What XML-IR Users May Want

Alan Woodley[1], Shlomo Geva[1], Sylvia L. Edwards[2]

[1]School of Software Engineering and Data Communication,[2]School of Information Systems
Faculty of Information Technology, Queensland University of Technology
GPO Box 2434, Brisbane, Queensland, Australia
{ap.woodley@student.qut.edu, s.geva@qut.edu.au}

Abstract. It is assumed that by focusing on retrieval at a granularity lower than documents that XML-IR systems will better satisfy users' information need than traditional IR systems. Participates in INEX's Ad-hoc track develop XML-IR systems based upon this assumption, using an evaluation methodology in the tradition of Cranfield. However, since the inception of INEX, debate has raged on how applicable some of the Ad-hoc tasks are to real user tasks. The purpose of the User-Case Studies track from to explore the application of XML-IR systems from the users' perspective. This paper outlines QUT's involvement in this task. For our involvement we conducted a user experiment using an XML-IR system (GPX) and three interfaces: a standard keyword interface, a natural language interface (NLPX) and a query-by-template interface (Bricks). Following the experiment we interviewed the users about their experience and asked them - in comparison with a traditional XML-IR system - what type of tasks would they use an XML-IR system for, what extra information they would need to interact with an XML-IR system and how would they want to see XML-IR results presented. It is hoped that the outcomes of this study will bring us closer to understanding what users want from XML-IR systems.

1 Introduction

XML-IR systems differ from traditional IR systems by returning results to the user at the sub-document (that is element) level. The assumption is that XML-IR system will be able to better fulfil users' information needs since they only return the relevant parts of documents to users, rather than whole documents that will undoubtedly contain both relevant and irrelevant material. Most of the INEX tracks and tasks have been developed based upon this assumption, each which a slightly different user model in mind. INEX participating systems are evaluated in the Cranefield tradition involving sets of: source documents, end-user queries (topics), relevance judgements and metrics. Despite the progress made by INEX participants, debate has raised as to how applicable some of the tracks and task really are to potential end-users of XML-IR systems [3]. The aim of the User-Case is to examine this question and to investigate situations where XML-IR is suitable for end-users.

This paper details QUT's participation in the User-Case Studies track. Our participation stems from previous work in the Ad-hoc and NLP tracks. In previous years we

have developed a laboratory XML-IR system [1] and natural language interface [4] for participation in both of those tracks. This year for the first time we were able to test our systems with real users. Following the experiment we interviewed some of the participants asking them two sets of questions. The first set of questions focussed on their experiences using the natural language interface and an alternative template-by-query interface to formulate structured queries. The second set of the questions were more general, asking them how they felt about XML-IR overall and if there were situations where XML-IR would be more beneficial than traditional IR. The answers to the second set of questions forms the basis of this paper.

The rest of this paper is organised as follows. It begins with a description of the user experiment. Then it discusses the interviews with the participates following the experiment. Finally, it outlines how, based upon the information gathered from the interviews, ways that INEX can facilitated user-centred XML-IR tasks.

2 The Experiment

The experiment simulated the task of users interacting with an academic retrieval system. Sixteen participants took part in the experiment. The participants acted as academic researchers, for example: post-graduate research students, corporate researchers or academics. The participants searched a the INEX IEEE collection, a set of academic IEEE journal articles from 1995 to 2002. The journals had a broad range of focus, ranging from general journals such as Computing to specific journals such as Neutral Networks.

The participants were post-graduate information technology students who were uninitiated in the domain of XML-IR. While this may not a representative sample of possible XML-IR users, it was necessary to have such participants since understanding the technical nature of the information needs and source collection was beyond casual users. Also since the participants were uninitiated in the domain of XML-IR, it is valid for us for us to extrapolate the results of this experiment into the wider area of XML-IR. The participants were given six information needs that simulated those of a real user. The information needs contained both a detailed explanation of the information sought and a condition of relevance that described the motivation behind the information need. The information needs were sampled from the narrative elements of INEX Topics 253 – 284.

The system used in the experiment was separated into two parts: the front-end interfaces and the backend retrieval system. Two different interfaces were used: NLPX, that accepted queries written in natural language (English) [4], and Bricks, a query by template interface that allowed users to enter queries via a graphical user interface [5]. Examples of the input screen used for both interfaces appear in Figures 1 and 2. These examples capture the type of queries entered by the participants. The same backend search engine, GPX, was used for both interfaces. For each result retrieved by GPX, users were presented with the option of selecting to view the entire document or just the element. Since GPX only accepted formal language queries, both interfaces trans-

lated their user input into NEXI before submitting them to GPX. Below we describe NLPX, Bricks and GPX in more detail.

2.1 Interface A – NLPX

NLPX accepts natural language queries (NLQs) and produces formal queries written in the NEXI language. The NLPX translation process involves four steps. First, NLPX tags words either as special connotations (for instance structures) or by their part of speech. Second, NLPX divides sentences into atomic, non overlapping segments (called chunks) and then classifies them into grammatical classes. Third, NLPX matches the tagged NLQs to query templates that were derived from the inspection of previous INEX queries. Finally, NLPX outputs the query in NEXI format. Batch testing of a single backend search engine that used both natural language queries parsed through NLPX and formal NEXI queries has shown comparable results [4]. This is the first time that NLPX has been tested in a usability experiment.

2.2 Interface B – Bricks

Bricks is a query-by-template interface that allows users to input structured queries via a graphical user interface (GUI). Users enter their content needs via text boxes and their structural needs via drop-down boxes. To aid users, structural needs are indicated via conceptual rather than physical names, for example “a section” rather than sec. Bricks allows users to develop queries in several steps (”blocks”) starting with their desired unit of retrieval and then by adding any additional information needs. Blocks are also added as the user traverses the hierarchy of the documents. Upon completion of input, the data in the Bricks GUI is translated to formal NEXI expression, however, due to the constraints of the GUI, users are unable to enter malformed expressions. Usability testing has shown that users find Bricks superior to keyword only and NEXI interfaces [5].

Fig. 1. The NLPX search interface

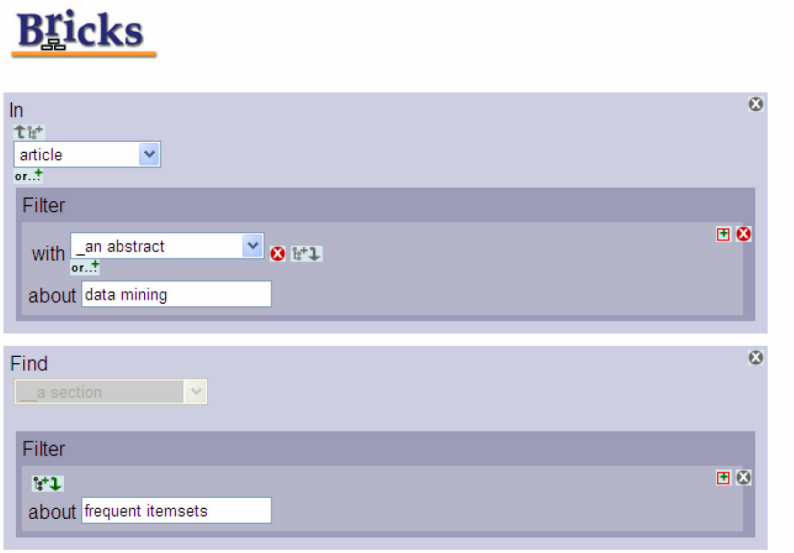

Fig. 2. The Bricks search interface

2.3 Backend Retrieval System – GPX

The backend retrieval system for this experiment was Gardens Point X (GPX) [1]. GPX was chosen since it has performed strongly at the annual INEX conference since 2002 - consistently among the top three systems. GPX stores the information about each leaf element in the collection as an inverted list. Upon retrieval, GPX matches query terms to all leaf elements that contain the term and then dynamically creates their ancestors. Elements are ranked according to their predicted relevance in GPX's ranking scheme. GPX rewards leaf elements that contain phrases and specific, rather than common, terms. It also rewards ancestors with multiple relevant children, rather than a single relevant child. For this experiment, the results list was filtered so that "overlapping elements" (that is, elements whose ancestors or descendants appear higher ranked on the results list) were removed before been being presented to users. This decision was made because users have been known to react negatively to overlapping elements [2].

3 Interviews

After the experiment 12 out of the 16 participants were interviewed. Some of the he questions asked were specifically about the experiment, in particular their experience using the query formulation interferences. A discussion on these questions is outside the scope of this paper. However, another set of questions were about their thoughts on the area of XML information retrieval as a whole in comparison traditional information retrieval. These questions, and some of their responses are presented here.

The main difference between XML-IR systems and traditional IR systems is that XML-IR systems returns elements rather than documents. We assume that since elements are more specific than documents that they will be more useful to users. However, this assertion has only been very limitedly tested with users, mainly in the con-

text of INEX's interactive track. Here we ask our participates in which situation would element-retrieval be more useful than document retrieval.

The first observation was that element retrieval would be more useful than document retrieval in situations where there was a lot of, largely irrelevant, information in the source documents. Or alternatively, situations where the user was searching for very precise or specific information. This is summarised in the responses made by participants 8,1 and 12 shown in Figures 3 – 5.

> **Participant 8**: "Technical forum, and you want to find a solution, and sometimes has hundred of pages, and each page has hundred of discussion, in that one it might help, just look at one of them and might help you to find the one you need."

Fig. 3. Participant 8's response regarding XML-IR uses

> **Participant 12**: "So any thing that is not free from, anything that has these logical sections would be beneficial. Specifically if you're just looking for, to just focus your search results on these specific categories, which obviously you can't do in an unstructured manner."

Fig. 4. Participant 1's response regarding XML-IR uses

> **Participant 1**: "You get much more precise searches by using markup. And all the information is all in the one place in the document, or rather gotten easily from the document, which can't be done with free text. If you take free text with no mark up and then you take text with all the author details, abstracts, bibliography all marked up you're going to be able to find stuff a lot quicker on the marked up one than the free text one. So XML would be a great improvement on free text."

Fig. 5. Participant 12's response regarding XML-IR uses

Others felt that the use of structure enabled them to write very detailed queries, particular those that may be about more than one topic. These types of queries are common in the CAS tasks, where a query may specify to a retrieve a particular about a topic, but also wishes the document to contain information about another topic. This opinion was expressed by participant 11.

Participant 11: “So yeah, I could say I wanted to particularly search for these keywords in the abstract of the paper or been able to express that kind of thing. Or saying that I needed a paper that was generally about a certain topic, but then I wanted a section in that paper that was about a particular kind of subtopic so I didn’t end up with these other papers that were about the right kind of general topic but not about the specific subtopic that I wanted.”

Fig. 6. Participant 11’s response regarding complex queries

An observation made by some participants was that the users of XML-IR system would need to know the structure of the document that they were searching, and possibly be domain experts. This was a point raised by participants 1 and 5.

Participant 1: “I'm guessing that every time you open up a new document, there's different ways of representing the structure, so I think that would make it quite difficult to use on a daily basis. If you were using the same file structure then Bricks would be great, but if you were using different structures or DTD then it would be really difficult to figure out how to use it.”

Fig. 7. Participant 1’s response regarding document knowledge

Participant 5: Its probably true that you need a bit of experience with the domain or at least in research to know where you have to look for a particular research type document. Similarly if your looking for a publicity or news type article you might want to have some idea how they're structured, and that’s obviously a bit of domain knowledge that you need to have, but once you've got it, it makes a lot to sense to use it because if I want the title to have something in it that I'm searching for then its good to be able to query that way.

Fig. 8. Participant 5’s response regarding document knowledge

Another point made by some participants was that element retrieval could be used in conjunction with document retrieval. Specifically, when documents are retrieved their most relevant elements could be highlighted. This opinion was expressed by participants 2 and 6 shown in Figures 9 and 10.

Participant 2: “Why not just group those different components together so that the user can have a choice. I see this document, and this document has 3 or 5 components relevant to the information need. And the other document has 2 parts [related to the] information needs.”

Fig. 9. Participant 2’s response regarding highlighting

Participant 6: "I mean its [element retrieval is] ok , as long as you can retrieve the actual whole document when you get the section back, if you could retrieve the whole document and see where it fits in that would be fine, and if you could retrieve the whole document and then see the section that you pulled up, that would be fine as well. But as long as you could see both the document and section that would be fine... so the whole document presentation with the section highlighted would work for me."

Fig. 10. Participant 6's response regarding document knowledge

The next section discusses how the outcomes of these interviews can be used to developed more user-orientated tasks at INEX.

4 Discussion

The results of the user experiment and interviews are of great value for the INEX community. Particularly pleasing was that the participants found merit in the use of XML-IR systems to fulfil their information need. The challenge for INEX participants and organisers is to use the information derived from these types of experiments to help focus our research efforts. An immediate way that we can put this into practise is by re-examining the tasks we preform each year, particularly in the Ad-hoc track, to see how well they correspond to tasks that users want. As a reference the following tasks are currently performed by INEX participants:

1. **Thorough Retrieval**: the aim of the thorough retrieval task is to retrieve all relevant elements matching an information request.
2. **Focussed Retrieval**: the aim of the focussed retrieval task is to retrieve the most relevant elements along an XPath matching an information request.
3. **All in Context**: the aim of the all in context retrieval task is to first, retrieve the most relevant documents matching an information request and second, to highlight all relevant information within those documents.
4. **Best in Context**: the aim of the best in context retrieval task is to first retrieve the most relevant documents matching an information request and second, to find the best entry point for those documents.

In this experiment results were returned as a single list ranked list of relevant elements. Overlapping elements were removed, and therefore, the presentation is analogous to the output of systems in focussed retrieval task. Users were, at least initially, confused by the presentation of results as a single list of ranked elements. However, this is not too surprising since the users experience with retrieval systems has been solely with document retrieval systems, and hence the idea of receiving back elements would have appeared "unnatural" to some. Users seemed to find the retrieval of elements with little or no context particular confusing. It is important to note that overlapping elements were removed from the presentation since previous experiments

have shown that users react adversely to them. If they were included in this experiment then the user reaction may have been even more negative. At first, this seems alarming for proponents of the Thorough and Focussed tasks since these tasks are based on returning lists of elements. However, even if the tasks are not suitable for end users they might still be worthwhile perusing since they could be used a precursor for other XML-IR tasks, such as Best or All in Context, or other information seeking tasks, such as question and answering.

During the post-experiment interviews it was discovered that users reacted positively to the idea of highlighting relevant elements within a document. This is a positive sign for INEX since it correlates well to the All in Context task. The users felt that highlighting passages would be beneficial to deciding if the document they are browsing is relevant. It would also help them when browsing large documents, particularly for documents that contain a lot of irrelevant information. This presents an interesting opportunity for INEX since it opens the possibility of having a document retrieval task at further workshops. This would allow participants to examine if the if techniques specifically designed for XML-IR are able to find more relevant documents (not elements) or even documents that are more relevant than traditional IR techniques. This task could be run in conjunction with one of the other document evaluation forums such as TREC or CLEF.

Expanding on this issue users also commented that they liked the idea of a best "entry point" into documents. Against this is pleasing for INEX since it directly correlates to the Best in Context task. Some users also commented that they would like to see the elements within the documents ranked according to relevance. This presents the opportunity to extend the All in Context task to measure the retrieved elements within each document as a ranked rather than unordered list.

We have already discussed how XML-IR could help users when they wish determine if their document is relevant. However, there are other information seeking tasks where XML-IR could be useful. One such task is when a single user query has multiple information requests. Often, in this scenario the user wish to retrieve a particular item for instance *sections about information retrieval* inside of articles that will have a second information item such as *paragraphs about compression* even if they don't wish to retrieve items matching the second request. This type of "complex" information request would be encapsulated in the NEXI expression **//article[about(//p,compression)]//sec[about(.,information retrieval)]** and is typical of one of the more complex CAS queries. This is a validation that this type of query is suitable for users, particularly when accessing documents that are about multiple topics, and that INEX should continue to use these types of queries in the future

A final comment made by interviews was the XML-IR system enabled them to find more specific results than traditional IR systems. INEX could capitalise on this situation in several ways. First, it strengthens the motivation for INEX's named entity task, since information need for that task is very specific, and inherently requires some sort of sub-document retrieval. Another interesting area of research that the INEX community would be to examine how users' information needs change as they interact with the retrieval system. One could assume that their information needs would start vague and then become more specific as they interact with the system. And as their needs become more specific one could assume that an XML-IR system would become more

useful than a traditional IR system. Some of the INEX tracks, particularly the interactive track, could examine if this is true.

5 Conclusion

This paper outlined QUT's involvement in this years User-Case Studies track. Out participation stems from our work in two of the other INEX tracks, namely, the Ad-hoc and NLP tracks. This paper detailed an experimentation we performed, and user interviews following the experiment. It then discussed, using information derived from the interviews, ways in which the INEX community can focuses on user-centred research.

6 Acknowledgements

The authors would like to acknowledge Utrecht University especially Jeroen Bass and Roelof van Zwol for the development of and permission to use Bricks, as well as the anonymous participants of the experiments, in particular those participants who were interviewed following the experiment. Without their contribution this work could not been conducted.

References

1. Geva, S.: GPX - Gardens Point XML Information Retrieval INEX 2004. In: Fuhr, N., Lalmas, M., Malik, S., Szlavik Z. (eds.): Advances in XML Information Retrieval: Third International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2004, Dagstuhl, Germany, December 6–8, 2004, Revised Selected Papers. Lecture Nodes in Computer Science, Vol 3493. Springer-Verlag, Berlin Heidelberg New York (2005) 221–222
2. Pehcevski, J., Thom, J. A., Tahaghoghi, S. M. M., Vercoustre, A-M.: Hybrid XML Retrieval Revisited, In: Fuhr, N., Lalmas, M., Malik, S., Szlavik Z. (eds.): Advances in XML Information Retrieval: Fourth International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2005, Dagstuhl, Germany, November 28 – 30, 2005, Revised Selected Papers. Lecture Nodes in Computer Science, Vol 3977. Springer-Verlag, Berlin Heidelberg New York (2005), 153 - 167
3. Trotman, A.: Wanted: Element Retrieval Users. In Proceedings of the INEX 2005 Workshop on Element Retrieval Methodology, Glasgow, Scotland, 30 July (2005) 58 -64
4. Woodley, A., Geva, S.: NLPX at INEX 2005, In: Fuhr, N., Lalmas, M., Malik, S., Szlavik Z. (eds.): Advances in XML Information Retrieval: Fourth International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2005, Dagstuhl, Germany, November 28 – 30, 2005, Revised Selected Papers. Lecture Nodes in Computer Science, Vol 3977. Springer-Verlag, Berlin Heidelberg New York (2005) 358–372

5. van Zowl, R., Bass, J., van Oostendorp, H., Wiering, F.: Query Formulation for XML Retrieval with Bricks. In Fuhr, N., Lamas, M., Trotman, A. (eds.): In Proceedings of INEX 2005 Workshop on Element Retrieval Methodology, Glasgow, Scotland , 30 July (2005) 75-83