Overview of the INEX 2009 Interactive Track

Nils Pharo[1], Ragnar Nordlie[1], Norbert Fuhr[2], Thomas Beckers[2] and Khairun Nisa Fachry[3]

[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
norbert.fuhr@uni-due.de, tbeckers@is.inf.uni-due.de
[3] Archives and Information Studies, University of Amsterdam, the Netherlands
k.n.fachry@uva.nl

Publised as: Pharo, N., Nordlie, R., Fuhr, N., Beckers, T. & Fachry K.N. (2010). Overview of the INEX 2009 Interactive Track. Lecture Notes in Computer Science, 6203, 303-311.

Abstract. In the paper we present the organization of the INEX 2009 *interactive track*. For the 2009 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. Thus the data are more structured than in previous years' experiments, consisting of traditional bibliographic metadata, user-generated tags and reviews and promotional texts and reviews from publishers and professional reviewers. Through monitoring searches based on three different task types the experiment aims at studying how users interact with highly structured data. We describe the methods used for data collection and the tasks performed by the participants. Some preliminary results of the interaction analysis are reported.

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 (iTrack) was run for the first time in 2004 [3], repeated in 2005 [4], in 2006/2007 [5] (due to technical problems the tasks scheduled for 2006 were actually run in early 2007), and in 2008 [14]. Although there has been variations in task content and focus, some fundamental premises has been in force throughout:

- 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

This has ensured that through a manageable effort, participant institutions have had access to a rich and comparable set of data on user background and user behavior, of sufficient size and level of detail to allow both qualitative and quantitative analysis. This has already been the source of a number of papers and conference presentations ([6], [7], [8], [9], [10], [11], [12], [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. We decided to base the experiments on a new data collection with richer structure and more semantic markup than has previously been available, and have 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 cataloguing tool LibraryThing (a more specific description of the database 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. The database intended 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 ..)?

2 Tasks

For the 2009 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. The two categories of tasks were intended to reflect the most common purposes a searcher would have for visiting a database of primarily bibliographic data, a broad, explorative task and a narrower, more specific, purpose-driven task. The self-selected task was intended to force the searcher to perform a more quality-driven search than the two others.

The broad tasks

These task were designed to investigate thematic exploration, aiming to provide data on query development, metadata type preference and navigation patterns. The tasks were as follows:

1. You are considering to start studying sociology. In order to prepare for the course you would like to get acquainted with some good and recent introductory texts within the field as well as some of its classics.
2. You are interested in taking a course on environmental friendly energy. In order to prepare for the course you would like to get acquainted with some good introductory texts on the field.
3. You are considering to start studying existentialism. In order to prepare for the course you would like to get acquainted with some good introductory texts within the field as well as some of its classics.

The narrow tasks

These tasks represent relatively narrow topical queries where the purpose was to allow us to study the basis for relevance decisions and compare the searchers' preference of different document representations. The following tasks were provided:

1. Find trustworthy books discussing the conspiracy theories which developed after the 9/11 terrorist attacks in New York.
2. Find books which present documentation of the specific health and/or beauty effects of consuming olive oil.
3. The Kabbalah is an esoteric religious tradition which has inspired works of fiction. Find novels where the plot is inspired by the Kabbalah, and a factual treatment of the origins and development of this tradition.

The semi self-selected task

For one of the courses you are currently attending, you need an additional textbook. You have only money for one book (assuming they all have about the same price). You are free to select the course topic yourself.

3 Participating groups

Due to unfortunate delays in the preparation of the experimental system, the experiments were launched late in the INEX 2009 research cycle, and only 3 research groups were able to submit experiment data by the deadline for this report: Oslo University College, University of Glasgow, and University of Duisburg-Essen. Data from a total of 123 searches performed by 41 test subjects were collected, in addition to 36 searches by 12 subjects using Duisburg's alternative system (see below).

4 Research design

4.1 Search system

The experiments were conducted on a java-based retrieval system built within the Daffodil framework [13], which resides on a server at and is maintained by the University of Duisburg-Essen. The collection was indexed with Apache Solr 1.3, which is based on Apache Lucene. Lucene applies a vector space retrieval model. The system is also partially based on the *ezDL* (http://www.is.inf.uni-due.de/projects/ezdl/). The basis of the search system is the same as have been 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.

Figure 1 shows the interface of the system. The main features available to the user are

- When a search term is entered, the searcher can choose to search on "content", "reviews", or both together. "Content" searches all the "formalized" text connected to each book – title, keywords, publisher's description etc. "Reviews" allows search in the text of any user reviews of the book. In both cases the search index bases result rankings on term occurrence. In addition, there is field-based search available on author, title or publication year.

- 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 people have rated the quality of the books).

- The system will show results twenty titles at a time, with features to assist in moving further forwards or backwards in the result list.

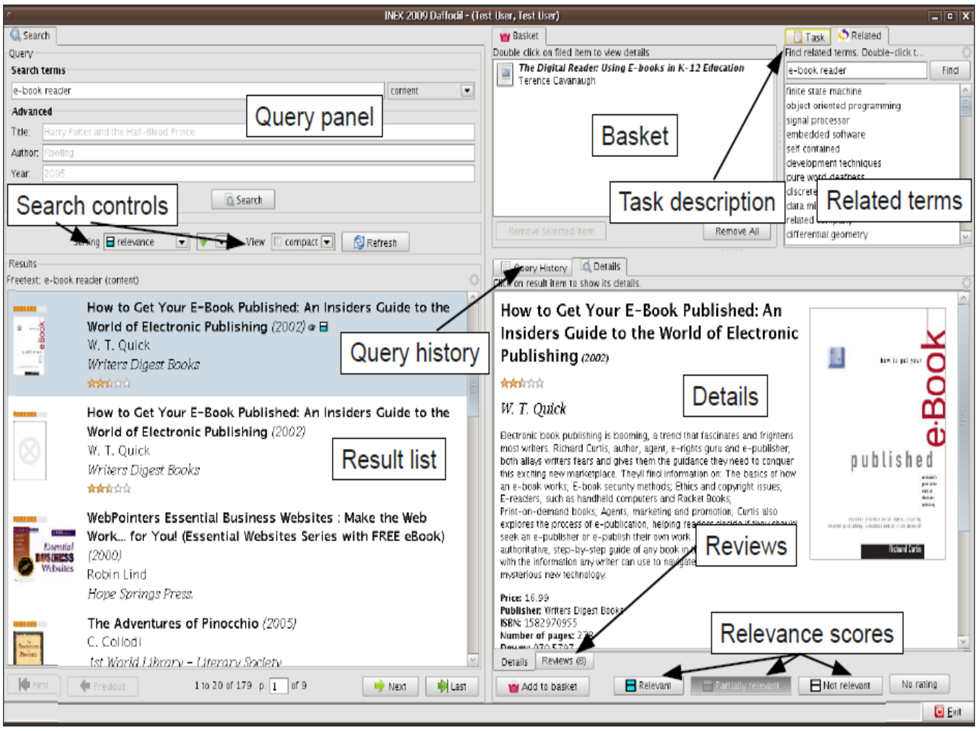
- A double click on an item in the result list will show the book details in the “Details” window. If the book has been reviewed, the reviews can be seen by clicking the “Reviews” tab at the bottom of this window.

- The relevance of any which is examined should be determined, 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 should be moved to a result basket by clicking the “Add to basket” button next to the relevance buttons.

- When the first search term has been entered, the system will use the task window to suggest search terms which might be relevant to the task. A double click on a term in this list will move it to the search term window.

- A “Query history” button in the middle of the screen displays the search terms used so far in the search session.

- 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.



INEX 2009 Daffodil - (Test User, Test User)

Search terms: e-book reader (content)

Title: Harry Potter and the Half-Blood Prince

Author: Rowling

Year: 2005

Sorting: relevance

View: compact

Freetext: e-book reader (content)

How to Get Your E-Book Published: An Insiders Guide to the World of Electronic Publishing *(2002)*
W. T. Quick
Writers Digest Books

How to Get Your E-Book Published: An Insiders Guide to the World of Electronic Publishing *(2002)*
W. T. Quick
Writers Digest Books

WebPointers Essential Business Websites : Make the Web Work... for You! (Essential Websites Series with FREE eBook) *(2000)*
Robin Lind
Hope Springs Press.

The Adventures of Pinocchio *(2005)*
C. Collodi
1st World Library - Literary Society

1 to 20 of 179 p. 1 of 9

Double click on filed item to view details

The Digital Reader: Using E-books in K-12 Education
Terence Cavanaugh

Find related terms. Double-click t...

e-book reader

finite state machine
object oriented programming
signal processor
embedded software
self contained
development techniques
pure word deafness
discrete
data mi
related
differential geometry

Click on result item to show its details.

How to Get Your E-Book Published: An Insiders Guide to the World of Electronic Publishing *(2002)*

W. T. Quick

Electronic book publishing is booming, a trend that fascinates and frightens most writers. Richard Curtis, author, agent, e-rights guru and e-publisher, both allays writers fears and gives them the guidance they need to conquer this exciting new marketplace. Theyll find information on: The basics of how an e-book works; E-book security methods; Ethics and copyright issues; E-readers, such as handheld computers and Rocket Books; Print-on-demand books; Agents, marketing and promotion; Curtis also explores the process of e-publication, helping rea[illegible] seek an e-publisher or e-publish their own work. [illegible] authoritative, step-by-step guide of any book in [illegible] with the information any writer can use to na[illegible] mysterious new technology.

Price: 16.99
Publisher: Writers Digest Book[illegible]
ISBN: 1582970955
Number of pages: 2[illegible]

Fig. 1. Daffodil interface

4.2 Document corpus

The collection contains metadata of 2 780 300 English-language books. The data has been crawled from the online bookstore of *Amazon* and the social cataloging web site *LibraryThing* in February/March 2009 by the University of Duisburg-Essen. The MySQL database containing the crawled data has size of about 190 GB. Cover images are available for over one million books (100 GB of the database). Several millions of customer reviews were crawled.

The XML-coded 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 data was 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, searchers were issued brief online questionnaires to support the analysis of the log data. Before the search tasks were introduced, 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 is to learn about the searchers' use of and their opinion on various features of the search system, in relation to the just completed task. The experiment sessions were closed with a post-experiment questionnaire, which elicited the searchers' general opinion of the search system.

4.4 Relevance assessments

The users' task was partly to indicate the relevance of any item in the result list found sufficiently interesting for them to view in detail, partly to collect a result set which they considered to constitute an answer to their task. A three-part relevance scale of "relevant", "partly relevant" and "not relevant" was used.

4.5 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. In addition to system logs, some participating institutions have been logging additional data through eye-tracking, screen image capture etc.

4.6 System comparison

A modified version of the search system (the B version) was developed at the University of Duisburg-Essen. This special version was less interactive and powerful due to missing reviews, tools (related terms, query history) and search options (content & review, review).

12 of the 24 participants in Duisburg used the B version, while the other 12 used the A version (the standard version employed by all other participants in this track). Additionally, the experiments were also recorded by an eyetracking system. It is expected that users behave differently with a more traditional, less interactive search system.

5 Experimental Procedure

Each experiment has been performed following the standard procedure outlined below. Steps 7 to 10 were repeated for each of the three tasks performed by the searcher.

1. Experimenter briefs the searcher, and explains format of study. The searcher reads and signs the Consent Form.
2. The experimenter logs the searchers into the experimental system. Tutorial of the system is given with a training task provided by the system. The experimenter hands out and explains the system features document.
3. Any questions answered by the experimenter.
4. The experimenter administers the pre-experiment questionnaire.

5. Topic descriptions for the first task category administered, and a topic selected.
6. Pre-task questionnaire administered.
7. Task begins by clicking the link to the search system. Maximum duration for a search is 15 minutes, at which point the system issues a “timeout” warning. Task ended by clicking the “Finish task” button.
8. Post-task questionnaire administered.
9. Steps 5-8 repeated for the second and third task.
10. Post-experiment questionnaire administered.

6 Data analysis

As the experiment phase was delayed, only a preliminary analysis of the questionnaire data is available at the deadline for this report. Log analysis, combined with further questionnaire analysis, will continue and will be reported elsewhere.

The questionnaires included open-question invitations for comments by the participants on both the system and the search experience. The positive comments include the following items:

+ well arranged interface
+ everything fits on the screen, no scrolling
+ reviews are very useful

Some users experienced technical problems. Also, missing highlighting and filtering as well as too many books without enough metadata were points of negative criticism:

- technical problems (search, query syntax, drag and drop)
- “related terms” are not always useful
- no highlighting of query terms in results
- some books do not have enough details
- no filtering

From the quantitative questionnaire data we have attempted to analyze the effect of the different types of search task on searchers’ use of the various types of metadata available.

Table 1. The influence of task category on searchers’ preferences of metadata field

	Task category 1	Task category 2	Task category 3	Overall
Title	3.79	3.81	3.96	3.85
Author	1.62	1.57	2.17	1.78

Year	2.55	1.91	2.83	2.43
Publisher’s name	1.75	1.51	1.83	1.70
Keywords	3.28	3.29	2.82	3.13
User tags	2.65	2.75	2.58	2.66
Reviews	3.23	3.34	3.38	3.32
Publisher’s description	3.45	3.64	3.42	3.50
Image	2.36	2.85	2.45	2.55
Relevance score	2.98	2.98	2.98	2.98

The searchers were asked to indicate on a five point scale how useful (5 for very useful) different types of metadata were for solving their search tasks. From Table 1 we see that document titles, publishers’ book descriptions and reviews (by users) were the three most popular metadata fields. It is also worth noting that the searchers found keywords (from Amazon) to be more useful than the user-created tags. It seems that searchers put more trust in authoritative sources that use a controlled vocabulary than users’ idiosyncratic tagging.

We see that the variation between the different categories of task only differs significantly with respect to the usefulness of “year”. We believe the reason that searchers find year to be more important for the textbook tasks (category 1 and 3) is the sheer number of relevant documents generated by these queries. The Category 2 tasks are more specific and the relevant documents are probably easier to select from the result list without reference to additional distinguishing factors such as publication year.

We have also looked at the searchers’ familiarity with the topics and seen how this correlates with the usefulness of the metadata components. Our finding is that there is no systematic correlation between topic familiarity and metadata preference.

References

[1] Malik, S., Trotman, A., Lalmas, M. & Fuhr, N. (2007): Overview of INEX 2006. In: Fuhr, N., Lalmas, M. and Trotman, A. eds. *Comparative Evaluation of XML Information Retrieval Systems, 5th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2006, Dagstuhl Castle, Germany, December 2006*. Berlin: Springer, p. 1-11.

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

[3] Tombros, A., Larsen, B. and Malik, S. (2005): The Interactive Track at INEX 2004. In: Fuhr, N., Lalmas, M., Malik, S. and Szlávik, Z. eds. *Advances in XML Information Retrieval: Third International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2004, Dagstuhl Castle, Germany, December 6-8, 2004*. Berlin: Springer, p. 410-423

[4] Larsen, B., Malik, S. and Tombros, A. (2006): The interactive track at INEX 2005. In: Fuhr, N., Lalmas, M., Malik, S. and Kazai, G. eds. *Advances in XML Information Retrieval and Evaluation, 4th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2005, Dagstuhl Castle, Germany, November 28-30, 2005*. Berlin: Springer, p. 398-410.

[5] Larsen, B., Malik, S. & Tombros, A. (2007): The Interactive track at INEX 2006. In: Fuhr, N., Lalmas, M. and Trotman, A. eds. *Comparative Evaluation of XML Information Retrieval Systems, 5th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2006, Dagstuhl Castle, Germany, December 2006*. Berlin: Springer, p. 387-399.

[6] Pharo, N. & Nordlie, R. (2005): Context Matters: An Analysis of Assessments of XML Documents. In: F. Crestani and I. Ruthven eds. *Information Context: Nature, Impact, and Role: 5th International Conference on Conceptions of Library and Information Sciences, CoLIS 2005, Glasgow, UK, June 4-8, 2005*. Berlin: Springer, p. 238-248.

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

[8] Malik, S., Klas, C.-P., Fuhr, N., Larsen, B. and Tombros, A. (2006): Designing a user interface for interactive retrieval of structured documents: lessons learned from the INEX interactive track? In: Gonzalo, J. et al. eds. *Research and Advanced Technology for Digital Libraries, 10th European Conference, ECDL 2006*. Alicante, Spain, September 17-22, 2006, Proceedings. Berlin: Springer,

[9] Kim, H. & Son, H. (2006): Users Interaction with the Hierarchically Structured Presentation in XML Document Retrieval. In: Fuhr, N., Lalmas, M., Malik, S. & Kazai, G. eds. *Advances in XML Information Retrieval and Evaluation: 4th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2005, Dagstuhl Castle, Germany, November 28-30, 2005.* Berlin: Springer, p. 422-431.

[10] Kazai, G. & Trotman, A (2007): 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*. Budapest: Foundation for Information Society, p. 247-260.

[11] Larsen, B., Malik, S & Tombros, A. (2008): A Comparison of Interactive and Ad-Hoc Relevance Assessments. In: Fuhr, N., Kamps, J., Lalmas, M. & Trotman, A. eds. *Focused Access to XML Documents: 6th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2007 Dagstuhl Castle, Germany, December 17-19, 2007.* Berlin: Springer, p. 348-358.

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

[13] Fuhr, N., Klas, C.P., Schaefer, A. & Mutschke, P. (2002): Daffodil: An integrated desktop for supporting high-level search activities in federated digital libraries. In *Proceedings of the 6th European Conference on Research and Advanced Technology for Digital Libraries (ECDL),* p. 597-612.

[14] Pharo, N., Nordlie, R. & Fachry, K. N. (2009): Overview of the INEX 2008 Interactive Track. In: Geva, S., Kamps, J. and Trotman, A. eds. *Advances in Focused*

Retrieval, 7th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2008, Dagstuhl Castle, Germany, December 2008. Berlin: Springer, p. 300-313.

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