Run Description

This run is produced by Reciprocal Rank Fusion (Cormack et al., 2009) of 6 Anserini runs (https://github.com/castorini/anseri
covid.md) along with 10 Terrier runs using various combinations of (query, question) and (body, abstract)
fields. Top 50 RRF results are then re-ranked using LightGBM LambdaRank that takes as features (a) RRF
score (b) two best performing individual runs (Terrier using query + question with body or abstract fields)
(c) two TF-Ranking BERT models (https://arxiv.org/abs/2004.08476) trained on either MS-Marco or its
medical queries subset (https://arxiv.org/abs/2005.02365).

Summary Statistics	
Run ID	UPrrf16lgbert50-r2
Topic type	feedback
Contributed to judgment sets?	no

Overall measures	
Number of topics	35
Total number retrieved	35000
Total relevant	3002
Total relevant retrieved	2109
MAP	0.2964
Mean Bpref	0.4414
Mean NDCG@10	0.6077
Mean RBP(p=0.5)	0.6507 +0.0016

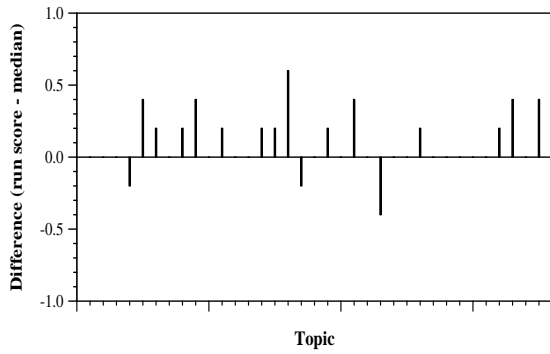| Document Level Averages	
	Precision
At 5 docs	0.6971
At 10 docs	0.6600
At 15 docs	0.6229
At 20 docs	0.5886
At 30 docs	0.5200
R-Precision	
Exact	0.3359

Document Counts at Rank 50
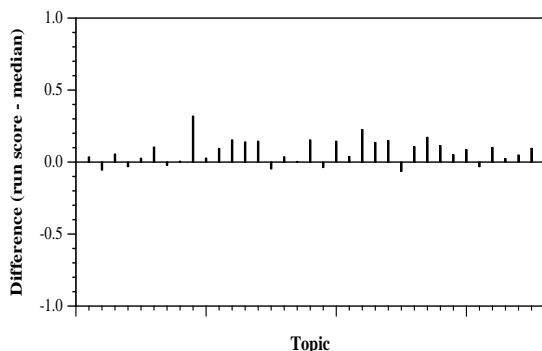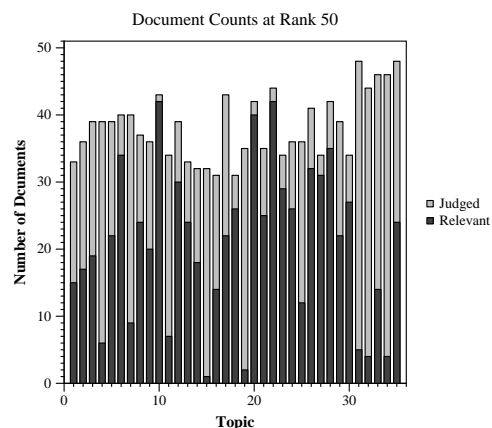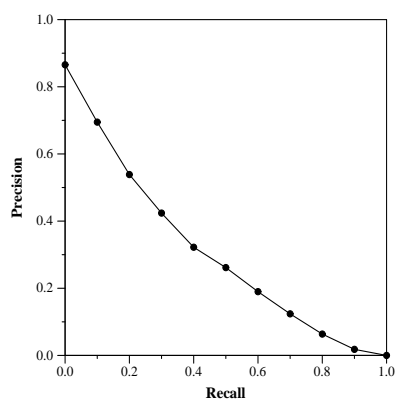
Per-topic difference from median bpref for all Round 2 runs
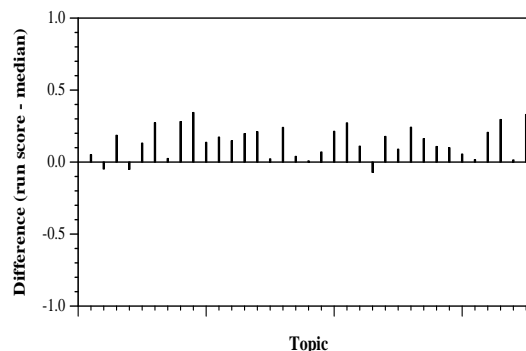
Per-topic difference from median NDCG@10 for all Round 2 runs

Per-topic difference from median P@5 for all Round 2 runs
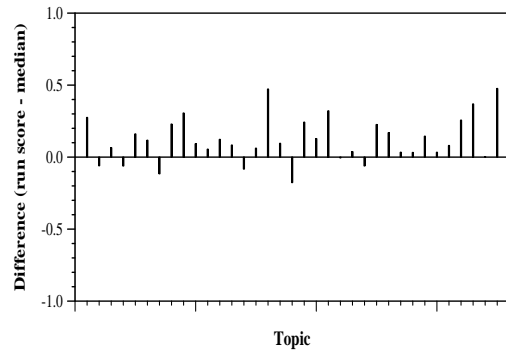
Per-topic difference from median RBP(p=0.5) for all Round 2 runs