

# The effect of granularity and order in XML element retrieval

Nils Pharo, Oslo University College, Pb 4 St. Olavs plass, N-0130 Oslo, Norway

**Abstract:** The article presents an analysis of the effect of granularity and order in an XML encoded collection of full text journal articles. 218 sessions of searchers performing simulated work tasks in the collection have been analysed. The results show that searchers prefer to use smaller sections of the article as their source of information. In interaction sessions during which articles are assessed, however, they are to a large degree evaluated as more important than the articles' sections and subsections.

**Keywords:** interactive information retrieval; XML; user studies

**Vitae:** Nils Pharo is an associate professor at the Oslo University College, his main research interests are interactive information retrieval and information behaviour studies.

Published as: Pharo, N. (2008). The effect of granularity and order in XML element retrieval. *Information Processing & Management*, 44 (5), 1732-1740.

## 1. *Introduction*

XML-coding of semi-structured documents facilitates the retrieval of document elements as an alternative to retrieving full documents (Luk et al, 2002). There are many arguments for partial document retrieval, e.g., it makes it possible to filter out only those parts of the documents that specifically treat the issues related to searchers' queries. In addition to explicating the logical structure of documents XML also makes it possible to add extra semantics to elements thus combining the markup and content for more precise retrieval. Little, however, is known about how searchers react to such features from the XML retrieval systems. For example, we do not know much about the effect on searchers of presenting them with document parts in addition to full documents.

This article presents an analysis of the granularity of documents and the order in which searchers prefer to examine parts of documents and whole document. The study is based on data collected in the INEX 2005 interactive track, of which there were three different sub tracks. The current study is based on Track A data, which was compulsory for all track participants. It was performed in a data set consisting of articles in computer science journals from the IEEE computer society using the Daffodil<sup>1</sup> retrieval system.

We have investigated two research questions in this article:

1. What level of document granularity do searchers judge to be the most relevant?
2. In what order do searchers interact with and judge the relevance of elements of different granularity?

The article is built up in the following way: in the next section we address previous research in XML retrieval, in particular the work being done as part of the INEX conferences, and relate it to our approach in the next section. Section 3 contains a description of our method. In Section 4 we present our findings and Section 5 contains the discussion and our conclusions.

## 2. *Background*

The Initiative for the Evaluation of XML retrieval (INEX) started in 2002 in order to provide “an infrastructure to evaluate the effectiveness of content-oriented XML retrieval systems” (Kazai, Lalmas, Fuhr & Gövert, 2004). INEX builds its experimental design on the TREC model, with a test collection which consists of topics/tasks (submitted by the participating groups), documents (approximately 12 000 articles from a selection of IEEE Computer society's journals<sup>2</sup>) and relevance assessments provided by the participants, thus making it possible to compute the retrieval effectiveness of different matching algorithms. Since its beginning several tracks have been introduced to the initiative in order to explore topics such

---

<sup>1</sup> <http://www.is.inf.uni-due.de/projects/daffodil/>

<sup>2</sup> From 2006 the collection consists of approximately 600 000 Wikipedia articles.

as relevance feedback, heterogeneous collections, natural queries, document mining, multimedia and interactive information retrieval (Inex 2005).

The interactive track was first introduced in 2004 (Tombros, Larsen & Malik, 2005) and aims at studying how end-users interact with XML-encoded data. The IEEE collection was used for the 2004 and 2005 (Larsen, Malik & Tombros, 2006) main tracks and has resulted in some new knowledge. In 2005 an additional data set was added to the track from the Lonely Planet travelling guides.

The data from the INEX 2004 and 2005 interactive track have been used to evaluate and redesign the user interface (Malik, Klas, Fuhr, Larsen & Tombros, 2006). The questions related to system design have been analysed showing that searchers in general were positive to both the 2004 and 2005 systems, but that the 2005 system had better “learnability”.

Pharo and Nordlie used the INEX 2004 (Pharo & Nordlie, 2005) IEEE data to investigate whether users, when presented with the opportunity to interact with relevant parts of documents instead of or in addition to the full document, prefer to work with the elements in or out of context with the full text. Their findings are not conclusive, but they indicate that giving users access to the most relevant text elements on lower levels of granularity is valuable, but only if the full article is present to provide context.

Hammer-Aebi, Christensen, Lund and Larsen (2006) used the INEX Lonely Planet dataset in a similar study. They found that users prefer elements of depth 2-4 rather than full documents. Their study also involved comparing systems with and without context, which, surprisingly, showed not much preference of searchers for the context providing system.

Kim and Son (2006) performed a small scale study to compare the presentation format of the search engines used at the 2005 and 2004 INEX interactive tracks. Their findings indicate that the searchers has good use of the embedded table of content in the 2005 interface.

The interactive track organisers (Larsen, Tombros & Malik, 2006; Malik, Larsen & Tombros, 2007) have performed initial studies of the 2005 interactive search logs which suggest that users seem to prefer document elements rather than whole documents and that elements of short length are often assessed as not relevant. They also suggest that searchers prefer to use document metadata as their initial entry point.

Ramírez and de Vries (2006) have analysed the effect of three different contextual factors on users interacting with the INEX 05 IEEE data. They found that users with higher factual knowledge about the topic were better able to perform compound tasks, and also that such users made more use of larger elements, such as whole articles, than users performing simpler tasks.

Our study is related to the study reported by Larsen, Tombros and Malik (2006), which it to a certain extent overlaps and also to the study by Hammer-Aebi, Christensen, Lund and Larsen (2006) who use a different data set, the Lonely Planet collection. Our study is, however, more specific in identifying the context of searching, i.e. how the order of interaction with elements of different granularity affects the relevance judgements of said elements.

### 3. Method

The data used in our analysis stems from the logs collected as part of the INEX 2005 interactive track, in this section we will present the data collection procedures used in the experiment. There were eleven institutions from around the world participating in the interactive track, each group was instructed to collect data from minimum six end users (hereafter called searchers). In our data we have analysed 218 sessions performed by 73 different searchers.

The experiments were conducted using the Daffodil IR system with an interface developed specifically for the interactive track.

The IR system is designed with a quite simple search interface where searchers can enter their query terms. In the result list relevant documents and elements are presented hierarchically with the relevant elements presented within the documents they belong to (see Figure 1 captured from Larsen, Malik & Tombros, 2006). The documents are presented at four levels; *sections*, *subsections* and *sub-subsections* which can be accessed directly and a link within the document's title which appears to be to the whole document, but which actually leads to the document's *metadata*.

[place Figure 1 here]

Having selected an element from the result list searchers are presented with an interface containing the requested element and the ability to browse the full document upwards and downwards (see Figure 2, also captured from Larsen, Malik & Tombros, 2006). Also the interface contained the relevance judgement tool. Relevance are judged on a three-point scale:

- 0. Not relevant
- 1. Partially relevant
- 2. Relevant

[place Figure 2 here]

The relevance scale had been simplified from the scale used in the previous year in order to make relevance judgements less difficult for searchers (Pechevski, Thom & Vercoustre, 2005; Pharo & Nordlie, 2005; Larsen, Tombros & Malik, 2005a).

Each searcher were asked to perform three search tasks. There were two categories of tasks, challenging and general, from which the searchers should select one of each (see Table 1). The third task was formulated by the individual searcher.

The search tasks were formulated as simulated work task situations using Borlund's (2000) method, which means that in addition to presenting the topic searchers were also given instructions on the context in which the information sought for was to be used. Searchers were given a maximum of 20 minutes to perform each task. In order to neutralise learning effects the order of task performance was permuted. The searchers were asked to judge the relevance of every element they looked at. There was, however, no system enforcement implemented to prevent searchers from not judging the documents/elements.

**Table 1 Simulated work tasks at Inex 2005**

The searchers also filled out questionnaires before and after each task as well as a pre- and post *session* questionnaire. The main data used stems from the sessions' transaction logs which were stored on a central server. The logs contain information on the events in the sessions, including both input from searchers and the system's response (ref Table 2).

**Table 2 Excerpt from transaction log**

The SearcherID identifies the unique searcher, stating the affiliated institution, a user number and the current task. The excerpt shows searcher number 1 from Oslo University College, performing his or her challenging task (c). The event explains the actions performed, “detailquery” means that the searcher has clicked a new document/element, “detailbrowsing” indicates that the searcher views/browses the element and “relevance-feedback” is logged when the searcher enters his/her relevance assessment. Each event is time stamped and the article and its elements are identified, the article elements by an XPath expression. In this example we learn that the searcher has selected first the fm[1] element, which stands for “front matter” and contains article metadata. The element “/article[1]/bdy[1]/sec[3]” is chapter number three in the article’s body. The relevance column states the searcher’s assessment, in this case the specific section is judged as “partially relevant”.

## **4. Findings**

### **4.1 Preferred level of granularity**

In order to identify the levels of document granularity favoured by searchers, i.e. our first research question, we performed an analysis of the transaction logs. We identified occurrences of relevance judgements in the logs and counted the number of elements at different granularity levels and their scores. The levels of granularity (from low to high) are represented as follows:

1. article = top level of document in which all other elements are included
2. fm = front matter, the metadata of the article
3. sec = section, the second level of granularity
4. ss1 = subsection level 1
5. ss2 = subsection level 2
6. bm = back matters, appendices

The results are presented in Table 3. The results show that sections are by far the most often relevance assessed document element, representing 50 percent of all assessments. Interestingly, only 7 percent of the assessments are on the article level. In all, the searchers interacted with 1471 documents, including cases where searchers returned to revisit a document. Thus only 13.5 percent of the documents searchers inspected were assessed on article level.

**Table 3 Relevance assessments by elements**

From our findings, which replicate those of Larsen, Tombros and Malik (2006; Malik, Larsen & Tombros, 2007) the conclusion that searchers prefer document elements to full documents

is easy to draw. There are, however, indications in the transaction logs that conclusions must be regarded with a certain amount of scepticism. A feature of the Daffodil system's interface was to lead the searcher directly to the fm-element, rather than the full article, when he/she clicked on the article's title in the result list. We assume that this caused many searchers to access the fm-element involuntarily. As we can see, the assessment of fm-elements represents an exception to the general trend, which is probably due to the content of the element. Since none of the tasks explicitly call for bibliographical information, and many fm-elements do not contain abstracts, the element is very seldom useful for the searchers. We can thus speculate that more searchers might have assessed the full documents if they had been directly accessible via the result list, but it is impossible to say whether these documents then would have been assessed as relevant or partially relevant. We agree with Larsen, Tombros and Malik (2006, p. 664) that searchers either thought "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". A third explanation could be, however, that searchers sometimes use search strategies that are seemingly irrational (Pharo & Järvelin, 2006). Such strategies can, e.g., be grounded in the searcher's erroneous model of the information system, but it may also be that searchers prefer general heuristic search strategies which they find more efficient than developing sophisticated strategies for each specific system.

Of course, the distribution of assessed elements depends on the total distribution of elements in the collection. The basic assumption of XML retrieval, that it facilitates access to the appropriate level of granularity, will be interesting to explore further in other data sets, such as the Wikipedia collection.

In Table 4 we look at the relative distribution of assessments across elements of different granularity. We can see that there is an inverse relationship between the relevance assessments and degree of granularity, i.e., the lower the granularity the higher the proportion of elements judged "relevant". This can be related to the results of the ad hoc track (e.g., Kamps, Marx, de Rijke & Sigbjörnsson, 2003) where full articles are recognized as having a higher prior probability of relevance. Note, however, that the distribution of partial relevance assessments does not differ much across elements.

**Table 4 Relevance assessments by elements, percent**

Our analysis shows that sections are the most frequently assessed document parts, and that although articles are relatively seldom assessed, they tend to be more relevant for those assessing them. Elements on a very high level of granularity, on the other hand, are often judged not relevant. This may be due to several factors such as the lack of context or the order in which elements are explored, we shall discuss this below. Our findings are quite similar to the Lonely Planet study (Hammer-Aebi, Christensen, Lund & Larsen, 2006), which also points to the fact that searchers prefer elements on medium-level granularity such as sections and subsections. In the latter study, however, a different relevance scale was used which also indicated to what extent an element needed the context of the surrounding elements to be understood. A significant difference between our studies is that documents and elements are more often assessed as partially relevant in our study whereas in the Lonely Planet study many documents/elements are considered too broad or too narrow and quite few (16 %) are considered partially relevant. On the other hand, documents and elements assessed as partially

relevant in the IEEE system *might* well have been assessed as such because the searchers felt they were too broad or narrow.

## 4.2 Interaction order

Our second research question deals with the order of searchers' interaction with and relevance assessments of elements. We will look into the following sub questions: a) what elements are inspected first?, b) when are the full articles inspected?, and c) in what order are elements relevance assessed?

### 4.2.1 Order of inspected elements

**Table 5 Distribution of first visits**

Table 5 presents the distribution of first inspections of article elements. We see that the metadata element (fm) is by far the element users choose first. Another significant finding is the missing appearance of the article element. As stated above, fm-elements rather than the full articles were retrieved when selecting the article's title from the result list. Thus we must expect that the searcher's intention in these cases was to retrieve the full article. Also this element was present for each individual article in the result list, which was not the case for elements of high granularity. Sections of articles were, on the other hand, often presented in the result list, and from Table 3 we see that they also represent the elements most often assessed. Unfortunately the logs do not contain information on the number of elements of different granularity in the result list which means we do not know whether the distribution of selected elements reflects the distribution of available elements.

### 4.2.2 Inspection of articles

We were interested in learning when in the process the searchers inspected the full article compared to other elements. This means that we narrowed down our analysis to a sub set of the transaction logs. We picked out from the logs those occurrences where interaction with the full document was identifiable. In other words, we analysed the user-document interactions during which the searcher chose to click on the full article link available in the document's table of contents. In our analysis we have thus used 263 such interaction processes stemming from 112 different search tasks (sessions) performed by 60 different searchers. This means that full articles were inspected at least once in every other search session, and by a large majority (82 percent) of all searchers.

We started by dividing the order of article inspection into three categories; the beginning, the middle, and the end of the interaction. Since the searchers were prevented from directly accessing the full article from the result list we defined the case when searchers first retrieved the fm-element and immediately thereafter the full article as the beginning of the interaction. When elements of another level of granularity are inspected before and after the article-element we have defined this as a case of the middle of the interaction. When the article is inspected as the final element it is defined as the end of the interaction. Table 6 contains the distribution of interaction order.

**Table 6 Distribution of article inspection order**

In 41 cases the article-element was inspected more than once, in addition to the 27 explicit occurrences in the right hand side of the table there are also some cases when the article elements were inspected more than once in the middle of the process.

We see that in half of the interactions the article element was inspected as the final element and in 42.6 percent of the cases this was the only inspection of the full article. This indicates that searchers want to look at the document parts before viewing the whole unit. In other words, that many searchers prefer to view the more specific parts of the document before they look at it in context. Above we have assumed that a large number of searchers have tried to access the article from the result list, therefore it is reasonable to believe that the beginning-category is somewhat underrepresented. From our data we can, however, claim that relatively seldom do searchers inspect the full article in the middle of the interaction period. In the middle of a session searchers prefer to inspect high granularity elements. In general, it is not surprising that searchers tend to choose either a “top-down”-strategy, from metadata/article to relevant document parts or a “bottom-up”-strategy where they first inspect sub-elements and then the full article.

### 4.2.3 Order of element assessments

The order in which searchers assess the individual elements indicates their shifts in focus during interaction. Also it provides us an opportunity to see what elements are considered more important and how, if any, learning effects influence the assessment of subsequent elements.

**Table 7 Relevance assessments of article-elements and their sub-elements**

We first present the relationship between the assessments of the article element and elements of higher granularity within the same document. There were, in all, 198 interaction sessions in which the article element was assessed. As already noted this represents only 7 percent of all assessments and 13.5 percent of all documents interacted with by the searchers. Nevertheless, they represent items in which the impact of granularity is particularly visible, more so than sections which can be rather small.

From Table 7 we see that in quite many cases no other elements in the document were assessed. In one third of the cases the article and its sub-elements were given identical assessments, e.g. both an article and one or more sections or subsections of the same article were assessed as partially relevant. In 53 cases, however, there was a discrepancy between assessment of the full article and one or more of its parts.

We shall inspect the case of similar assessments in more detail. If we use the article-element as our point of departure we can see *when* in the searcher-document interaction process the assessments were made. Our analysis found that in a large majority of cases the other elements were assessed prior to the assessment of the whole article, in only nine out of the 69 documents the user judges the sub-elements to be equally relevant after having assessed the full article, in 56 of the remaining cases (i.e. 81 percent of all cases) all sub-elements were assessed *before* the article assessment, for four documents sub-elements were assessed both before and after the article element assessment.

This indicates that the assessment value of article is influenced by the value set in preceding element assessments. We need, however, to contrast this with the cases where assessment values differ.

In the 69 documents a total of 139 relevance assessments of articles' sub-elements were performed, of these 109 were made before and only 30 after the article assessment. It is reasonable to believe that an increase in the number of element assessment per document increases the probability of discrepancy between article and sub-element assessment values. In this case we see that there on average 2 element assessments per article, in other words we have an assessment ratio of 2 to 1.

In the 53 documents where searchers had valued sub-elements different from the full article a total number of 173 sub-element assessments have been recorded, i.e. an assessment ratio of 3,26 to 1.

Table 8 shows the distribution of element assessments before and after article assessment and how the assessment values differ. In the table "higher"/"lower" is used to denote that the element has been assessed as more/less relevant than the article. If, e.g., a section has been assessed as relevant prior to the article being assessed as partially relevant, it will be recorded in the "higher"-column within the "before"-column.

Be aware that 46 percent of all the assessed article elements are perceived as relevant, and hence document parts cannot have a higher relevance assessment. This explains that relatively few sub-elements are assessed as more relevant than the full article.

#### **Table 8 Distribution of assessment values and time of assessment related to article assessment**

We can see from Table 8 that there are eighteen instances of a section being assessed as less relevant than the article it belongs to *before* the searcher has assessed the article itself. Similarly we see that sixteen sections are judged to be less relevant *after* the article is assessed.

In general searchers assess more sub-elements before they judge the full article than afterwards. This should come as no surprise since searchers probably tend to understand the task as being fulfilled having judged the full article.

For most of the sub-elements the distribution does not seem to differ significantly. The exception is the subsection 1 (ss1) element. We have found that almost all ss1 elements that are assessed after the article element have been judged to be of less relevance than the article. On the other hand, when ss1 elements are assessed before the article they are often given the same value as the full document. This supports the above findings that sub-elements that are assessed equally relevant to the article tend to be assessed before the article.

With respect to the order of element inspection and assessment our findings indicate that in those interaction processes when searchers actively retrieve the full article for assessment they prefer to look at the article early in the process. Thereafter they inspect and assess individual elements before they return to take a new look the article and assess its relevance.

## 5. *Discussion and conclusion*

Our findings show that, in general, full articles are seldom accessed nor assessed during the search processes. Searchers prefer to use smaller sections of the article as their source of information. Information system designers should thus take into account the preference of searchers to interact with sub-elements rather than full documents, if possible.

The full articles must, however, be directly accessible from the systems, since many searchers apparently try to access them from the results list. Also our study shows that *when* articles are evaluated they are to a large degree judged as more important than the articles' sections and subsections.

We have performed our analysis in the journal article domain and hence it should have consequences for the design of systems for e-journals, in particular this will be of significance when the e-journal providers' systems index the articles in full text. But also system containing other content that can be deconstructed into individual content parts could benefit from providing direct access to the sub elements. One domain could be institutional archives of full text research publications, such as theses and dissertations. These document types typically are structured in ways that make a difference to end users. For example, information searchers may prefer to get direct access to the (sub) chapters containing methodical considerations, literature reviews or results.

We also find that the searchers prefer rather large document parts, such as sections rather than sub-sections. This relates to a discussion in the INEX environment of whether passage retrieval is more suitable than element retrieval (Kaamps & Koolen, 2007). What is the optimal size of retrievable units in information retrieval? Probably it is dependent on the domain, in social sciences and humanities more text is probably needed to produce a statement compared to the "discourse" of the natural sciences, where meaningful content may be compressed to a single formula. Also texts from scientific disciplines with rigorous constraint on the document structure are probably better candidates for structural document mark-up.

Our second research question dealt with the order in which searchers interacted with elements of different granularity. Our data shows that when searchers interact with the full articles they tend to do so after they have looked at article sub-elements. Thus it might be an option in IR systems to design user tools that can be tuned to facilitate such preferences.

Although our findings support one of the core presuppositions of XML information retrieval, that searchers prefer element retrieval it must be followed up by more research to control the effect of the system interface.

The tasks performed by searchers in INEX experiment are divided into challenging and general tasks, thus further analysis may reveal different strategies prepared for different tasks. Both task types used in the current INEX experiments can be considered quite demanding on searchers, compared to fact finding queries. Further experiments should be performed where different types of tasks and queries are emphasised.

Further studies should use the INEX Wikipedia collection to test whether the pattern of preferences we have observed is general, or if our findings are genre dependent. The Wikipedia collection differs much from the IEEE collection. The documents are smaller, of

another genre, have other kinds of authors and are probably used for solving other kinds of information needs. This will be the subject of future research.

## **6. Acknowledgements**

INEX is partly funded by DELOS, a network of excellence in digital libraries. We would like to thank the participants of the INEX 2005 interactive track. We want to acknowledge the invaluable effort invested by Kristin Strømme in this project and also the comments from Ragnar Nordlie.

Finally we would like to thank the reviewers for good suggestions that have helped make this paper better.

## **7. References**

Borlund, P. (2000). Evaluation of interactive information retrieval systems. Åbo: Åbo Akademi University Press.

Hammer-Aebi, B., Christensen, K. W., Lund, H. & Larsen, B. (2006). Users, structured documents and overlap : interactive searching of elements and the influence of context on search behaviour. In Ruthven, I. et al. (Eds.) Proceedings of the 1st Symposium on Information Interaction in Context (IliX 2006) (80-94). Copenhagen: Royal School of Library and Information Science.

Kamps, J. & Koolen, M. (2007). On the relation between relevant passages and XML document structure. In SIGIR 2007 Workshop on Focused Retrieval (28-32). Dunedin: University of Otago.

Kamps, J., Marx, M., de Rijke, M. & Sigurbjörnsson, B. (2003). XML retrieval: what to retrieve? In Proceedings of SIGIR 2003 (409-410). New York: ACM Press (poster).

Kazai, G., Lalmas, M., Fuhr, N., & Gövert, N. (2004). A report on the first year of the initiative for the evaluation of XML retrieval (INEX 2002). Journal of the American society for information science and technology, 55(6), 551-556.

Kim, H. & Son, H. (2006). Users interaction with the hierarchically structured presentation in XML document retrieval. Lecture notes in computer science, 3977, 422-431.

Larsen, B., Malik, S. & Tombros, A. (2006). The Interactive track at INEX 2005. Lecture notes in computer science, 3977, 398-410.

Larsen, B., Tombros, A. & Malik, S. (2005). Obtrusiveness and relevance assessment in interactive XML IR experiments. In Trotman, A., Lalmas, M. and Fuhr, N. eds. Proceedings of the INEX 2005 Workshop on Element Retrieval Methodology (39-42). Dunedin: University of Otago.

Larsen, B., Tombros, A. & Malik, S. (2006). Is XML retrieval meaningful to users? Searcher preferences for full documents vs. Elements. In Proceedings of SIGIR 2006 (663-664). New York: ACM Press (poster).

- Luk, R.W.P., Leong, H.V., Dillon, T.S., Chan, A.T.S., Croft, W.B. & Allan, J. (2002). A Survey in indexing and searching XML documents. *Journal of the American Society for Information Science and Technology*, 53(6), 415-437
- Malik, S., Klas, C.P., Fuhr, N., Larsen, B & Tombros, A. (2006). Designing a user interface for interactive retrieval of structured documents :. lessons learned from the INEX interactive track. *Lecture Notes in Computer Science*, 4172, 291-302.
- Malik, S. Larsen, B & Tombros, A (2007). Report on the INEX 2005 interactive track. *SIGIR Forum*, 41(1), 66-73.
- Pehcevski, J., Thom, J.A., & Vercoustre, A.-M. (2005). Users and assessors in the context of INEX: are relevance dimensions relevant? In Trotman, A., Lalmas, M. and Fuhr, N. eds. *Proceedings of the INEX 2005 Workshop on Element Retrieval Methodology* (47-62). Dunedin: University of Otago.
- Pharo, N. & Järvelin, K. (2006). 'Irrational' searchers and IR-rational researchers. *Journal of the American Society for Information Science and Technology*, 57(2), 222-232.
- Pharo, N. & Nordlie, R. (2005). Context Matters: An Analysis of assessments of XML Documents. *Lecture Notes in Computer Science*, 3507, 238-248.
- Ramirez, G. & de Vries, A.P. (2006). Relevant Contextual Features in XML Retrieval . In Ruthven, I. et al. (Eds.) *Proceedings of the 1st Symposium on Information Interaction in Context (IiX 2006)*. Copenhagen: Royal School of Library and Information Science.
- Tombros, A., Larsen, B. & Malik, S. (2005). The Interactive track at INEX 2004. *Lecture notes in computer science*, 3493, 410-423.

## Figures

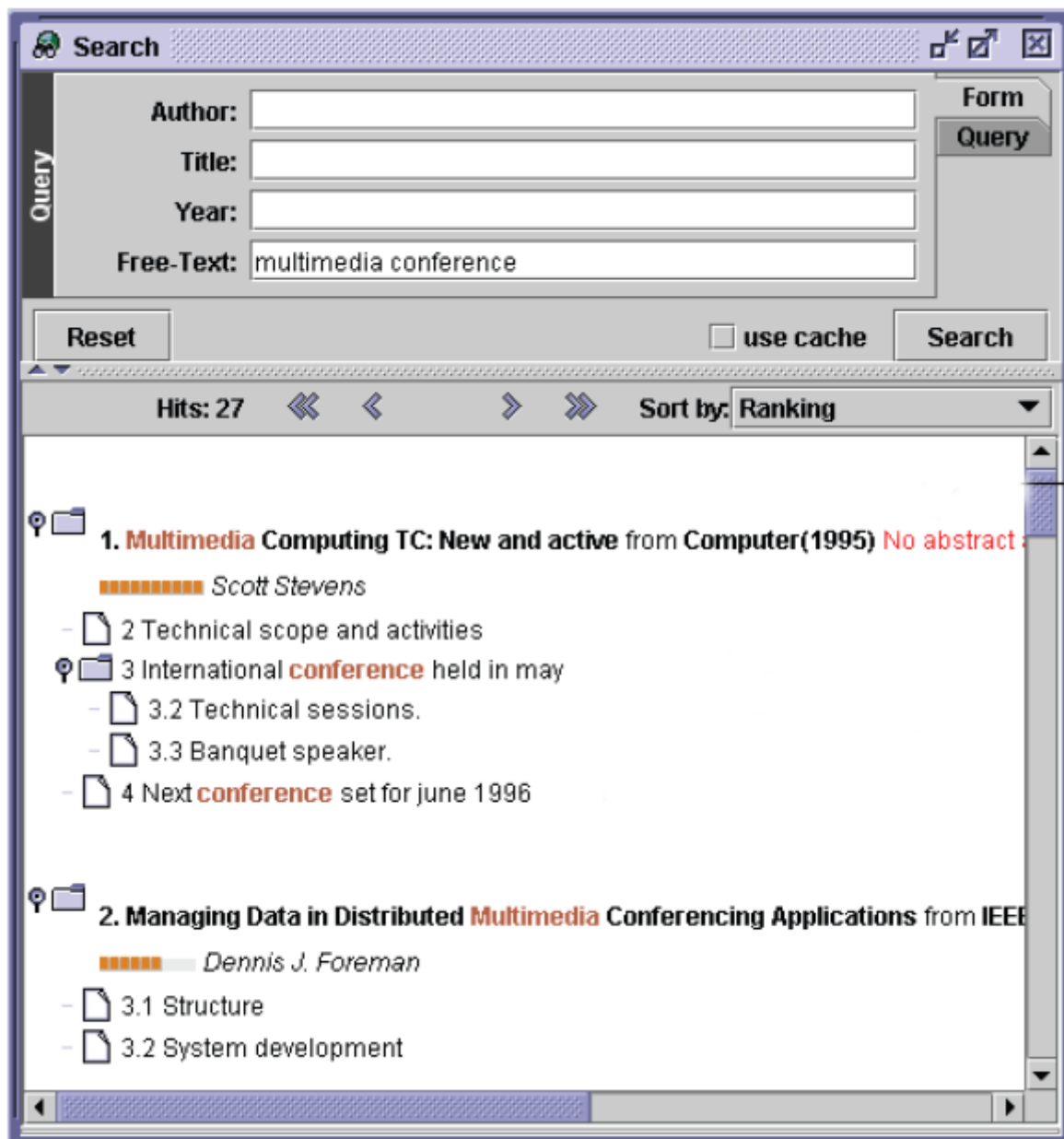


Figure 1. The search interface with result list

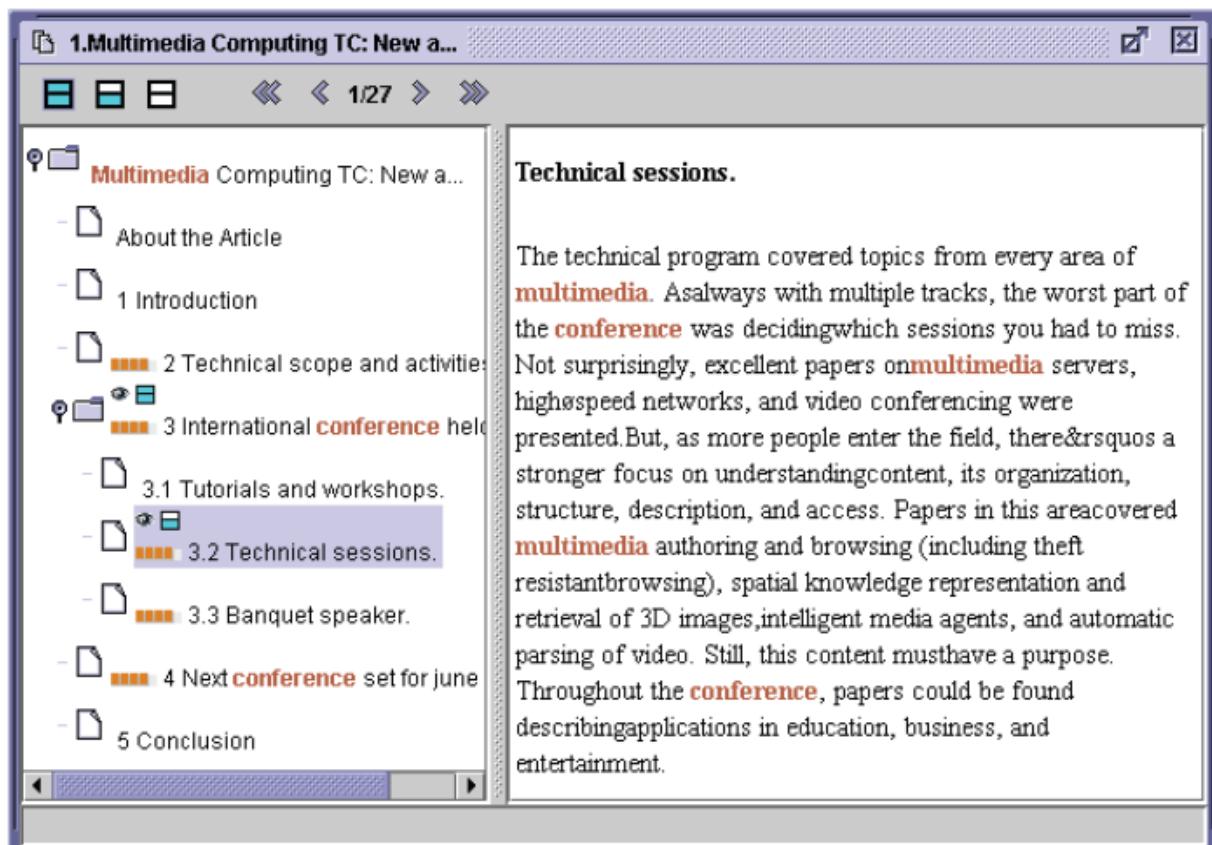


Figure 2. The document view

Tables  
Table 1

|  |  |  |
|--|--|--|
| <p><b>Task ID: C1:</b> One of your friends has recently bought a small handheld Global Positioning System (GPS) unit, and the possibilities offered by this technology have caught your interest. You would like to explore new killer applications for mobile devices. Therefore, you are looking for examples and descriptions of applications that use GPS, for devices such as mobile phones, PDAs (Personal Desktop Assistants) and other wireless and mobile devices.</p> <p>Find, for instance, information that discusses examples of how applications that use GPS can be used to accomplish new tasks or provide new services.</p> | <p><b>Task ID: C2:</b> In your daily work you sign on to a range of different systems both locally and remotely. On many of them you have different user IDs and different passwords, and you find it annoying to have to verify your identity again and again. In addition, you find it demanding to maintain all these IDs and passwords and to keep them secure.</p> <p>You have heard about LDAP (Lightweight Directory Access Protocol) and other single sign-on procedures, and wish to learn more about them to assess the potentials for creating a single sign-on procedure for your local network (with both Unix, Linux, PC and Mac platforms).</p> <p>Find, for instance, information that discusses single sign-on procedures, or state of the art user-authentication methods.</p> | <p><b>Task ID: C3:</b> Data security and authenticity is an important issue at your work place. One approach to ensure data authenticity is the so-called “steganography” where data is embedded in various media files like images, sound files, video files and so on. A commonly used data embedding technique is Watermarking where data can be effectively hidden in a file without the changes being visible to the common person. You want to learn more about Watermarking as a technique for data embedding that will enable you to verify the authenticity of a file.</p> <p>Find, for instance, information that discusses the use of Watermarking techniques to hide information that will allow later validation of a files authenticity.</p> |
| <p><b>Task ID: G1:</b> New anti-terrorism laws allow intelligence agencies like the FBI (Federal Bureau of Investigation) and CIA (Central Intelligence Agency) to monitor computer communications to spot suspected criminals and terrorists. You would like to find information about how this affects your own and other people’s privacy and to know what concerns have been raised.</p> <p>Find, for instance, information that discusses the Carnivore or Echelon projects or other similar surveillance of computer communication.</p>  | <p><b>Task ID: G2:</b> Your department has produced a Linux-program and it is being discussed whether to release it under a public license such as GNU or GPL (General Public License). Therefore, you have been asked to find information about the implications of releasing the code under a public license as an open source program.</p> <p>Find, for instance, information that discusses different licensing schemes or articles about the impact of open source programs.</p>  | <p><b>Task ID: G3:</b> Video games are being played by an ever increasing number of people of all ages, and the game industry is becoming a major economic player. You would therefore like to find non-technical information about how video games have affected people’s lives as well as how the games have changed the entertainment industry.</p> <p>Find, for instance, information discussing the concerns that playing video games may lead to a rise in violent behaviour, or information about the effect of video games on the film industry.</p>   |

Table 2

| SearcherID   | Time         | Event              | ArticleID         | Article element           | Relevance |
|--------------|--------------|--------------------|-------------------|---------------------------|-----------|
| uoslo-a001-c | 15:47:38:321 | detailquery        | mu/1999/u4088.xml | /article[1]/fm[1]         |           |
| uoslo-a001-c | 15:47:40:561 | detailbrowsing     | mu/1999/u4088.xml | /article[1]/fm[1]         |           |
| uoslo-a001-c | 15:47:47:720 | detailbrowsing     | mu/1999/u4088.xml | /article[1]/bdy[1]/sec[3] |           |
| uoslo-a001-c | 15:48:08:840 | relevance-feedback | mu/1999/u4088.xml | /article[1]/bdy[1]/sec[3] | 1         |

Table 3

| Relevance level           | Article      | fm            | sec            | ss1           | ss2         | bm          | Total          |
|---------------------------|--------------|---------------|----------------|---------------|-------------|-------------|----------------|
| relevant 2                | 91           | 70            | 551            | 229           | 19          | 17          | 977            |
| partially relevant 1      | 56           | 95            | 501            | 221           | 27          | 33          | 933            |
| not relevant 0            | 51           | 234           | 401            | 206           | 32          | 45          | 969            |
| Total<br>(share of total) | 198<br>(7 %) | 399<br>(14 %) | 1453<br>(50 %) | 656<br>(23 %) | 78<br>(3 %) | 95<br>(3 %) | 2879<br>(100%) |

Table 4

|                      | Article | fm   | sec  | ss1  | ss2  | bm   |
|----------------------|---------|------|------|------|------|------|
| relevant 2           | 46.0    | 17.5 | 37.9 | 34.9 | 24.4 | 17.9 |
| partially relevant 1 | 28.3    | 23.8 | 34.5 | 33.7 | 34.6 | 34.7 |
| not relevant 0       | 25.8    | 58.7 | 27.6 | 31.4 | 41.0 | 47.4 |
| %                    | 100.1   | 100  | 100  | 100  | 100  | 100  |

Table 5

| fm     | sec    | ss1    | ss2   | /app[1]/sec[] | /app[1]/sec[]/ss1[] |
|--------|--------|--------|-------|---------------|---------------------|
| 71.3 % | 14.8 % | 11.5 % | 0.7 % | 1.5 %         | 0.2 %               |

Table 6

| Beginning     | Middle        | End            | Beginning & middle | Beginning & end | Middle & end | Total         |
|---------------|---------------|----------------|--------------------|-----------------|--------------|---------------|
| 72<br>(27.3%) | 52<br>(19.8%) | 112<br>(42.6%) | 5<br>(1.9%)        | 11<br>(4.2%)    | 11<br>(4.2%) | 263<br>(100%) |

Table 7

| Only article assessed | Similar assessment | Different assessment | Total      |
|-----------------------|--------------------|----------------------|------------|
| 76 (38.4%)            | 69 (34.8%)         | 53 (26.8%)           | 198 (100%) |

Table 8

|         | before |      |       | after  |      |       |
|---------|--------|------|-------|--------|------|-------|
|         | higher | same | lower | higher | same | lower |
| fm      |        | 2    | 17    |        | 2    | 6     |
| sec     | 7      | 30   | 18    | 7      | 25   | 16    |
| ss1     | 2      | 12   | 5     | 1      | 2    | 12    |
| ss2     | 1      |      |       |        |      |       |
| app/sec |        | 1    | 5     |        | 1    | 1     |
| sum     | 10     | 45   | 45    | 8      | 30   | 35    |
|         | 100    |      |       | 73     |      |       |