

Round 5 results — Run UPrf93-wt-r5 submitted from unique_ptr

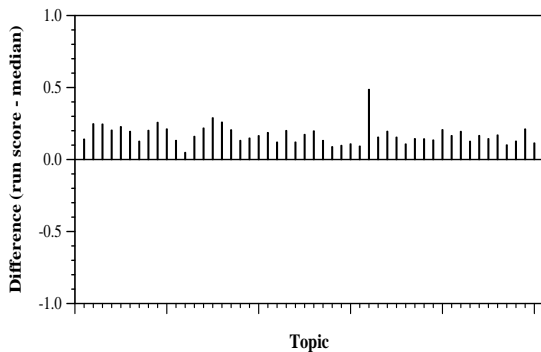
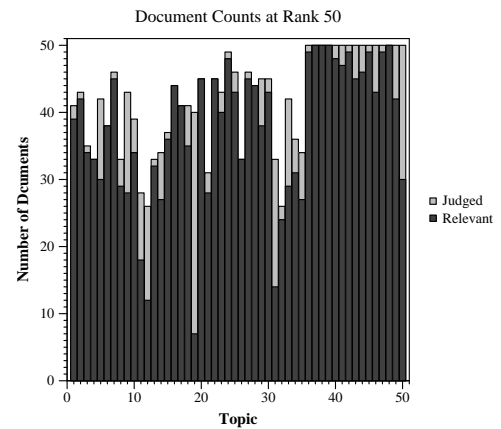
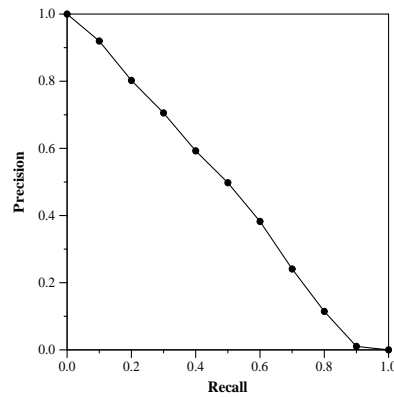
Run Description

A weighted reciprocal rank fusion of (a) Anserini & Terrier runs based on different indexing schemes and topic variants, including relevance feedback (b) Neural retrieval runs based on synthetic query generation (<https://arxiv.org/abs/2004.14503>) (c) TF-Ranking + multiple BERT variants with softmax loss (<https://arxiv.org/abs/2004.08476>) trained on MS-Marco (d) TF-Ranking + multiple BERT variants using the question topic field, fine-tuned with softmax loss on the relevance judgments from Rounds 1 - 4. For weighting the runs, we use a simple scheme which doubles the weight of any runs that have access to relevance judgments in the reciprocal rank fusion.

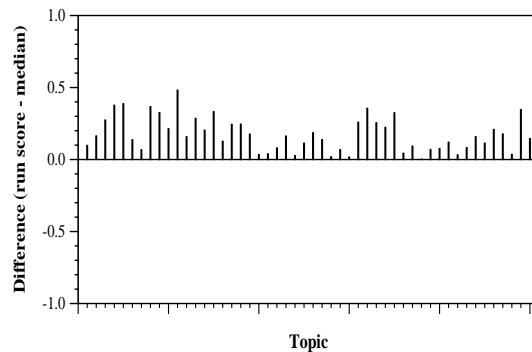
Summary Statistics	
Run ID	UPrf93-wt-r5
Topic type	feedback
Contributed to judgment sets?	yes

Overall measures	
Number of topics	50
Total number retrieved	49887
Total relevant	10910
Total relevant retrieved	8085
MAP	0.4718
Mean Bpref	0.6372
Mean NDCG@20	0.8496
Mean RBP(p=0.5)	0.9197 +0.0004

Document Level Averages	
	Precision
At 5 docs	0.9720
At 10 docs	0.9460
At 15 docs	0.9040
At 20 docs	0.8760
At 30 docs	0.8220
R-Precision	
Exact	0.4912

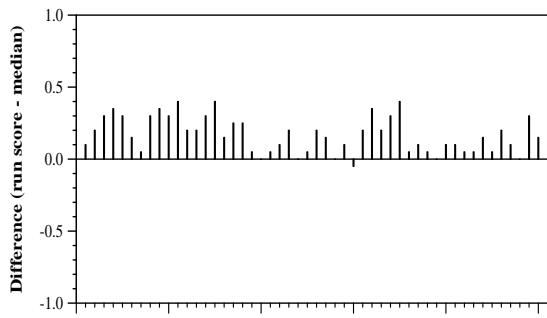


Per-topic difference from median bpref for all Round 4 runs

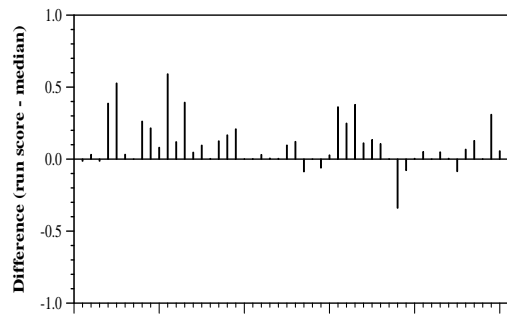


Per-topic difference from median NDCG@20 for all Round 4 runs

Round 5 results — Run UPrrf93-wt-r5 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