

Round 3 results — Run OHSU_Rerank submitted from OHSU

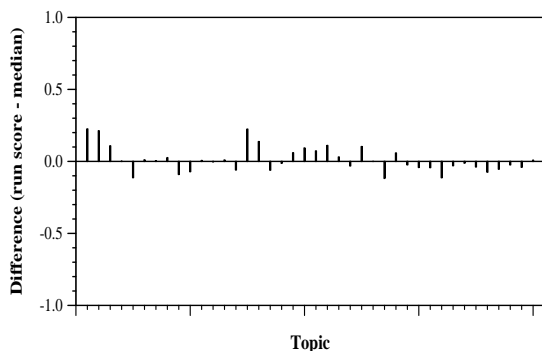
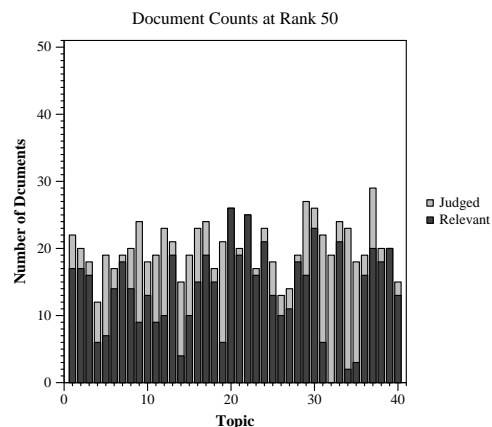
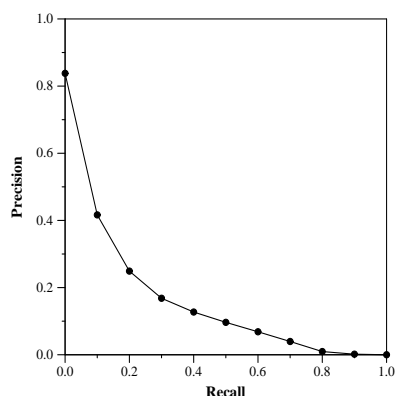
Run Description

Tokenized combinations of narrative, query, and question fields for each topic were input into an Anserini BM25 searcher for 3 indexes: abstract only, full-text only, and paragraphs+abstracts. RRF was performed to combine these 3 runs for the top 2000 documents/topic. BioBERT trained on pubmed was used as a reranker.

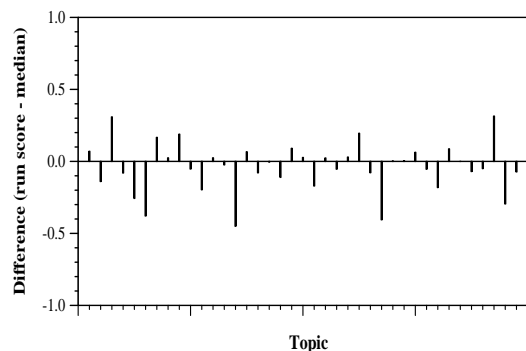
Summary Statistics	
Run ID	OHSU_Rerank
Topic type	manual
Contributed to judgment sets?	yes

Overall measures	
Number of topics	40
Total number retrieved	39923
Total relevant	4698
Total relevant retrieved	2893
MAP	0.1544
Mean Bpref	0.4771
Mean NDCG@10	0.5515
Mean RBP(p=0.5)	0.5668 +0.0297

Document Level Averages	
	Precision
At 5 docs	0.6500
At 10 docs	0.6550
At 15 docs	0.5217
At 20 docs	0.4462
At 30 docs	0.3558
R-Precision	
Exact	0.1899

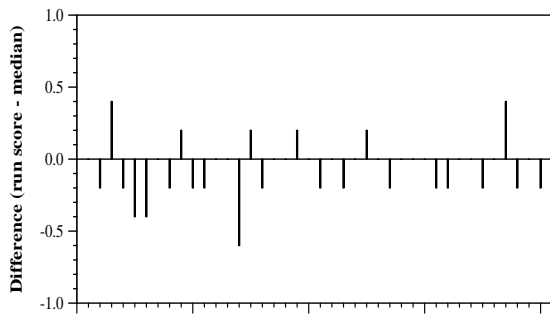


Per-topic difference from median bpref for all Round 3 runs

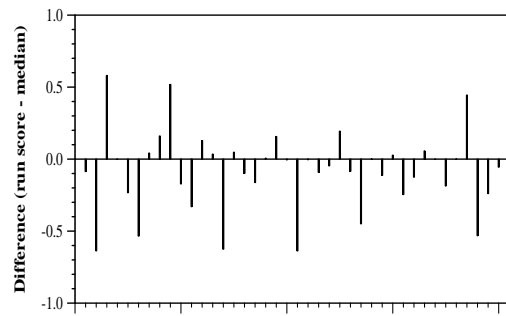


Per-topic difference from median NDCG@10 for all Round 3 runs

Round 3 results — Run OHSU_Rerank submitted from OHSU



Per-topic difference from median P@5 for all Round 3 runs



Per-topic difference from median RBP(p=0.5) for all Round 3 runs