1. INTRODUCTION

RMIT submitted four systems for the NTCIR-13 We Want Web English subtask. For this round we focused on improving effectiveness via classic retrieval techniques:

- Markov Random Field Models (MRFs)
- Field-based weighting (body, title, anchor)
- Query Expansion
- Static document features.

For these tasks we made extensive use of the Indri search system and the flexible query language it provides.

2. SYSTEMS

1. RMIT-1: A field-based Sequential Dependence Model with RM3 Query Expansion
2. RMIT-2: Linear combination of RMIT-1 + 0.25 × PageRank priors
3. RMIT-3: Full Dependence Model + RM3 Query Expansion
4. RMIT-4: n-gram fields + RM3 Query Expansion

All system configurations made use of Query Expansion and Term Dependency Models. Identical smoothing parameters were used for all systems, with $\mu = 2000$ and $\mu_{prox} = 2000$. Post-retrieval spam filtering was applied to systems RMIT-2, RMIT-3, and RMIT-4. Documents with a spam score less than 70 were removed from retrieved results.

3. RESULTS

<table>
<thead>
<tr>
<th>System</th>
<th>ERR@5</th>
<th>ERR@10</th>
<th>NDCG@5</th>
<th>NDCG@10</th>
<th>RBP@0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMIT-3</td>
<td>0.5065</td>
<td>0.5207</td>
<td>0.3977</td>
<td>0.3968</td>
<td>0.7670+0.0242</td>
</tr>
<tr>
<td>RMIT-2</td>
<td>0.5285</td>
<td>0.5378</td>
<td>0.4186</td>
<td>0.4069</td>
<td>0.7533+0.0228</td>
</tr>
<tr>
<td>RMIT-4</td>
<td>0.5635</td>
<td>0.5728</td>
<td>0.4402</td>
<td>0.4249</td>
<td>0.7422+0.0270</td>
</tr>
<tr>
<td>RMIT-1</td>
<td>0.5548</td>
<td>0.5712</td>
<td>0.4670</td>
<td>0.4783†</td>
<td>0.8438+0.0221†</td>
</tr>
</tbody>
</table>

Post-hoc analysis of submissions:

- BM25: 0.4760, 0.4879, 0.3718, 0.3713, 0.6509+0.1919†
- R3-NOE: 0.4595, 0.5096, 0.3884, 0.3879, 0.7560+0.0348
- R2-NOE: 0.5279, 0.5403, 0.4161, 0.4125, 0.7537+0.0408
- R4-NOE: 0.5533, 0.5637, 0.4276, 0.4071, 0.7238+0.0456
- RBC-14: 0.5819, 0.5951, 0.4817†, 0.4776†, 0.8263+0.0025†
- R1-NOE: 0.5743, 0.5884, 0.4723†, 0.4877†, 0.8220+0.0453†

4. POST-HOC BM25 ANALYSIS

Comparison of ERR for RMIT-1 and RMIT-4 across the topics that were identified as navigational.

5. OVERALL RESULTS DISCUSSION

- Use of document fields resulted in improved effectiveness that is commonly employed on large web collections.
- It can be seen that RMIT-1 significantly outperforms the other submissions for NDCG@10 and RBP (also NDCG@20 not shown).
- Spam filtering was not applied to RMIT-1.
- RMIT-2 performance did not align with our expectations – static document features such as PageRank and Spam are known to improve effectiveness on ClueWeb collections.

6. NAVIGATIONAL QUERIES

Comparison of ERR for RMIT-1 and RMIT-4 across the topics that were identified as navigational.

7. QUERY EXPANSION ANALYSIS

RMIT-1 shows improved effectiveness when Query Expansion is not used across both ERR and NDCG.

For NDCG@10 the n-gram fields configuration (RMIT-4) is significantly improved by Query Expansion, as is FDM (RMIT-3).

RMIT-4 has a significantly larger total of NDCG@10 wins over RMIT-3, suggesting that the field-based model is better at leveraging Query Expansion to additively improve effectiveness for this situation.

The above table depicts NDCG@10 wins, ties and losses comparing the four submitted systems to their without Query Expansion equivalent, all other configuration settings remain the same. Statistical significance between each pair (Sys. A, Sys. B) is denoted † indicating significance at $p = 0.05$ and ‡ indicating significance at $p = 0.01$. Scores are tied for NDCG@10 $\Delta$ ± 0.025, and ignored in the $\sum$ Win and $\sum$ Loss columns.

ACKNOWLEDGMENTS

This work was supported by the Australian Research Council’s Discovery Projects Scheme (DP170102231), an Australian Government Research Training Program Scholarship and a grant from the Mozilla Foundation.