# University of Amsterdam at INEX 2005: Interactive Track

Jaap Kamps[1,2], Maarten de Rijke[2], and Börkur Sigurbjörnsson[2]

[1] Archives and Information Science, Faculty of Humanities, University of Amsterdam
[2] Informatics Institute, Faculty of Science, University of Amsterdam

**Abstract.** This is a preliminary report on the University of Amsterdam's participation in the INEX 2005 Interactive Track. We participated in Task A, a common baseline system with the IEEE collection, as well as in Task B, in which the baseline system is compared to a home-grown XML element retrieval system, xmlfind.

## 1 Introduction

This paper documents the University of Amsterdam's participation in the INEX 2005 Interactive Track. We conducted two experiments. First, we took part in the concerted effort of Task A, in which a common baseline system, Daffodil/-HyREX, is used to study test-persons searching the IEEE collection, Second, as part of the Interactive Track's Task B, we conducted a comparative experiment, in which the baseline retrieval system, Daffodil/HyREX, is contrasted with our home-grown XML element retrieval system, xmlfind.

The rest of the paper is organized as follows. Next, Section 2 documents the XML retrieval systems used in the experiment. Then, in Section 3, we detail the set-up of the experiments. The preliminary results of the experiments are discussed in Section 4. Finally, in Section 5, we draw some initial conclusions.

## 2 XML Retrieval Systems

### 2.1 Baseline System: Daffodil

The Daffodil system supports the information seeking process in Digital Libraries [2]. As a back-end, the HyREX XML retrieval system was used [3]. For details, see [4].

### 2.2 Home-grown System: xmlfind

The xmlfind system provides an interface for a XML information retrieval search engine [1]. It runs on top of a Lucene search engine [5]. The underlying index contains individual XML element in the IEEE collection [6].

Figure 1(top) shows the search box and the result list. The results are grouped per article, where (potentially) relevant elements are shown. Clicking on any of

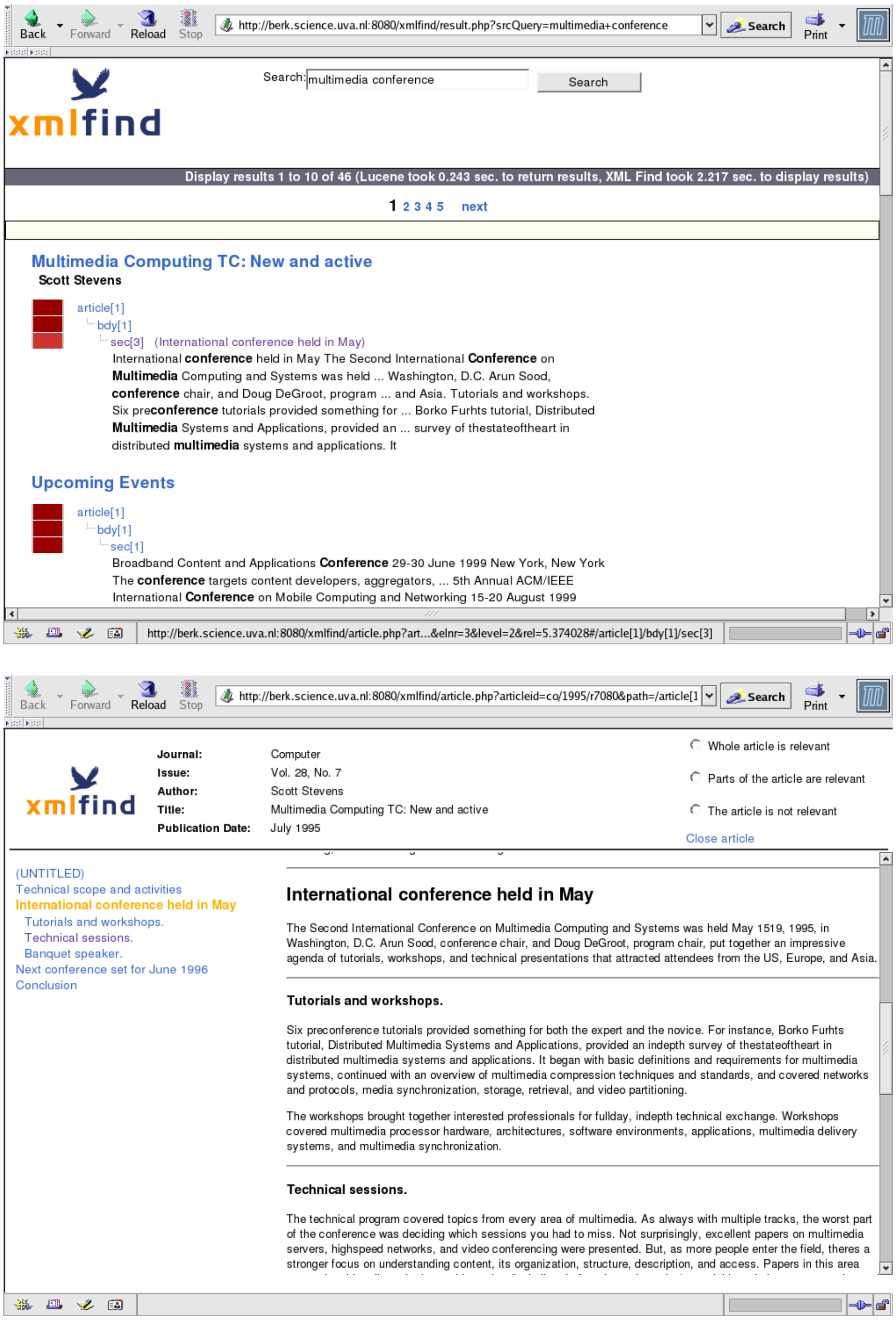

**Fig. 1.** Screen shots of xmlfind: (top) result list, (bottom) detailed view.

**Table 1.** Experimental matrix for the comparative experiment.

| # | Rotation | Task 1 | | Task 2 | | Task 3 | |
|---|---|---|---|---|---|---|---|
| | | Task | System | Task | System | Task | System |
| 1 | 1 | G-1 | Daffodil | C-1 | xmlfind | Own | choice |
| 2 | 2 | C-1 | Daffodil | G-1 | xmlfind | Own | choice |
| 3 | 3 | G-1 | xmlfind | C-1 | Daffodil | Own | choice |
| 4 | 4 | C-1 | xmlfind | G-1 | Daffodil | Own | choice |
| 5 | 1 | G-2 | Daffodil | C-2 | xmlfind | Own | choice |
| 6 | 2 | C-2 | Daffodil | G-2 | xmlfind | Own | choice |
| 7 | 3 | G-2 | xmlfind | C-2 | Daffodil | Own | choice |
| 8 | 4 | C-2 | xmlfind | G-2 | Daffodil | Own | choice |
| 9 | 1 | G-3 | Daffodil | C-3 | xmlfind | Own | choice |
| 10 | 2 | C-3 | Daffodil | G-3 | xmlfind | Own | choice |
| 11 | 3 | G-3 | xmlfind | C-3 | Daffodil | Own | choice |
| 12 | 4 | C-3 | xmlfind | G-3 | Daffodil | Own | choice |
| 13 | 1 | G-1 | Daffodil | C-1 | xmlfind | Own | choice |
| 14 | 2 | C-1 | Daffodil | G-1 | xmlfind | Own | choice |

the elements will open a new window displaying the result. Figure 1(bottom) shows the full article with the focus on the selected element. The results display window has three planes. On the left plane, there is a Table of Contents of the whole article. On the right plane, the article is displayed with the selected part of the document in view. On the top plane, the article's title, author, etc. are displayed, as well as a menu for assessing the relevance of the result.

## 3 Experimental Setup

The whole experiment was run in a single session where test persons for both Task A and Task B worked in parallel. The test persons were first year Computer Science students.

### 3.1 Task A: Community Experiment

Task A is the orchestrated experiment in which all teams participating in the Interactive Track take part [4]. We participated in Task A with six test persons, who searched the IEEE Collection with the Daffodil/HyREX baseline system. There were three tasks: two simulated work tasks (a 'general' task and a 'challenging' task) and the test person's were asked to think up a search topic of their own. The experiment was conducted in close accordance with the guidelines, for further details we refer to [4].

### 3.2 Task B: Comparative Experiment

Task B is a comparison of the home-grown xmlfind system with the Daffodil/-HyREX baseline system. We participated in Task B with fourteen test persons.

**Table 2.** Topic created by test person.

| |
|---|
| A. *What are you looking for?*<br>Who build the first computer and what did it look like? |
| B. *What is the motivation of the topic?*<br>I would like to know how the history of the computer began and what the first computer looked like, was it very big or very small, did it have a monitor? |
| C. *What would an ideal answer look like?*<br>The name of the inventor and a picture of how the first computer looked. |

The experimental setup is largely resembling the setup of Task A. Again, test persons do two simulated work tasks (a 'general' and a 'challenging' task) as well as search for a topic they were asked to think up themselves. The experimental matrix is shown in Table 1. Every test person searches for two simulated tasks, each one with a different system. Next, the test persons search for their own topic with a system of their choice.

Due to the number of test persons involved, we were unable to conduct individual exit interviews. Instead, we used an extended post-experiment questionnaire.

## 4 Preliminary Results

We have only started to process the massive amount of data collected during the experiment. Each test person searched with four different accounts, one for each task, plus one or two additional accounts for training. This generated in total 94 search logs. Additional to this, each person filled in questionnaires before and after each task, and before and after the experiment, resulting in, in total, 160 questionnaires. Below, we will just give some preliminary results.

**Own topics** As part of the experiments, test persons were asked to think up a search topic of their own interest, based on a short description of the IEEE collection's content. Some topics created by test persons were excellent. Table 2 shows an example of a topic being (i) within the collection's coverage, (ii) reflecting a focused information need, and (iii) even containing potential structural retrieval cues. However, most topic were not so perfect. Even though test persons were asked to think up two different topics, almost half of the test persons did not create a very suitable topic. Frequently, topics addressed very practical advice on computer components or software, such as addressed in FAQs, and some were simply off-topic. Perhaps more positively, the own topics were for the vast majority focused, asking for very specific information that could, in principle, be contained in a relatively short piece of text.

**Table 3.** Responses by test persons.

| 13. *Did you like the idea that the search engine takes into account the structure of the documents? Why?* | 14. *Do you find it useful to be pointed to relevant parts of long articles? Why?* |
|---|---|
| Yes, you will have a good overview of the total article/document. | Yes, because you are able to see which articles are worth reading and which are not. |
| Yes, for specific information this is very useful. | Yes, gives the user an idea about the article in question. |
| Yes, easier to see how long the article is. | You don't need to see other parts. |
| Yes, its easier to see the contents of the document, better navigation. | Yes, you don't have to dig into the article yourself. |
| Yes, it didn't bother me. | Yes, it's more easy to find what you're looking for. |
| Yes, less reading time, clear overview. | Yes, saves time. |
| Yes, it shortens search time. | Yes, because if scan-read long articles, you easily miss some relevant parts. |
| Yes, saves work. | Yes, works faster. |
| Yes, because its much faster. | Yes, its faster. |
| Yes, this way of finding information takes less time. | Yes, now you don't have to read the whole article. You can get straight to the part where the information is. |
| Yes, its easier to see where relevant information is located. | Yes, it takes less time to find the relevant parts. |
| Yes, it makes it easier to find specific paragraphs. | Yes, if programmed right it can save time. |
| Yes, it makes it a lot easier to find what you are looking for. | Yes, it is lots more easier. |
| Yes, because makes me have to search less. | Yes, to search less. |

**System of Choice** Test persons in Task B were free to select with which of the two system they searched for the third topic. Out of the 14 test persons, 4 (28.6%) choose to search with the Daffodil/HyREX system, the other 10 (71.4%) choose to search with the xmlfind system.

**General Views** Test persons in Task B were, as part of the extended post-experiment questionnaire, asked a number of questions about their opinions on the concept of an XML retrieval engine. Table 3 lists the responses to two of the questions, where each row represents the same test person. The responses where equivocally positive, and the responses highlight many of the hoped advantages of an XML retrieval system.

## 5 Discussion and Conclusions

This paper documents the University of Amsterdam's participation in the INEX 2005 Interactive Track. We participated in two tasks. First, we participated in the concerted effort of Task A, in which a common baseline system, Daffodil/-HyREX, was used by six test-persons to search the IEEE collection, Second, we conducted a comparative experiment in Task B, in which fourteen test persons searched alternately with the baseline retrieval system, Daffodil/HyREX, and our home-grown XML element retrieval system, xmlfind.

We detailed the experimental set-up of the comparative experiment. Both experiments, involving in total twenty test persons, were conducted in parallel in a single session. This ensured that the experimental conditions for all test persons are very equal. Unplanned external causes, such as the down-time of the Daffodil/HyREX system were affecting all test persons equally. Due to the large number of test persons present at the same time, we had to minimize the need for experimenter assistance. This was accomplished by generating personalized protocols for all test persons. In these protocols, test persons were guided through the experiment by means of verbose instructions on the transitions between different tasks. Four experimenters were available, if needed, to clarify the instructions or provide other assistance. This worked flawlessly, and allowed us to handle the large numbers of test persons efficiently.

Although we are still in the process of analyzing the massive amount of data collected during the experiments, we discussed a few initial results. The general opinion on the XML retrieval systems was equivocally positive. Departing from earlier systems that return ranked lists of XML elements, both the Daffodil/HyREX and xmlfind are grouping the found XML elements per article (similar to the Fetch & Browse task in the Adhoc Track). Test persons seem to conceive the resulting system as an article retrieval engines with some additional features—yet with great appreciation for the bells and whistles!

**Acknowledgments**

This research was supported by the Netherlands Organization for Scientific Research (NWO) under project numbers 017.001.190, 220-80-001, 264-70-050, 612-13-001, 612.000.106, 612.000.207, 612.066.302, 612.069.006, and 640.001.501.

## References

1. T. Bakker, M. Bedeker, S. van den Berg, P. van Blokland, J. de Lau, O. Kiszer, S. Reus, and J. Salomon. Evaluating XML retrieval interfaces: xmlfind. Technical report, University of Amsterdam, 2005.
2. Daffodil. Distributed Agents for User-Friendly Access of Digital Libraries, 2005. `http://www.is.informatik.uni-duisburg.de/projects/daffodil/`.
3. HyREX. Hyper-media Retrieval Engine for XML, 2005. `http://www.is.informatik.uni-duisburg.de/projects/hyrex/`.
4. B. Larsen, S. Malik, and T. Tombros. The interactive track at INEX 2005. In *This Volume*, 2005.
5. Lucene. The Lucene search engine, 2005. `http://lucene.apache.org/`.
6. B. Sigurbjörnsson, J. Kamps, and M. de Rijke. An Element-Based Approch to XML Retrieval. In *INEX 2003 Workshop Proceedings*, pages 19–26, 2004.