Report on the INEX 2005 Interactive Track

Saadia Malik[1], Birger Larsen[2] and Anastasios Tombros[3]
[1]University of Duisburg-Essen, Duisburg, Germany
malik@uni-duisburg.de
[2]Royal School of Library & Information Science, Copenhagen, Denmark
blar@db.dk
[3]Queen Mary, University of London, United Kingdom
tassos@dcs.qmul.ac.uk

1 Introduction

The overall motivation for the Interactive Track (iTrack) at INEX is 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. One of the major outcomes of the track in 2004 was the need to investigate methods that can be supportive during the search process based on features extracted from the XML formatting[1][2]. Problems that might be solved using such methods include overlapping components and the presentation of retrieved elements in the result list. In the experimental system that was offered in 2005, these two issues were addressed. This offered the opportunity to study how overall user search behaviour was affected by these changes when compared to the behaviour observed in 2004. In addition, the following issues were addressed in 2005, following the recommendations of the INEX Methodology Workshop at the Glasgow IR Festival[1]:

- The elicitation of user perceptions of what is needed from an XML retrieval system and the identification of applications for element retrieval. The aim is to see whether element retrieval is what users need: Does element retrieval make sense at all to users, do users prefer longer components, shorter components or whole documents, would they rather have passages than elements, etc. In addition, a mixture of topics that were simulated work tasks [3] and information needs formulated by the test persons themselves were used. The aim of including the latter was to enable studies of what characterises the tasks users formulate, and to see what kinds of applications users might need an element retrieval system for.

- The introduction of an alternative document collection with the Lonely Planet collection as an optional task, in order to broaden the scope of INEX and to allow test persons with different backgrounds (e.g. educational) to participate.

[1]http://www.cs.otago.ac.nz/inexmw/

The format of the track in 2005 was deliberately of an exploratory nature, and has relatively broad aims rather than addressing very specific research questions. Element retrieval is still in its infancy and many basic questions remain unanswered, as shown by the discussions at the IR Festival. Aside from the automatic and detailed logging of test persons as used last year, emphasis was also placed on producing qualitative results through the use of interviews and questionnaires.

2 Interactive track tasks

A total of three tasks were available to the track participants: one compulsory task that all participants had to fulfill with a minimum number of test persons (Task A), and two optional tasks (Tasks B and C). These tasks combined several element retrieval systems, topic types and XML collections.

2.1 Task A: Common search system using the IEEE collection

The minimum requirement for sites to participate in the 2005 iTrack was to provide runs using 6 searchers on the baseline version of the desktop-based XML retrieval system provided. In Table 1, the 11 participating sites are shown.

University of Duisburg-Essen (Germany)	Kyungpook National University (South Korea)
Queen Mary University of London (UK)	RMIT (Australia)
Rutgers University (USA)	University of Amsterdam (The Netherlands)
University of Tampere (Finland)	Utrecht University (The Netherlands)
Oslo University College (Norway)	Royal School of LIS (Denmark)
CWI & University of Twente (Netherlands)	

Table 1: List of participating sites at the INEX 2005 Interactive Track

2.1.1 Evaluation Methodology

An extended version of the INEX IEEE document collection was used (comprising 16,819 documents). We used content only (CO) topics from the INEX collection that refer to document contents but do not contain any constraints with regards to document structure. In order to make the search tasks comprehensible by other people besides the topic author, it was required to add why, and in what context, the information need had arisen. Thus the topics are in effect simulated work task situations as developed by Borlund [3]. Six of the 2005 CO topics were used in the study. In addition, test persons were asked to supply two examples of their own information needs. Depending on the coverage in the collection, one of these tasks was selected by the experimenter.

In total, each searcher performed three tasks. The goal for each searcher was to locate sufficient information towards completing a task, in a maximum time frame of 20 minutes per task. Test persons had to fill in questionnaires at various points in the study (before the start of the experiment, before each task, after each task, and at the end of the experiment). A semi-structured

interview and debriefing of the subjects concluded the experiment. The collected data comprised questionnaires completed by the test persons, the logs of searcher interaction with the system, the notes experimenters kept during the sessions and interviews, and the informal feedback provided by searchers at the end of the sessions.

2.1.2 Search system

The common search system used in Task A is a Java-based element retrieval system built within the Daffodil framework[2]. The HyREX retrieval engine[3] was used as backend in the baseline system. Figure 1 shows the query and results list interface of the search system. The results are presented as documents and the system indicates which elements within a document may be most closely related to the query.

A detailed document and element view (Fig. 2) can be accessed by double-clicking on an element title at the result list. This view is split in two panes: one with a Table of Contents (ToC) of the whole document, and one with the full text of the selected element on the right. The Table of Contents also indicates the currently viewed element, other retrieved elements, and viewed and assessed elements (searchers can assess elements on a three-scale grade). When double-clicking on a document title at the result list, searchers were presented with document metadata (document title, authors, abstract, etc.).

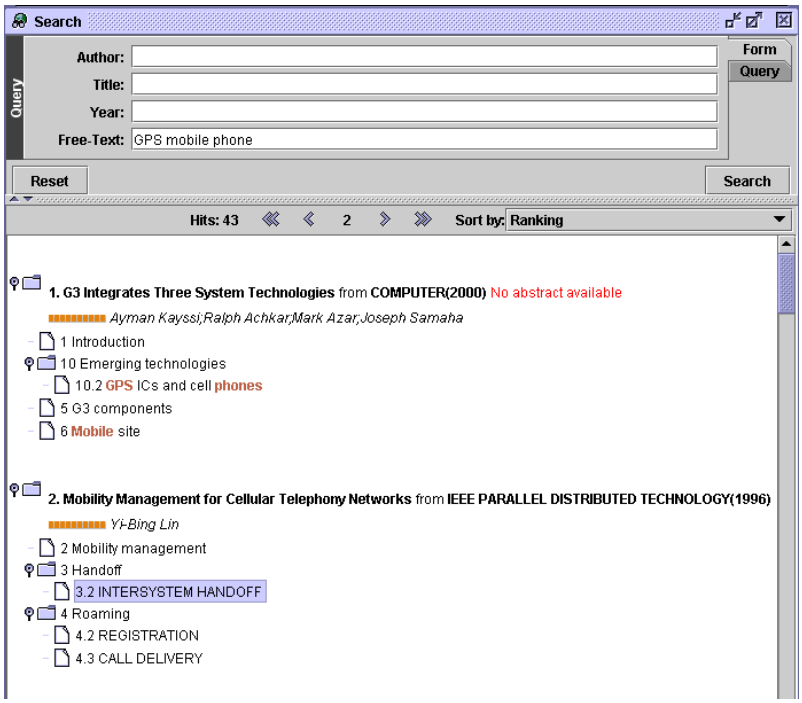

Figure 1: Result list view for the Task A retrieval system

[2]http://www.is.informatik.uni-duisburg.de/projects/daffodil/index.html.en
[3]http://www.is.informatik.uni-duisburg.de/projects/hyrex/

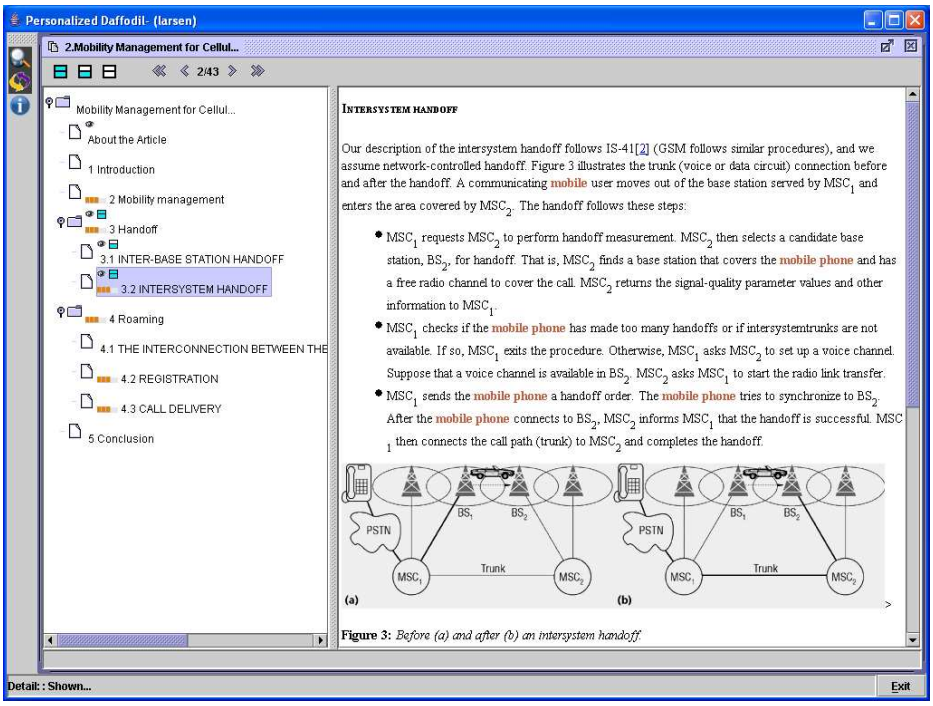

Figure 2: Detail view for the Task A retrieval system

2.2 Task B: Participation with own element retrieval system

This task allowed groups that have a working element retrieval system to test their system against the baseline system used in Task A. Groups participating in Task B were free to choose between the IEEE collection or the Lonely Planet collection, and had a large degree of freedom in setting up the experiment to fit the issues they wanted to investigate in relation to their own system. The recommended experimental setup was very close to that of Task A, with the main difference that simulated work tasks should be assigned to test persons rather than freely chosen, in order to allow for valid comparisons between systems. The University of Amsterdam (The Netherlands) participated in this task. In [7] their participation in Task B and some initial findings are presented.

2.3 Task C: Searching the Lonely Planet collection

This task allowed groups to carry out experiments with the Lonely Planet collection. The Lonely Planet collection consists of 462 XML documents with information about destinations (e.g. transport, culture, major events, facts, etc.). The collection is called the "WorldGuide" and has been provided by the publishers of the Lonely Planet guidebooks.

The aim in Task C was to allow test persons with backgrounds other than that of computing to take part in the study. A search system (B3-SDR) was provided by Utrecht University for this task. Additional test persons needed to be engaged for Task C. The participating sites in this task were: RMIT (Australia), Royal School of LIS (Denmark), Rutgers University (USA), and Utrecht

University (The Netherlands). In [8] an analysis of data from this task is presented.

3 Main findings

Here we give an overview of the main findings and issues from the 2005 Interactive Track.

User preference for elements vs. documents. The results of the track [5] suggest that searchers found document elements more useful than whole documents for tackling their information seeking tasks. On the whole, searchers tended to view and assess a large number of sections and subsections of documents when browsing documents, and a large proportion of the viewed elements was assessed as relevant or partially relevant to the task at hand. Overall, document elements were much more frequently visited, and were also assessed as more relevant, than full documents.

What was also evident, was that searchers tended to select document metadata (e.g. document title, authors, abstract, etc.) as their entry point for accessing the retrieved documents. This corresponded to searchers clicking on the title of the documents at the result list (Fig. 1), which might have led them to believe that they could access the full text of the document. However, they seldom chose to access the full text of documents from the metadata. Overall, searchers tended to find document metadata useful at providing an initial view of the document contents, and there was a preference for searchers to not click on retrieved elements at the ranked list level.

The results also indicated that elements up to the size of 200 words were predominantly marked as not relevant. For elements with size over 200 words, the elements assessed as not relevant are less than those marked as relevant or partially relevant. Elements marked as relevant or partially relevant were, in the majority of cases, between 10% and 40% of the size of full documents.

Ramirez and de Vries [9] discovered a dependency of preferred element size on the type of task that searchers were performing. For example, searchers performing broad search tasks that ask for information on many topics tended to prefer longer elements than those searchers performing tasks on topics with more practical applications.

The general trend of the results for preference of elements over full documents was also confirmed by other participants of the track in all three tasks [6, 7, 8]. These findings offer support to the usefulness of XML retrieval.

Dealing with element overlap. One of the critical issues of element retrieval is the possible retrieval of overlapping elements, i.e. components from the same document where one includes the other (due to the hierarchic structure of XML documents). Typically these elements are shown at non-adjacent ranks in the hit list. In the 2004 interactive track the presence of element overlap in the result list induced a specific, limited, interaction style by searchers [2]. In the 2005 track we tackled overlap by presenting a hierarchical grouping of elements from a single document in the result list (Fig. 1). Retrieved elements from a given document were also indicated at the table of contents at the detail view (Fig. 2).

Data from user interaction (i.e. logs), as well as from the questionnaires and the interviews, suggest that dealing with overlap in this way at the interface level tackles the issues that were raised in last year's track. A specific issue tackled was the observed "one-click interaction": searchers did not interact much with other components of a given XML document after the selection of the initial component from the result list. This behaviour was mainly caused by the fact that searchers would recognise a document that they had already visited by another retrieved element from that same document. This year, this behaviour was significantly reduced, indicating a much smoother interaction with the element retrieval system.

Participants of Task B [7] also chose to deal with overlap at the interface level in a similar manner. In their comparative study of the two XML retrieval systems, none of the searchers regarded the presence of overlapping elements as problematic. Similar findings were reported in Task C [8], where searchers consciously chose to view some overlapping elements while at the same time viewing less overlapping elements than expected.

Search task analysis. Ramirez and de Vries [9] performed an analysis of the searchers' own information needs. They analysed the descriptions that searchers provided, and classified search tasks along two dimensions: the specificity (narrow or broad) and complexity (simple or compound) of the task. They further analysed the intentions of the the searchers' information needs, and were able to classify them into five categories: application, decision, explanation, study and personal interest. The relationship between task characteristics and relevance assessments provided by searchers was also investigated.

Interface issues. A number of issues at the interface level were also highlighted [4]. First, the presence of the logical structure of the documents alongside the contents of the accessed components was a feature that searchers commented positively on. The Table of Contents of each document (see Fig. 2) seemed to provide sufficient context to searchers in order to decide on the usefulness of the document. Many users found the ToC of the whole article useful because it provided easy browsing, navigation, less scrolling and a quick overview of which elements may be relevant and which may be not.

In addition, query term highlighting at the result list and detail levels was found useful by searchers in directing their attention to potentially relevant parts of documents. Finally, marking the identified entry points (i.e. retrieved elements) at the detail level, and indicating the elements already visited and/or assessed were also found to help users tackling their tasks.

4 Conclusion and outlook

In its second year, the Interactive Track at INEX focused on addressing some fundamental issues of interactive XML retrieval. These include the validation of the element retrieval approach, user preference over varying element size, and the handling of element overlap. In addition, the track also expanded by including two additional tasks and by attracting more participating groups. A total of 11 research groups and 108 test persons participated in the three different tasks that were

included in the track.

The analysis of the 2005 data carried out so far supports the usefulness of element retrieval, and we believe that there is scope for further research, as well as for element retrieval applications. The collected data is rich and allows for further analysis, e.g. of the questionnaire and interview data and the relations of this with log data. An initial analysis of questionnaire and interview data in [7] for Task B demonstrates the usefulness of element retrieval. A user case track was also included in INEX 2006 to identify actual users of XML retrieval systems, how they might use XML retrieval systems, and for which realistic tasks.

The INEX 2006 interactive track is still ongoing at the time of writing this report. The track is using documents from the on-line encyclopaedia Wikipedia, making participation of a wider searcher population more feasible. In addition, three directions are explored in the track. First, element retrieval is compared against passage retrieval by allowing participants to use two variations of a search system. Second, a systematic categorisation of search tasks was made, allowing us to look into differences between different task types. Third, we investigate the effect that incorporating some further features on the interface level (e.g. summarisation at the result list level, use of related terms for query refinement, etc.) may have on the search process. The track will be completed in February 2007.

References

[1] Tombros, A., Larsen, B., Malik, S. The interactive track at INEX 2004. *In Proceedings of the Third International INEX Workshop (INEX 2004)*. Lecture Notes in Computer Science, Volume 3493, pp. 410-423. Springer-Verlag (2005).

[2] Tombros, A., Malik, S., Larsen, B. Report on the INEX 2004 interactive track. *SIGIR Forum* **39(1)**:43-49 (2005).

[3] Borlund, P. The IIR evaluation model: a framework for evaluation of interactive information retrieval. *Information Research*, 8(3), paper no. 152. [Available at: http://informationr.net/ir/8-3/paper152.html] (2003).

[4] 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 *Proceedings of the 10th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2006)*, pp. 291-302. Alicante, Spain (2006).

[5] Larsen, B., Tombros, A., Malik, S. Is XML retrieval meaningful to users? Searcher preferences for full documents vs. elements. In: *Proceedings of the 29th ACM SIGIR Conference*, pp. 663-664. Seattle, USA (2006).

[6] Kim, H. and Son, H. Users interaction with the hierarchically structured presentation in XML document retrieval. *In Proceedings of the Fourth International INEX Workshop (INEX 2005)*. Lecture Notes in Computer Science, vol. 3977, pp. 422-431. Springer-Verlag (2006).

[7] Kamps, J., Sigurbjrnsson, B. What do users think of an XML element retrieval system? *In Proceedings of the Fourth International INEX Workshop (INEX 2005)*. Lecture Notes in Computer Science, vol. 3977, pp. 411-421. Springer-Verlag (2006).

[8] Hammer-Aebi, B., Christensen, K.W., Lund, H., Larsen, B. Users, structured documents and overlap: interactive searching of elemenbts and the influence of context on search behavior. In *Proceedings of the First Symposium on Information Interaction in Context*, pp. 80-94. Copenhagen, Denmark (2006).

[9] Ramirez, G. and de Vries, A.P. relevant contextual features in XML retieval. In *Proceedings of the First Symposium on Information Interaction in Context*, pp. 95-110. Copenhagen, Denmark (2006).