

Round 2 results — Run COMBSUM submitted from UB_BW

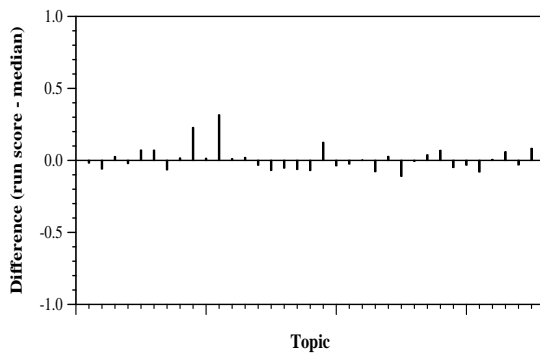
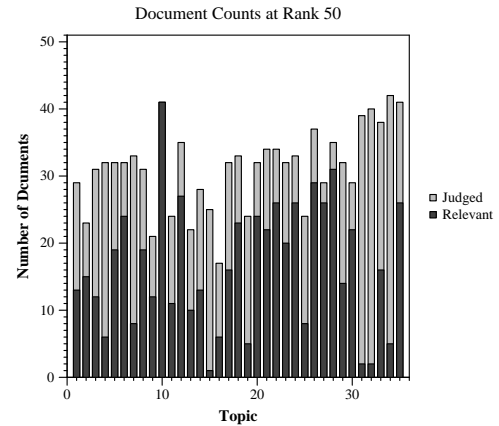
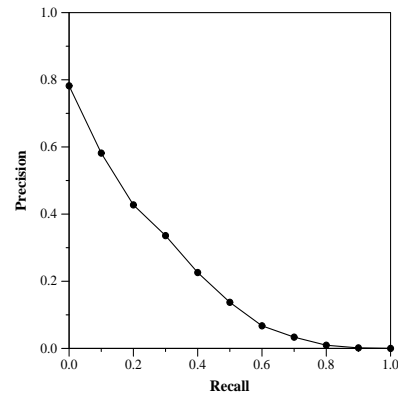
Run Description

We indexed only the title and the abstract using Terrier-v5.2. For the final document ranking, we deployed CombSUM to combine the results from two different systems. The first system used PL2 term weighting model while the second system used TF-IDF term weighting model. During retrieval, we used both the query and the question tags, which were then expanded with terms selected using the Bo1 model for query expansion.

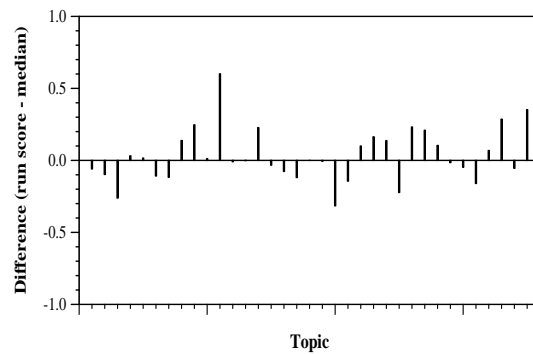
Summary Statistics	
Run ID	COMBSUM
Topic type	automatic
Contributed to judgment sets?	yes

Overall measures	
Number of topics	35
Total number retrieved	31601
Total relevant	3002
Total relevant retrieved	1687
MAP	0.2135
Mean Bpref	0.3784
Mean NDCG@10	0.5028
Mean RBP(p=0.5)	0.5471 +0.0013

Document Level Averages	
	Precision
At 5 docs	0.6114
At 10 docs	0.5514
At 15 docs	0.5200
At 20 docs	0.4814
At 30 docs	0.4048
R-Precision	
Exact	0.2736

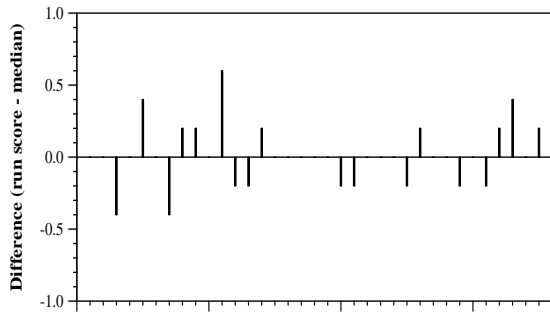


Per-topic difference from median bpref for all Round 2 runs

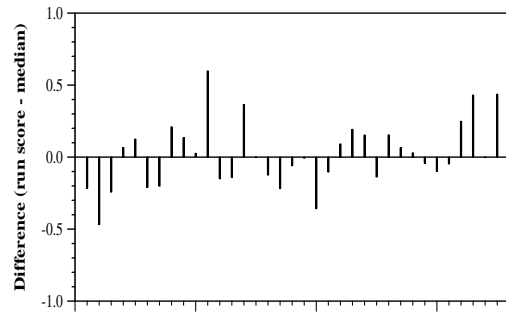


Per-topic difference from median NDCG@10 for all Round 2 runs

Round 2 results — Run COMBSUM submitted from UB_BW



Per-topic difference from median P@5 for all Round 2 runs



Per-topic difference from median RBP(p=0.5) for all Round 2 runs