

Round 1 results — Run run1 submitted from GUIR_S2

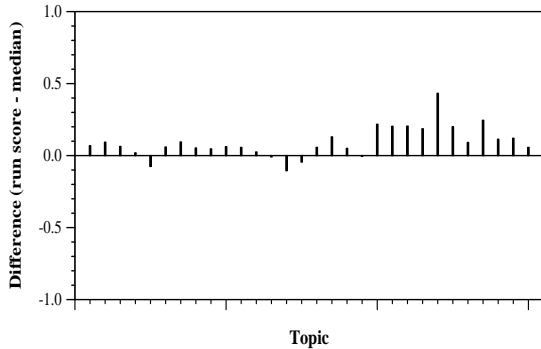
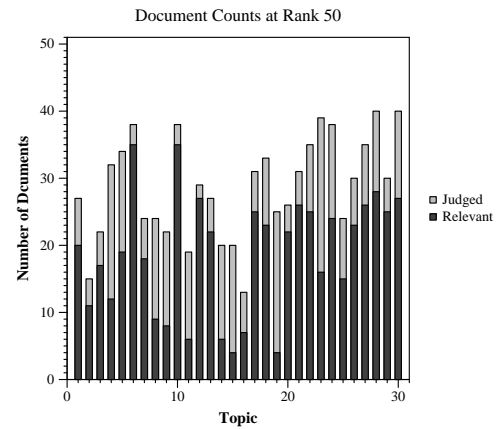
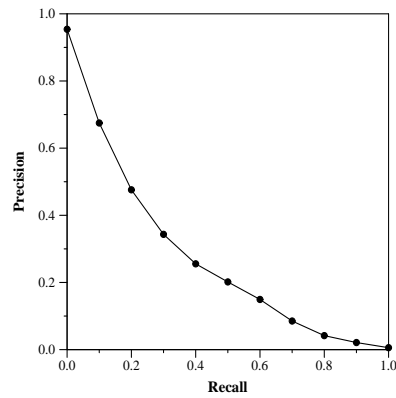
Run Description

Initial ranking with BM25 and query over fulltext, re-ranking using SciBERT (trained on MS-MARCO) using only title and abstract with the question.

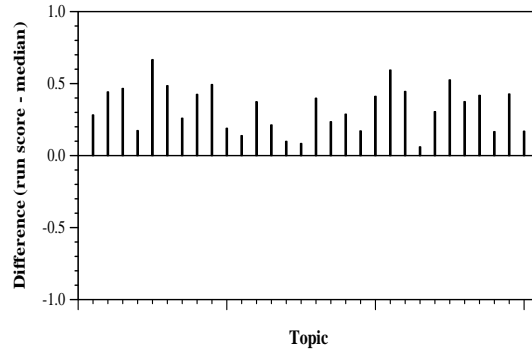
Summary Statistics	
Run ID	run1
Topic type	manual
Contributed to judgment sets?	yes

Overall measures	
Number of topics	30
Total number retrieved	14983
Total relevant	2352
Total relevant retrieved	1203
MAP	0.2663
Mean Bpref	0.4159
Mean NDCG@10	0.6844

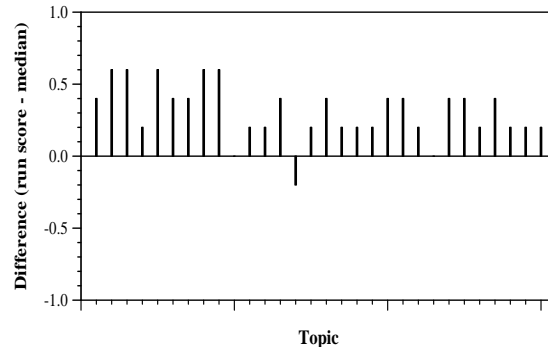
Document Level Averages	
	Precision
At 5 docs	0.7933
At 10 docs	0.7067
At 15 docs	0.6200
At 20 docs	0.5700
At 30 docs	0.4811
R-Precision	
Exact	0.3202



Per-topic difference from median bpref for all Round 1 runs



Per-topic difference from median NDCG@10 for all Round 1 runs



Per-topic difference from median P@5 for all Round 1 runs