The Use of Summaries in XML Retrieval

Zoltán Szlávik, Anastasios Tombros, and Mounia Lalmas

Department of Computer Science,
Queen Mary University of London

Abstract. The availability of the logical structure of documents in content-oriented XML retrieval can be beneficial for users of XML retrieval systems. However, research into structured document retrieval has so far not systematically examined how structure can be used to facilitate the search process of users. We investigate how users of an XML retrieval system can be supported in their search process, if at all, through summarisation. To answer this question, an interactive information retrieval system was developed and a study using human searchers was conducted. The results show that searchers actively utilise the provided summaries, and that summary usage varied at different levels of the XML document structure. The results have implications for the design of interactive XML retrieval systems.

1 Introduction

As the *eXtensible Markup Language (XML)* is becoming increasingly used in digital libraries (DL), retrieval engines that allow search within collections of XML documents are being developed. In addition to textual information, XML documents provide a markup that allows the representation of the logical structure of XML documents in content-oriented retrieval. The logical units, called elements, are encoded in a tree-like structure by XML tags. The logical structure allows DL systems to return document portions that may be more relevant to the user than the whole document, e.g. if a searcher wants to read about how Romeo and Juliet met, we do not return the whole play but the actual scene about the meeting. This content-oriented retrieval has received large interest over the last few years, mainly through the INEX initiative [6].

As the number of XML elements is typically large (much larger than that of documents), we believe it is essential to provide users of XML information retrieval systems with overviews of the contents of the retrieved elements. One approach is to use summarisation, which has been shown to be useful in interactive information retrieval (IIR) [9, 7, 15].

In this paper, we investigate the use of summarisation in XML retrieval in an interactive environment. In interactive XML retrieval, a summary can be associated with each document element returned by the XML retrieval system. Because of the nature of XML documents, users can, in addition to accessing any retrieved element, browse within the document containing that element. One method to allow browsing XML documents is to display the logical structure

of the document containing the retrieved elements [13]. This has the benefit of providing (sometimes necessary) context to users when reading an element. Therefore, summaries can also be associated with the other elements of the document, in addition to the returned elements themselves.

The aim of our work is to investigate how users of an XML retrieval system can be supported in their search process, if at all, through summarisation. To answer this question, an interactive information retrieval system was developed and a study using human searchers was conducted.

The paper is organised as follows. In Section 2 we present the background of our work, then we describe the experimental system and methodology that was used in Section 3. The analysis of our data is described in Section 4, which is followed by the conclusions and future work.

2 Background

In recent years, interactive aspects of the IR process have been extensively investigated. Major advances have 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 following links to another document part.

The interactive aspect of XML IR has recently been investigated through the interactive track at INEX (iTrack) [13, 10, 8]. A major result from iTrack 2004 was that searchers did not interact enough with the elements of retrieved XML documents [14]. Searchers seemed to appreciate the logical structure of XML documents as a means of providing context for identifying interesting XML elements within a document, but they did not browse much within XML documents. Tombros et al. suggest that this behaviour may have been due to limitations of the interactive XML IR system used. Among these limitations was that XML element (or document) summarisation capabilities were few, and therefore searchers did not have enough relevance clues to decide which elements to visit [14]. In this paper, we focus on the presentation of the document structure as a hierarchical table of contents, and on the use of summarisation to facilitate the users' search process.

Text summarisation has attracted attention primarily after the information explosion on the Internet; however, significant work was done as early as the 1950's and 1960's. Edmundson proposed extraction methods considering various sentence features, e.g. location, title words [5]. In recent summarisation systems, users' query terms are also considered in generating summaries [15]. Few researchers recently have investigated the summarisation of information available in XML format (e.g. [1, 2]). In our work, we considered a simple summarisation algorithm that takes advantage of the sentence location and the query (referred to as query-biased), as our main aim is to study how users "interact" with summaries.

The use of summaries in interactive IR has been shown to be useful for various information seeking tasks in a number of environments such as the web (e.g. [16, 4]). However, in the context of interactive XML retrieval, summarisation has not yet been investigated extensively. Our main focus in this paper is to study how searchers behave in an environment that provides them with structural documents, and how they use summaries of document elements that are presented to them. To do so, we created and tested, through user-based studies, an interactive XML retrieval system with XML element summarisation capabilities. We describe the system and the setup of our study in the next section.

3 Experimental Setup

In this section, we describe the system and method that was used in our study. We include only the necessary details for the presentation of the analyis and results reported in this paper. A more detailed description can be found in [12].

User Interface The user interface is a web based system which passes the query to a retrieval module, processes and displays the retrieved list of elements and shows each of these elements. The system allows users to enter a search query and start the retrieval process by clicking on the search button. The display of the list of retrieved elements is similar to standard search interfaces (Figure 1).

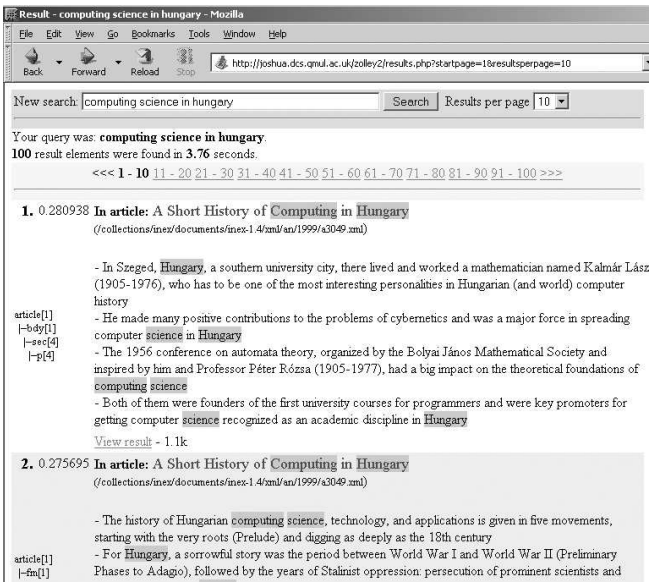


Fig. 1. The list of the result elements.

Once searchers follow the link to a particular result element, the element is displayed in a new window (Figure 2). The frame on the right shows the content of the target element. The structure is displayed on the left as an automatically generated table of contents (ToC) where each *structural item* is a hyperlink that will show the corresponding XML element in the right window when clicked. For this user study, four levels of structural items were displayed. Level one always refers to the whole article; level two contains the body, front and backmatters; level three usually contains the abstract, sections and appendices; and level four usually means subsections or paragraphs, depending of the inner structure of articles. The number of levels could be changed by searchers. For each item in the ToC, summaries were generated and displayed as 'tool tips', i.e. when users moved the mouse pointer over an item in the ToC, the summary of the target element was shown. Query terms in the summaries, as well as in the displayed documents, were highlighted.

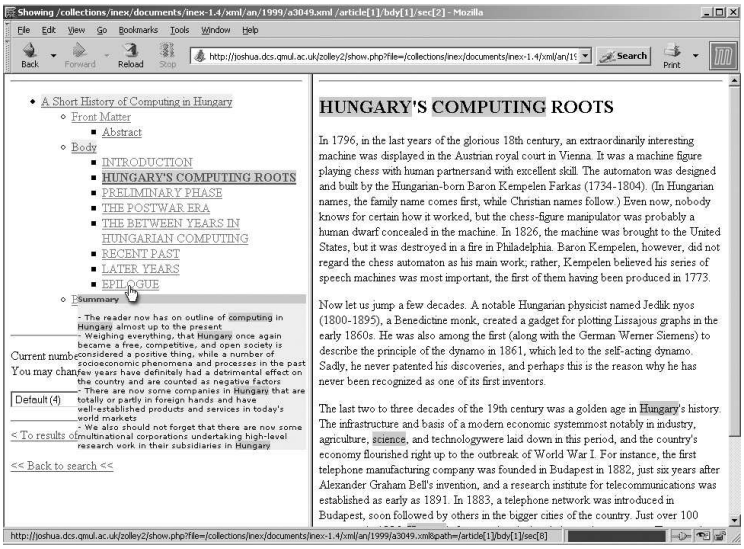


Fig. 2. On the left, the structure of the XML document with a summary; on the right, the content of a section element displayed.

Summarisation Summaries were displayed in the result list view for each result element and for the displayed elements in the ToC in element view. Since our aim at this stage of the research was not to develop sophisticated summarisation methods, but to investigate summary usage in XML retrieval, we implemented and used a simple query-biased algorithm. Four sentences with the highest scores were presented as extracts of the source XML elements, in order of appearance in the source element (for further details, see [12]).

Document Collection The document collection we used was the IEEE collection (version 1.4) which contains 12,107 articles, marked up in XML, of the IEEE Computer Society's publications from 12 magazines and 6 transactions, covering the period of 1995-2002. On average, an article contains 1532 XML nodes and the average depth of a node is 6.9. These properties provided us with a suitably large collection of articles of varying depth of logical structure.

XML Retrieval Engine The retrieval was based on the HySpirit retrieval framework [11]. To be able to examine the relation between the structural display and the use of summaries, only paragraphs were returned as retrieval results. This strategy ensured that elements deeply nested in a document logical structure were returned, so as to "force" searchers to browse through the structural display on the left panel of Figure 2 (instead of simply scrolling down the right window).

Searchers Twelve searchers (9 males and 3 females) were recruited for this study. All of them had computer science background as the collection used contained articles from the field of computer science.

Experimental and Control Systems Two versions of the developed system were used in this study. The control system (S_c) had all the functionalities we described in previous sections, whereas the experimental system (S_e) differed in the display mode of summaries: System S_e displayed summaries only at high levels in the hierarchical structure, i.e. the upper three levels had associated summaries, the fourth level did not. The rationale behind this is that we wanted to see whether searchers' behaviour is affected by the different display. To avoid bias towards the use of the hierarchical structure and summarisation, we employed a blind study, i.e. searchers were not told what the purpose of the study was.

Tasks Four search tasks were used in the experiments. The tasks described simulated work task situations [3]. We used modified versions of the INEX 2005 ad-hoc track topics which ensured that the tasks were realistic, and that relevant documents could be found in the document collection. Two types of search tasks were chosen. Background type tasks instructed searchers to look for information about a certain topic (e.g. concerns about the CIA and FBI's monitoring of the public) while List type tasks asked searchers to create a list of products that are connected to the topic of their tasks (e.g. a list of speech recognition software). From each group of tasks, searchers could freely choose the one that was more interesting to them. Searchers had a maximum of 20 minutes for each task. This period is defined as a *search session*. Search sessions of the same searcher (i.e. one searcher had two search sessions) are defined and used in this paper as a *user session*.

Search Design To rule out the fatigue and learning effects that could affect the results, we adopted Latin square design. Participants were randomly assigned

into groups of four. Within groups, the system order and the task order were permuted, i.e. each searcher performed two tasks on different systems which involved two different task types. We made an effort to keep situational variables constant, e.g. the same computer settings were used for each subject, the same (and only) experimenter was present, and the place of the experiments was the same.

Data Collected Two types of events were logged. One type was used to save the users' actions based on their mouse clicks (e.g. when users clicked on the 'search' button, or opened an element for reading). The other type corresponds to the summary-viewing actions of users, i.e. we logged whenever a summary was displayed (users moved the mouse pointer over an item in the ToC).

During the analysis of summary log files, calculated summary-viewing times that were shorter than half a second or longer than twenty seconds were discarded, because the former probably corresponds to a quick mouse move (without users having read the summary), and the latter may have recorded user actions when the keyboard only was used (e.g. opening another window by pressing CTRL+N).

4 Analysis

In this section, the analysis of the recorded log files is described. To investigate whether summarisation can be effectively used in interactive XML retrieval, we formed four groups of research questions. The first group (Section 4.1) is about summary reading times. The second group (Section 4.2) is about the number of summaries searchers read in their search sessions. Section 4.3 investigates the relation between summary reading times and number of summaries read (the third group). The fourth group (Section 4.4) looks into the relation between the multi-level XML retrieval and traditional retrieval.

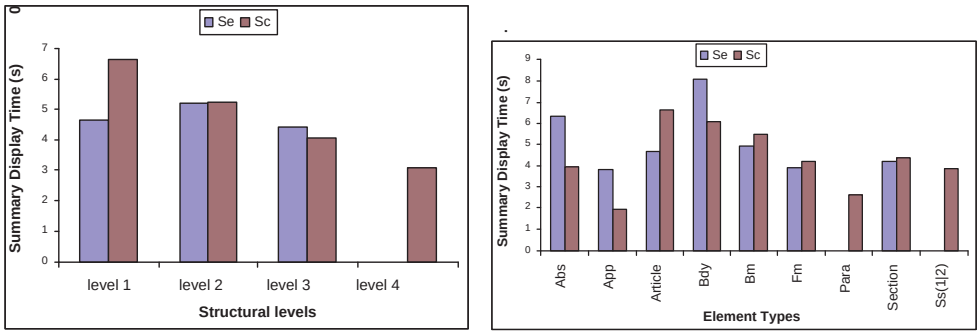
4.1 Summary Times

In this section, we examine how long searchers read an average summary; whether there are differences in reading times for summaries that are associated with elements at various structural levels and element types; and whether the average summary reading time changes when summaries are not shown at all structural levels in the ToC.

Taking into account both systems S_e and S_c, an average summary was displayed for 4.24 seconds with a standard deviation of 3.9. The longest viewed summary was displayed for 19.57s, while the shortest accounted summary was viewed for 0.51s.

Figure 3a shows the distribution of summary display times by structural levels for each system. Display times of S_c tend to be shorter when users read summaries of deeper, i.e. shorter elements, although the length of summaries

were the same (i.e. four sentences). For S_e, times are more balanced. This indicates that if there are summaries for more levels and the lowest level is very short (sometimes these paragraphs are as short as the summary itself), people trust summaries of larger, i.e. higher, elements more. If the difference in size between the deepest and highest elements is not so big, times are more balanced.



(a) **(b)**

Fig. 3. Summary times by structural levels (a) and XML element types (b)

Figure 3b shows the display time distribution by XML element types (tags). We can see, for example, that the *bdy* (body) element has high summary viewing times; this is the element that contains all the main text of the article. We can also see that paragraphs (*para*) and subsections ($ss1$ and $ss2$) have low summary reading times for S_c and, obviously, none for S_e (as they are not displayed at these levels). These three element types appear on the lowest, i.e. fourth, structural level.

We compared the two systems (S_e and S_c) to find out whether significant differences in summary reading times can be found. The comparison of the overall summary-viewing times showed significant difference between S_e and S_c, i.e. the average summary viewing time for system S_e (4.58s) is significantly higher than that of system S_c (3.98s). To examine where this difference comes from, we compared the two systems by tag types (e.g. whether summary reading times for sections are different for the two systems). However, we did not find significant differences at comparable tag types[1]. We also compared the two systems with respect to structural levels (e.g. whether average summary reading time at level one is significantly different for the two systems). We did not find significant difference for level one (article), two (body, front and back matters) and three (abstract, sections, appendix) elements.

[1] Tag types for which summaries were not displayed for any of the systems were not compared as one of the sample groups would contain zero samples.

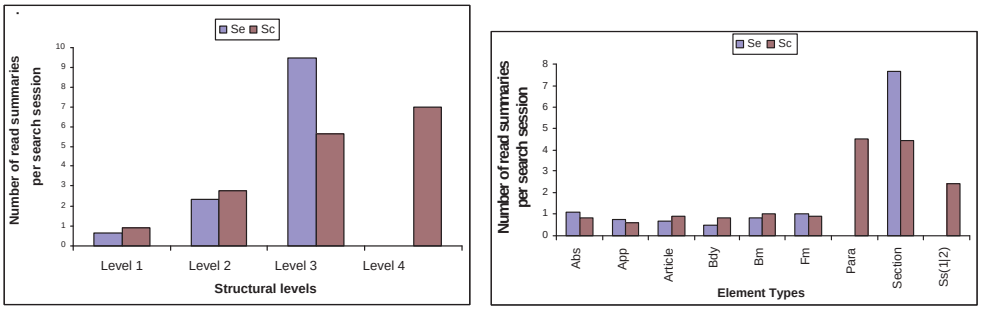
To sum up, our results showed that users of system S_e read summaries 0.5s longer than that of system S_c. However, we could not find significant difference at levels or element types between the two systems. An interpretation of this result is that since S_e searchers had less available summaries to examine, they were less confused and overloaded by the information available and could take their time reading a particular summary.

4.2 Number of Summaries Read

This section looks into the number of summaries that were read by searchers. We first examine the average number of summaries seen by users in a search session, and then we look into the distributions of the number of read summaries at different structural levels and element types. Differences between the two systems with respect to the number of read summaries are also discussed in this section.

Considering both systems together, an average user read 14.42 summaries in a search session (20 minutes long), with a standard deviation of 10.77. This shows a considerable difference in user behaviour regarding summary reading. The least active summary reader read only one summary in a search session, while the most active saw 52 summaries for at least half a second.

Figure 4a shows that the deeper we are in the structure of the ToC, the more summaries are read, on average, in a search session. This is consistent with the nature of XML, and all tree-like structures: the deeper we are in a tree, the more elements are available on that level. However, our data shows that the difference between the two systems is not only based on this structural property, because when only three levels of summaries were displayed, reading of third level summaries (usually summaries of sections) showed higher activity than when four levels of summaries were displayed, i.e. the third level seems to be more interesting than the first and second.



(a) **(b)**

Fig. 4. Number of read summaries per search sessions, by structural levels (a) and XML element types (b)

The next step is to find out whether this interest is only at these deeper levels, or connected to some element types. Contents of the same element types are supposed to have the same amount and kind of information, e.g. paragraphs are a few sentences long; front matters usually contain author, title information and the abstract of the paper. Our log analysis shows that summaries of sections, subsections and paragraphs are those most read (Figure 4b), although users take less time to read them (see previous section). Other tag types are less promising to users according to their summary usage. We can also see in Figure 4b that when paragraph and subsection summaries are not available (S_e), section summary reading increases dramatically. We interpret these results as indication that for the IEEE collection, sections, that appear mostly at level three, are the most promising elements to look at when answering an average query.

The comparison of the overall number of viewed summaries showed that an average user of system S_e read 12.5 summaries per search session, and of system S_c 16.33 summaries per session. In other words, our test persons read more summaries where more summaries were available. However, this difference is not statistically significant.

We compared the two systems using the same categories (i.e. tag types and levels) as previously for summary reading times. T-tests did not show significant differences at comparable levels and element types between S_e and S_c in number of read summaries.

4.3 Reading Times vs. Number of Read Summaries

In this section, we examine the relationships between the data and findings of the previous two sections. One question we are looking into is whether searchers with higher summary reading times read less summaries in a search session.

Users of system S_e read less summaries than those who used system S_c. This is in accordance that they had less summaries available. However, users of system S_e also read summaries for longer. This shows that if there are less available summaries, users can focus more on one particular summary, and vice versa, if there are many summaries to view, reading can become superficial.

Considering both systems and tag types, we found negative correlation between the summary reading time and the number of read summaries. In other words, it is true for users of both systems that the more summaries they read on a particular level the shorter the corresponding reading times are. However, this is only an indication as the correlation coefficient (-0.5) does not indicate significance. Also, since the number of summaries read increases when going deeper in the structure, we view this as an indication that, for searchers, summaries of higher level elements are more indicative to the contents of the corresponding elements than those of lower, and also shorter, elements.

4.4 Usage of the ToC and Article Level

XML retrieval has the advantage of breaking down a document into smaller elements according to the document's logical structure. We investigated whether

searchers take advantage of this structure: do they click on items in the ToC, do they use the article (unstructured) level of a document, and how frequently, do they alternate between full article and smaller XML element views?

Regarding the usage of the XML structure in terms of the ToC, 58.16% of the displayed elements were results of at least "second" clicks, i.e. more than half of the elements were displayed by clicking on an element in the ToC. This shows that searchers actively used the ToC provided (unlike those in [14]), and that they used the logical structure of the documents by browsing within the ToC.

The log files show that only 25% of the searchers clicked on article elements to access the whole document, and none of these searchers clicked on an article type link more than three times in a search session. The distribution of viewing whole articles did not depend on the system, i.e. we observed three article clicks for each system. This result follows naturally, since the display of the article level was not different in S_e and S_c.

Article level clicks show that articles were only 3.56% of all the displayed elements. This may be misleading as the retrieval system did not return article elements in the result list. We therefore compared article usage to elements that were displayed when users were already in the document view, i.e. we excluded elements that were shown right after a searcher clicked on a link in the result list. The updated number shows that article elements were displayed in 6.12% of these clicks. This suggests that searchers of an XML retrieval system do use the structure available in terms of the ToC, and although it was the first time they had used an XML retrieval system, they did not simply prefer to see the whole document as they were accustomed to.

Our results from the previous sections suggest that searchers still want to have access to, and use, the full-article level. For example, searchers read summaries of articles and read them for longer but, they did not necessarily want to use the full-articles directly, i.e. looking at the full-article summary may be enough to decide whether reading any part of the article is worthwhile.

5 Conclusions and Future Work

In this paper, we presented a study of an interactive XML retrieval system that uses summarisation to help users in navigation within XML documents. The system takes advantage of the logical structure of an XML document by presenting the structure in the form of a table of contents (ToC). In this ToC, summaries of corresponding XML elements were generated and displayed.

Searchers in our study did indeed use the provided structure actively and did not only use the whole article in order to identify relevant content. In addition, searchers made good use of the XML element summaries, by spending a significant amount of time reading these summaries. This implies that our system, by the use of summarisation, facilitated browsing in the ToC level more than that at INEX 2004 interactive track [13].

Regarding the use of element summaries, in our study searchers tended to read more summaries that were associated with elements at lower levels in the structure (e.g. summaries of paragraphs), and at the same time summaries of lower elements were read for a shorter period of time. Our results also suggest that if more summaries are made available, searchers tend to read more summaries in a search session, but for shorter time.

In our experiment, the ToC display and summary presentation were highly connected (i.e. no summary can be displayed without a corresponding item in the ToC). Based on the close relation between them, for such an XML retrieval system it is important to find the appropriate ToC, and summary, presentation. If the ToC is too deep, searchers may lose focus as the reading of many summaries and short reading times at low levels indicated. Nevertheless, if the ToC is not detailed enough, users may lose possibly good links to relevant elements. Our results suggest, that for the used collection, a one or two-level ToC (containing reference to the whole article, body, front and back matter) would be probably too shallow, while displaying the full fourth level (normally to paragraph-level) is sometimes too deep.

We view our results as having implications for the design of interactive XML IR systems that support searchers by providing element summaries and structural information. One implication of the results is that the display of the ToC in XML IR systems needs to be carefully chosen (see previous paragraph). Our results also showed that summarisation can be effectively used in XML retrieval. A further implication of the results is that XML retrieval systems should consider the logical structure of the document for summary generation, as searchers do not use summaries in the same way at all levels of the structure.

Based on the outcomes of this study, our future work includes the development of an improved interactive XML retrieval system that includes adaptive generation of summaries at the ToC level, where the structural position and estimated relevance of the element to be summarised will also be considered (some initial work is done in [2]). We also plan to take into account structural IR search task types (e.g. fetch and browse), that are currently being investigated at INEX [6]. The aim of the fetch and browse retrieval strategy is to first identify relevant documents (the fetching phase), and then to identify the most relevant elements within the fetched articles (the browsing phase). We believe that summarisation can be particularly helpful in the browsing phase, where finding relevant elements within a document is required.

6 Acknowledgments

This work was partly funded by the DELOS Network of Excellence in Digital Libraries, to which we are very grateful.

References

1. A. Alam, A. Kumar, M. Nakamura, A. F. Rahman, Y. Tarnikova, and C. Wilcox. Structured and unstructured document summarization: Design of a commercial

summarizer using lexical chains. In *ICDAR*, pages 1147–1152. IEEE Computer Society, 2003.
2. M. Amini, A. Tombros, N. Usunier, M. Lalmas, and P. Gallinari. Learning to summarise XML documents by combining content and structure features (poster). In *CIKM'05*, Bremen, Germany, October 2005.
3. P. Borlund. The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research*, 8(3), 2003.
4. S. Dumais, E. Cutrell, and H. Chen. Optimizing search by showing results in context. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 277–284, New York, NY, USA, 2001. ACM Press.
5. H. P. Edmundson. New methods in automatic extracting. *J. ACM*, 16(2):264–285, 1969.
6. N. Fuhr, M. Lalmas, S. Malik, and Z. Szlávik, editors. *Advances in XML Information Retrieval: Third International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2004*, volume 3493. Springer-Verlag GmbH, may 2005.
7. J. Goldstein, M. Kantrowitz, V. Mittal, and J. Carbonell. Summarizing text documents: sentence selection and evaluation metrics. In *SIGIR'99*, pages 121–128. ACM Press, 1999.
8. H. Kim and H. Son. Interactive searching behavior with structured xml documents. In Fuhr et al. [6], pages 424–436.
9. J. Kupiec, J. Pedersen, and F. Chen. A trainable document summarizer. In *SIGIR'95*, pages 68–73. ACM Press, 1995.
10. N. Pharo and R. Nordlie. Context matters: An analysis of assessments of XML documents. In Fabio Crestani and Ian Ruthven, editors, *CoLIS*, volume 3507 of *Lecture Notes in Computer Science*, pages 238–248. Springer, 2005.
11. T. Rölleke, R. Lübeck, and G. Kazai. The hyspirit retrieval platform. In *SIGIR'01*, page 454, New York, NY, USA, 2001. ACM Press.
12. Z. Szlávik, A. Tombros, and M. Lalmas. Investigating the use of summarisation for interactive XML retrieval. In *Proceedings of the 21st ACM Symposium on Applied Computing, Information Access and Retrieval Track (SAC-IARS'06), to appear*, pages 1068–1072, 2006.
13. A. Tombros, B. Larsen, and S. Malik. The interactive track at INEX 2004. In Fuhr et al. [6].
14. A. Tombros, S. Malik, and B. Larsen. Report on the INEX 2004 interactive track. *ACM SIGIR Forum*, 39(1):43–49, June 2005.
15. A. Tombros and M. Sanderson. Advantages of query biased summaries in information retrieval. In *SIGIR'98*, pages 2–10. ACM Press, 1998.
16. R. W. White, J. M. Jose, and I. Ruthven. A task-oriented study on the influencing effects of query-biased summarisation in web searching. *Inf. Process. Manage.*, 39(5):707–733, 2003.