USERS' PERSPECTIVES ON THE USEFULNESS OF STRUCTURE FOR XML INFORMATION RETRIEVAL

Gabriella Kazai

Microsoft Research, Cambridge, UK
gabkaz@microsoft.com

Andrew Trotman

Department of Computer Science, University of Otago, New Zealand
andrew@cs.otago.ac.nz

Keywords: XML IR, Passage Retrieval, Element Retrieval, Interactive Experiments, Qualitative User Study

Abstract: The widespread use of the eXtensible Markup Language (XML) on the Web and in digital libraries has led to a drastic increase in the number of XML Information Retrieval (IR) systems being developed. XML IR approaches exploit the logical structure of documents for their querying, retrieval and presentation to the user. Despite their abundance, there remains uncertainty regarding the advantages that structural information may bring to IR. In this paper we report on a user study exploring questions around the potential benefits of structure to users, such as: Is structural information useful when searching for relevant information? Can the structure of a document help to locate relevant information when browsing inside a document? Does the role of structural information depend on the length of a document? Our investigation was conducted as part of the INEX 2006 interactive track experiment, which we supplemented with questionnaires. Our qualitative analysis of the data collected from seven participants aims to identify how users will interact with XML IR systems. We do this by drawing parallels with paper based information searching, Web searching, and digital library searching. What we find is that XML IR users are unlike Web users – they use advanced search facilities, they prefer a list of results supplement with branch points into the document, and they need better methods of navigation within long documents.

1. INTRODUCTION

Since its approval by the World Wide Web Consortium in 1998, XML (Bray et al., 1998) has become a de facto data representation and exchange format on the Web and the adopted document representation format in digital libraries. The continuous growth in the amount of XML documents available in various repositories is closely matched by increasing efforts in the development of systems for searching and browsing XML documents, e.g., (Baeza-Yates et al., 2000; Baeza-Yates et al., 2002; Blanken et al., 2003; Fuhr and Großjohann, 2001; Malik et al., 2006c). These so called content-oriented XML IR approaches focus on the document-centric view of XML, where markup serves mainly as a means for exposing the logical structure of documents.

XML IR has attracted a lot of attention in the IR community as it offers the potential to change the way we interact with textual information. By breaking away from the notion of a document as the fixed unit of retrieval, XML IR represents a radical departure from traditional document retrieval. Exploiting the explicitly available document structure, XML IR approaches implement a more focused retrieval strategy and return document components (instead of whole documents) to the user. This focused retrieval approach is seen of particular benefit for collections containing long documents or documents covering a wide variety of topics (e.g. books, user manuals, legal documents, etc.), where the user's effort to locate relevant content can be reduced by directing them to relevant parts of the documents.

In addition to harnessing a document's structure to return relevant information in response to a traditional

keyword based query, users of XML IR systems may enrich their queries by adding structural conditions. This increased expressive power of XML query languages has been recognized as a means that can lead to performance enhancement (Schlieder and Meuss, 2002). For example, the difference between the queries "smith" and "smith as author" can be clearly stated. Due to the inherent ambiguity in the first query, and the disambiguation of the second, a reduction in the number of non-relevant documents can result in an increase in precision. So there are theoretical advantages to using structure in a query.

Furthermore, the presentation of retrieval results to the user could be improved and post-query navigation (Olston and Chi, 2003) inside a relevant document or a collection of documents could be better supported by exploiting structural information (Malik et al., 2006a). For example, relevant paragraphs can be highlighted or table of contents generated.

Despite these potential benefits, there remains uncertainty about the advantage of XML IR over plain-text IR. The expressiveness of XML query languages may, for example, be of little value if users are unable to make use of it – a very real problem as demonstrated in (O'Keefe and Trotman, 2003). O'Keefe and Trotman examined the ability of experienced XML users to accurately write queries in XPath and found that the vast majority could not. Another question is with regard to the use of structure in post-query navigation. What evidence do we have, for example, about the usefulness of a simple table of contents in helping users navigate electronic documents? If the problem of within-document information foraging is to be addressed, it is necessary to know how people navigate electronic documents.

Although much progress has been made, especially as part of the Interactive Track at the INitiative for the Evaluation of XML retrieval (INEX) (Malik et al., 2006b), there is still little known about what users consider useful features of XML IR systems. Relatively little research has been carried out to study user interactions with IR systems that take advantage of the additional features offered by XML documents, and so little is known about how users behave in the context of XML IR systems. Thus, several research questions remain regarding the usefulness of structure when it comes to querying, retrieval, and result presentation in XML IR. These are important questions to answer in order to guide user interface design for future systems.

Questions surrounding the information seeking behavior of XML IR users also bear important implications over the assumptions made when considering the evaluation of such systems. For example, the employed metrics at INEX make explicit assumptions on user preferences. Some of the assumptions behind the metrics include: users browse through retrieved components in a linear order; and they "jump" with a given probability from one XML element to another within the document's structure (Kazai and Lalmas, 2006; Piwowarski and Dupret, 2006). The understanding of user behavior and the motivations behind these is, hence, crucial for facilitating progress in XML IR.

In this paper we investigate the user perspective on research questions central to XML IR: Can the use of structural information in querying, retrieval, and result presentation improve retrieval effectiveness? We elicited general user opinion on questions including: Can users make use of more complex query languages? Can users benefit from implicit or explicit structural knowledge to help them locate relevant information or to help them in deciding if a document is relevant?

To investigate these issues, we took part in the Interactive Track (iTrack) at INEX 2006. In this experiment, passage and XML element retrieval were compared by users performing specific simulated work tasks (Malik et al., 2006b). We extended the experiment by administering a questionnaire designed to explore user opinion on the usefulness of structural information.

The paper is structured as follows. In Section 2, we review previous work on interactive XML IR. In Section 3, we discuss the INEX 2006 interactive experiment in which we participated; we present our research questions, our questionnaire questions, and our findings. We conclude in Section 4.

2. WHAT DO XML IR USERS WANT?

2.1 Assumptions and intuitions

Most of the early work in XML IR built on limited evidence from past user studies, e.g., (Chiaramella et al., 1996), as well as assumptions regarding the behavior of potential users of XML IR systems, and intuitions on what features may be beneficial.

The main assumption is that users will benefit from a more focused retrieval approach. For instance, Dao et al. in (Dao et al., 1996) state that "Retrieval of components at all levels of the document hierarchy should be supported as the unit of retrieval appropriate

for answering a user query varies with the user needs. Sometimes a whole book or a whole chapter is required, at other times only a page, a paragraphs or a line is requested".

Allowing structural conditions within a suitable query language is seen as a functionality users require as well as a precision enhancing method that users should be able to exploit. For example, (Baeza-Yates and Navarro, 1996) envisions the following search task: "Suppose, for example, a typical situation of visual memory: a user remembers that what he/she wants was typed in italics, short before a figure that said something about earth." It is argued that searching, in the traditional IR sense, for the word "earth" or for any text in italics would prove a wasteful tactic, while an appropriate structural query language would provide the solution. On the other hand, this need should be balanced against another whereby "The user should not be aware of details about how the structure of the document is internally represented." (Navarro and Baeza-Yates, 1997). The example used to demonstrate this point is that of two words which appear as a contiguous sequence of text, but which may (or may not) have markup between them. It cannot be expected of the user to be aware of such structure. This has been resonated by many researchers, including (Abiteboul et al., 1997): "When querying semi structured data, users cannot be expected to be fully aware of the complete structure, especially if the structure evolves dynamically". The balance is to provide structural querying facilities as an optional resource for users who require it: "A user having a better knowledge about the structure of the XML document should be able to put some structural constraints into the query and therefore limit the number of uninteresting results" (Al-Khalifa et al., 2003).

2.2 User studies

Studies of user behavior in the context of XML IR have been carried out by (Finesilver and Reid, 2003; Lalmas and Reid, 2003; Reid et al., 2006; Crestani et al., 2004). The small-scale study in (Finesilver and Reid, 2003) investigated searchers' information seeking behavior, comparing two variants of the same interface: one highlighting relevant objects and one highlighting best entry points (BEPs). The more detailed study of (Reid et al., 2006) examined aspects that influence users' BEP selection strategies with the aim to support automating BEP identification. Further small-scale studies exploring particular aspects of the issue of automatic BEP identification were carried out in (Lalmas and Reid, 2003). Crestani et al. in (Crestani et al., 2004) focused on evaluating a specific interface for XML retrieval.

A study aiming to elicit user expectations regarding their possible interaction in the context of XML retrieval was carried out in (Betsi et al., 2006). Through a series of interviews and using low-fidelity prototypes, such as paper mock-ups, the opinions of 10 employees of a small software company were collected on the advantages and disadvantages of their current IR system as well as of a potential XML IR system. The findings showed that all subjects would prefer to gain access to relevant document parts directly, but would want such components displayed to them in their context. It seemed important that they would be able to interact with the system and able to browse around a relevant component, while being aware of its wider context, e.g. what document it occurs in. Regarding the presentation of retrieval results, a main finding was that of users' disliking of long lists of retrieved items and showing clear preference toward a possible interface, which grouped elements by documents. It was not clear, however, if the documents should be ranked according to their most relevant component or some overall score of relevance or by some other mechanism. With respect to a query language, interviewees suggested that the use of structure could be beneficial provided a simple grammar for expressing such query constraints. However, most users stated that they would find it very difficult to write such queries successfully.

The largest user studies have been carried out as part of the interactive track at INEX (Tombros et al., 2005; Hammer-Aebi et al., 2006).

The aim of the INEX 2004 interactive track was to study the behavior of users when interacting with components of XML documents, and investigate which approaches for XML retrieval are effective in user-based environments. In response to a user query, the search engine returned a ranked list of components which searchers were free to explore. When users selected a result component, its content was displayed to them together with the table of contents for the document containing the component. Searchers had access to other components within the same document either by using the table of contents, or by linear navigation (i.e. previous or next buttons). The granularity of all components on the interface was limited to the article, section, subsection and

subsubsection levels.

The main findings of the study included the general observation that overlapping components, i.e., nested components from the same document, returned at different ranks in the ranked list, frustrated many users (Malik et al., 2006a). This may then contributed to the finding that searchers did not interact much with other components of a given XML document after the selection of the initial result element from the ranked list. Recognizing that they have accessed the same document through a different result component, searchers would normally return to the ranked list and access a different result instead of browsing again within an already visited document. An important finding regarded the use of document structure as contextual information that users often consulted in order to decide on the usefulness of a document. Furthermore, the analysis of users' browsing behavior indicated that they tend to browse to more specific information rather than to more exhaustive information.

In its second year, the interactive track aimed to elicit user perceptions of what is needed from an XML retrieval system and identify applications for element retrieval (Malik et al., 2007; Tombros et al., 2005). The main questions investigated were: "Is element retrieval useful for searchers?", "What granularity of elements do searchers find more useful?", "What applications for element retrieval can be viable in interactive environments?". Two versions of a baseline search system were used for the experiments: one presented the retrieval results in the context of the full text (i.e., highlighted); the other presented the results in isolation.

The findings showed that the majority of users make use of direct access to individual XML elements as their entry points. The obtained relevance assessments showed clear advantages for focused retrieval techniques, with users labeling only parts of documents as relevant in most cases. This was confirmed by other track participants (Larsen et al., 2006; Kim et al., 2005). The study also showed that the issue of overlapping elements could be naturally resolved by appropriate user interface design: by presenting a hierarchical grouping of elements from a single document in the result list, and indicating retrieved elements from a given document in the table of contents in the detail view. Similar findings were reported in (Kamps and Sigurbjörnsson, 2005). In general, the presence of the logical structure of the documents alongside the contents of the accessed components was a feature that searchers found useful as it provided easy navigation, and a quick overview of which elements may be relevant.

We took part in the interactive experiments at INEX 2006 in order to investigate some of the less explored questions concerning user experiences with structure for querying or result navigation in their daily search routines. We summarize our participation and findings in the next section.

3. INEX INTERACTIVE TRACK EXPERIMENTS

The interactive track (iTrack) at INEX 2006 (Malik et al., 2006b) examined the user interaction with XML documents in an experimental laboratory environment. The format of the track was of an exploratory nature and had relatively broad aims rather than addressing very specific research questions. The task focused on pitting XML element retrieval against passage retrieval to explore potential benefits and trade-offs.

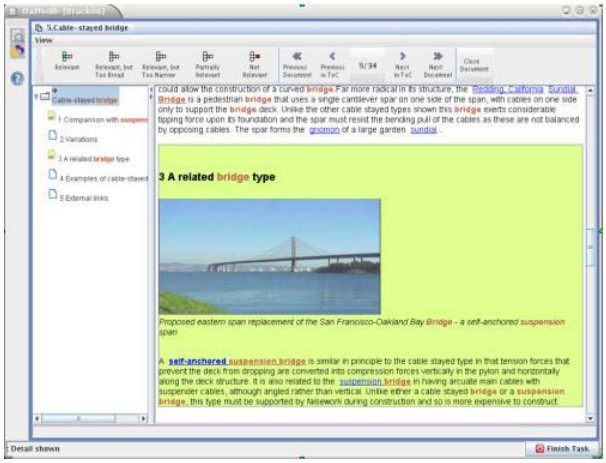

Figure 1: TopX XML element retrieval system

The experiment used the INEX Wikipedia collection consisting of 659,388 documents totaling about 4.6GB of data (Denoyer and Gallinari, 2006). Twelve topics were chosen by the interactive track organizers.

Participants were exposed to two different search engines that shared (where possible) the same user interface: an XML element retrieval search engine, and a passage retrieval search engine. The difference between the two systems was subtle. In both, as a response to a user query, the search engine returned an ordered list of documents and within each document an ordered list of non-overlapping relevant passages. The

difference was that in the element retrieval system the passages were restricted to starting and ending on the boundary of a single XML element (of granularity chosen by the search engine).

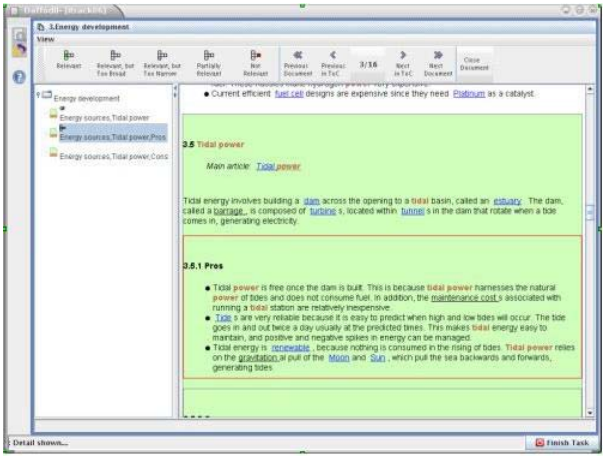

Figure 2: Panoptic passage retrieval system

The element retrieval engine was TopX (Theobald et al., 2005) and the passage retrieval system was Panoptic™/Funnelback™ provided by CSIRO. In order to remove any experimental bias due to user interface differences, both search engines were interfaced with a consistent look and feel. The similarity of the two systems can be seen in Figures 1 and 2. Figure 1 shows the front end to the TopX search engine, while Figure 2 shows the front end to Panoptic. There are two differences between the systems; first the relevance ranking is different, with one identifying elements, the other passages within the documents. Second, when a document is displayed, either elements or passages are highlighted as relevant.

3.1 Experimental Setup

Each participant conducted the experiment on their own, but with one of the authors present. Each had to complete four search tasks, having a choice of three topics for each task. Fifteen minutes were allocated to complete a task using one of the search engines.

Initially the participants were presented with a sample topic and given as long as necessary to familiarize themselves with each search systems. The experiment commenced with a pre-experiment questionnaire, which collected demographics such as age, gender, and searching experience. Each task was preceded with a pre-task questionnaire, collecting information regarding participants' familiarity and level of interest in their chosen search topic. The post-task questionnaire gathered user feedback on the completed task and the usefulness of certain features of the interface. The experiment closed with a post-experiment questionnaire, which asked comparative questions over the two search systems.

Table 1: Topic categorizations

Category	**Topics**
A1	1,2,3
A2	5,6,7
A3	9,10,11
A4	4,8,12
B1	2,3,4
B2	6,7,8
B3	10,11,12
B4	1,5,9

Table 2: Permutations of the categories across participants. Shown shaded is the passage retrieval system and un-shaded is the element retrieval system. The order of the two systems was also permuted.

Participant	Search Task			
P1	A1	A2	A3	A4
P2	A2	A1	A4	A3
P3	A3	A4	A1	A2
P4	A4	A3	A2	A1
P5	B4	B3	B2	B1
P6	B3	B4	B1	B2
P7	B2	B1	B4	B3
P8	B1	B2	B3	B4

To avoid biasing the experiment two different categorizations of the 12 topics were used: A and B as shown in Table 1. Furthermore, the order in which each category was presented to a user, and the system on which the task was assigned changed between users as shown in Table 2.

In addition to the track experiments, we devised our own, extensive survey focusing on eliciting general opinions on how structure can be useful for querying, retrieval and presentation of results. In the following sections, we summarize our findings both from our participation in the iTrack and from the feedback we collected in our own survey. We focus on reporting the findings from our own questionnaire as the 2006 iTrack continues throughout 2007 and thus it is too early to present general results from it.

3.2 Participants

Of our eight participants only seven completed the experiment, the eighth was unable to finish due to difficulties. The gender breakdown of the seven was: six male and one female. The average age group was 39.7, the youngest being 24, the oldest 63 years old. On average, our users had 12 years of search experience on the Web. Additional information is given in Table 3.

Table 3: Participants. UD=Undergraduate degree, Lec=lecturer, SE=Software engineer, UC=Usability consultant, Res=Researcher, Stu=Student.

Participant	P1	P2	P3	P4	P5	P6	P7
Gender	M	M	M	M	M	M	F
Age	44	63	42	45	29	24	31
English is first lang.	Y	Y	Y	Y	Y	N	N
Education	PhD	PhD	UD	PhD	PhD	UD	PhD
Occupation	Lec	Lec	SE	UC	Res	Stu	Res
Web search exp. (years)	18	10	10	10	12	12	12

3.3 Querying

Two different interaction models for querying XML documents have been explored at INEX: querying by Content-Only (CO) and querying by Content-And-Structure (CAS). The former is the traditional keyword based search. The latter allows combining structural constraints within a query: either by specifying the context where a search term should appear (support element) or stating what structures should be returned (target element).

The interpretation of structural constraints in a query has been subject to some debate. A strict interpretation supports a more database oriented view, but allows for users to express precise information needs, e.g. "<author>smith</author>". Following a more amorphous, IR view, structural constraints within a query can be considered merely as hints (Trotman and Lalmas, 2006). Should the user specify that they are looking for a section of a document about, say, "medieval archery", it does not mean that a sub-section cannot be relevant. This assumption was examined in (Betsi et al., 2006) who found that most users expected all relevant elements to be found irrespective of the structural constraint.

Studies (Trotman and Lalmas, 2006) into the interpretation of structural hints have shown that with support elements the interpretation can always be considered a hint, but with target elements this is not the case. The search engine must know if the user considers the constraint to be a hint or a strict constraint.

Woodley et al. (Woodley et al., 2006) conducted a questionnaire in which they asked users when element retrieval would be more useful than document retrieval. They concluded that there were two situations, the first was when relevant information was surrounded by irrelevant information, and the second was when the query was on two or more topics.

The study of O'Keefe and Trotman (O'Keefe and Trotman, 2003) highlighted that constructing queries with structural constraints is far from obvious to users. Trotman and Sigurbjörnsson went on to specify a simple query language called NEXI (Trotman and Sigurbjörnsson, 2004) and showed that it could be used effectively by (essentially) the same user group (Trotman and Lalmas, 2006).

Several methods have been proposed for supporting users in specifying structural queries (see (Amer-Yahia and Lalmas, 2006) for a detailed survey of XML query languages), but two prevail. The first is the traditional text-box query input; the second is the advanced form query such as that of Van Zwol et al. (van Zwol et al., 2006). Their advanced query builder called Bricks was compared to the traditional text-box style querying and Bricks was found to be preferred.

3.3.1 Research questions

Although there remain many unanswered questions about querying, in our study, we focus on the following research question:

R1. Do users specify structural hints in queries?

Many online search engines allow users to specify structural hints. Google, for example, has an advanced search form, which allows users to specify (amongst other things) Internet domain and file formats as well as where in a document the search terms should be found (e.g. title, body, URL, anywhere).

Online digital libraries also typically provide advanced search forms, allowing users to narrow their search by specifying structural hints in their queries. For example, sites such as Wiley InterScience allow searching by author, article title, and publication information.

To answer questions about XML IR querying, we surveyed our users about their querying behavior on the Web and in digital libraries. Specifically we asked:

Q1. Please state which online digital libraries you use.
Q2. When searching in digital libraries, do you use:
- a) Keyword search
- b) Advances search
  - i) Author
  - ii) Title
  - iii) Published data
  - iv) Publisher
  - v) Classification
  - vi) Abstract
  - vii) Introduction
  - viii) Conclusion
  - ix) Chapter Title
  - x) Search in table of contents (ToC) of a document

Q3. When searching the Web, do you use:
- a) Keyword search
- b) Advances search
  - i) Language
  - ii) File format
  - iii) Date
  - iv) Domain
  - v) Title only
  - vi) Other

Q4. Do you use any book /thesis searching services (please state)?
Q5. Would you want to search by chapter/section title in a:
- a) Book database
- b) Thesis database
- c) On the Web

Q6. Do you use "exact phrase" searching on the Web?

3.3.2 Study findings

All, but one of our participants (p3) use digital libraries (DLs) on a regular basis. Enlisted DLs (Q1) include the ACM Digital Library, IEEE, SpringerLink, company specific libraries, University libraries, Ingenta, Web of Science, as well as Google Scholar. All six use keyword search as the most common form of querying a digital library (Q2). Five of our participants also make use of the advanced search facilities with varying frequency (p1, p4 and p7 use it often, p5 and p6 sometimes). From the advanced options, author and title searches are used most often, while publication date, classification and abstract search are accessed sometimes by two of the participants. Publication source, e.g. conference name, is also often used by one participant (p7). Advanced search was described as having "more power and flexibility"; "[useful] if I am looking for a specific article [..] to obtain the best results quickly".

By comparison, all participants use keyword search on the Web (Q3). Advanced Web search facilities are used rarely by five and not at all by two of the participants (p2 and p7). The rarely used advanced search options include file type filters (e.g. ".jpg"), domain restrictions (e.g. "nz") or searching on dates.

The difference in behavior between a digital library and Web search is expected due to the underlying nature of the different collections being searched: an online digital library is more structured and homogenous than the Web. Users of a DL are more likely to be familiar with the structural properties of the collection and are more likely to make use of this knowledge in order to narrow their searches and improve precision.

Focusing on book services (Q4), we found that only two of the participants use book search (p1 and p3): they work with various library catalogs and Amazon. They would both look for general books (e.g. fiction) on Amazon but use library catalogs to locate specific, work related books.

None of the participants had any experience with thesis databases (Q4).

When asked if they would search a book database by chapter or section title (Q5), two participants answered yes (p1 and p6). Examples of when such a search facility would be useful focused on querying reference works and textbooks. In comparison, three of the participants would search by chapter title in a thesis database, commenting that they would find this useful when searching for overview information or a chapter on related work, or looking for results. None of the participants would search by section title on the Web.

Exact phrase searching (Q7) is used by six of the seven participants. Furthermore, both p1 and p3 added that they often employ Boolean query constructs. Three of the participants (p1, p3 and p6) would also find it useful if they were able to specify proximity constraints within a query (i.e. when search terms occur in the same section or paragraph).

Even though our user population is small, we can clearly see an emerging trend regarding user perception of the usefulness of structure for querying. Although users typically use keyword searching, most also use advanced searching. The more structured the collection, the more useful are considered the structural constraints of a query.

Our users claim to use advanced searching between often and sometimes. This is in stark contrast to the known frequency of such queries in Web query logs. Spink et al. (Spink et al., 2001), for example, analyzed a query log of over a million Web queries from Excite and found that phrase searching was used in only 5% of queries. In total they found that "advanced search features" (including Boolean) were used by a very small number of users and for an "overwhelming" number of users these advanced search features don't exist. Our sample must come from those that do – and for them the advanced search is important functionality.

Spink et al., go on to suggest that the use is so low as to raise questions regarding their desirability in a search engine. Perhaps for a general-purpose Web search engine this is a reasonable conclusion, but in our study we are do not have general-purpose users searching the Web – we have a domain specific search engine and advanced users. For them the advanced search interfaces, and the ability to add structure to their queries is desirable. As six of our users claim to use this facility on the Web, they are amongst that very small class of users Spink et al. identified.

In answer to our research question (R1): the users we sampled use facilities provided by advances search interfaces in digital libraries and on the Web. In other words, they (either directly or indirectly) specify structural hints in their queries. It is thus essential to provide such a facility, and to continue to research how to take advantage of structural hints when present in a query.

Spink et al. also identify many instances of the incorrect use of advanced search facilities on the Web. O'Keefe and Trotman (O'Keefe and Trotman, 2003) did the same for text-based XML IR queries, and (Trotman and Sigurbjörnsson, 2004) for the query language NEXI. In all cases users continue to make mistakes in their query specification. It is not clear how to eliminate the potential for error, however we believe that advanced query paradigms such as Bricks (van Zwol et al., 2006) may hold the answer.

3.4 Finding relevant information

Trotman et al. (Trotman et al., 2006) reviewed a range of XML IR retrieval paradigms and concluded that document-centric retrieval was the most plausible search scenario from the user's point of view. Betsi et al. in (Betsi et al., 2006) concluded that users expected to interact with whole documents. Woodley et al. (Woodley et al., 2006) identified users suggesting that element retrieval could be used in conjunction with document retrieval to identify relevant parts of a document. They also identify situations where element retrieval would be more useful than document retrieval; however, these are not incompatible with the first observation. These recent findings mirror the recommendations of (Chiaramella et al., 1996) in support of a fetch and browse strategy to structured document retrieval: the task of a retrieval system is to fetch relevant documents and then support the user in browsing through its content in search of relevant text fragments.

INEX identifies several retrieval tasks including: focused, thorough, relevant-in-context, and best-in-context (Malik et al., 2006c). The credibility of each has been examined (Trotman et al., 2006). The relevant-in-context task is considered the most credible direct end-user task. In this task an XML IR search engine must rank whole XML documents on topical relevance, and then within a document highlight the relevant text fragments for the user. An alternative, which has been explored at INEX in the 2005 fetch and browse task, is to provide a ranking of XML elements contained within the document based on their topical relevance. A result list of this nature represents first an ordering of relevant documents, then for each document a "heat-map" of relevance. Both of these solutions rely on the same principle: to first guide users to discover relevant documents through the medium of a ranked result list, and then to aid users to locate relevant information inside the documents.

We accept the work of (Trotman et al., 2006) that the most credible XML IR task is relevant-in-context and we aim to explore questions around the usefulness of structural information in supporting users to find relevant information.

3.4.1 Research questions

When it comes to estimating the relevance of XML documents and/or elements, there are plenty of proposed ranking algorithms to choose from (Amer-Yahia and Lalmas, 2006; Malik et al., 2006c). Such algorithms are typically tuned based on relevance judgments data over a given corpus, such as the INEX test collection. It is well known that relevance is a complex decision that builds on multiple aspects (Mizzaro, 1997). Our current interest lies in investigating the following research questions:

R2. What mechanisms do users find useful in deciding about the relevancy of a document?

R3. What mechanisms do users employ in locating relevant information inside a document?

To answer these research questions, we asked our participants the following questions at the end of the interactive experiments:

Q1. Summarize the differences between the two iTrack systems.
Q2. Give an overall rating for the usefulness of the following iTrack system features:
a) Paragraph highlighting
b) Query term highlighting
Q3. Do you find Web "cached" documents with highlighted search terms useful?
Q4. Do you use the "find in web page" feature of your web browser?
Q5. How do you decide if an academic paper/thesis or book/web page is relevant to your information need?[1]
Q6. Can a table of contents help you in finding relevant information in a document? Why?
Q7. Can a site map help you in finding relevant information on the Web? Why?

3.4.2 Study findings

Regarding the difference between the two systems (Q1) none of our participants were able to identify any difference between the element and passage retrieval systems. This suggests that the use of structure (as exposed by the two systems) did not play a role in user's ability to find relevant documents or relevant information inside the returned documents.

Participants used a scale of 1 (not at all useful) to 5 (extremely useful) to rate the usefulness of paragraph highlighting and query term highlighting (Q2). The average rating for paragraph highlighting was 1.75 (low). That is, users didn't find this functionality useful. One of the main reasons for this was that the use of background color as a highlighting mechanism was associated with standard web page background and was hence automatically ignored by participants: "Because on the Web lots of documents have different color background, I trained myself to ignore it". Only one participant reported this feature as very useful and commented that it was especially helpful "when searching for facts, where reading the whole article would have been annoying".

On the same scale, query term highlighting scored 3 (average). One participant commented that query term highlighting interfered with the paragraph highlighting, while another said that they would only "look at query term highlighting on the snippets, but on the full text it was a bit much".

These findings of the iTrack are further supported by an analogy from the Web domain explored in our survey. Of our participants, four make use of the cached documents feature of their web browser (Q3) which highlights search terms on a web page. P3 commented that the usefulness of query term highlighting depends on the search: it is "useful for finding collections of specific words in a complex Boolean query." On the other hand, all our participants use the "Find in web page" function of their browser (Q4). Several also commented that this was an absolutely vital function and p3 even mentioned that it is in fact preferred over the highlighting feature of a cashed page.

Q5 aimed at eliciting high-level strategies that users employ when deciding about the relevance of varied information items.

Participants reported that title, author, keywords and abstract are their most prominently used features when weighing up the relevancy of academic papers. The role of abstract was specifically emphasized by most participants. For example, in the words of p7: "Title grabs my attention during search, but abstract is what lets me assess the relevance of the document". An interesting angle was offered by p5: "I look at the pictures and read the abstract/conclusions/references, and browse the main technical part".

Participants' strategy to assess the relevance of a thesis was slightly different from academic papers: Although abstract still served as the main reference for relevance decisions, the table of contents was mentioned by three users. Other sources of useful information included the preface (p4), introduction (p7), online summaries (p3) and other people's reviews (p3 and p7).

For web pages, users rely on less structured methods and usually simply scan the page to decide if it is relevant, although the title of the page can be given more weight in the decision process. Other factors include "the number of words highlighted" (p2, p5) and "site maps" (p4).

When asked about the usefulness of a table of contents in finding relevant information in a document (Q6), participants had mixed views and suggested that the usefulness would depend on the length and

[1] As separate questions

complexity of the document. The table of contents can be useful to provide an overview of the structure and aid browsing by being able to directly skip to a relevant part, but it was considered inferior to the much more direct find-within-document search function. An insightful comment was made by p7, who stated "For me, it serves more as a context/structure tool and quick browsing tool. I think it is not necessarily good at pointing out relevant information".

Feedback regarding site maps (Q7) was similar in vein. The usefulness very much depends on the quality of the site map as well as on the quality of the entry page: "[Sitemaps are useful when] the other forms of navigation are awful"; "Sometimes quicker than in-page links, if done well, but usually not". For p7, site maps "serve for contextualizing the content in a high-level view of the structure, not necessarily to indicate relevance".

Our participants have indirectly provided us with a list of features useful in determining information relevance, and the order in which they prefer to use them. The find-within-document feature of a web browser (that was not provided at iTrack) is vital for navigation. After that an abstract or summary of the information content of the page, then query term highlighting (perhaps only within the summary). The table of contents was considered useful for outlining the content but not so useful for navigation. Finally passage highlighting was not considered desirable.

Cross-referencing the use of hit-term highlighting against Web cached document use showed that about the same number of participants considered each to be useful. This is reassuring as Web cached page display highlights hit terms.

When entering this experiment we expected our participants to identify passage highlighting to be of value. After all, one of the advantages of structure within a document is the ability to use it when rendering. Matched with the ability to identify where within a document the relevant content can be found, the result is (and was) a yellow-highlighting like method identifying for the user where and what to read. There are many reasons it might not have been valued.

Users are familiar with find-within-document searching, and not with yellow-highlighting. As one of our participants stated, users are even pre-programmed to ignore background colors when searching. A third reason could be the quality of the search engine. Yellow-highlighting is only valuable if it is correct; if it is not, then users will quickly learn to ignore it. In short, the precision of the search engine in within-document searching my not be high enough to be considered reliable by the users.

This realization leads to two further research questions that we leave for further work. First, if the precision is high enough, then is the computational cost of identifying relevant passages justified by use? Second, is the structure of value when identifying relevant segments within a document?

In answer to our research questions (R2, R3): users refer to titles, abstracts, and summaries to decide if a document is relevant and they prefer searching within a document to the other methods examined.

3.5 Navigating in a document

3.5.1 Research questions

In an electronic environment, the navigation of a document presents unique challenges. A user cannot just flip through the pages and stop when their attention is caught by an interesting page. New and novel forms of navigation unique to the medium are needed. Furthermore, users may have established mechanisms for paper documents, which may transfer naturally to the digital medium. For example, a table of contents is typically seen as a useful means of navigation both in paper and electronic forms (Malik et al., 2007).

The element retrieval engine at the iTrack experiments provided a table of contents for within-document navigation, while the passage retrieval system simply listed entry points to the highlighted passages within a document. We were interested in gathering feedback on user experience on the usefulness of these navigation mechanisms. Specifically our research questions were:

R4. How do users navigate within a document?

R5. How do users combine searching and browsing?

To explore these issues, we asked several questions:

Q1. Give an overall rating for the usefulness of the following iTrack system features:

a) Result presentation
b) Table of contents

Q2. Do you think a table of contents is useful in a document?

a) Why?
b) For what type of documents?
c) What information does a table of contents give you?

d) What information should it give you?

Q3. Do you typically browse or search inside an electronic document? What does your decision depend on?

3.5.2 Study findings

The scale of 1 (not at all useful) to 5 (extremely useful) was used to rate the usefulness of system features.

The results list at iTrack included a sub-list, which listed between one to three entry points into the document, leading to relevant XML elements or passages. The usefulness of the result list (Q1) was rated at 3.75. This was partly as participants felt that the result list was an essential component, but also because participants liked the ability to "go directly into the relevant section". On the down side, this result presentation strategy could prove risky as users expect high levels of retrieval accuracy, which would not only be able to locate relevant information, but would also be able to balance the user's need for context. For example, p6 commented, "Sometimes, however, it would jump into the middle of a section, but I would then have to scroll up to read from the top to make sense of it. Also there was not always enough information displayed to decide if a section was relevant."

Participants rated the usefulness of the table of contents at an averaged 2.5. Most of them did not make much use of this feature. Several commented that the scrollbar was easier to use for navigating the document. Although the table of contents was useful to get an overall feel of the document's structure, it was deemed too cumbersome to be used for navigation compared with the simplicity of the scrollbar. Another reason why the table of contents was not so useful was that participants relied on the result list's sub-list to directly access relevant text fragments inside the document, hence reducing the need for further navigation inside the document.

All participants answered yes to Q2. All participants agreed that the reason why a table of contents is useful is that it provides an overview of the structure of the document and improves navigation. Opinions differed on the type of documents that a table of contents is useful for. Three users (p1, p2 and p5) thought that any documents longer than a page can benefit from a table of contents. The others (p3, p4, p6 and p7) felt that a table of contents is only useful for long documents (e.g. books, theses, reports) or naturally structured documents (p7).

Participants typically employed a mixed strategy, which combined searching and browsing inside a document (Q3). The type of information need (specific vs. broad) is quoted as the main influencing factor. Searching is seen as a more direct and faster means of locating specific information, whereas browsing has more of an exploratory function. The comments by p1 also suggest that there is qualitative intentional difference: "Depends on how much I care, or think this document might be relevant. If I care, then I am inclined to browse". An unexpected technology related issue was mentioned by p5: "If I have a scroll wheel I'm more likely to browse than if I don't".

In answer to our research questions (R4, R5): users search, browse, and use the table of contents to navigate through a document. The presence of entry points in the list of results did, however, reduce the amount of table of contents use because the user could navigate directly to a part of the document from the results list. Exactly which strategy they use depends on the nature of their information need.

We were not expecting the use of the table of contents to conflict with the elements presented in the results list. This does, though, suggest a natural extension to the results list in a structured document environment – the presence of the table of contents. Perhaps such a document outline would only list the main document element (sections but not subsections) in order to avoid overwhelming the results list. It might even be used in conjunction with query specific document synopses presenting the location within the document from where the synopsis was extracted.

3.6 Searching in digital vs. paper documents

3.6.1 Research questions

Typical web pages are very short, in the order of a few kilobytes. By contrast, a typical academic paper might be 8 pages in length, and a chapter from a textbook more like 30 or 50 pages in length.

For long documents a web browser will provide a scroll bar. We believe there remains substantial room for improvement in navigation methods for long electronic documents. In 2007 a book search track has started at INEX and we are interested in identifying document navigation methods that can be used with such a collection. Specifically, our research question is:

R6. How do users navigate within short and long documents?

We expected methods used on paper and electronically to be substantially different, so we asked:

Q1. What strategy do you use to find relevant information in a short (e.g. up to 12 pages) printed document?
Q2. What strategy do you use to find relevant information in a long (100+ pages) printed document?
Q3. What strategy do you use to find relevant information in a short (up to 12 screens) electronic document?
Q4. What strategy do you use to find relevant information in a long (100+ pages) electronic document?

3.6.2 Study findings

For short printed documents (Q1), five participants said they skim documents, and also five stated that they look for a table of contents and read that. Some stated both, with one stating that they "Skim reading, looking at headings/diagrams".

For long printed documents (Q2), all participants made use of the table of contents. Four stated that they would also use the index, although this was less popular even with those who did use it, with one acknowledging its use as "very rarely".

With short electronic documents (Q3), all our participants used within-document keyword searching as their primary method. This was in fact so ingrained in the information finding experience of our participants that one participant simply referred to it as "Ctrl+F". Two participants stated that they glance at section headings while scrolling, while one said they "glanced" at the content without reading it thoroughly.

Long electronic document navigation methods (Q4) were less variable. Keyword searching within the document was universal, additionally two participants stated table of contents use.

Navigation of long documents makes it possible for users to satisfy their information needs more quickly. Should those documents be paper based and divided into sections and subsection, the table of contents is the preferred navigation aid. If the same documents were housed electronically then keyword search would be preferred. This difference suggests either keyword search is better than table of contents searching, or that the table of contents navigation of electronic document is cumbersome.

In answer to our research question (R6): for short documents skimming is a popular method for both paper and electronic documents. For electronic documents, however, the primary mechanism to locate relevant content is via the use of within-document search functions. For long documents table of contents is preferred in paper, but electronically keyword search is preferred.

Further investigation is needed to determine how the table of contents navigation method could be made of value for electronic documents. Should the table of contents be at the top of the document (as seen in ScienceDirect) then constant scrolling might detract from its use. On the other hand, if present at the beginning of each section (as seen in BioMedCentral) of a document it might be useful.

4. CONCLUSIONS

While the small sample size of users is a limitation of our study, our findings add further evidence to the growing pool of studies on users of XML IR systems.

In this paper, we have identified several important differences between typical Web use, paper based information foraging, and XML IR use. The typical Web user uses keyword search and does not use advanced search features. XML IR users, on the other hand, will use advanced search facilities such as structural constraints. This raises questions as to how to best solicit the users information need from them when structured documents are available.

The participants of this study liked the ability to jump directly into a part of the document from the list of results – something not possible if the documents are not structured. The section titles in the list of results also acted as an outline of the document making the table of contents less useful. Further investigation is needed to see how to best take advantage of the list of results to help direct navigation into a document.

We found that our participants did not like the hit-section highlighting. One possible reason for this is the accuracy of the technique; another is that Web users are self-trained to ignore background colors.

Together these two findings suggest that research effort should be spent in identifying good reduced document outlines and query specific snippets rather than accurately identifying the exact location within the document for later highlighting.

Keyword searching within a document ("Ctrl-F" as one participant called it) is an important navigation mechanism for finding information in both short and

long documents. The table of contents in a long document was not a popular navigation tool. It is not clear why, but it might be because excessive scrolling is needed if it is presented at the top of a document. Further research is needed to determine if it is simply not wanted or is currently awkward to use.

We have identified that the behavior of XML IR users is quite different from Web users, and the way they interact with structured electronic documents is quite different from how they interact with paper documents. We know more about how they use structured documents and have opened avenues to address when it comes to improving such systems. Most specifically we believe that our investigation suggests much about a suitable user interface for an XML IR system and hope to build such a system for searching books as part of the INEX book search track.

REFERENCES

Abiteboul, S., Quass, D., McHugh, J., Widom, J., and Wiener, J. L. (1997). The Lorel query language for semistructured data. *International Journal on Digital Libraries*, 1(1):68–88.

Al-Khalifa, S., Yu, C., and Jagadish, H. V. (2003). Querying structured text in an XML database. Proceedings of the 2003 ACM SIGMOD *International Conference on Management of Data*, pages 4–15, ACM Press.

Amer-Yahia, S. and Lalmas, M. (2006). XML search: Languages, INEX and scoring. *SIGMOD Record*, 35(4):16–23.

Baeza-Yates, R., Fuhr, N., and Maarek, Y., editors (2002). Proceedings of the SIGIR 2002 Workshop on XML and Information Retrieval.

Baeza-Yates, R., Fuhr, N., Sacks-Davis, R., and Wilkinson, R., editors (2000). Proceedings of the SIGIR 2000 Workshop on XML and Information Retrieval.

Baeza-Yates, R. and Navarro, G. (1996). Integrating contents and structure in text retrieval. *SIGMOD Record*, 25(1):67–79.

Betsi, S., Lalmas, M., Tombros, A., and Tsikrika, T. (2006). User expectations from XML element retrieval. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 611–612.

Blanken, H. M., Grabs, T., Schek, H.-J., Schenkel, R., and Weikum, G., editors (2003). Intelligent Search on XML Data, Applications, Languages, Models, Implementations, and Benchmarks, volume 2818 of *LNCS*. Springer.

Bray, T., Paoli, J., and Sperberg-McQueen, C. (1998). Extensible Markup Language (XML) 1.0., W3C (World Wide Web Consortium) Recommendation. Technical report. http://www.w3.org/TR/1998/REC-xml-19980210

Chiaramella, Y., Mulhem, P., and Fourel, F. (1996). A model for multimedia information retrieval. Technical report FERMI ESPRIT BRA 8134, University of Glasgow.

Crestani, F., Vegas, J., and de la Fuente, P. (2004). A graphical user interface for the retrieval of hierarchically structured documents. *Inf. Proc. Man.*, 40(2):269–289.

Dao, T., Sacks-Davis, R., and Thom, J. A. (1996). Indexing structured text for queries on containment relationships. *Proceedings of the 7th Australasian Database Conference*, pages 82–91.

Denoyer, L. and Gallinari, P. (2006). The Wikipedia XML corpus. In (Fuhr et al., 2007), pages 12–19.

Finesilver, K. and Reid, J. (2003). User behaviour in the context of structured documents. *Advances in Information Retrieval: 25th European Conference on IR Research (ECIR),* pages 104–119.

Fuhr, N. and Großjohann, K. (2001). XIRQL: a query language for information retrieval in XML documents. *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 172–180. ACM Press.

Fuhr, N., Lalmas, M., Malik, S., and Kazai, G., editors (2006). Advances in XML Information Retrieval and Evaluation, 4th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2005, Dagstuhl, Germany, November 28-30, 2005, Revised Selected Papers, volume 3977 of LNCS. Springer.

Fuhr, N., Lalmas, M., and Trotman, A., editors (2007). Comparative Evaluation of XML Information Retrieval Systems, 5th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2006, Dagstuhl, Germany, December 17-20, 2006, Revised and Selected Papers, volume 4518 of LNCS. Springer.

Hammer-Aebi, B., Christensen, K. W., Lund, H., and Larsen, B. (2006). Users, structured documents and overlap: interactive searching of elements and the influence of context on search behaviour. *Proceedings of the 1st International Conference on Information Interaction in Context*, pages 46–55. ACM Press.

Kamps, J. and Sigurbjörnsson, B. (2005). What do users think of an XML element retrieval system? In (Fuhr et al., 2006), pages 411–421.

Kazai, G. and Lalmas, M. (2006). eXtended Cumulated Gain measures for the evaluation of content-oriented XML retrieval. *ACM Trans. Inf. Syst*., 24(4):503–542.

Kim, H. and Son, H. (2005). Users interaction with the hierarchically structured presentation in XML document retrieval. In (Fuhr et al., 2006), pages 422–431.

Lalmas, M. and Reid, J. (2003). Automatic identification of best entry points for focused structured document retrieval. *Proceedings of the 12th International Conference on Information and Knowledge Management (CIKM)*, pages 540–543. ACM Press.

Larsen, B., Tombros, A., and Malik, S. (2006). Is XML retrieval meaningful to users?: Searcher preferences for full documents vs. elements. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 663–664. ACM Press.

Malik, S., Klas, C.-P., Fuhr, N., Larsen, B., and Tombros, A. (2006a). Designing a user interface for interactive retrieval of structured documents - lessons learned from the INEX interactive track. *European Conference on Digital Libraries (ECDL),* volume 4172 of LNCS, pages 291–302. Springer.

Malik, S., Larsen, B., and Tombros, A. (2007). Report on the INEX 2005 interactive track. *SIGIR Forum*, 41(1):67–74.

Malik, S., Tombros, A., and Larsen, B. (2006b). The interactive track at INEX 2006. In (Fuhr et al., 2007), pages 387–399.

Malik, S., Trotman, A., Lalmas, M., and Fuhr, N. (2006c). Overview of INEX 2006. In (Fuhr et al., 2007), pages 1–11.

Mizzaro, S. (1997). Relevance: The whole history. *J. Am. Soc. Inf. Sci.*, 48(9):810–832.

Navarro, G. and Baeza-Yates, R. A. (1997). Proximal nodes: A model to query document databases by content and structure. *ACM Trans. Inf. Syst*., 15(4):400–435.

O'Keefe, R. A. and Trotman, A. (2003). The simplest query language that could possibly work. *Proceedings of the 2nd Workshop of the Initiative for the Evaluation of XML Retrieval (INEX'02).* Dagstuhl, Germany, December 15–17, 2003.

Olston, C. and Chi, E. H. (2003). ScentTrails: Integrating browsing and searching on the Web. *ACM Trans. Comput.-Hum. Interact.,* 10(3):177–197.

Piwowarski, B. and Dupret, G. (2006). Evaluation in (XML) information retrieval: expected precision-recall with user modelling (EPRUM). *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 260–267. ACM Press.

Reid, J., Lalmas, M., Finesilver, K., and Hertzum, M. (2006). Best entry points for structured document retrieval - part I: Characteristics. *Inf. Process. Manage*., 42(1):74–88.

Schlieder, T. and Meuss, H. (2002). Querying and ranking xml documents. *J. Am. Soc. Inf. Sci. Technol.*, 53(6):489–503.

Spink, A., Wolfram, D., Jansen, M. B. J., and Saracevic, T. (2001). Searching the Web: the public and their queries. *J. Am. Soc. Inf. Sci. Technol.*, 52(3):226–234.

Theobald, M., Schenkel, R., and Weikum, G. (2005). An efficient and versatile query engine for TopX search. *Proceedings of the 31st International Conference on Very Large Data Bases (VLDB)*, pages 625–636. ACM Press.

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

Trotman, A. and Lalmas, M. (2006). Strict and vague interpretation of XML-retrieval queries. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 709–710. ACM Press.

Trotman, A., Pharo, N., and Lehtonen, M. (2006). XML IR users and use cases. In (Fuhr et al., 2007), pages 400–412.

Trotman, A. and Sigurbjörnsson, B. (2004). Narrowed Extended XPath I (NEXI). *Proceedings of the 4th Workshop of the Initiative for the Evaluation of XML Retrieval (INEX'04),* volume 3493 of LNCS, pages 16–40. Springer.

van Zwol, R., Baas, J., van Oostendorp, H., and Wiering, F. (2006). Bricks: The building blocks to tackle query formulation in structured document retrieval. *Advances in Information Retrieval: 28th European Conference on IR Research (ECIR),* volume 3936 of LNCS, pages 314–325. Springer.

Woodley, A., Geva, S., and Edwards, S. L. (2006). What XML IR users may want. In (Fuhr et al., 2007), pages 423–431.