

Round 2 results — Run factum-hybrid submitted from Factum

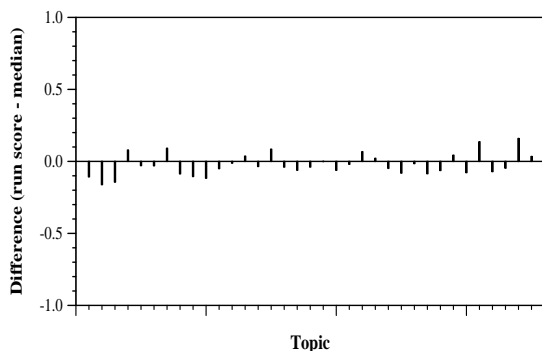
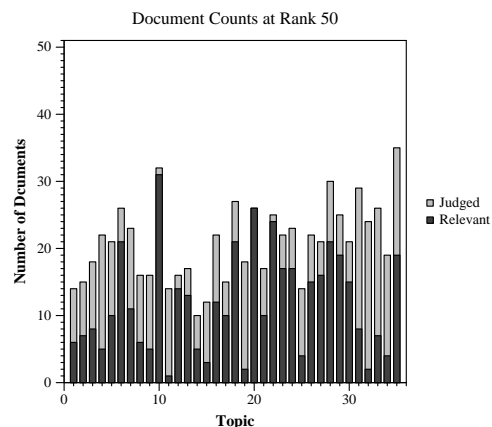
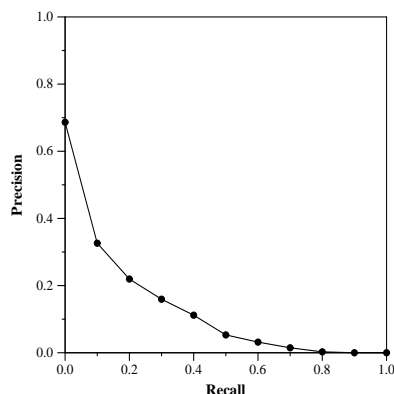
Run Description

We used a siamese BERT encoder trained on NLI and QA data with sampled softmax loss to encode snippets of up to 5 sentences from the paragraphs of the articles. The final score is derived by a combination of cosine similarity of the dense vectors and BM25 score of the sparse vectors (unigrams + bigrams with lemmatisation). We removed articles published before December 2019.

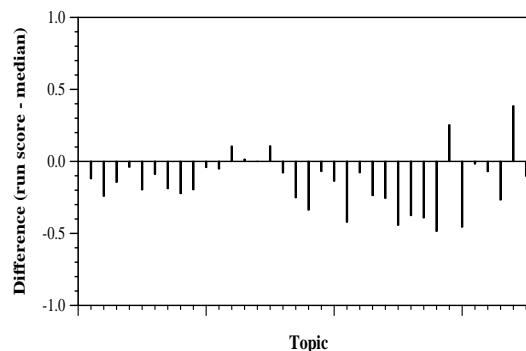
Summary Statistics	
Run ID	factum-hybrid
Topic type	automatic
Contributed to judgment sets?	no

Overall measures	
Number of topics	35
Total number retrieved	29409
Total relevant	3002
Total relevant retrieved	1471
MAP	0.1174
Mean Bpref	0.3463
Mean NDCG@10	0.3253
Mean RBP(p=0.5)	0.3713 +0.0491

Document Level Averages	
	Precision
At 5 docs	0.4343
At 10 docs	0.3714
At 15 docs	0.3390
At 20 docs	0.3086
At 30 docs	0.2762
R-Precision	
Exact	0.1941

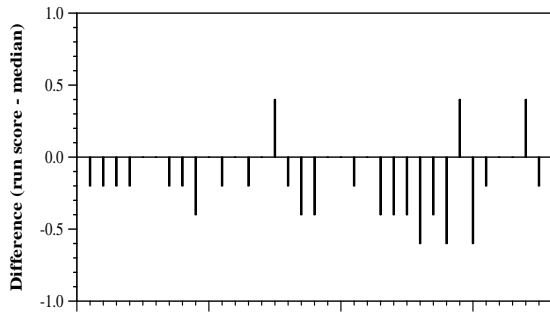


Per-topic difference from median bpref for all Round 2 runs

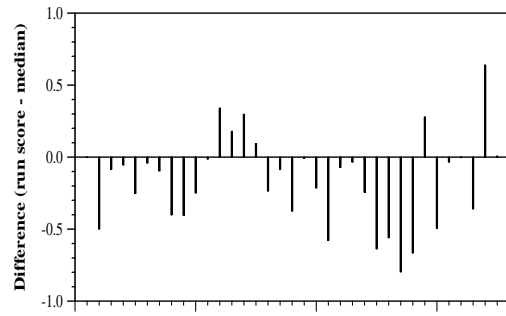


Per-topic difference from median NDCG@10 for all Round 2 runs

Round 2 results — Run factum-hybrid submitted from Factum



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