

Round 4 results — Run UPrrf38rrf3v2-r4 submitted from unique\_ptr

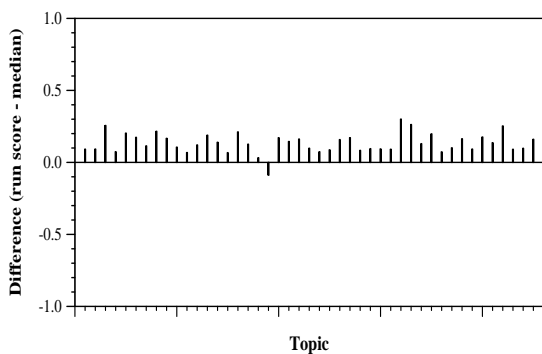
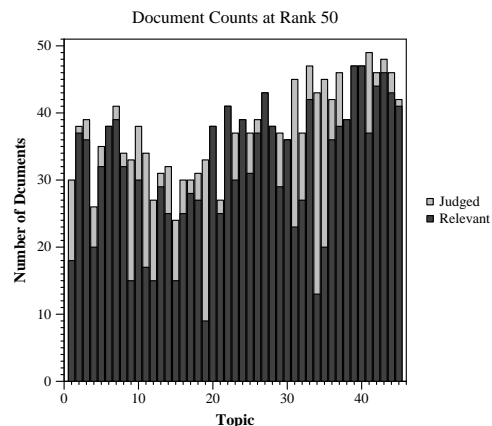
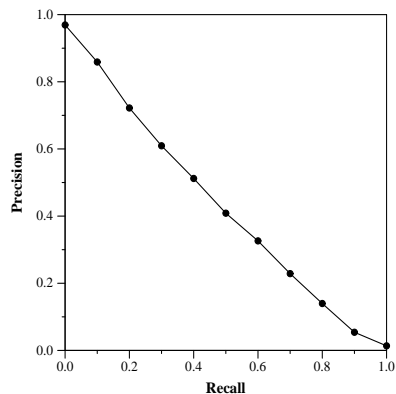
Run Description

This run is produced by Reciprocal Rank Fusion (Cormack et al., 2009) of (a) Terrier & Anserini retrieval runs and neural retrieval runs based on synthetic query generation (<https://arxiv.org/abs/2004.14503>) with various combinations of query, question and narrative (b) TF-Ranking + ELECTRA with softmax loss (<https://arxiv.org/abs/2004.08476>) trained on MS-Marco dataset (c) TF-Ranking + BERT with softmax loss fine-tuned on the relevance judgments from Rounds 1 - 3.

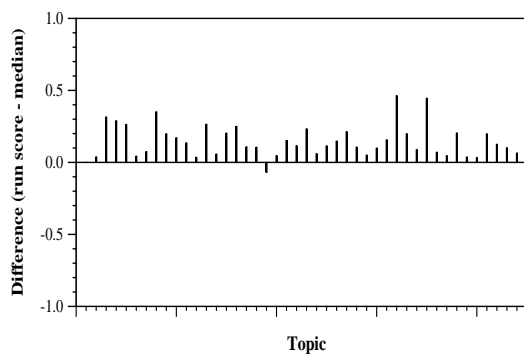
Summary Statistics	
Run ID	UPrrf38rrf3v2-r4
Topic type	feedback
Contributed to judgment sets?	yes

Overall measures	
Number of topics	45
Total number retrieved	44980
Total relevant	5824
Total relevant retrieved	4474
MAP	0.4310
Mean Bpref	0.6514
Mean NDCG@20	0.7706
Mean RBP(p=0.5)	0.8718 +0.0004

Document Level Averages	
	Precision
At 5 docs	0.9378
At 10 docs	0.8778
At 15 docs	0.8311
At 20 docs	0.7856
At 30 docs	0.7274
R-Precision	
Exact	0.4504



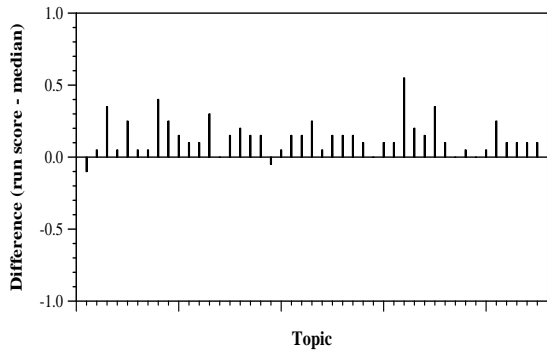
Per-topic difference from median bpref for all Round 4 runs



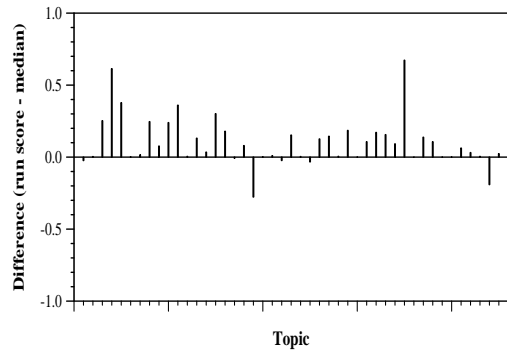
Per-topic difference from median NDCG@20 for all Round 4 runs

Round 4 results — Run UPrrf38rrf3v2-r4 submitted from unique\_ptr

---



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