Is XML Retrieval Meaningful to Users? Searcher Preferences for Full Documents vs. Elements

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
London, U.K.
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 aim of this study is to investigate whether element retrieval (as opposed to full-text retrieval) is meaningful and useful for searchers when carrying out information-seeking tasks. Our results suggest that searchers find the structural breakdown of documents useful when browsing within retrieved documents, and provide support for the usefulness of element retrieval in interactive settings.

Categories and Subject Descriptors
H.3 [**Information Storage and Retrieval**]: H.3.3 Information Search and Retrieval - Search process; H.3.7 Digital Libraries.

General Terms
Design, Experimentation, Human Factors.

Keywords
XML retrieval, element retrieval, interactive studies, INEX

1. INTRODUCTION
A substantial research effort is put into XML retrieval, with the Initiative for the Evaluation of XML Retrieval (INEX) as the main driving force. Noteworthy advances have been made in the investigation of the possible benefits of document structure in Information Retrieval (IR) (see, e.g., [1]). At the present state we may draw on this knowledge to design and test IR techniques that can index and retrieve elements from XML documents that have a high likelihood of being relevant. We have, however, little knowledge about whether users would at all opt for this feature if implemented in, e.g., a digital library search engine [3].

A first question to ask, is whether making elements retrievable is worth the added effort: Are elements valuable to users in a retrieval situation, or are users just as well served by IR systems that retrieve whole documents? In this paper, we examine indications of searcher preferences for whole documents versus elements from their behaviour in an interactive experiment. More specifically, we adress the following research questions:

1. Do searchers opt for whole documents or elements in the hitlist of an XML IR system?
2. Do searchers view the full text of, and assess as relevant, whole documents or elements?

SIGIR'06, August 6–11, 2006, Seattle, Washington, USA.
ACM 1-59593-369-7/06/0008.

To appear as:
Larsen, B., Tombros, A. and Malik, S. (2006): Is XML retrieval meaningful to users? Searcher preferences for full documents vs. elements. In: Proceedings of SIGIR 2006 : the twenty-ninth annual international ACM SIGIR conference on research and development in information retrieval, August 6-11, 2006, Seattle, Washington, USA. New York: ACM Press (Poster paper - accepted for publication)

2. INTERACTIVE EXPERIMENT
This study was part of the Interactive Track at INEX 2005 (see [2] for details), where 73 test persons performed 219 tasks: each searched two given work tasks (selected from two categories) and one of their own (11 of these tasks had to be discarded due to logging problems). The corpus consisted of articles from the IEEE Computer Society's journals, and a maximum of 20 minutes were given to complete a task. The tasks were so-called content only tasks; that is, they did not refer to the XML structure or require the answers to be either elements or full documents.

In response to a free-text query, the XML IR system returned a hitlist of selected high ranking elements (represented by their titles), grouped by the containing documents (represented by title, author, journal and year). Both the elements and the document titles provided access to the full-text view: clicking a document title displayed document metadata (including an abstract) but not the full document. Clicking an element title displayed the full text of the element directly in a new view. We show the full-text view in Figure 1 (the hitlist is not shown for lack of space). The full-text view always showed a table of contents (ToC) of elements in the document, and the full text of the selection. The following document levels could be viewed: **article**, **metadata**, sections (**sec**), sub-sections (**ss1**) and sub-sub-sections (**ss2**). Searchers were instructed to assess all viewed elements, but not forced to do so by the system. Relevance assessments could be given on a 3-grade scale: Relevant, Partially relevant and Not relevant.

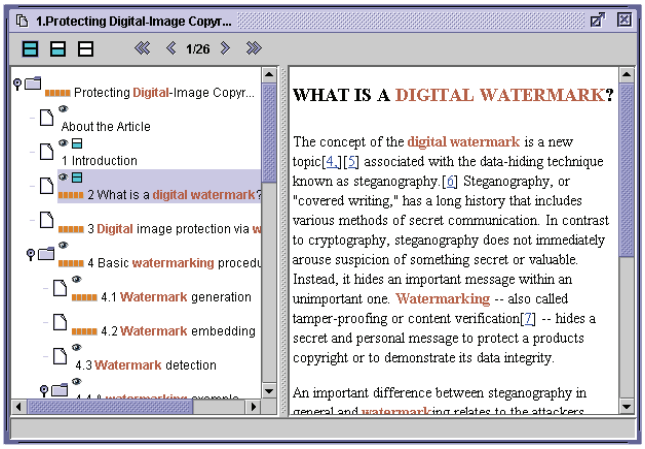

Figure 1. Full-text view with Table of Contents (ToC)

Searchers were given a full system tutorial before the start of search sessions. All interactions with the system were logged in detail. In this paper, we analyze the log data for aspects of searcher preference for whole documents vs. elements.

3. EXPERIMENTAL RESULTS

Below we give an overview of the main results of the experiment. In the analysis we do not take the possible overlap between elements into consideration (i.e. a subsection and its containing section are both counted independently).

A total of 1371 documents were accessed in the experiment. In the hitlist these documents were each represented by the document title, authors, journal and year, and an additional 3.2 clickable elements on average, e.g., sections and subsections. Searchers predominantly clicked the title of the whole document as their entry point to the full text: 71% of the available documents were accessed this way displaying metadata in the full-text view, even though a large number of sections and subsections also could have been used as direct access. Sections accounted for 17% of the entry points, sub-sections 11% and sub-sub-sections only 1%.

Table 1 shows data for the full-text view (see Figure 1). Here more elements per document were available, because all elements (from the levels described) were shown in the ToC: 15.3 on average (including one set of metadata per document). Percentages of *Viewed* are in relation to *Available*, and percentages of *Assessed* are in relation to *Viewed*.

Table 1. Available, viewed and assessed elements in the full text view (includes entry points from the hitlist)

	Available	Viewed	Assessed
article	1371 -	251 (18%)	189 (75%)
metadata	1371 (7%)	1007 (73%)	383 (38%)
sec	9372 (45%)	1960 (21%)	1455 (74%)
ss1	7910 (38%)	906 (11%)	644 (71%)
ss2	2376 (11%)	121 (5%)	81 (67%)
Sum	21029 (100%)	4245 (20%)	2752 (65%)

The difference between the actually viewed elements (including whole articles) in Table 1 with the ones accessed from the hitlist is noticeable: of the 4245 viewed elements, only 1007 were metadata (24%) and almost all of these (987) were the entry points clicked in the hitlist. Note that, in contrast to the hitlist, whole articles were accessible in the full-text view; this was requested in 251 of the 1371 documents accessed (18%). Overall, sections and elements smaller than sections accounted for 2987 or 70% of all viewed items. On average, per 20 minute task only 6.6 documents were examined, but within these documents 14.4 sections and smaller elements were inspected per task.

The total number of assessments (including Not relevant) is given in Table 1. As searchers were not forced to assess all viewed elements, only 65% were explicitly assessed. Overall, a notably smaller proportion of metadata (38%) were assessed compared to other elements (and many of these as Not relevant – see Figure 2).

Figure 2 shows the distribution of assessed as well as un-assessed elements over relevance grades. The total number of assessments for the 3 grades is quite alike. For the Relevant and Partially relevant, the distribution is very similar: sections accounted for the largest proportion, followed by sub-sections. Comparing articles and metadata, more articles were assessed Relevant and more metadata were assessed Partially relevant. Examining the Not relevant and Not assessed, metadata stands out: a much larger proportion consisted of metadata in both cases; especially in the Not assessed. However, sections and subsections also accounted for a large proportion of the Not relevant and the Not assessed.

On the whole, searchers tended to view and assess a relatively large number of sections and subsections when browsing the full text, and a large proportion of these were assessed as Relevant or Partially relevant; of the 2987 viewed representations of elements (sec, ss1, ss2) 51% were Relevant or Partially relevant.

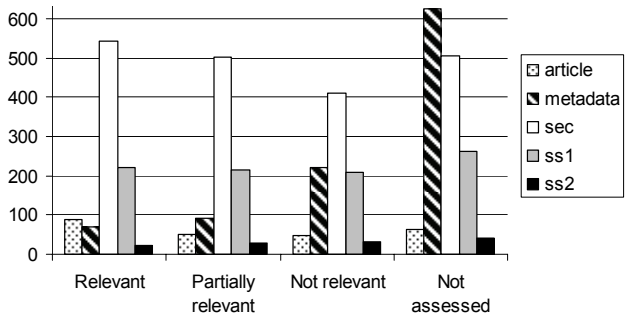

Figure 2. Distribution of the 2752 assessments over elements, as well as the 1493 viewed but un-assessed elements.

4. DISCUSSION AND FUTURE WORK

Our results suggest that searchers predominantly selected metadata as their entry point for accessing the retrieved documents. This corresponded to searchers clicking on the title of the documents, which might have led them to believe that they could access the full text of the document. The insistence of searchers to select this entry point from the ranked list, even when it becomes evident that it does not provide them with access to the full text, can be attributed to two causes: either that the information given by metadata was useful, or that they expected that at some point they may be given access to the full text by this action. In either case, there is a strong preference for searchers to not click on retrieved elements at the ranked list level.

This picture changes significantly when searchers are presented with the full-text view (Figure 1). Elements are much more frequently visited, and the proportion of relevant items is at the same level as that of full documents. In conclusion, this suggests that searchers do find elements useful for their tasks, and that they find a lot of the relevant information in specific elements rather than full documents. Sections, in particular, appear to be helpful. Our results thus support the ongoing effort invested in XML retrieval in INEX and elsewhere.

As this is only an initial analysis, we plan to analyse the data in greater detail, e.g., by examining possible differences between individual tasks and tasks types as some tasks may be better served by XML retrieval than others. The results can be correlated with qualitative data that were also collected through questionnaires and interviews. Such results may aid in establishing a much needed user model for element retrieval [3], and inform decisions about, e.g., document representation, matching models and performance metrics.

5. REFERENCES

[1] Fuhr, N., Lalmas, M., Malik, S. and Szlávik, Z. (2005): *Advances in XML Information Retrieval : INEX 2004 Proceedings*. LNCS 3493.

[2] Larsen, B., Malik, S. and Tombros, A. (2005): The interactive track at INEX 2005. In: Fuhr, N. et al. eds. *INEX 2005 Pre-proceedings,* p. 313-326. [http://inex.is.informatik.uni-duisburg.de/2005/]

[3] Trotman, A. (2005): Wanted: element retrieval users. In: *Proceedings of the INEX 2005 Workshop on Element Retrieval Methodology, Glasgow, 30 July 2005*, p. 63-69. [http://www.cs.otago.ac.nz/inexmw/Proceedings.pdf]