

Round 2 results — Run IRIT_markers_all submitted from IRIT_markers

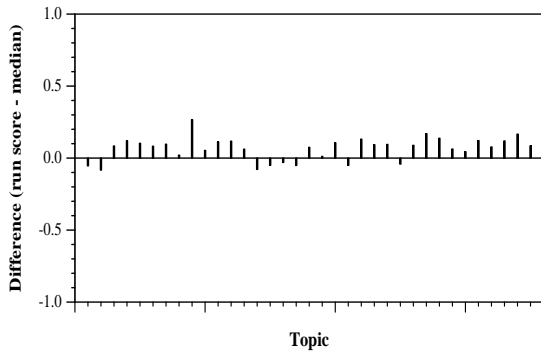
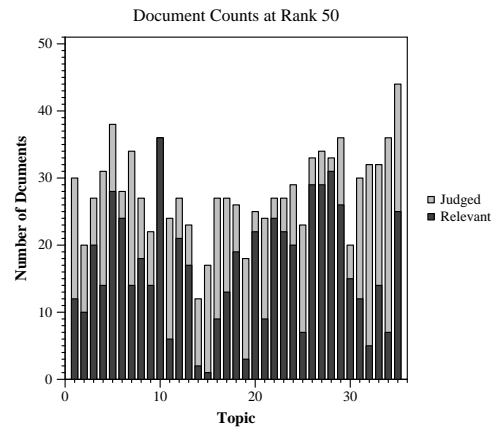
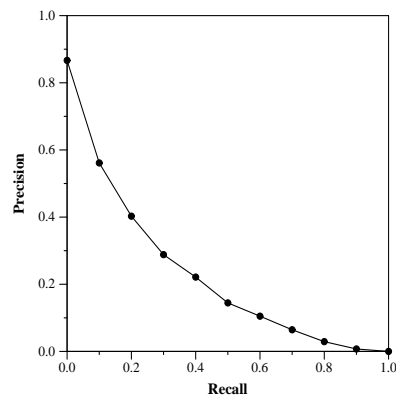
Run Description

We used a two-step ranker, an initial list of the top-1000 documents are retrieved using Anserini BM25, we use UDel queries and produce a run for each index: abstract, full text and paragraphs then merge the three to get a fusion run(<https://github.com/castorini/anserini/blob/master/docs/experiments-covid.md>). Then we use a BERT_base re-ranker finetuned on Msmarco passages. We use 3 models based on BERT, an original BERT(no modifications) and the 2 others use augmented entries with markers that highlight exact term matching. This final run is obtained by a fusion of the three re-ranking results.

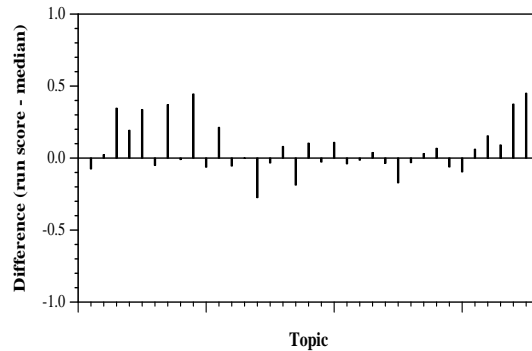
Summary Statistics	
Run ID	IRIT_markers_all
Topic type	automatic
Contributed to judgment sets?	yes

Overall measures	
Number of topics	35
Total number retrieved	30840
Total relevant	3002
Total relevant retrieved	1934
MAP	0.2174
Mean Bpref	0.4348
Mean NDCG@10	0.5366
Mean RBP(p=0.5)	0.6089 +0.0020

Document Level Averages	
	Precision
At 5 docs	0.6629
At 10 docs	0.5771
At 15 docs	0.5010
At 20 docs	0.4514
At 30 docs	0.3990
R-Precision	
Exact	0.2680

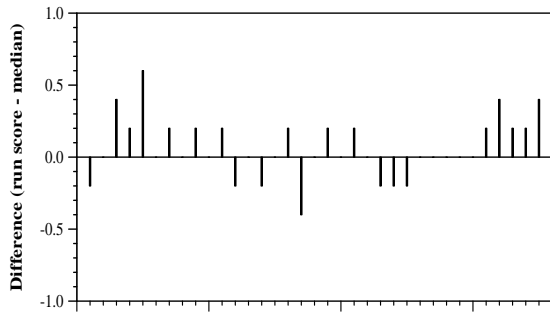


Per-topic difference from median bpref for all Round 2 runs

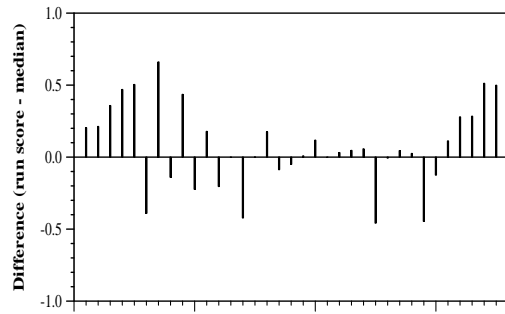


Per-topic difference from median NDCG@10 for all Round 2 runs

Round 2 results — Run IRIT_markers_all submitted from IRIT_markers



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