

Round 2 results — Run FullTxt_R2_Time submitted from OHSU

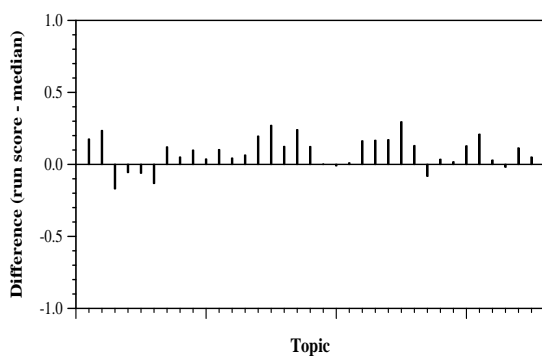
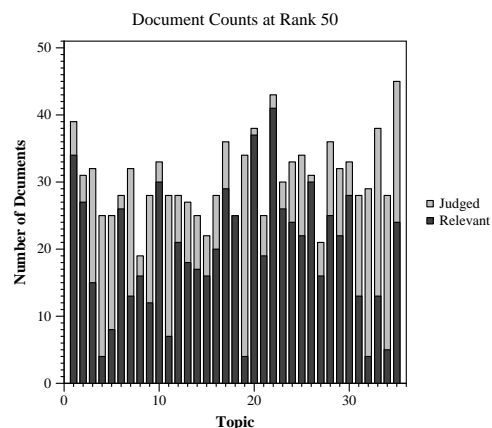
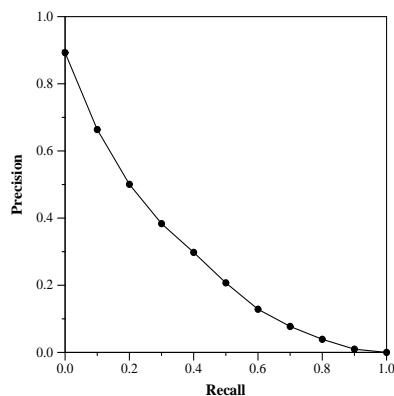
Run Description

Queries, questions, and narratives were tokenized and combined to form a query for each topic. Additional manual review was performed for identification of relevant query terms. These queries were inputted into a Lucene full-text index was searched using Pyserini using BM25, LDA, and RM3 reranking to identify the top 2000 documents. Results were filtered such that only documents published after 1/1/20 were considered for inclusion in the final 1000 documents per topic.

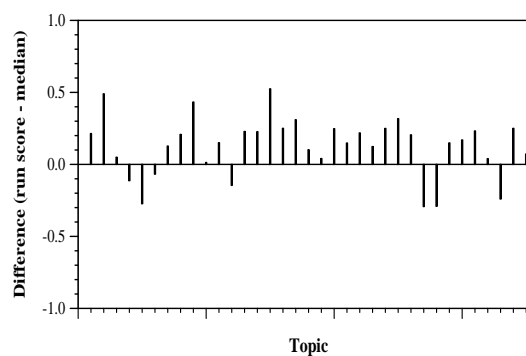
Summary Statistics	
Run ID	FullTxt_R2_Time
Topic type	manual
Contributed to judgment sets?	yes

Overall measures	
Number of topics	35
Total number retrieved	21069
Total relevant	3002
Total relevant retrieved	1691
MAP	0.2680
Mean Bpref	0.4525
Mean NDCG@10	0.5969
Mean RBP(p=0.5)	0.6193 +0.0012

Document Level Averages	
	Precision
At 5 docs	0.7029
At 10 docs	0.6657
At 15 docs	0.6019
At 20 docs	0.5429
At 30 docs	0.4838
R-Precision	
Exact	0.3264

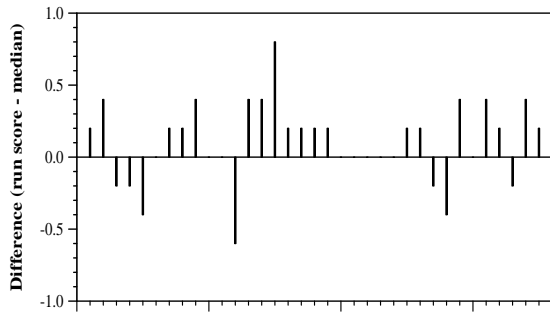


Per-topic difference from median bpref for all Round 2 runs

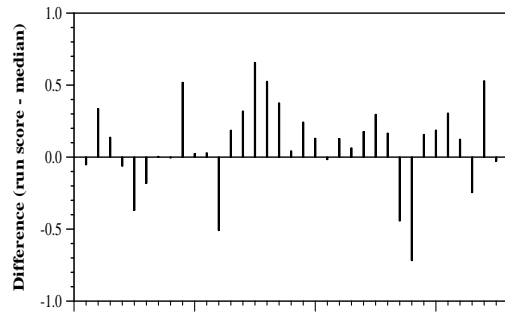


Per-topic difference from median NDCG@10 for all Round 2 runs

Round 2 results — Run FullTxt_R2_Time submitted from OHSU



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