

Round 4 results — Run run3_Crf_A_SciB_N submitted from CIR

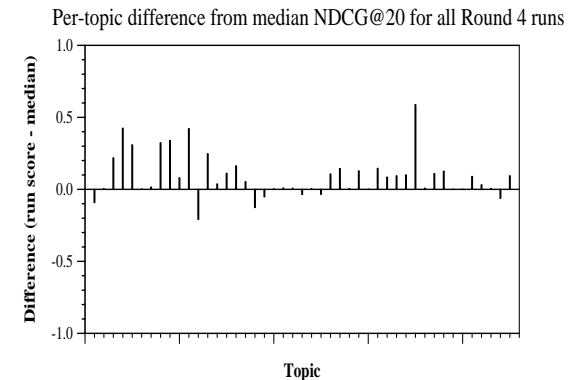
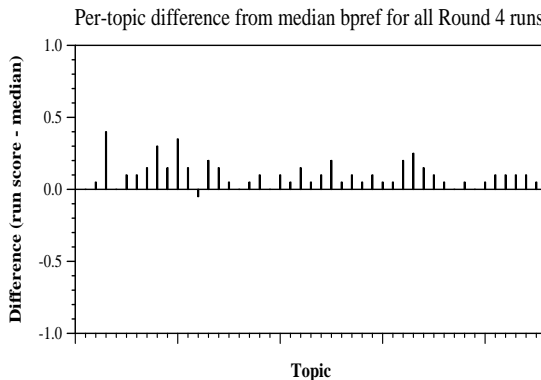
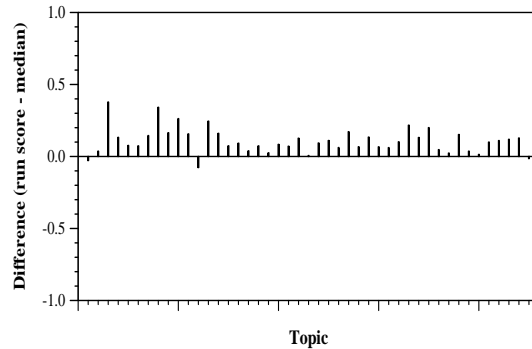
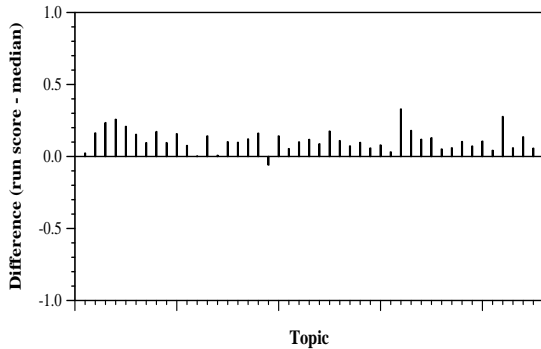
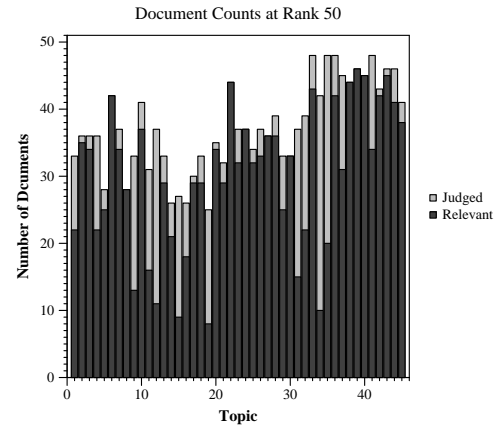
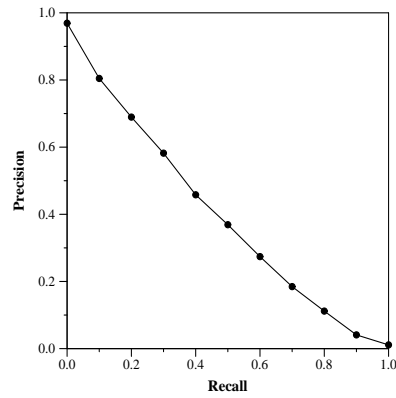
Run Description

Fusion of: (1) Anserini BM25 baseline(r4.rf), (2) Med-MARCO ANN dense embedding paragraph retrieval with PRF based on embedding similarity, (3) MedMARCP SciBERT paragraph reranker(top 1000). Fusion weights tuned based on NDCG@10. Language model finetuned on Round-4 documents using MLM task.

Summary Statistics	
Run ID	run3_Crf_A_SciB_N
Topic type	feedback
Contributed to judgment sets?	yes

Overall measures	
Number of topics	45
Total number retrieved	44979
Total relevant	5824
Total relevant retrieved	4365
MAP	0.3967
Mean Bpref	0.6299
Mean NDCG@20	0.7345
Mean RBP(p=0.5)	0.8598 +0.0003

Document Level Averages	
	Precision
At 5 docs	0.8978
At 10 docs	0.8467
At 15 docs	0.7941
At 20 docs	0.7522
At 30 docs	0.7015
R-Precision	
Exact	0.4213



Per-topic difference from median P@20 for all Round 4 runs

Per-topic difference from median RBP(p=0.5) for all Round 4 runs