

Round 2 results — Run IRIT_markers_base_mu submitted from IRIT_markers

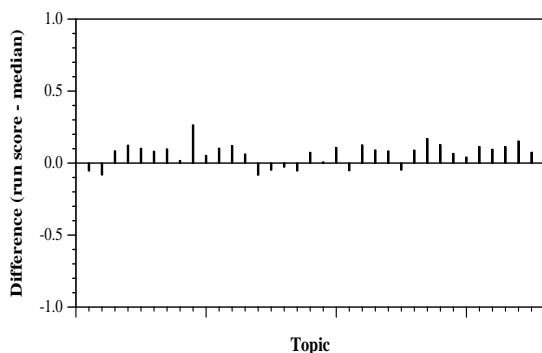
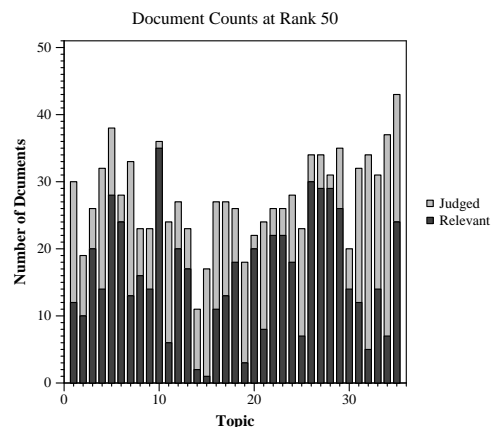
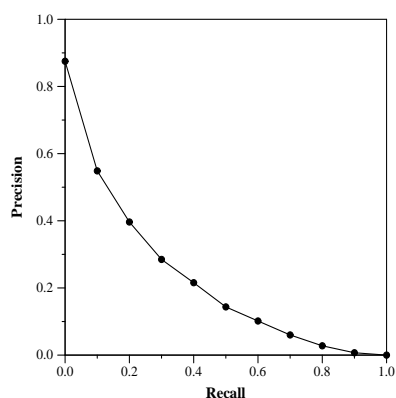
Run Description

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 the question as the query and only the title and the abstract of each document. We use 2 models based on BERT, an original BERT(no modifications) and the other uses augmented entries with markers that highlight exact term matching using identifier tokens ([ei][\ei]) to surround the term i. This final run is obtained by a fusion of the two re-ranking results.

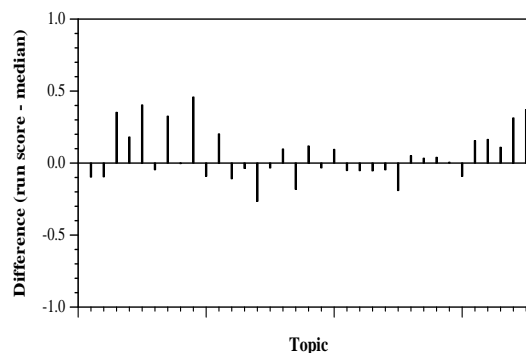
Summary Statistics	
Run ID	IRIT_markers_base_mu
Topic type	automatic
Contributed to judgment sets?	no

Overall measures	
Number of topics	35
Total number retrieved	30840
Total relevant	3002
Total relevant retrieved	1934
MAP	0.2140
Mean Bpref	0.4327
Mean NDCG@10	0.5292
Mean RBP(p=0.5)	0.6025 +0.0052

Document Level Averages	
	Precision
At 5 docs	0.6514
At 10 docs	0.5657
At 15 docs	0.4895
At 20 docs	0.4529
At 30 docs	0.3905
R-Precision	
Exact	0.2662

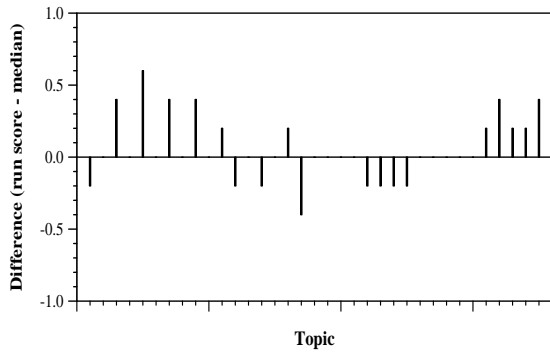


Per-topic difference from median bpref for all Round 2 runs

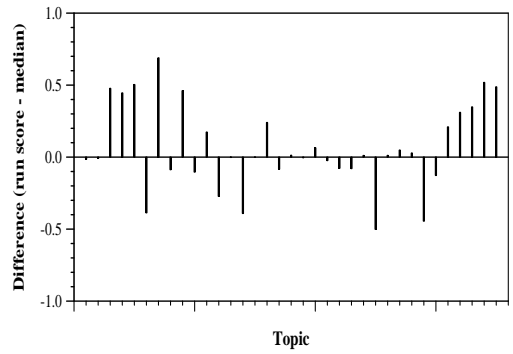


Per-topic difference from median NDCG@10 for all Round 2 runs

Round 2 results — Run IRIT_markers_base_mu 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