

Overview and Results of the INEX 2009 Interactive Track

Thomas Beckers¹, Norbert Fuhr¹, Nils Pharo², Ragnar Nordlie², and
Khairun Nisa Fachry³

¹ University of Duisburg-Essen, Germany
{tbeckers, fuhr}@is.inf.uni-due.de

² Oslo University College, Norway
{nils.pharo, ragnar.nordlie}@jbi.hio.no

³ University of Amsterdam, The Netherlands
k.n.fachry@uva.nl

Abstract. We present results of the INEX 2009 Interactive Track which focussed on how users behave in interactive search systems. Three types of working tasks based on a collection of book metadata were regarded. The results show differences with respect to the task types and point out improvements and new research questions for the next track in 2010.

1 Introduction and Research Questions

The INEX Interactive Track (iTrack) is a cooperative research effort run as part of the INEX Initiative for the Evaluation of XML retrieval¹. The overall goal of the iTrack is to investigate how users behave in interactive search systems. In the 2009 run of this track [1] the focus was on what aspects of documents the users are interested in, how they make use of various search tools, and finding out new challenges for the coming iTracks. We created a collection based on a crawl of 2.7 million records from the book database of the online bookseller Amazon.com, consolidated with corresponding bibliographic records from the cooperative book cataloging web site LibraryThing.

In this paper, we provide an analysis of the collected logging and questionnaire data and point out challenges for the next run in 2010.

2 System Description

The search system (see fig. 1) was developed at the University of Duisburg-Essen. It is based on Daffodil [2] and partially on ezDL² while the retrieval component was implemented using Apache Solr. The client application part is based on a tool metaphor. Six tools (rounded rectangles in fig. 1) are available and are described in more detail below.

¹ <http://www.inex.otago.ac.nz/>

² <http://www.ezdl.de>, <http://www.is.inf.uni-due.de/projects/ezdl/>

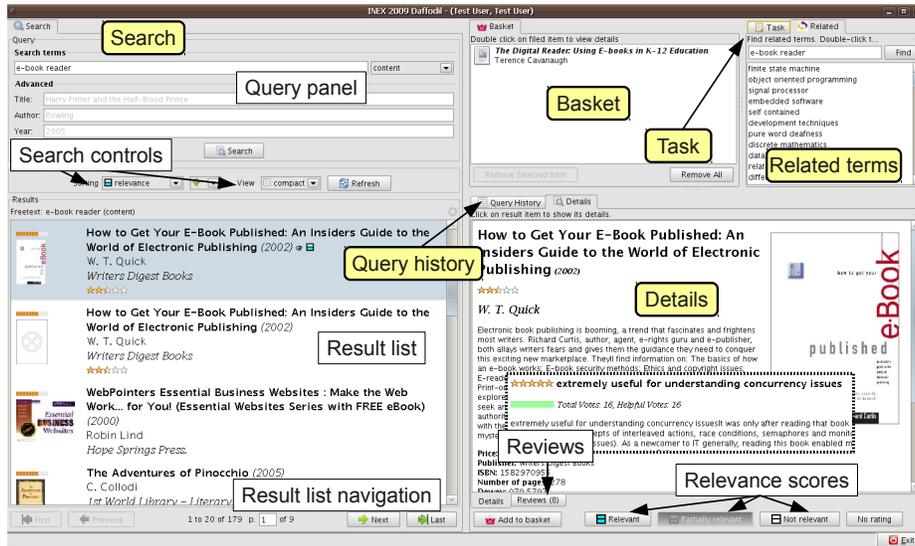


Fig. 1: The client desktop application of the iTrack search system

The **search tool** offers a Google-like search field as well as advanced search fields for title, author, and year. Below the query panel, the user can select fields for sorting or changing the display style of the result list. The lower half of the search tool contains the result list. The default surrogate contains the title, authors, year, publisher, average customer rating and a thumbnail of the book front cover. Each page of the result list contains 20 result items. The user can use the buttons at the bottom to navigate to other pages of the result list.

A double-click on a result item shows book details in the **detail tool**. Users can indicate the relevance of any book which is examined either as *Relevant*, *Partially relevant*, or *Not relevant*, by clicking markers at the bottom of the tool. A second tab shows reviews of the selected book.

The user can select any book as part of the answer to the search task by moving it to the **basket tool**. This can be done by drag-and-drop or by clicking the *Add to basket* button next to the relevance buttons.

A history of performed search queries is stored in the **query history tool**. The **related terms tool** presents terms related to those used in the search query. A search for related terms can also be triggered manually by the user. Finally, the **task tool** shows the current working task.

3 Evaluation Design

For this track, 41 volunteers were recruited mostly from students of computer science, cognitive and communication science, library science and some other related fields. 24 of them were male and 17 of them female. Their average age was

about 28. On average, they have used the Internet for 9.5 years. All participants had experiences with web search engines, searching in digital libraries or digital bookstores.

There are three different task groups, namely *broad tasks* (I), that require exploratory search behaviour, *narrow tasks* (II), that are about a relatively narrow topic, and a *self-selected* (III) task about finding a single book for a course the volunteers were currently attending. The first two task groups consist of three concrete working tasks of which one had to be chosen by the participant. Pharo et al. [1] provide a more detailed description of the working tasks.

At the beginning, the participants were asked to fill out a pre-experiment questionnaire. Each task was preceded by a pre-task and concluded by a post-task questionnaire. After all three working tasks had been worked on, a post-experiment questionnaire was answered. Actions by the system and the participants were recorded by the system and stored in a database.

4 Results

The participants were given the possibility to express positive as well as negative general comments on the questionnaires. The user interface was praised because it is well arranged and everything fits on a single screen without the need to scroll up or down. The inclusion of another document aspect, namely the reviews, was also pointed out positively by the participants.

Technical problems of the search system and sometimes useless related terms suggestions due to topical limitations of the data source were points of criticism. Participants also missed highlighting of query terms and filtering options for the result list. Also, the heterogeneity of the data was disliked, that is, some books have lots of metadata while other have just very little of it.

Table 1 shows the average event count per session for each task group. In task group I the details of books were requested more often than for other task groups while only slightly more searches were performed. The participants looked at the reviews for only a small part of the books they viewed details of. The related terms tool and the query history tool were used rarely. Less often than once per session a related term was searched/added to the query or a previous search was executed again.

The average length of a session decreased from 829s (I) and 725s (II) to 622s (III). The average duration of a search (that is, the time between two queries) was 99s (I), 92s (II), and 83s (III). For more explorative working tasks, longer and more searches were performed.

Event	I	II	III
Search	8.07	7.46	7.13
Next Search Results	2.66	1.93	2.63
View Details	23.78	15.44	16.05
View Review Details	3.66	4.61	3.35
Relevance Rating	15.66	9.39	9.33
Added To Basket	7.02	5.34	2.38
Sorting and View Changed	1.68	0.56	1.15
Related Terms Search	0.44	0.27	0.08
Related Term Selected	0.85	0.17	0.25
Query from History Selected	0.41	0.61	0.65

Table 1: The average count per task session of the most important events

At the end of a session the participants had 6.61 (I), 4.96 (II), and 1.15 (III) books respectively in their basket. The participants collected more books for the broad tasks than for the narrow tasks. The average query length (number of terms in the simple query field) was 1.99 (I), 2.46 (II), and 2.21 (III). The broadest tasks resulted in the shortest average query length.

The most frequent event transitions are similar for each task groups. The most frequent transition is *Visible Results Changed* → *View Details* which means that after results have been displayed or after the participant has scrolled the result list he/she views details of a book in the results. The second most frequent transition *View Details* → *Relevance Rating* was generated if the participants requested details and then assessed the relevance of the book with regard to the working task.

Overall, advanced tools such as the *related terms tool* or the *query history tool* were used less often than expected. It was often not necessary for the participants to use these tools to work successfully on the tasks. In our opinion the main reason is that the working tasks focussed too much on problems obvious from the title of a book. This was especially the case for task group III. Many participants understood this task group as a known-item search instead of a thoughtful search for a single unknown book as it was actually intended. The reviews were judged as a useful aspect but the participants mostly concentrated on the details only.

5 Conclusion and Outlook

The search system was well received by the participants. The use of metadata of real and recent books (until March 2009) from Amazon/LibraryThing was pointed out positively. Differences in the user behaviour in respect to different types of working tasks could be observed. This run of the iTrack has also shown some advices and challenges for the next run.

For the upcoming iTrack in 2010 we plan to design the working task such that they are expected to require more specific queries that cannot be matched by a whole book. The users will then be encouraged to not only rely on well-known metadata, such as title and author but also more on e. g. reviews or book covers. The book data is to be cleaned-up to increase the homogeneity, that is, entries with sparse metadata will be removed from the collection. Additionally, the search mode (*search* vs. *browse*) of users for the working tasks will be another research question. To investigate our hypotheses, we plan to perform a comparison of multiple systems.

References

1. Pharo, N., Nordlie, R., Fuhr, N., Beckers, T., Fachry, K.N.: Overview of the inex 2009 interactive track. In: Proceedings of the 8th International Workshop of the Initiative for the Evaluation of XML Retrieval (INEX 2009). (2010)
2. Fuhr, N., Klas, C.P., Schaefer, A., Mutschke, P.: Daffodil: An integrated desktop for supporting high-level search activities in federated digital libraries. In: ECDL 2002 Proceedings, Heidelberg et al., Springer (2002) 597–612