Run Description

A per-query ensemble of (a) a reciprocal rank fusion of 20 retrieval runs using Anserini and Terrier (b) LightGBM learning-to-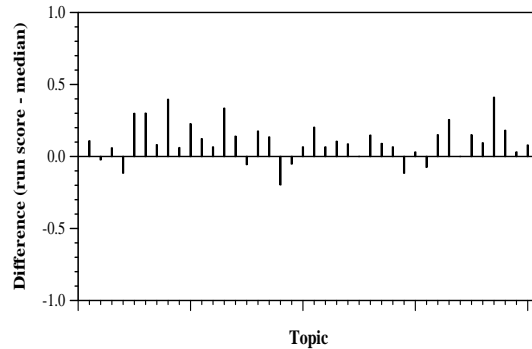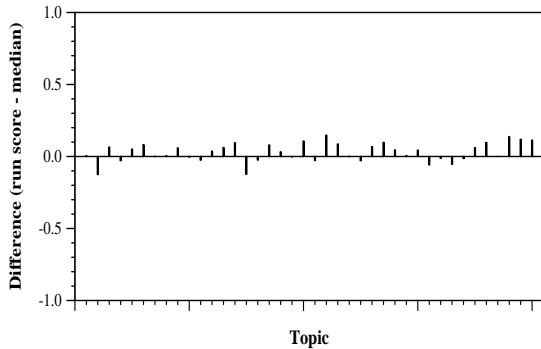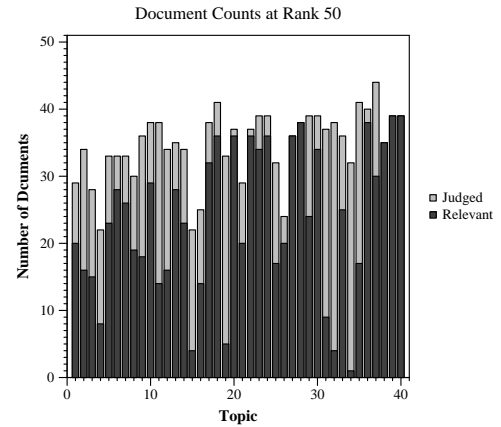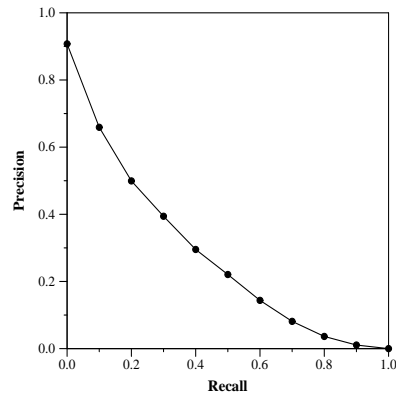rank (c) TF-Ranking + BERT fine-tuned on the relevance judgments from Rounds 1 & 2. The ensemble was created by weighting each method by its NDCG@10 for a given query.

Summary Statistics	
Run ID	UPrrf20maxens-r3
Topic type	feedback
Contributed to judgment sets?	yes

Overall measures	
Number of topics	40
Total number retrieved	39905
Total relevant	4698
Total relevant retrieved	3135
MAP	0.2740
Mean Bpref	0.4952
Mean NDCG@10	0.6954
Mean RBP(p=0.5)	0.7420 +0.0274

Document Level Averages	
	Precision
At 5 docs	0.7650
At 10 docs	0.7400
At 15 docs	0.6833
At 20 docs	0.6300
At 30 docs	0.5642
R-Precision	
Exact	0.3236

Document Counts at Rank 50

Per-topic difference from median bpref for all Round 3 runs

Per-topic difference from median NDCG@10 for all Round 3 runs
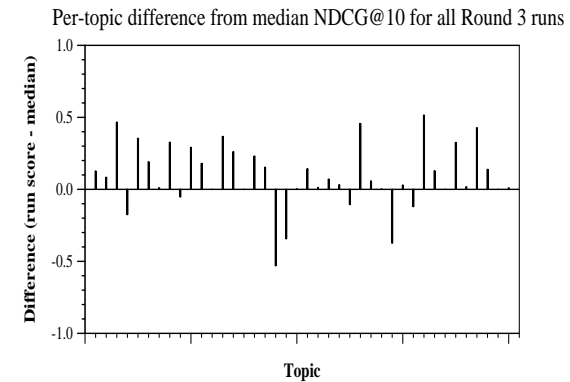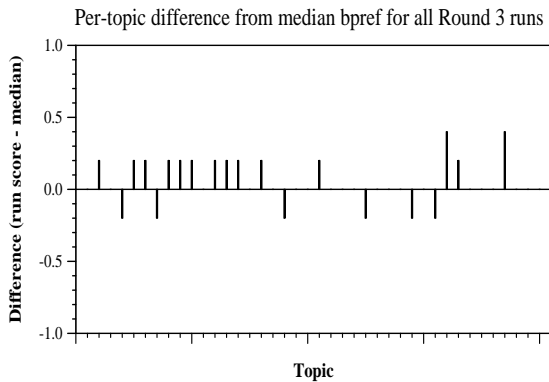
Per-topic difference from median P@5 for all Round 3 runs

Per-topic difference from median RBP(p=0.5) for all Round 3 runs