Overview of the INEX 2010 Interactive Track

Nils Pharo[1], Thomas Beckers[2], Ragnar Nordlie[1], and Norbert Fuhr[2]

[1] Faculty of Journalism, Library and Information Science, Oslo University College, Norway
nils.pharo@jbi.hio.no, ragnar.nordlie@jbi.hio.no
[2] Department of Computer Science and Applied Cognitive Science,
University of Duisburg-Essen, Germany
tbeckers@is.inf.uni-due.de, norbert.fuhr@uni-due.de

Abstract. In the paper we present the organization of the INEX 2010 *interactive track*. For the 2010 experiments the iTrack has gathered data on user search behavior in a collection consisting of book metadata taken from the online bookstore Amazon and the social cataloguing application LibraryThing. The collected data represents traditional bibliographic metadata, user-generated tags and reviews and promotional texts and reviews from publishers and professional reviewers. In this year's experiments we designed two search task categories, which were set to represent two different stages of work task processes. In addition we let the users create a task of their own, which is used as a control task. In the paper we describe the methods used for data collection and the tasks performed by the participants.

1 Introduction

The INEX interactive track (iTrack) is a cooperative research effort run as part of the INEX Initiative for the Evaluation of XML retrieval [1]. The overall goal of INEX is to experiment with the potential of using XML to retrieve relevant parts of documents. In recent years, this has been done through the provision of a test collection of XML-marked Wikipedia articles. The main body of work within the INEX community has been the development and testing of retrieval algorithms. Interactive information retrieval (IIR) [2] aims at investigating the relationship between end users of information retrieval systems and the systems they use. This aim is approached partly through the development and testing of interactive features in the IR systems and partly through research on user behavior in IR systems. In the INEX iTrack the focus over the years has been on how end users react to and exploit the potential of IR systems that facilitate the access to *parts* of documents in addition to the full documents.

The INEX interactive track was run for the first time in 2004, in the first two years the collection consisted of journal articles from IEEE computer science journals [3, 4]. In 2006/7 [5] and 2008 [6] the Wikipedia corpus was used. In 2009 [7] the iTrack switched to a collection consisting of book metadata compiled from the bookstore Amazon and the social cataloguing application LibraryThing.

S. Geva et al. (Eds.): INEX 2010, LNCS 6932, pp. 227–235, 2011.

Throughout the years the design of the iTrack experiments has been quite similar:

- a common subject recruiting procedure
- a common set of user tasks and data collection instruments such as interview guides and questionnaires
- a common logging procedure for user/system interaction
- an understanding that collected data should be made available to all participants for analysis

In this way the participating institutions have gained access to a rich and comparable set of data on user background and user behavior, with a relatively small investment of time and effort. The data collected has been subjected to both qualitative and quantitative analysis, resulting in a number of papers and conference presentations ([8], [9], [10], [11], [12], [13], [14], [15]).

In 2009, it was felt that although the "common effort" quality of the previous years was valuable and still held potential as an efficient way of collecting user behavior data, the Wikipedia collection had exhausted its potential as a source for studies of user interaction with XML-coded documents. It was therefore decided to base the experiments on a new data collection with richer structure and more semantic markup than had previously been available. The collection was 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 cataloguing tool LibraryThing. A sub-set of the same collection was used in this year's experiments, with a change to a new IR system, of which two alternative versions were made available (a more specific description of the system and collection is given below). The records present book descriptions on a number of levels: formalized author, title and publisher data; subject descriptions and user tags; book cover images; full text reviews and content descriptions. New this year is that more emphasis is given to the distinction between publisher data and user-generated data. The two systems differ in that it is not possible to query the reviews nor the book abstracts in one of the two versions. The database was chosen with the intention to enable investigation of research questions concerning, for instance

- What is the basis for judgments on relevance in a richly structured and diverse material? What fields / how much descriptive text do users make use of / chose to see to be able to judge relevance?
- How do users understand and make use of structure (e.g. representing different levels of description, from highly formalized bibliographic data to free text with varying degrees of authority) in their search development?
- How do users construct and change their queries during search (sources of terms, use and understanding of tags, query development strategies ..)?
- How do users' search strategies differ at different stages of their work task processes?

2 Tasks

For the 2010 iTrack the experiment was designed with two categories of tasks constructed by the track organizers, from each of which the searchers were instructed

to select one of three alternative search topics. In addition the searchers were invited to perform one semi-self-generated task, which would function as a control task. The two task categories were designed to be presented in contexts that reflect two different stages of a work task process [16]. The theory underlying our choice of tasks is that searchers at an early stage in the process will be in a more explorative and problem-oriented mode, whereas at a later stage they will be focused towards more specific data collection.

The first set of tasks was designed to let searchers use a broad selection of metadata, in particular combining topical searches with the use of review data. The tasks were thus designed to inspire users to create "polyrepresentative" [17] search strategies, i.e. to use explorative search strategies, which should give us data on query development, metadata type preference and navigation patterns.

At the second stage we have attempted to simulate searchers that are in a rather mechanistic data gathering mode. The tasks have also been designed to focus on non-topical characteristics of the books. Information should typically be found in publisher's texts and possibly in user-provided tags.

The self-selected task was intended to function as a "control" task for comparison with the performance of two others.

The task groups are introduced in the following way:

Task Group 1: The Explorative Tasks

You are at an early stage of working on an assignment, and have decided to start exploring the literature of your topic. Your initial idea has led to one of the following three research needs:

1. Find trustworthy books discussing the conspiracy theories which developed after the 9/11 terrorist attacks in New York.
2. Find controversial books discussing the climate change and whether it is man-made or not.
3. Find highly acclaimed novels that treat issues related to racial discrimination.

Task Group 2: The Data Gathering Tasks

You are in a data gathering stage of an assignment and need to collect a series of books for further analysis. This has led to one of the following three research needs:

4. Find novels that won the Nobel Prize during the 1990's.
5. Find bestseller crime novels by female authors.
6. Find biographies on athletes active in the 1990's.

The Semi Self-selected Task

7. Try to find books about a specific topic or of a certain type, but do not look for a specific title you already know.

3 Participating Groups

3 research groups participated in this year's track: Oslo University College, University of Duisburg-Essen, and University of Glasgow. Data from a total of 147 sessions performed by 49 test subjects were collected from October 2010 to January 2011. The participation was compensated for some participant with a EUR 12 Amazon voucher.

4 Research Design

4.1 Search System

The experiments were conducted on a Java-based retrieval system built within the ezDL framework[1], which resides on a server at and is maintained by the University of Duisburg-Essen. The collection was indexed with Apache Solr 1.4, which is based on Apache Lucene. Lucene applies a variation of the vector space retrieval model. The basis of the search system is similar to the interfaces used for previous iTracks, but the interface has been modified extensively to accommodate the new data set, and a set of new functionalities have been developed. Two versions (A and B) were developed for the experiments.

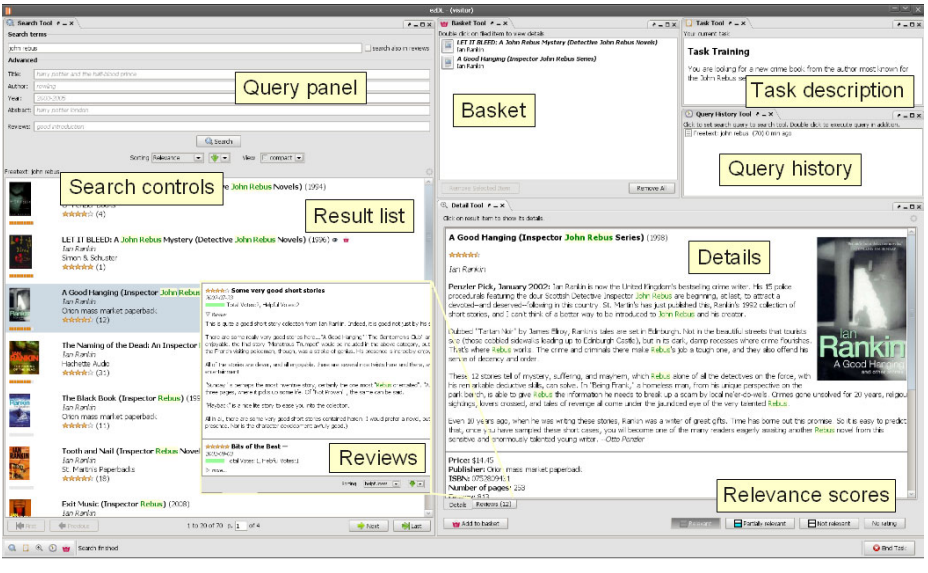

Fig. 1. The search system interface

[1] http://ezdl.de, http://www.is.inf.uni-due.de/projects/ezdl/

Figure 1 shows the interface of the system (A version). The main features available to the user are:

- The query interface provides a Google-like query field as well as additional query fields for title, author, year, abstract and reviews. When a search term is entered, the searcher can choose if he wants to search also in the reviews.
- The system can order the search results according to "relevance" (which books the system considers to be most relevant to your search terms), "year" (publication year of the book), or "average rating" (in the cases where quality ratings from readers were available).
- The system will show results twenty titles at a time, with features to assist in moving further forwards or backwards in the result list.
- A double click on an item in the result list will show the book details in the "Details" window.
- If the book has been reviewed, the reviews can be seen by clicking the "Reviews" tab at the bottom of this window. Each review shows the title, the rating, the date and the helpfulness rating. A simple click on a review extends the review by the full review text
- The users are instructed to determine the relevance of any examined book, as "Relevant", "Partially relevant" or "Not relevant", by clicking markers at the bottom of the screen. Any book decided to constitute part of the answer to the search task can be moved to a result basket by clicking the "Add to basket" button next to the relevance buttons.
- A "Query history" button in the right of the screen displays the query terms used so far in the current search session. A single click sets a query to the search tool. A double-click also executes this query
- A line of yellow dots above an item in the result list is used to indicate the system's estimate of how closely related to the query the item is considered to be.
- Query terms are highlighted in the result list and the detail tool.

The B version of the search system did not allow the user to search in reviews or abstracts, i.e. no query fields for abstract and reviews were available to the user.

4.2 Document Corpus

The collection contains metadata of 2 780 300 English-language books. The data has been crawled from the online bookstore Amazon and the social cataloging web site LibraryThing in February and March 2009 by the University of Duisburg-Essen. The MySQL database containing the crawled data has a size of about 190 GB.. Several millions of customer reviews were crawled. For this year's run of the track we cleaned up the data by removing all records that do not have an image of the book cover. This was thought to be a good heuristic for removing records that only have very sparse data. After the clean-up, metadata from approximately 1.5 million books remained in the database.

The records present book descriptions on a number of levels: formalized author, title and other bibliographic data; controlled subject descriptions and user-provided

content-descriptive tags; book cover images; full text reviews and publisher-supplied content descriptions. The following listing shows what items has been crawled from either Amazon or LibraryThing:

Amazon
ISBN, title, binding, label, list price, number of pages, publisher, dimensions, reading level, release date, publication date, edition, Dewey classification, title page images, creators, similar products, height, width, length, weight, reviews (rating, author id, total votes, helpful votes, date, summary, content) editorial reviews (source, content).

LibraryThing
Tags (including occurrence frequency), blurbs, dedications, epigraphs, first words, last words, quotations, series, awards, browse nodes, characters, places, subjects.

4.3 Online Questionnaires

During the course of the experiment, the system presented the searchers with online questionnaires to support the analysis of the log data. The searchers were given a pre-experiment questionnaire, with demographic questions such as searchers' age, education and experience in information searching in general and in searching and buying books online. Each search task was preceded with a pre-task questionnaire, which concerned searchers' perceptions of the difficulty of the search task, their familiarity with the topic etc. After each task, the searcher was asked to fill out a post-task questionnaire. The intention of the post-task questionnaire was to learn about the searchers' use of and their opinion on various features of the search system, in relation to the just completed task. Each experiment sessions were closed with a post-experiment questionnaire, which elicited the searchers' general opinion of the search system.

4.4 Relevance Assessments

The searchers were instructed to indicate the relevance of the items in the result list, using a three-part relevance scale of "relevant", "partly relevant" and "not relevant".

4.5 "Shopping" Basket

To simulate the purchase of relevant books the system provides a shopping basket feature in which searchers were asked to add books they would have purchased for solving the task. Books can be added and removed from the basket.

4.6 Logging

All search sessions were logged and saved to a database. The logs register and time stamp the events in the session and the actions performed by the searcher, as well as the responses from the system.

5 Experimental Procedure

The experimental procedure for each searcher is outlined below.

1. When recruiting searchers for the experiment, the experimenter gives the searchers the instructions for the self-selected task.
2. Experimenter briefs the searcher, and explains format of study. The searcher reads and signs the Consent Form.
3. The experimenter logs the searchers into the system. This presents the searcher with the task assignments and the questionnaire. The experimenter hands out and explains the User guidelines document. It is important to take good time to demonstrate and explain how the system works. A tutorial of the system with a training task is provided.
4. The experimenter answers questions from user.
5. The searcher selects his/her tasks from each of the two categories. In addition the self-selected task is input into the appropriate form. Tasks are rotated by the system, thus any of the three tasks may be the first to be solved by the searcher.
6. The searcher answers the Pre-experiment questionnaire provided by the system.
7. The searcher answers the Pre-task questionnaire provided by the system.
8. The task is started by clicking the link to the IR system. Each task has a duration of 15 minutes, at which point the system will tell the user time has run out. The IR system is closed by clicking the "End task" button.
9. The searcher answers the Post-task questionnaire provided by the system.
10. Steps 6-9 repeated for the two other tasks.
11. The searcher answers the Post-experiment questionnaire provided by the system.
12. At the end of the evaluation session the user presses the "Finish" button in the evaluation/questionnaire system to store his data into the database.

6 Results and Future Plans

Table 1 shows the distribution of systems and tasks. As can be seen very few searchers chose task 4 (Nobel Prize winning novels), the other tasks were fairly evenly distributed. For some unknown reason one searcher performed two tasks in system A and one task in system B, although our distribution system was programmed to allocate three system B task for this user. This explains the system distribution being slightly skewed.

We are currently analyzing our data, which will be presented in forthcoming conference and journal papers. A primary focus of the analysis will be searchers' choice of sources of information for the completion of the tasks. Amongst the issues that we plan to look at is the effect of searchers' topic knowledge and the influence of task types.

Table 1. Distribution of systems and tasks

	Task							Total
	1	2	3	4	5	6	7	
System A	14	10	2	6	8	11	26	77
B	11	10	2	13	5	6	23	70
Total	25	20	4	19	13	17	49	147

References

[1] Malik, S., Trotman, A., Lalmas, M., Fuhr, N.: Overview of INEX 2006. In: Fuhr, N., Lalmas, M., Trotman, A. (eds.) INEX 2006. LNCS, vol. 4518, pp. 1–11. Springer, Heidelberg (2007)

[2] Ruthven, I.: Interactive Information Retrieval. Annual Review of Information Science and Technology 42, 43–91 (2008)

[3] Tombros, A., Larsen, B., Malik, S.: The Interactive Track at INEX 2004. In: Fuhr, N., Lalmas, M., Malik, S., Szlávik, Z. (eds.) INEX 2004. LNCS, vol. 3493, pp. 410–423. Springer, Heidelberg (2005)

[4] Larsen, B., Malik, S., Tombros, A.: The Interactive Track at INEX 2005. In: Fuhr, N., Lalmas, M., Malik, S., Kazai, G. (eds.) INEX 2005. LNCS, vol. 3977, pp. 398–410. Springer, Heidelberg (2006)

[5] Malik, S., Tombros, A., Larsen, B.: The Interactive Track at INEX 2006. In: Fuhr, N., Lalmas, M., Trotman, A. (eds.) INEX 2006. LNCS, vol. 4518, pp. 387–399. Springer, Heidelberg (2007)

[6] Pharo, N., Nordlie, R., Fachry, K.N.: Overview of the INEX 2008 Interactive Track. In: Geva, S., Kamps, J., Trotman, A. (eds.) INEX 2008. LNCS, vol. 5631, pp. 300–313. Springer, Heidelberg (2009)

[7] Pharo, N., Nordlie, R., Fuhr, N., Beckers, T., Fachry, K.N.: Overview of the INEX 2009 Interactive Track. In: Geva, S., Kamps, J., Trotman, A. (eds.) INEX 2009. LNCS, vol. 6203, pp. 303–311. Springer, Heidelberg (2010)

[8] Pharo, N., Nordlie, R.: Context Matters: An Analysis of Assessments of XML Documents. In: Crestani, F., Ruthven, I. (eds.) CoLIS 2005. LNCS, vol. 3507, pp. 238–248. Springer, Heidelberg (2005)

[9] Hammer-Aebi, B., Christensen, K. W., Lund, H., Larsen, B.: Users, structured documents and overlap: interactive searching of elements and the influence of context on search behaviour. In: Ruthven, I. et al. (eds.) Proceedings of Information Interaction in Context : International Symposium on Information Interaction in Context : IIiX 2006, Copenhagen, Denmark, October 18-20, pp. 80-94, Royal School of Library and Information Science (2006)

[10] Pehcevski, J.: Relevance in XML retrieval: the user perspective. In: Trotman, A., Geva, S. (eds.) Proceedings of the SIGIR 2006 Workshop on XML Element Retrieval Methodology: Held in Seattle, Washington, USA, August 10, pp. 35–42. Department of Computer Science, University of Otago, Dunedin, New Zealand (2006)

[11] Malik, S., Klas, C.-P., Fuhr, N., Larsen, B., Tombros, A.: Designing a user interface for interactive retrieval of structured documents: lessons learned from the INEX interactive track? In: Gonzalo, J., Thanos, C., Verdejo, M.F., Carrasco, R.C. (eds.) ECDL 2006. LNCS, vol. 4172, pp. 291–302. Springer, Heidelberg (2006)

[12] Kim, H., Son, H.: Users Interaction with the Hierarchically Structured Presentation in XML Document Retrieval. In: Fuhr, N., Lalmas, M., Malik, S., Kazai, G. (eds.) INEX 2005. LNCS, vol. 3977, pp. 422–431. Springer, Heidelberg (2006)

[13] Kazai, G., Trotman, A.: Users' perspectives on the Usefulness of Structure for XML Information Retrieval. In: Dominich, S., Kiss, F. (eds.) Proceedings of the 1st International Conference on the Theory of Information Retrieval, pp. 247–260. Foundation for Information Society, Budapest (2007)

[14] Larsen, B., Malik, S., Tombros, A.: A Comparison of Interactive and Ad-Hoc Relevance Assessments. In: Fuhr, N., Kamps, J., Lalmas, M., Trotman, A. (eds.) INEX 2007. LNCS, vol. 4862, pp. 348–358. Springer, Heidelberg (2008)

[15] Pharo, N.: The effect of granularity and order in XML element retrieval. Information Processing and Management 44(5), 1732–1740 (2008)

[16] Kuhlthau, C.C.: Seeking meaning: a process approach to library and information services, 2nd edn. Libraries Unlimited, Westport (2004)

[17] Larsen, B., Ingwersen, P., Kekäläinen, J.: The Polyrepresentation continuum in IR. In: Ruthven, I., et al. (eds.) Information Interaction in Context: International Symposium on Information Interaction in Context, IIiX 2006, pp. 148–162. ACM Press, New York (2006)

[18] Fuhr, N., Klas, C.-P., Schaefer, A., Mutschke, P.: Daffodil: An integrated desktop for supporting high-level search activities in federated digital libraries. In: Agosti, M., Thanos, C. (eds.) ECDL 2002. LNCS, vol. 2458, pp. 597–612. Springer, Heidelberg (2002)