CISR at INEX 2006

Wei Lu[1], Stephen Robertson[2,3], Andrew Macfarlane[3]

[1] Center for Studies of Information Resources, School of Information Management
Wuhan University, China and City University
sa713@soi.city.ac.uk
[2] Microsoft Research, Cambridge, U.K
ser@microsoft.com
[3] Centre for Interactive Systems Research, Department of Information Science
City University London
andym@soi.city.ac.uk

Abstract. In this paper, we describe the Centre for Interactive Systems Research's participation in the INEX 2006 adhoc track. Rather than using a field-weighted BM25 model in INEX 2005, we revert back to using the traditional BM25 weighting function. Our main research aims in this year are to investigate the effects of document filtering by result record cut-off, element filtering by length cut-off and the effect of using phrases. The initial results show the latter two methods did not do well, while the first one did well on FOCUSED TASK and RELEVANT IN CONTEXT TASK. Finally, we propose a novel method for BEST IN CONTEXT TASK, and present our initial results.

1. Introduction

This is the second year that the CISR has participated in INEX. In INEX 2005, we used a field-weighted BM25 model and submitted runs for two adhoc CO tasks [1]. Our results suggest that the method is promising. Subsequent to this, we investigated XML retrievable units and element inheritance in [2] and the average element length in [3]. This year, rather than further exploiting the field-weighted method, our work focused on investigating the effects of document filtering, element filtering and using phrases.

In traditional text retrieval systems, a document is usually treated as an independent unit. But for XML element retrieval, elements in the same document are usually semantic relevant and are not independent units themselves, e.g., article title, abstract, and section title to section text in IEEE's data collection. This raises an issue of how context elements impact on the effectiveness of XML element retrieval e.g. the impact a parent element has on a child. Some work has been done in this area. Lu et al [1] and Robertson et al [2] used a field-weighted method to exploit the inheritance from context elements; Abolhassani et al [4], Geva et al [5] and Ogilive et al [6] used two different methods to compute the parent element's weight by merging its sub-elements weight. Both of these methods consider the element weight inheritance from context elements but without evidence from the whole document. Sigurbjornsson et al

[7] and Mass et al [8] investigated document weight's contribution to element retrieval by using an interpolation method of merging the document weight into element weight. The results show this method is beneficial and has yielded good results at INEX 2004 and INEX 2005.

In this paper, we use a different approach to element retrieval. That is, we divided element retrieval into two phases: we conducted document level retrieval and set a cut-off for the retrieved results and then used the filtered results to further execute element level XML retrieval. Our aim is to investigate whether using top weighted documents could produce good results. Because of time limitations and issues with the newly adopted test collection, we did not compare the two methods directly in our experiments this year.

In order to avoid the too-small element problem, we restrict our set of retrievable units to article, body, section and paragraph, and set a cut-off for element length to abandon those elements which are shorter than the cut-off value. We also use phrases instead of single words to see if it could improve retrieval effectiveness.

In section 2, we describe the BM25 model used in our experiment. Section 3 presents our results in INEX 2006. In Section 4, we evaluate and compare our results. We conclude the paper with closing remarks and future research directions to extend our work..

2. BM25 model

In this work we use the original BM25 model. This is in contrast to our previous work in the area at INEX 2005 [1]. We reverted back to the BM25 model so that we could use it in the first phase of the method we deployed in our INEX 2006 experiments. For adhoc retrieval, and ignoring any repetition of terms in the query, BM25 can be simplified to:

$$wf_j(d,C) = \frac{(k_1+1)tf_j}{k_1((1-b)+b\frac{dl}{avdl})+tf_j}\log\frac{N-df_j+0.5}{df_j+0.5} \qquad (1)$$

where C denotes the document collection, tf_j is the term frequency of the *j*th term in document *d*, df_j is the document frequency of term *j*, *dl* is the document length, *avdl* is the average document length across the collection, N is the total number of documents in the collection and k_1 and *b* are tuning parameters.

Formula (1) uses a logarithmic function to compute term's collection weight. For frequently occurring terms, this function will produce negative weight values. To avoid this, we used an alternative weight function (command "w fn=3" in Okapi bss system) instead of the logarithmic function.

3. Description of the Experiments

Within the adhoc XML retrieval task there are four sub-tasks: BEST IN CONTEXT TASK, THOROUGH TASK, FOCUSED TASK and RELEVANT IN CONTEXT TASK . For each sub-task, we submitted 3 or 4 runs only for CO queries but not for CAS queries. Our purpose is twofold: to meet the experimental criteria of INEX 2006 and to test our stage approach taken this year. The details of these experiments are as follows:

3.1 BEST IN CONTEXT TASK

BEST IN CONTEXT TASK is a new adhoc task which aims at locating the best entry point of XML retrieval. We used two methods for this task and submitted four runs. In the first method, we take the element with the highest weight score (best-match element) in each document as the best entry point. This method was used in the two submitted runs BEST-BM25-cutoff400 and BEST-BM25-filter1500.

In the second method, we propose a novel way of selecting the best entry point. The distribution of element weight scores in the document is considered. Our basic idea is that, given an element, if more than one of its sub-elements has a good score, then this element should be chosen as the candidate best entry point rather than using its sub-element as the candidate best entry point. A problem for this particular method is how to determine a good score. In implementation, we set half the score of the best-match element in each document as the cut-off for determining a good score and use a bottom up method for selecting the best entry point. For each document, we find the best-match element in the document. Then we consider all other elements which do not overlap with this one. If any of these elements scores higher than half the score of the best-match element, then it should be included in the scope implied by the entry point. That is, we move the entry point up to the start of a higher-level element, such that the higher-level element includes all the high-scoring elements.

For example, in Fig. 1, the best-match element is E and the best-match element weight score is 2.11, so the cut-off value is 1.055. Using this method, we get the best entry point B.

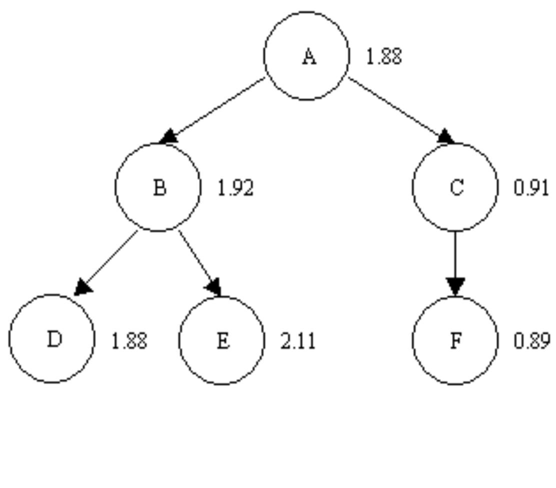

Fig.1 XML element tree with element weight score

In INEX 2006, the two submitted runs BEST-BM25-400-1500-level-p and BEST-BM25-Level-filter1500 used this method. For the initial evaluation of the effect of element length cut-off, we tuned element length cut-off between 0 to 550 (character length) on the INEX 2005's data collection by using metrics Manx(10), Manx(25), Manx(50). The effect of the cut-off on "gen" Manx measure is shown in Fig.2. From this figure, we can see that using element length cut-off is beneficial especially to Manx(10) and Manx(25). We found an increase of 8% for the best tuned Manx(10) value over the non-tuned value.

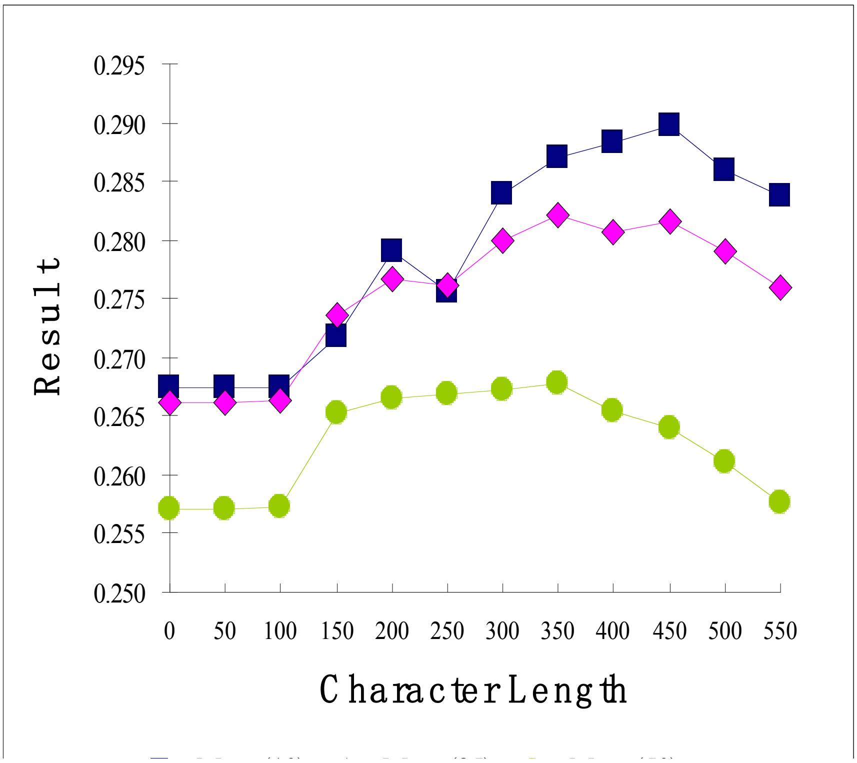

Fig. 2 Tuning results of element length cut-off on INEX 2005's data collection

3.2 THOROUGH TASK

We submitted 3 runs for THOROUGH TASK. They are THOR-BM25-nobody, THOR-BM25-nobody-cutoff400 and THOR-BM25-400-1500-phrase:

- THOR-BM25-nobody directly uses BM25 to compute the element weight score;

- THOR-BM25-nobody-cutoff400 is much the same as THOR-BM25-nobody except the element length cut-off, which filters out elements shorter than a fixed length, is set to 400;
- THOR-BM25-400-1500-phrase uses the same element length cut-off, and it also set document result cut-off (1500) and uses phrases instead of single words.

3.3 FOCUSED TASK

The 3 submitted runs for FOCUSED TASK are as follows:

- FOCU-BM25-cutoff400 uses 400 characters length as element length cut-off;
- FOCU-BM25-cutoff400-filter1500 uses the same element length cut-off and also uses document result cut-off (1500);
- FOCU-BM25-cutoff400-filter1500-phrase is similar to FOCU-BM25-cutoff400-filter1500 (above) except it uses phrases instead of single terms

3.4 RELEVANT IN CONTEXT TASK

For this task, we submitted runs All-BM25-cutoff400, All-BM25-cutoff400-filter1500 and All-BM25-cutoff400-filter1500-phrase. These runs use the same conditions as the ones for FOCUSED TASK. The difference is that the results in the runs are grouped by articles and without overlap elements.

4. Evaluation

The evaluation results of our runs are shown in Tables 1 to 4. In Table 1, the run using the basic BM25 model was the best. In Tables 2 and 3, the runs using element length cut-off do best, which also rank them at the top of all INEX 2006 s corresponding submitted runs. In contrast, the runs using document filtering and phrases did not do well. Table 4 shows the results of our runs for locating document's best entry point. Although the run BEST-BM25-cutoff400 does best among the 4 runs, it can be seen that the two runs BEST-BM25-400-1500-level-p and BEST-BM25-Level-filter1500 using the derived novel method do better than the run BEST-BM25-filter1500. These three runs use the same document result set for locating the best entry point. Further investigation into why this happens is merited.

Table 1: Results for THOROUGH Task

Runs	**Metric: ep/gr**	
	un-filtered	**filtered**
THOR-BM25-nobody	**0.0228**	**0.0431**

THOR-BM25-nobody-cutoff400	0.0215	0.0407
THOR-BM25-400-1500-phrase	0.0118	0.0217

Table 2: Results for FOCUSED Task (Using un-filtered assessments)

Runs	**Metric: nxCG (Overlap=on)**			
	5	**10**	**25**	**50**
FOCU-BM25-cutoff400	**0.3961**	**0.3428**	**0.2638**	**0.2001**
FOCU-BM25-cutoff400-filter1500	0.3054	0.2557	0.1873	0.1335
FOCU-BM25-cutoff400-filter1500-phrase	0.2849	0.2452	0.1836	0.1332

Table 3: Results for RELEVANT IN CONTEXT Task

Runs	**MAgP**	**Metric: gP**			
		5	**10**	**25**	**50**
All-BM25-cutoff400	**0.1161**	**0.2936**	**0.2456**	**0.1622**	**0.1109**
All-BM25-cutoff400-filter1500	0.0583	0.2227	0.1725	0.1147	0.0741
All-BM25-cutoff400-filter1500-phrase	0.0602	0.2298	0.1734	0.1155	0.0746

5 Conclusion

Rather than using field-weighted BM25 model in INEX 2005, we reverted back to using the basic BM25 model. We exploited the effects of element filtering by length cut-off, document filtering by result record cut-off and the effects of using phrases. The results show that the latter two methods did not do well, while the first one did very well on FOCUSED TASK and RELEVANT IN CONTEXT TASK. Finally, in the THOROUGH TASK, the results were inconclusive as to whether or not the method was effective. We also utilized a novel method for BEST IN CONTEXT TASK. However we did not consider the number of sub-elements and the adjacency of the relevant elements. These issues need to be investigated further. Given more time and resources, it would be useful to undertake a full scale study comparing the field weighing element retrieval used in last years INEX and the two stage method utilized for our experiments this year. In this context we could investigate the issue of sub-element cardinality and adjacency of element relevant to the information need.

Table 4: Results for BEST IN CONTEXT Task (Using un-filtered assessments)

Runs	Metric: BEPD				
	A=0.01	A=0.1	A=1	A=10	A=100
	Metric: EPRUM-BEP-Exh-BEPDistance				
	A=0.01	A=0.1	A=1	A=10	A=100
BEST-BM25-cutoff400	**0.0860**	**0.1311**	**0.1984**	**0.3175**	**0.4532**
	0.0221	**0.0435**	**0.0760**	**0.1431**	**0.2349**
BEST-BM25-filter1500	0.0655	0.1071	0.1706	0.2621	0.3441
	0.0139	0.0311	0.0547	0.0956	0.1384
BEST-BM25-400-1500-level-p	0.0664	0.1071	0.1710	0.2626	0.3429
	0.0147	0.0319	0.0567	0.0989	0.1420
BEST-BM25-Level-filter1500	0.0610	0.1087	0.1749	0.2632	0.3413
	0.0153	0.0367	0.0629	0.1014	0.1395

Acknowledgements

This work is supported in part by National Social Science Foundation of China 06CTQ006.

References

[1] W. Lu, S. Robertson, A. Macfarlane. Field-Weighted XML Retrieval Based on BM25. Proceedings of INEX 2005. LNCS. 2006 126-137

[2] S. Robertson, W. Lu, A. Macfarlane. XML-structured documents: retrievable units and inheritance. FQAS 2006. Springer LNCS. 2006 121-132

[3] W. Lu, S. Robertson, A. Macfarlane. Investigating Average Element Length for XML Retrieval by Using BM25. (submitted to review)

[4] M. Abolhassani, N. Fuhr, S. Malik. HyREX at INEX 2003. Proceedings of the Second Workshop of the Initiative for The Evaluation of XML Retrieval (INEX), 15-17 December 2003, Schloss Dagstuhl, Germany.

[5] S. Geva. GPX - Gardens Point XML Information Retrieval at INEX 2004. INEX 2004, Lecture Notes in Computer Science, Springer-Verlag GmbH Volume 3493 / 2005.

[6] P. Ogilvie, J. Callan. Hierarchical Language Models for XML Component Retrieval. INEX 2004, Lecture Notes in Computer Science, Springer-Verlag GmbH Volume 3493 / 2005.

[7] B. Sigurbjornsson, J. Kamps, M. Rijke, An element based approach to XML Retrieval, Proceedings of the Second Workshop of the Initiative for The Evaluation of XML Retrieval (INEX), 15-17 December 2003, Schloss Dagstuhl, Germany.

[8] Y. Mass, M. Mandelbrod, Component Ranking and Automatic Query Refinement for XML Retrieval, INEX 2004, Lecture Notes in Computer Science, Springer-Verlag GmbH Volume 3493 / 2005.