# RMIT University at TREC 2009: Web Track

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

## 1 Introduction

RMIT participated in the 2009 Web Track tasks. Our submissions utilised the Zettair search engine[1] to index and search the Category B subset of the ClueWeb collection used by the Web Track.

The Web Track was composed of two tasks, a traditional adhoc retrieval task, and a new diversity task where participants attempted to retrieve documents covering a range of sub topics for each query. Sub topics were not provided with the queries.

Our experiments utilised the well known measures Okapi BM25 and language modeling with Dirichlet smoothing for the adhoc task. For the diversity task we attempted to improve the diversity of query results by minimising the number of documents returned for a single domain.

## 2 Description of Runs

Runs were generated using a customised version of the Zettair search engine which was adapted to deal with the large scale ClueWeb collection. All runs used the default *light* stemming option in Zettair that removes the suffixes: *-e -es -s -ed -ing -ly -ingly*, and replaces the suffixes *-ies*, *-ied* with *-y*. No stopping was used.

### 2.1 Adhoc Task

For the adhoc task two runs where submitted:

- *RmitOkapi*: The top 1,000 retrieved documents ranked by the Okapi BM25 similarity measure (Sparck Jones et al., 2000). Parameters were left at the default Zettair values with *K1*, *K3*, and *B* set to 1.2, $\infty$, and 0.75 respectively.

- *RmitLm:* The top 1,000 retrieved documents ranked by a Dirichlet smoothed language modeling measure (Ponte and Croft, 1998, Zhai and Lafferty, 2004), with the $\mu$ parameter set to 1,500.

[1] Zettair is available under a BSD License from: `http://www.seg.rmit.edu.au/zettair`

Table 1: Stat mean average precision and stat mean nDCG for adhoc runs.

| Run Label | statMAP | statMnDCG |
|---|---|---|
| RmitOkapi | 0.1558 | 0.3222 |
| RmitLm | 0.1686 | 0.3113 |
| Mean per topic median | 0.1539 | 0.2956 |
| Mean per topic best | 0.4304 | 0.6091 |

Table 2: Alpha nDCG and intent aware precision scores for submitted and baseline diversity run.

| | alpha-ndcg | | | IAP | | |
|---|---|---|---|---|---|---|
| Run | @5 | @10 | @20 | @5 | @10 | @20 |
| RmitLm | 0.103 | 0.147 | 0.188 | 0.055 | 0.075 | 0.084 |
| RmitDiv | 0.097 | 0.157 | 0.193 | 0.046 | 0.070 | 0.072 |

### 2.2 Diversity Task

For the diversity task, one run was submitted (*RmitDiv*) where the top 1,000 retrieved documents ranked by a Dirichlet smoothed language model similarity measure were returned after filtering the results such that the domain name of a document appeared no more than once in a document list. Where a document from the same Internet domain appeared more than once, the highest ranked result was retained in the list, and all other documents were removed.

The rationale behind this approach is that documents returned from the same domain have a higher likelihood of discussing the same sub topics, as opposed to documents of differing domains that may cover different sub topics of the query.

## 3 Results

Global measures are not yet available for the Web Track limiting the analysis of our results. Our system performance in comparison to the median and best results for individual topics is discussed below.

### 3.1 Adhoc Task Results

For the adhoc task, the stat mean average precision (StatMAP) and stat mean normalised discounted cumulative gain (StatMnDCG) results for each of the submitted runs are presented in Table 1. The two similarity measures resulted in runs of similar accuracy.

For each topic, the best, median, and worst values of stat AP and stat nDCG over all submitted runs were available. By averaging these results, the mean median and mean best values were calculated and presented in Table 1. As expected, having submitted baseline system runs, the overall accuracy of our submissions is close to that of the median per topic average score, and significantly lower than the best score average.

### 3.2 Diversity Task Results

A single run was submitted for the diversity task. Table 2 shows the mean scores for alpha-nDCG and intent aware precision for our *RmitDiv* submission, as well as the scores for a baseline run *RmitLm* on which the submitted run was based.

The table shows mixed results with a minor improvement in accuracy measured only with alpha-nDCG@10 and alpha-nDCG@20. For the primary measure alpha-nDCG@10, our approach of eliminating duplicate domains from the result lists resulted in an improvement for 14 of the 50 query topics, while a decrease in accuracy was observed for 8 of the query topics, with the remaining 28 unaffected.

Overall the benefits of the naive approach taken is questionable given the mixed results over the varying measures. We hypothesis that in combination with other techniques such as document clustering such an approach would prove more beneficial.

## 4 Conclusions

RMIT submitted several runs to the various Web Track tasks. The two "out of the box" Zettair adhoc run submissions achieved median like accuracy which is to be expected for a simple term matching based approach. This result forms a reasonable baseline for more advanced techniques in future submissions using the newer large collection introduced this year.

The unique domain name approach taken for the diversity task produced minimal changes to the accuracy of the run when compared to the baseline run.

## References

Ponte, J. M. and Croft, W. B. (1998), A language modeling approach to information retrieval, *in* W. B. Croft, A. Moffat, C. J. van Rijsbergen, R. Wilkinson and J. Zobel, eds, 'Proc. 21th ACM SIGIR conf. on Research and Development in Information Retrieval', ACM Press, Melbourne, Australia, pp. 275–281.

Sparck Jones, K., Walker, S. and Robertson, S. E. (2000), 'A probabilistic model of information retrieval: development and comparative experiments. Parts 1&2', *Information Processing and Management* **36**(6), 779–840.

Zhai, C. and Lafferty, J. (2004), 'A study of smoothing methods for language models applied to information retrieval', *ACM Transactions on Information Systems* **22**(2), 179–214.