

Round 2 results — Run GUIR_S2_run2 submitted from GUIR_S2

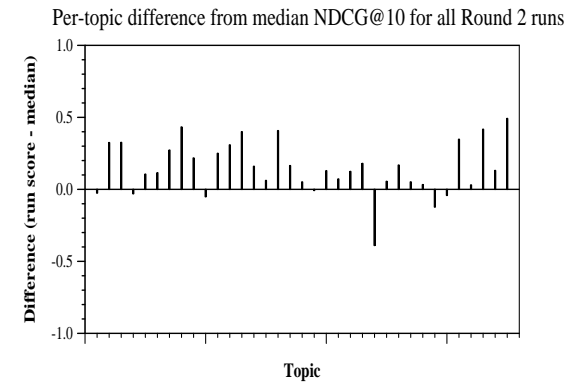
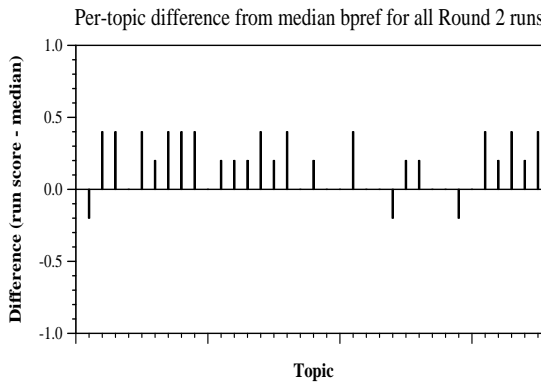
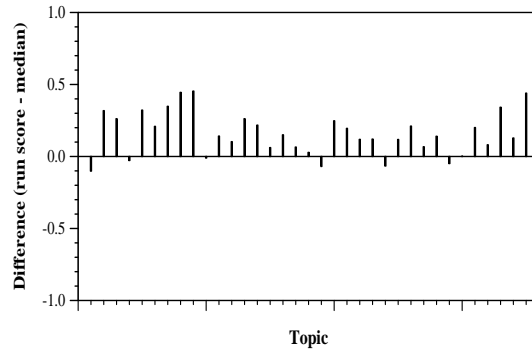
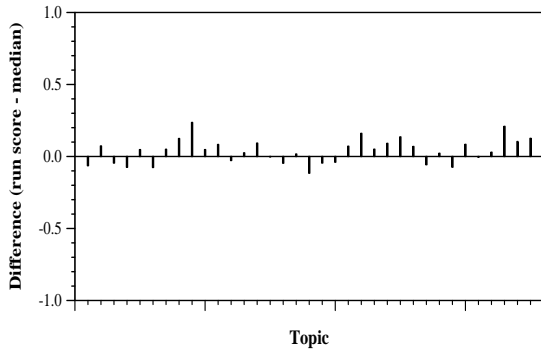
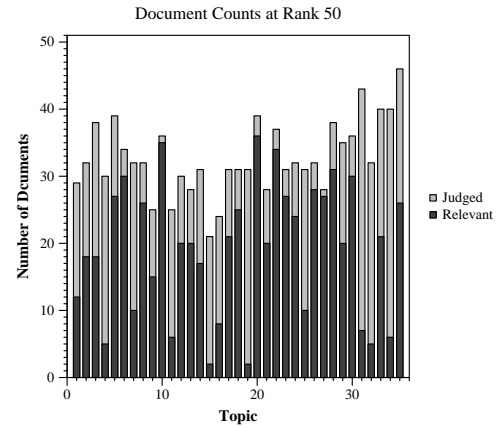
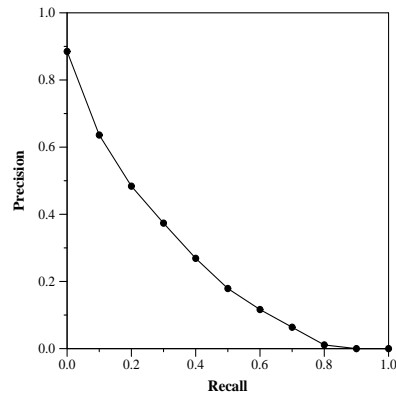
Run Description

Fusion of SciBERT over paragraphs, SPECTER feedback (top 5 SciBERT documents), and SPECTER query+document similarity. SciBERT model trained on MS-MARCO. Tuning conducted on judged documents from Round 1.

Summary Statistics	
Run ID	GUIR_S2_run2
Topic type	feedback
Contributed to judgment sets?	yes

Overall measures	
Number of topics	35
Total number retrieved	25150
Total relevant	3002
Total relevant retrieved	1637
MAP	0.2531
Mean Bpref	0.4067
Mean NDCG@10	0.6286
Mean RBP(p=0.5)	0.6805 +0.0014

Document Level Averages	
	Precision
At 5 docs	0.7771
At 10 docs	0.6714
At 15 docs	0.5943
At 20 docs	0.5429
At 30 docs	0.4800
R-Precision	
Exact	0.3054



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