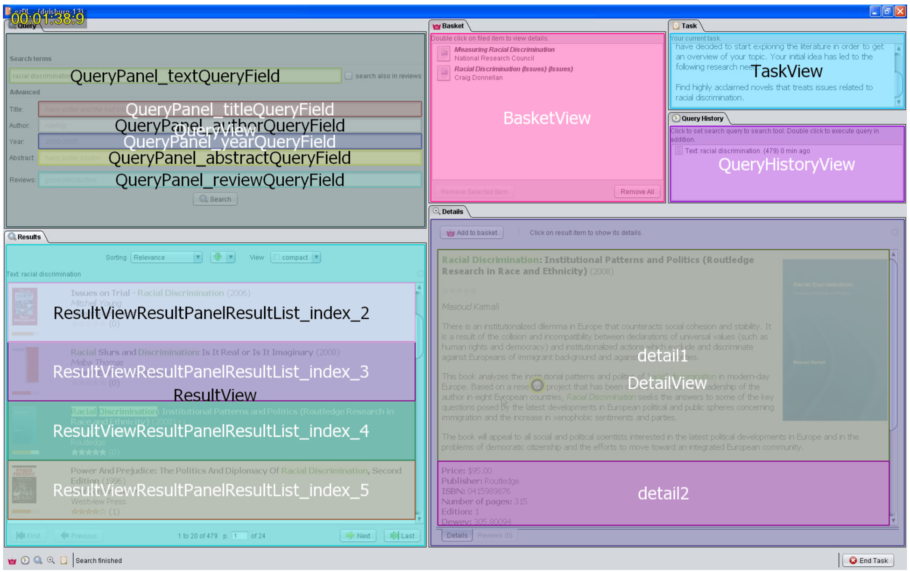

Fig. 1. User interface and areas of interest

software only allows for the monitoring of static AOIs. In contrast, users might like to resize the windows displayed on the screen or scroll through lists of items, and there also may be pop-up boxes. In order to be able to deal with these issues, we developed a framework called *AOILog*[2] which automatically keeps track of position, visibility and size of all user interface objects at any point in time. By combining this information with the gaze-tracking data, we always know the object the user is currently looking at. As an example, see the AOIs depicted in Figure 1; in the left lower window, each result item corresponds to one (dynamic) AOI, where the system keeps track of the items even during scrolling.

3 An example evaluation

For our experiments, we used a collection based on a crawl of 2.7 million records from the book database of the online bookseller Amazon.com. The data was crawled from February to March 2009 and indexed with Apache Solr 1.4. Besides the bibliographic data and the abstracts, a document record also contains the user reviews and a thumbnail of the book cover.

As test subjects, we recruited 12 students of computer science, cognitive and communication science and related fields. After an introduction into the system, users had to work on two tasks from each of the two categories, with a time limit of 15 minutes per task. The ‘complex’ tasks defined searches where users had to

[2] *AOILog* is open source software based on Java; since components only have to register for using it, it can be easily integrated into any user interface based on Java Swing.

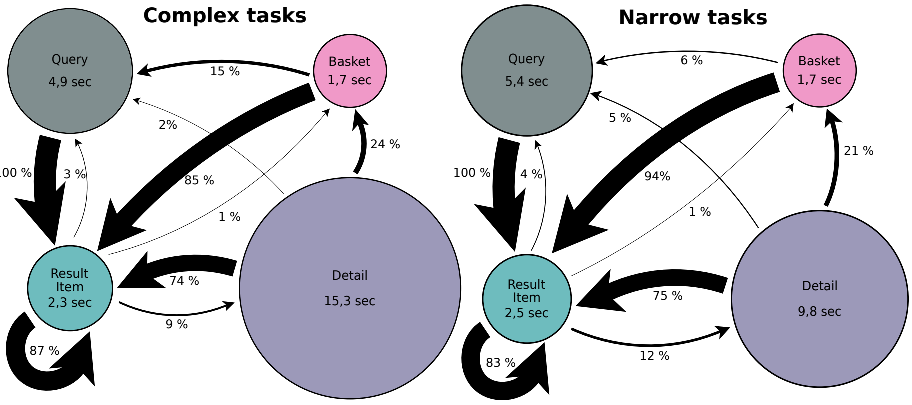

Fig. 2. Transition probabilities and user efforts

look into the user reviews of a book in order to judge about its relevance, while the 'narrow' tasks usually could be solved by reading the abstracts.

The user interface of our system consists of four major areas (see Figure 1): a query input field, a list of result items, a detail area showing all available data about the currently selected document, and a basket where users should place documents they deemed relevant. There are also two small areas showing the search history and the current working task. All windows are resizeable.

As mentioned above, our analysis is based on the combination of logging and gaze tracking data. For the latter, as in other studies, we focus on the so-called fixations, and also consider them only is they last for at least 80 ms, since this is the minimum time required for reading anything on the screen [2].

Corresponding to the four areas of the user interface, we can distinguish four types of user actions: formulating *queries*, looking at a *result item*, regarding document *details*, and looking at the *basket*. Thus, a one-to-one mapping of gaze tracking data and user actions seems to be straightforward. However, a closer analysis showed that this is too simplistic.

First, we noted that users were frequently looking back and forth between the details and the basket. This behavior was due to the fact that the book database often contains very similar entries (e.g. different editions of a book), thus forcing users to check if the current result item was substantially different from the books already placed in the basket. Occasionally, this also happened for items in the result list. In a similar way, when formulating a new query, users did not only look at the query field, but also checked the details of the current document as well as the items in the basket.

In order to deal with these problems, we separate between two levels, the gaze tracking level and the action level. Then we define a mapping from the former to the latter, which also considers the logging data. By default, the area the user looks at defines the current action, with the following exceptions:

- When the user looks from a result item to the basket and back, without moving the item to the basket, this is counted as part of regarding the item.
- The same holds for looking back and forth between details and the basket.
- Query formulation starts when the user's gaze wanders from the details or the basket to the query field for the first time, even when it returns to the basket/details several times before the query is actually submitted.

Applying these rules resulted in the transition probabilities and average times spend for the different actions displayed in Figure 2. On the left hand side, we show the Markov model for the complex tasks. One can see that users spend 4.9 s for formulating a query, after which they go through the result list. Each result item takes 2.3 s, and for only 9% of these items, users also look at the corresponding details, which takes another 15.3 s on average. In 24% of all details considered, the document is judged relevant and put into the basket, in 74% of the cases the document is considered not relevant and users go to the next result item, while with a probability of 2%, a new query is formulated. From the basket, users most often go to the next result item (85%), or they formulate a new query (15%).

Looking closer at this model, a number of observations can be made:

- Retrieval quality is surprisingly low—only about 3 % of the items in the result list are relevant. This is mainly due to the complexity of the retrieval tasks.
- After putting a document into the basket there is a probability of 15% that a new query is formulated. Most of the time, users go to the next result item, this may indicate that they hesitate to formulate new queries. A query expansion function could help here.
- Since only 24 % of the details regarded are judged as relevant, the quality of the result entries could be improved, in order to increase this rate (while the probability of the transition from result to details would decrease, thus leading to a reduction of the overall effort). Furthermore, we also see that click-through data is a poor indicator of relevance.

When we compared the statistics of complex vs. narrow tasks, we see that the number of queries (7.1 vs. 8.7) as well as the number of items placed in the basket (8.1 vs. 7.1) were about the same for the two task categories; for the complex tasks users looked at more items (88 vs. 78) but fewer details (17 vs. 25) than for the narrow tasks. However, users spend 50% more time per detail for the complex tasks (15.3 s vs. 9.8 s). This can be explained by the fact that complex tasks require the users to investigate more details about potential answers.

4 Conclusion

In this paper, we first presented the ezDL experimental framework for user-oriented evaluation of DL systems. Then we described a new quantitative analysis method which extends current methodologies for analyzing interactive IR by combining system logs with gaze tracking, which we enhanced by dynamic areas of interest. In order to relate this observation data to the user's cognitive actions, we separate between the gaze tracking and the action level, and introduce a corresponding mapping. Our example study points to possible system improvements, and shows that click-through rates are a poor indicator of relevance.

References

1. G. Buscher, E. Cutrell, and M. R. Morris. What do you see when you're surfing?: using eye tracking to predict salient regions of web pages. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 21–30, New York, NY, USA, 2009. ACM.
2. G. Buscher, A. Dengel, and L. van Elst. Eye movements as implicit relevance feedback. In *CHI '08 : extended abstracts on Human factors in computing systems*, pages 2991–2996, New York, NY, USA, 2008. ACM.
3. Z. Guan and E. Cutrell. An eye tracking study of the effect of target rank on web search. In *CHI '07 : Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 417–420, New York, NY, USA, 2007. ACM.
4. C. Hoare and H. Sorensen. Application of session analysis to search interface design. In *Proc. of the 14th European conference on Research and advanced technology for digital libraries*, ECDL'10, pages 208–215, Berlin, Heidelberg, 2010. Springer-Verlag.
5. P. Ingwersen and K. Järvelin. *The turn: integration of information seeking and retrieval in context.* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.
6. B. J. Jansen. Search log analysis: What it is, what's been done, how to do it. *Library & Information Science Research*, 28(3), 2006.
7. T. Joachims, L. Granka, B. Pan, H. Hembrooke, F. Radlinski, and G. Gay. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Trans. Inf. Syst.*, 25(2), 2007.
8. S. Jones, S. J. Cunningham, R. Mcnab, and S. Boddie. A transaction log analysis of a digital library. *International Journal on Digital Libraries*, 3:152–169, 2000.
9. C.-P. Klas, N. Fuhr, and A. Schaefer. Evaluating strategic support for information access in the DAFFODIL system. In *Proc. ECDL 2004*, pages 476–487, 2004.
10. S. Kumpulainen and K. Järvelin. Information interaction in molecular medicine: integrated use of multiple channels. In *Proc. of the third symposium on Information interaction in context*, IIiX '10, pages 95–104, New York, NY, USA, 2010. ACM.
11. J. Zhang and J. Kamps. A search log-based approach to evaluation. In *Proc. of the 14th European conference on Research and advanced technology for digital libraries*, ECDL'10, pages 248–260, Berlin, Heidelberg, 2010. Springer-Verlag.