

Round 4 results — Run UPrrf38rrf3-r4 submitted from unique_ptr

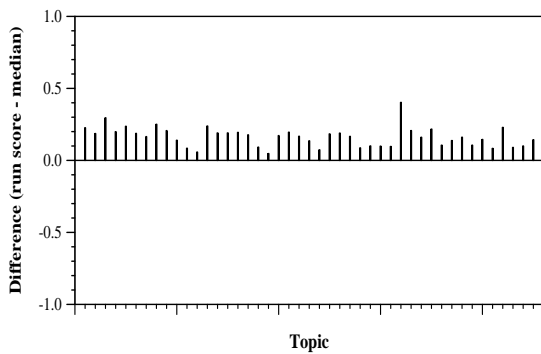
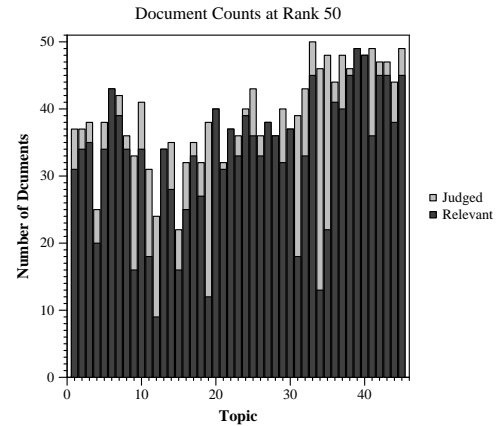
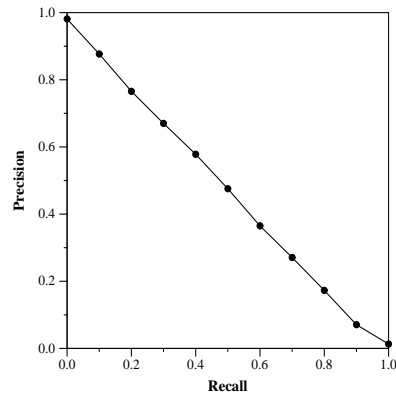
Run Description

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

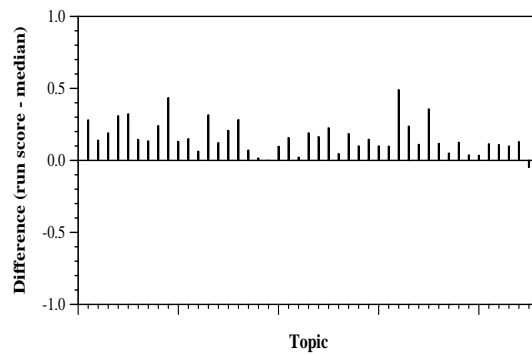
Summary Statistics	
Run ID	UPrrf38rrf3-r4
Topic type	feedback
Contributed to judgment sets?	yes

Overall measures	
Number of topics	45
Total number retrieved	44979
Total relevant	5824
Total relevant retrieved	4673
MAP	0.4681
Mean Bpref	0.6801
Mean NDCG@20	0.7843
Mean RBP(p=0.5)	0.8599 +0.0003

Document Level Averages	
	Precision
At 5 docs	0.9333
At 10 docs	0.9000
At 15 docs	0.8607
At 20 docs	0.8211
At 30 docs	0.7504
R-Precision	
Exact	0.4772

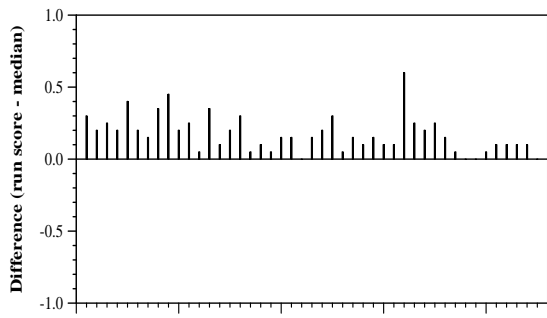


Per-topic difference from median bpref for all Round 4 runs

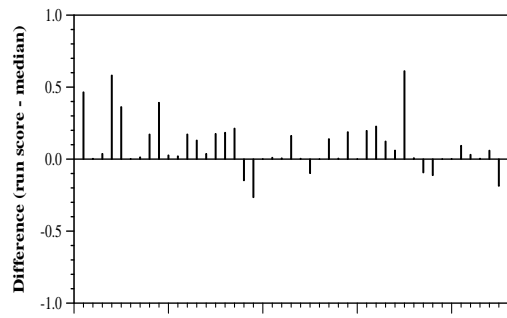


Per-topic difference from median NDCG@20 for all Round 4 runs

Round 4 results — Run UPrrf38rrf3-r4 submitted from unique_ptr



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



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