

Round 4 results — Run ILPS_UvA_allrounds_c submitted from ILPS_UvA

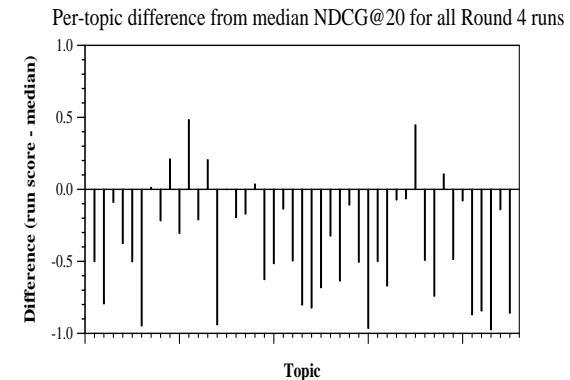
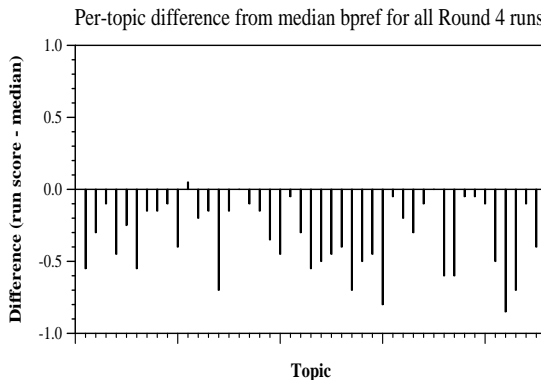
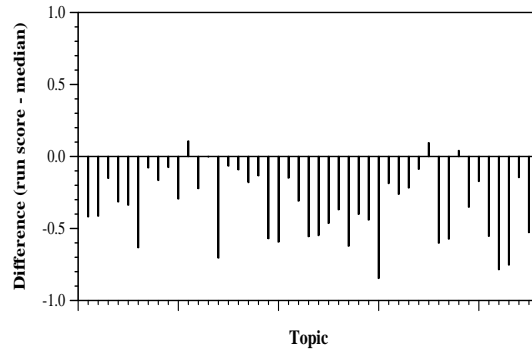
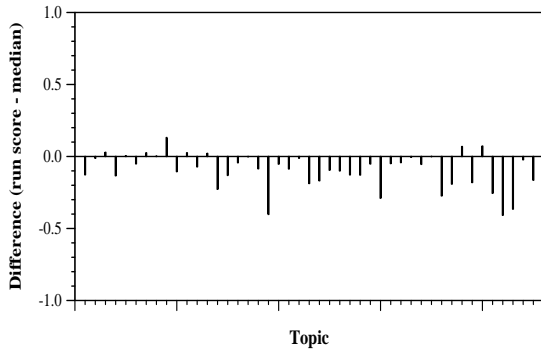
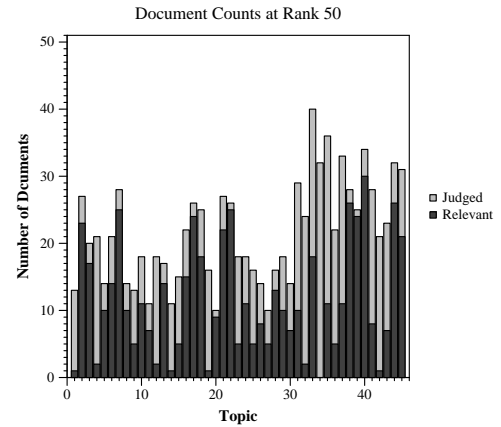
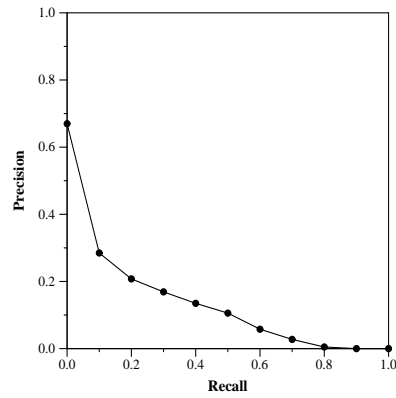
Run Description

We used a weak supervision technique to rerank with covidBERT (only abstracts). This run was selected due to promising results using our soft evaluation protocol and because the performance of the reranker did not seem to drop after fine-tuning for more than 3 epochs.

Summary Statistics	
Run ID	ILPS_UvA_allrounds_c
Topic type	feedback
Contributed to judgment sets?	yes

Overall measures	
Number of topics	45
Total number retrieved	38010
Total relevant	5824
Total relevant retrieved	3297
MAP	0.1244
Mean Bpref	0.4211
Mean NDCG@20	0.2916
Mean RBP(p=0.5)	0.3892 +0.0020

Document Level Averages	
	Precision
At 5 docs	0.4711
At 10 docs	0.4133
At 15 docs	0.3630
At 20 docs	0.3278
At 30 docs	0.2807
R-Precision	
Exact	0.1812



Per-topic difference from median P@20 for all Round 4 runs

Per-topic difference from median RBP(p=0.5) for all Round 4 runs