

Round 3 results — Run factum-sparse-rerank submitted from Factum

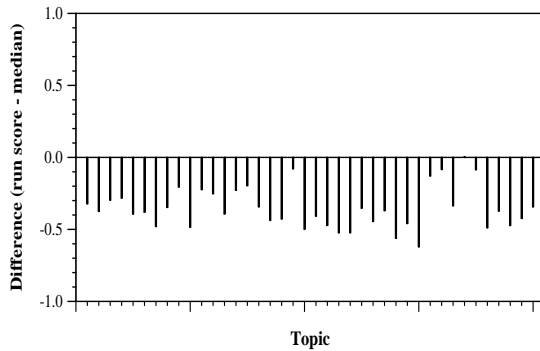
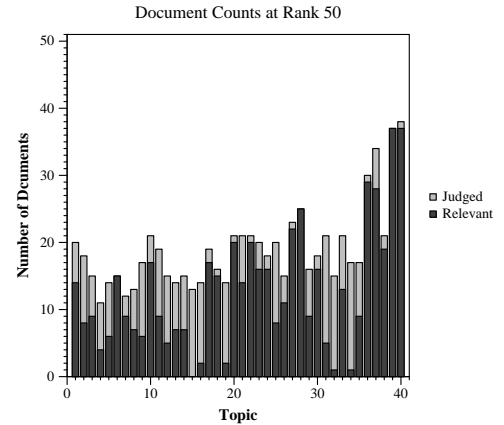
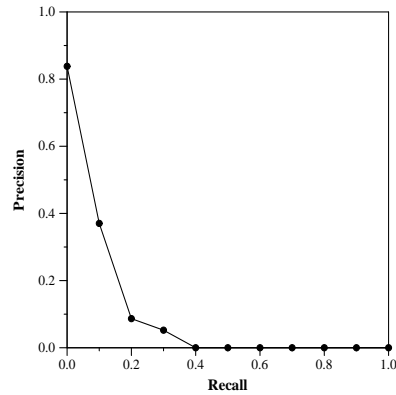
Run Description

Sparse retrieval (BM25) with fusion of scores from document, abstract and paragraph indexes. We used unigrams, with lemmatization and removal of stopwords. The top 100 documents are re-ranked using a BERT-base scoring model fine-tuned on MSMARCO. The final score of every document is given by the paragraph with the highest score from the ranking model.

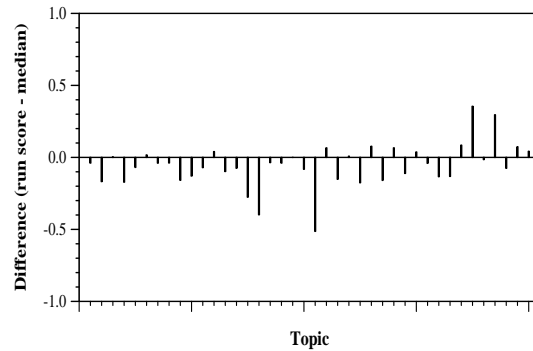
Summary Statistics	
Run ID	factum-sparse-rerank
Topic type	automatic
Contributed to judgment sets?	yes

Overall measures	
Number of topics	40
Total number retrieved	1413
Total relevant	4698
Total relevant retrieved	528
MAP	0.0868
Mean Bpref	0.1124
Mean NDCG@10	0.5373
Mean RBP(p=0.5)	0.5611 +0.0370

Document Level Averages	
	Precision
At 5 docs	0.6650
At 10 docs	0.6200
At 15 docs	0.5517
At 20 docs	0.4913
At 30 docs	0.3900
R-Precision	
Exact	0.1154

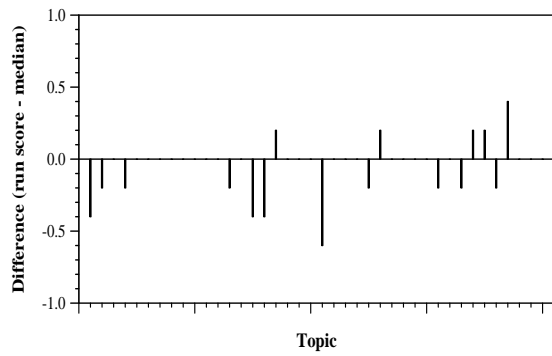


Per-topic difference from median bpref for all Round 3 runs

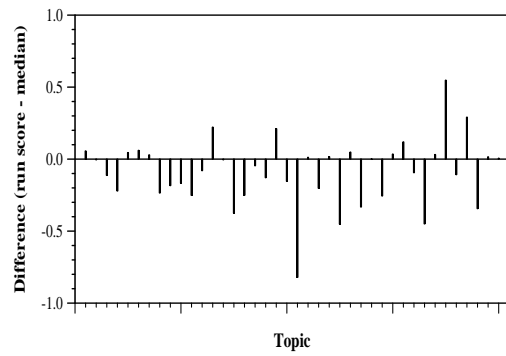


Per-topic difference from median NDCG@10 for all Round 3 runs

Round 3 results — Run factum-sparse-rerank submitted from Factum



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