

Round 5 results — Run rk_trf_brx_rrf submitted from risklick

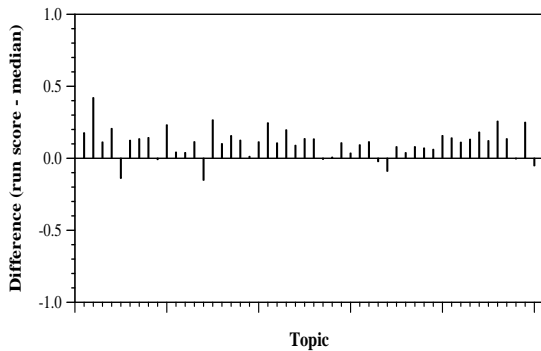
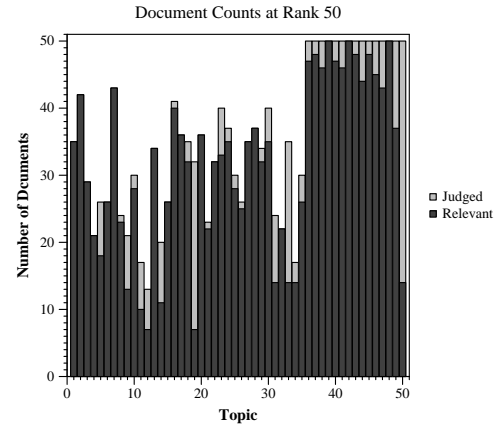
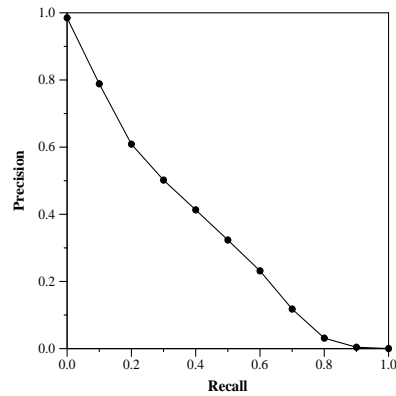
Run Description

A combination of deep neural language models (BERT, RoBERTa and XLNet) runs fine tuned on round 4 QREL and combined using reciprocal-rank fusion. Classic IR models (BM25, DFR, XLNet) was used as the baseline run.

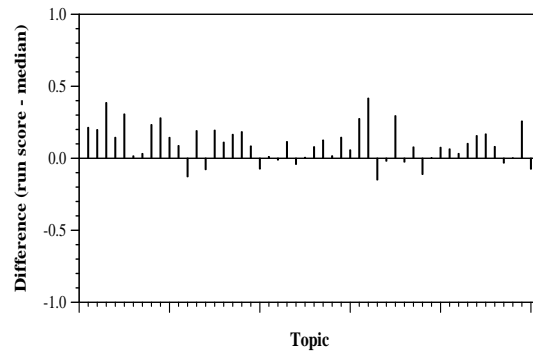
Summary Statistics	
Run ID	rk_trf_brx_rrf
Topic type	feedback
Contributed to judgment sets?	yes

Overall measures	
Number of topics	50
Total number retrieved	47699
Total relevant	10910
Total relevant retrieved	6963
MAP	0.3468
Mean Bpref	0.5680
Mean NDCG@20	0.7716
Mean RBP(p=0.5)	0.8706 +0.0008

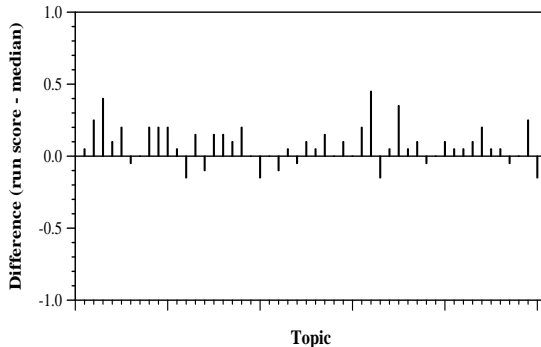
Document Level Averages	
	Precision
At 5 docs	0.9120
At 10 docs	0.8740
At 15 docs	0.8320
At 20 docs	0.7880
At 30 docs	0.7107
R-Precision	
Exact	0.3947



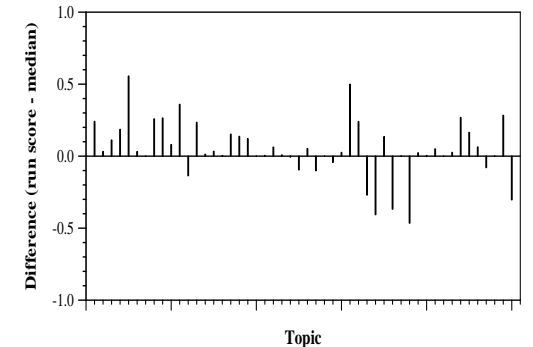
Per-topic difference from median bpref for all Round 4 runs



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



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