End users in the context of XML documents – setting up an interactive track at INEX

Birger Larsen
Department of Information Studies
Royal School of LIS,
Copenhagen, Denmark
blar@db.dk

Anastasios Tombros
Department of Computer Science
Queen Mary, University of London
United Kingdom
tassos@dcs.qmul.ac.uk

Saadia Malik
Fak. 5/IIS, Information Systems
University of Duisburg-Essen
Duisburg, Germany
malik@is.informatik.uni-duisburg.de

ABSTRACT
The widespread use of XML has stimulated research in developing appropriate methods for searching and browsing XML documents. However, relatively little research has been carried out that studies user interaction with IR systems that take advantage of the additional features offered by XML documents. In this paper, we describe the efforts to establish an interactive track at INEX and discuss its main motivation and aims.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval]**: Information Search and Retrieval – *search process*; H.3.7 [**Information Storage and Retrieval]**: Digital Libraries – *user issues.*

General Terms
Experimentation, Human Factors.

Keywords
XML documents; User Contexts; INEX; Interactive Track.

1. INTRODUCTION
The widespread use of XML and other high level mark-up languages has stimulated research in developing appropriate methods for searching and browsing XML documents. From the point of Information Retrieval (IR), highly structured XML documents are attractive because the mark up makes it possible to identify separate parts of the documents easily rather than to view them as a uniform bag of words.

The Initiative for the Evaluation of XML Retrieval (INEX) was initiated in 2002 as a large scale international effort to improve the efficiency of research in content-oriented XML retrieval, and to promote evaluation procedures that can assess the effectiveness of proposed methods [2; 3]. Evaluating the effectiveness of XML retrieval systems requires a test collection where the relevance assessments are provided according to a relevance criterion, which takes into account the imposed structural aspects. Such a test collection has been built as a result of INEX2002 and INEX 2003. The test collection consists of:

a) A corpus of 12,107 full text scientific documents from the IEEE Computer Society's 18 journals formatted in XML.

b) A set of topics each expressing an information need. The topics are developed by INEX participants.

c) Relevance assessments provided by the topic authors. Relevance assessments are non-binary and are provided on two dimensions (*coverage* and *topicality*).

In INEX 2002 and INEX 2003 two types of topics were considered: Content and Structure (CAS), which contain explicit references to the XML structure, and Content Only (CO), which disregard the XML structure in the queries.

Research under the INEX initiative has started to shed light in aspects of effectiveness of XML retrieval. However, relatively little research has been carried out to study user interaction with IR systems that take advantage of the additional features offered by XML documents and so little is known about how users behave in the context of such IR systems. One exception is the work done by Reid and associates, who studied end user interaction with a small test collection of Shakespeare's plays formatted in XML [4].

In order to learn more about end user interaction with XML-based IR systems an interactive track is being set up as part of INEX 2004. In this paper we describe this effort and discuss its main motivation and aims.

2. INTERACTIVE TRACK MOTIVATION
Issues relating to interactive IR have been extensively investigated in the last decade. A major advance in research has been made by co-ordinated efforts in the interactive track at TREC. These efforts have been in the context of unstructured documents (e.g. news articles) or in the context of the loosely-defined structure encountered in web pages. XML documents, on the other hand, define a different context, by offering the possibility of navigating within the structure of a single document, or of following links to another document.

This context is different to the one encountered in the conventional case of unstructured documents, and has provided the main motivation for the establishment of an interactive track

at INEX. The main aims for the interactive track are twofold. First, to investigate the behaviour of users when interacting with components of XML documents, and secondly to investigate and develop approaches for XML retrieval which are effective in user-based environments.

In the first year, we plan to address the first issue: to investigate the behaviour of searchers when presented with components of XML documents that have a high probability of being relevant (as estimated by an XML-based IR system). Presently, metrics that are used for the evaluation of system effectiveness in the INEX ad hoc track are based on certain assumptions of user behaviour [6]. These metrics attempt to quantify the effectiveness of IR systems at pointing searchers to specific relevant portions (or elements) of documents. Some of the assumptions behind the metrics include that users would browse through retrieved elements in a linear order, that they would "jump" with a given probability *p* from one element to another within the same document's structure, that they would not make use of links to another document, etc. These assumptions have not been formally investigated in the context of XML retrieval; their investigation forms the primary aim for the first year of the interactive track. Through appropriate logging software (see section 3) we aim to collect data which will allow us to examine user behaviour in this context.

In addition, the track will aim to investigate the effect that task type has on search behaviour in the context of XML documents. This forms a second type of context that we aim to take into account. The effect that the context determined by task type has on the behaviour of online searchers has been demonstrated in a number of studies [e.g. 7]. One way to categorise tasks is according to the "type" of information need they correspond to. In [7] the categorisation included background (find as much general information on a topic as possible), decision (make a decision based on the information found) and many-items tasks (compile a list of items related to the information need) types. It was shown that different task types promote the use of different criteria when assessing the relevance of web pages. It is likely that a similar effect, in terms of user behaviour within structured documents, may exist in the context of XML documents. Searchers may exhibit different browsing patterns and different navigational strategies for different task types.

In this way, the format of the track for the first year differs to that followed by e.g. the interactive track at TREC. The main difference is that a comparison between different interactive approaches is not the main focus for the first year. Instead, a more collective effort is planned, where the outcome of the studies will benefit the INEX initiative. Participating sites will have the option to develop and evaluate their own interactive approaches but this will not be required.

3. PLANNED EXPERIMENTAL SETUP

In this section we briefly outline the experimental set up for the first interactive track at INEX.

A number of the 2004 CO topics will be used in the study, along with the standard INEX collection. In order to make the topics comprehensible by other than the topic author, it is required that the INEX 2004 topics not only detail *what* is being sought for, but also *why* this is wanted, and in what *context* the information need has arisen. Thereby the INEX topics are in effect simulated work task situations as developed by Borlund [5]. Compared to the regular topics, more context on the motives and background of the topic is provided in the simulated work tasks. In this way, the test persons can better place themselves in a situation where they would be motivated to search for information related to the work tasks. The aim is to enable the test persons to formulate and reformulate their own queries as realistically as possible in the interaction with the IR system. Also, in order to examine the effect of task type, CO topics will be selected that represent different types of information needs.

The test persons, employed locally in each site, will need to identify documents which are useful for completing the requirements specified in the simulated work task. They can either identify these documents explicitly (e.g. by marking down a relevance score for each document) or implicitly (e.g. by saving or bookmarking useful documents). A time limit will be set for each simulated work task.

For the first year of the track, all participating sites will use the same system which will be made available. The system will provide a basic functionality which will be agreed upon. The present system under consideration is HyRex[1] with a web-based interface with a range of visualisation features. Additional systems may be employed locally if a participating site wishes to develop and compare their own interactive approach to the official "baseline".

Analysis of the collected data will be required in order to extract conclusions from the studies. The collected data will comprise (as minimum) of questionnaires completed by the test persons and the logging of searcher interaction with the system. The logged data will consist of the queries issued, the components returned by the system, the components actually viewed and the order in which they are viewed, relevance assessments of these, any browsing behaviour, as well as time stamps for each act of interaction between the test person and the system. Participating sites will also be given the opportunity to employ site-specific data collection methodologies (e.g. think-alouds, desktop monitoring software, etc.).

The results of the track will be reported in the next INEX workshop, to be held in December 2004 in Schloss Dagstuhl, Germany.

4. CONCLUSION AND FUTURE DIRECTIONS

The interactive track at INEX is an effort to systematically investigate issues relating to user behavior and effectiveness of interactive approaches in the context of XML retrieval. The track has so far received interest from 23 out of the 51 participating sites in INEX 2004. For the first year we expect that few of the participants will be evaluating their own interactive approaches. Instead, the anticipated benefits in this first effort are twofold: first, the instigation of an interest in the research community for interactive aspects of XML retrieval and second, the collection of sufficient data for the investigation of user behavior – this data will also be used for the validation of assumptions related to evaluation metrics. This is expected to be a significant contribution of the track to the wider INEX initiative.

[1] See http://www.is.informatik.uni-duisburg.de/projects/hyrex/

In the coming years, the aims of the track will shift towards the evaluation of the utility of various interactive approaches for XML retrieval. The context of the evaluation will also shift towards that of digital libraries, taking into account different users, document collections, systems and uses. This shift will also be facilitated by the expansion of the INEX initiative, with the inclusion of more tracks in INEX 2004.

5. ACKNOWLEDGMENTS

This work is done as part of the DELOS Network of Excellence on Digital Libraries. The authors wish to thank the participants of the 2nd INEX workshop in Schloss Dagstuhl, December 2003, for useful discussions on the format of the track.

6. REFERENCES

[1] Chiaramella, Y. (2001): Information retrieval and structured documents. In: *Lectures on information retrieval : third European summer-school, ESSIR 2000, Varenna, Italy: Revised Lectures*. Berlin: Springer, pp 286-309.

[2] Fuhr, N., Gövert, N., Kazai, G. and Lalmas, M. (2002): INEX : initiative for the evaluation of XML retrieval. In: *ACM SIGIR'2002 workshop on XML and information retrieval*, pp 62-70.

[3] Gövert, N. and Kazai, G. (2003): Overview of the initiative for the evaluation of XML retrieval (INEX) 2002. In: *Proceedings of the 1st workshop of the initiative for the evaluation of XML retrieval (INEX)*, pp 1-17.

[4] Finesilver, K. and Reid, J. (2003): User Behaviour in the Context of Structured Documents. In: *Advances in Information Retrieval: 25th European Conference on IR Research, ECIR 2003*, pp 104-119.

[5] Borlund, P. (2000): *Evaluation of interactive information retrieval systems*. Åbo: Åbo Akademi University Press. vi, 276 p. (PhD dissertation).

[6] Kazai, G. (2003): Report of the INEX 2003 metrics working group. In: *Proceedings of the 2nd workshop of the initiative for the evaluation of XML retrieval (INEX)*, pp 184-190.

[7] Tombros, A., Ruthven, I. and Jose, J (2004): Searchers criteria for assessing web pages. In: *Journal of the American Society for Information Science and Technology*. To appear.