

Round 2 results — Run UPrrf16lgbertd50-r2 submitted from unique_ptr

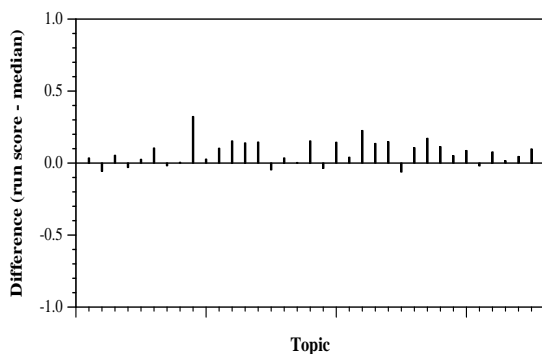
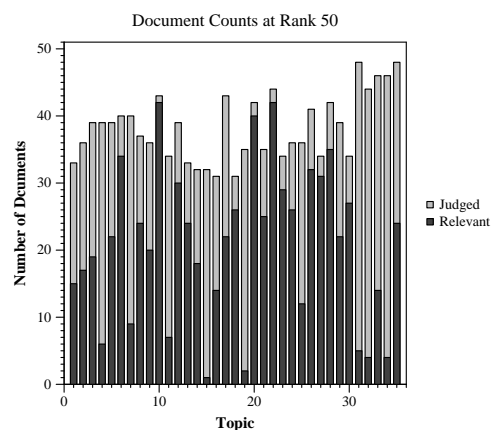
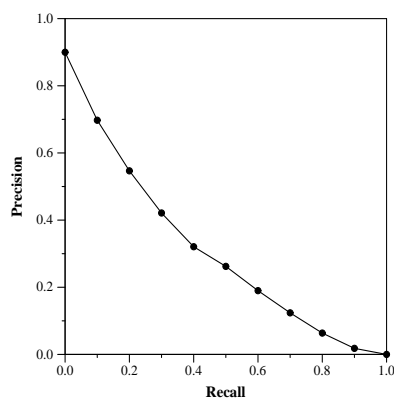
Run Description

This run is produced by Reciprocal Rank Fusion (Cormack et al., 2009) of 6 Anserini runs (<https://github.com/castorini/anserini> 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>) (d) publish date (e) containment of any COVID-19 related keywords.

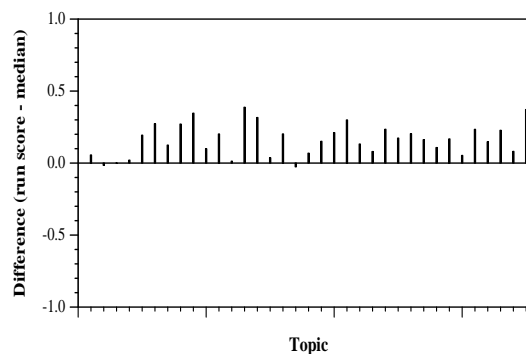
Summary Statistics	
Run ID	UPrrf16lgbertd50-r2
Topic type	feedback
Contributed to judgment sets?	yes

Overall measures	
Number of topics	35
Total number retrieved	35000
Total relevant	3002
Total relevant retrieved	2109
MAP	0.3000
Mean Bpref	0.4414
Mean NDCG@10	0.6320
Mean RBP(p=0.5)	0.7153 +0.0004

Document Level Averages	
	Precision
At 5 docs	0.7086
At 10 docs	0.6657
At 15 docs	0.6514
At 20 docs	0.5971
At 30 docs	0.5219
R-Precision	
Exact	0.3337

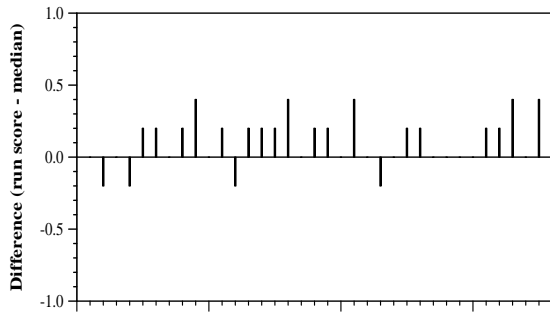


Per-topic difference from median bpref for all Round 2 runs

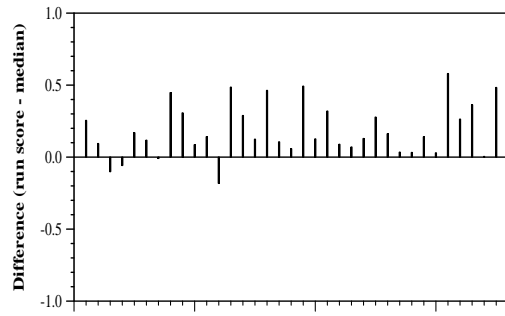


Per-topic difference from median NDCG@10 for all Round 2 runs

Round 2 results — Run UPrrf16lgbertd50-r2 submitted from unique_ptr



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