RMIT at TREC 2011 Session Track

Lorena Leal, Sadegh Kharazmi, Jasbir Dhaliwal,
Mark Sanderson, Falk Scholer, Sargol Sadeghi,
Fahad Alahmari

School of Computer Science and IT
RMIT University, GPO Box 2476
Melbourne 3001, Australia

1 Introduction

The 2011 Session track aims to study retrieval system performance by providing different components in a search session. We report on experiments and results based on query expansion techniques when lists of results are provided with or without clicked information. In contrast, a bag-of-words approach is employed as a baseline.

2 Experimental Setup

2.1 Indexing and Retrieval

We used Indri 5.1 for indexing and retrieving documents from the Clueweb Category B collection. Stopwords were removed and the Krovetz stemmer was used to pre-process the documents. The first 1000 results per query were retrieved using the Okapi BM25 similarity function without any expansion or relevance feedback setting.

2.2 Description of Runs

In all runs, the submitted query received the same pre-processing as documents in the collection.

The first task of the Session Track aims at evaluating system retrieval performance by using only the terms in a given query. In this run, the current query is supplied to the retrieval system in isolation due to the absence of

information about past interactions. The corresponding runs for the task are labelled as RMIT1 RL1, RMIT2 RL1 and RMIT3 RL1. In other words, all these RL1 submissions are standard bag-of-words queries.

In the second task (RL2), additional session information in the form of previously entered queries is made available. We refined the current query by concatenating query terms from other interactions in the session (RMIT1 RL2 and RMIT3 RL2). Another approach we tried was to use only the query words from all interactions (RMIT2 RL2). In cases where a term occurs multiple times across a session, it is only added to the query once, and no special weight is assigned.

For the third task (RL3), the ranked results lists from previous queries in the same search session are available. We employed a query expansion approach for the third task of the track (RMIT1 RL3, RMIT2 RL3 and RMIT3 RL3). Specifically, the snippet component of the top 5 results was taken to create a pseudo-relevance feedback setting [1]. Given that a session generally contained more than one interaction, the number of snippets for constructing the expanded query varies depending on the number of interactions. Stopwords were excluded from snippets, and then we proceed to apply Rocchio's approach [2] for selecting the 10 leading terms. It should be noted that the refined query does not contain terms from the current query or other interactions. This is achieved by modifying the influence of the original query in Rocchio's formulation. Other elements form the results such as titles and URLs were not employed. For these runs an unplanned bug was introduced as only expanded terms were supplied to the retrieval system, and no consideration was given to the current query.

In the last task of the track (RL4), in addition to the ranked results click information for prior searches in the session is provided. Clicked results within interactions were used as a source for relevance feedback in order to expand the current query. We followed two approaches for the relevance feedback. The first was constructed based on the document content (RMIT1 RL4); while the second only uses the snippet component (RMIT2 RL4). Where no information of clicked results was available in the interactions, we conducted a standard pseudo-relevance feedback by using the top 2 ranked results.

In order to exploit clickthrough data further, we assumed that positive feedback is provided by clicked results. Negative feedback, in contrast, can be taken from non-clicked data that surrounds a selected result. In the run RMIT3 RL4, the information in the snippet component from all interactions is used for applying Rocchio's formulation and retrieving the top 10 expansion terms.

For all runs in the fourth task, the pre-processing of documents and snip-

Table 1: Median nDCG@10 scores

Run	All participants			RMIT runs	
	allsubtopics	*lastquery*		*allsubtopics*	*lastquery*
RL1	0.3055	0.2207	RMIT1	0.3083	0.1939
RL2	0.3105	0.1922	RMIT1	0.2869	0.1420
			RMIT2	0.2878	0.1314
RL3	0.3084	0.1858	RMIT1	0.2982	0.1841
RL4	0.3262	0.1972	RMIT1	0.3506	0.1831
			RMIT2	0.3453	0.1715
			RMIT3	0.3565	0.1682

pets consisted of removal of stopwords. However stemming was not applied. The new expanded terms are not part of the current query or from other interactions. The same bug as in the third task occurred here, in that no original query terms were retained. Given the detected bug, we re-ran the experiments and report the correct results in the following section.

3 Results

In the Session track two approaches were followed to evaluate results. The first is based on whether the document is relevant for the current query or any subtopic of the query (*allsubtopics*). The second evaluates the document exclusively according to the subtopic that the current query is related to (*lastquery*).

For comparison of results we use nDCG@10 scores across multiple runs. Table 1 details median nDCG@10 scores for all partipants in order to compare to our results. In the *allsubtopics* evaluation approach, we obtained a slight improvement over the median when relevance feedback is used to expand the query, showing that session information can contribute to retrieval effectiveness (see Figure 1). However, these results are not statistically significant. The same technique did not lead to any improvement using the last query evaluation approach.

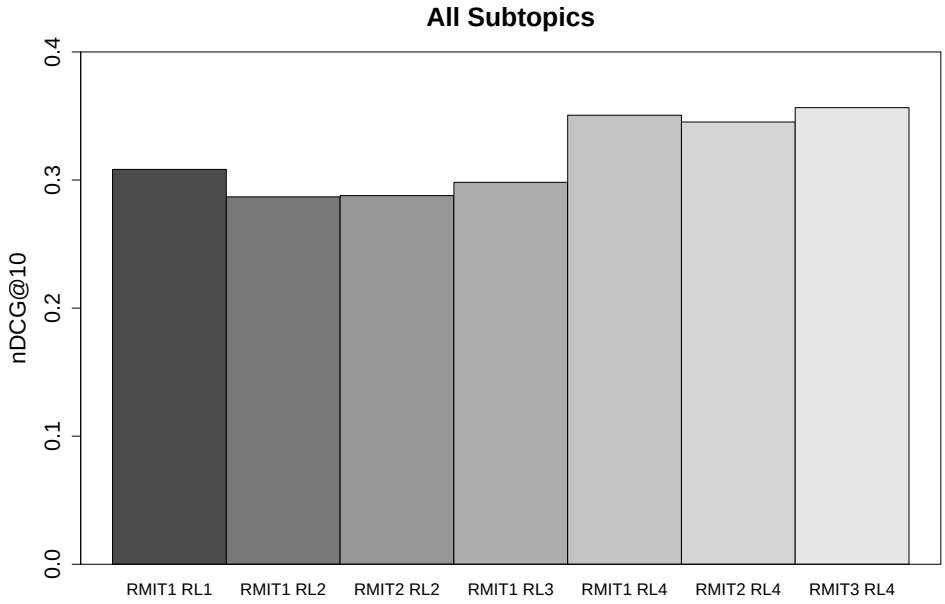
4 Future Work

Intuitively, one would hope that session-based informaton could improve the performance in retrieval systems. We plan to conduct a more through failure analysis to try and understand why our approaches, particularly in the RL2 and RL3 tasks, harmed retrieval performance.

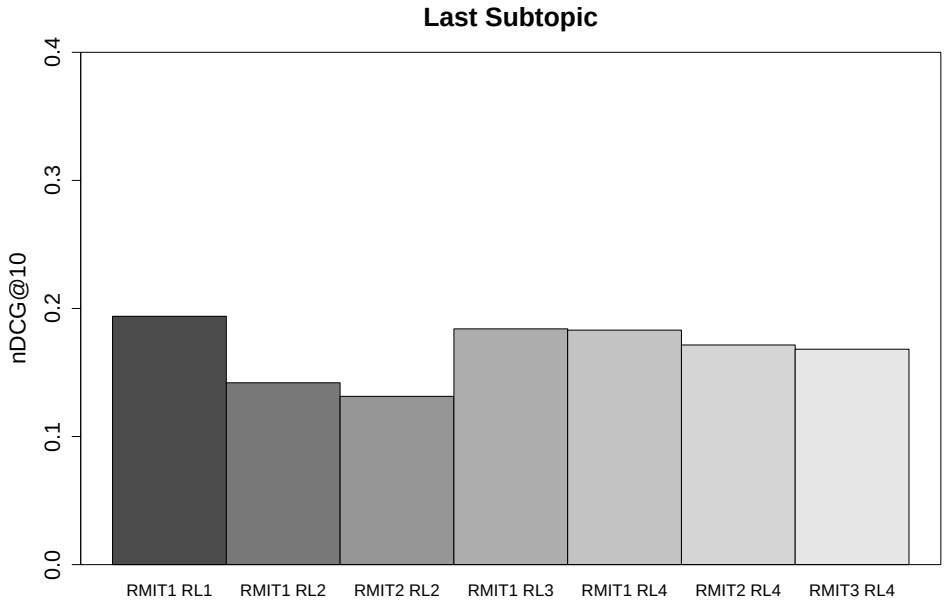
References

[1] A. M. Lam-Adesina and G. J. F. Jones. Applying summarization techniques for term selection in relevance feedback. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 1–9, New York, NY, USA, 2001. ACM.

[2] J. J. Rocchio. Relevance feedback in information retrieval. In *The SMART Retrieval System –Experiments in Automatic Document Processing*, pages 313–323. Prentice Hall Inc., Englewood Cliffs, NJ, 1971.



(a)



(b)

Figure 1: Median nDCG@10 scores for *allsubtopics* and *lastquery* evaluation approaches.