Why Structural Hints in Queries do not Help XML-Retrieval

Andrew Trotman
Department of Computer Science
University of Otago
Dunedin, New Zealand

andrew@cs.otago.ac.nz

Mounia Lalmas
Department of Computer Science
Queen Mary University of London
London, UK

mounia@dcs.qmul.ac.uk

ABSTRACT
For many years it has been commonly held that a user who adds structural "hints" to a query will improve precision in an element retrieval search. At INEX 2005 we conducted an experiment to test this assumption. We present the unexpected result that structural hints in queries do not improve precision. An analysis of the topics and the judgments suggests that this is because users are particularly bad at giving structural hints.

Categories and Subject Descriptors
H.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *query formulation, Search process.*

General Terms
Measurement, Performance, Experimentation.

Keywords
Element retrieval, XML retrieval, INEX.

1. INTRODUCTION
When explicit structure is present in a document, it is possible to ask retrieval questions of not only whole documents, but of those structures. If these structures are marked up as XML elements then a search engine that returns these structures is an element retrieval system.

We envisage the user of such a system entering a query that does not contain structure (a Content Only or CO query) such as `code signing verification`. This user, on discovering either too many results, results that are too large (perhaps a book), or too small (perhaps a single paragraph) adds a structural hint to the query in the expectation that this will increase precision. This added structure (+S) query (written in NEXI [6]) might read `//sec[about(., code signing verification)]` in which the user is targeting <sec> elements, (sections of documents). Of course, this structure is only a hint as the search engine, in fulfilling the information need, is not *required* to follow it.

It is generally assumed that a user who explicitly includes structure in their query will improve precision; indeed increased precision has been observed (e.g. [4]). The increase of expressive power should lead to better performance.

At INEX 2005 [2] we ran an experiment to measure the size of the improvement. We asked participants (our users) to provide CO queries, then to add structural constraints (+S) should the result of the CO query be inadequate. Unlike previous studies that *post-facto* remove structure [3], this experiment is representative of a user search session in which a user is unhappy with the initial result set so adds structural hints.

Our analysis shows that at best there is no significant difference between the best CO and CO+S runs. This is because users find it difficult identifying which structures in the collection contain relevant information. Additionally, these structures appear fixed for the given document collection.

2. Element Retrieval
Results under generalized quantization for Thorough and Focused retrieval were examined and MAep was chosen as the performance measure – this is the standard evaluation methodology used in element retrieval (see the appendix of the INEX 2005 proceedings for details [1]). Version 1.8 of the INEX IEEE collection consisting of 16,819 documents totaling 705MB was used. The DTD contains 194 elements. The 19 INEX 2005 topics containing CO and +S queries were used. Runs consisted of a ranked list of at most 1500 elements for each topic. 10 participants submitted runs for the Focused task and 14 for the Thorough task.

3. Results
For each of Focused and Thorough retrieval, Queensland University of Technology (QUT) produced the highest scoring CO run. QUT also submitted the best Focused CO+S run, while the University of Kaiserslautern the best Thorough CO+S run. The performance of those runs is shown in Table 1 where it can be seen that improvements are seen when structure is added to the query. Following the recommendation of Sanderson and Zobel [5], t-tests were performed on the results. None were significant even at the 5% level. That is, when structure was added to the query no significant improvement in MAep was seen.

We compared the results of the best CO to the CO+S run on a participant by participant basis. In Focused retrieval exactly half the participants showed an increase in precision. In Thorough retrieval 9 of the 14 showed improvements.

Using their own search engine Kamps *et al*. [3] compared the performance of queries containing structure to those with it subsequently removed. Using INEX 2004 topics under Thorough retrieval and strict quantization they observe a decrease in mean precision, but an increase at early recall levels. The 2005 University of Amsterdam runs do not exhibit this behavior under generalized quantization, perhaps due to the new assessment techniques

now used at INEX. None the less, we compared the performance of highest scoring runs at 1% and at 10% recall. No significant difference between CO and CO+S runs was seen in either Thorough or Focused retrieval.

Table 1: MAep scores for each of the 4 tasks

	Task	MAep
Focused	CO	0.10769
	CO+S	0.11859
	Improvement	10.12%
Thorough	CO	0.08029
	CO+S	0.08651
	Improvement	7.76%

4. Discussion

Adding structure to the query has not, on this query set, proven to increase precision. As it is reasonable to assume it should, the query set warrants further investigation.

The first row of Table 2 shows, for the query set, the elements the user specified as the preferred result element, and the number of queries with this target element (the other category consists of "*" two times, "bdy//*" and "(abs|sec)" one time each). It appears as though the query authors are not very creative in their choice of target elements even though the DTD was provided and, as INEX participants, they should have bee familiar with it. Because of the DTD, specifying <article> as a target element is tantamount to document retrieval. Specifying "*" or "bdy//*" is tantamount to adding no structural hint. None the less, the structural hints, if accurate, should still increase precision. This brings into doubt their accuracy.

Table 2: Target and Relevant Elements

Target	article (6)	sec (9)	other (4)
Actual	sec (3), p (3)	p (5), sec (3), ref (1)	p (4)

The second row of Table 2 shows the most frequent element judged as satisfying the information need for each target element. When the participant specified <article> as the target element (6 times), the assessments showed that for 3 queries <sec> was most highly represented in the relevant element set and for the other 3 queries <p> was most highly represented. The target element is not, in general the most common relevant element, regardless of the target. This suggests that the user finds it difficult to correctly specify the target elements.

It appears as though the same elements are relevant (<p> and <sec>) regardless of the target. In Table 3 all tags appearing in more than 5% of the relevant elements are listed in decreasing order of appearance. They lay in the same order regardless of the target element (<sec> as a Focused result to an <article> target is an exception; <tf> (a TeX equation) is anomalous). This suggests that the elements containing relevant information are a function of the document collection and not of the query.

Topic authors were asked to submit example relevant elements along with the topic at the topic creation time. Those elements appearing more than 5% of the time for each target element are given in Table 4. Again it appears as though regardless of the

specified target element, the relevant elements are the same. Our users think they want sections (<sec>), but in practice want paragraphs (<p>). This may be because they are looking for sequences of paragraphs, or passages, rather than elements. If this is the case then passage retrieval may be a better way to search the collection.

5. Conclusion

Within the context of INEX the assumption that adding structure to a query improves precision was investigated. Although improvements in some cases are seen, they are not significant. This is shown to be because the users are very bad at giving structural hints. The best structural elements appear to be a function of the document collection and not the query. It is not clear if this is true of only the given collection or all collections. Further investigation is required.

6. Acknowledgements

Table 3: Relevant Elements and Proportion of Occurrence

Tag	article		sec		other	
	Thoro %	Focus %	Thoro %	Focus %	Thoro %	Focus %
p	33.75	17.31	25.27	35.76	40.01	26.34
sec	28.69	43.59	23.99	24.09	14.51	21.48
tf					12.28	
article	12.30	15.38	10.50	9.79	5.48	8.95
bdy	11.05	15.06	9.28	9.28	5.07	8.18
b			6.29			

Table 4: Elements Identified Relevant with Topic

article	%	sec	%	other	%
sec	34.04	sec	65.85	sec	66.67
article	31.91	ss1	14.63	article	22.92
ss1	12.77	article	7.32	ss1	8.33
bdy	10.64	bdy	7.32		

7. References

[1] Fuhr, N., Lalmas, M., Malik, S., & Kazai, G. (2005). Appendix, *INEX 2005 workshop pre-proceedings*.

[2] INEX. (2005). Initiative for the evaluation of XML retrieval. Available: http://inex.is.informatik.uni-duisburg.de/2005/ [2005, 12 January].

[3] Kamps, J., Marx, M., Rijke, M. d., & Sigurbjörnsson, B. (2005). Structured queries in XML retrieval. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, (pp. 4-11).

[4] Pehcevski, J., Thom, J., & Tahaghoghi, S. M. M. (2005). RMIT university at INEX 2005. In *Proceedings of the INEX 2005*, (pp. 217-233).

[5] Sanderson, M., & Zobel, J. (2005). Information retrieval system evaluation: Effort, sensitivity, and reliability. In *Proceedings of the 28th ACM SIGIR Conference on Information Retrieval*, (pp. 162-169).

[6] Trotman, A., & Sigurbjörnsson, B. (2004). Narrowed Extended XPath I (NEXI). In *Proceedings of the INEX 2004 Workshop*, (pp. 16-40).