Run Description

A learning-to-rank model using the LambdaMART algorithm with features from classic information retrieval models (BM25, DFR an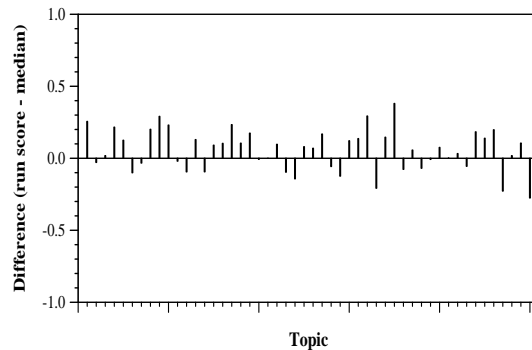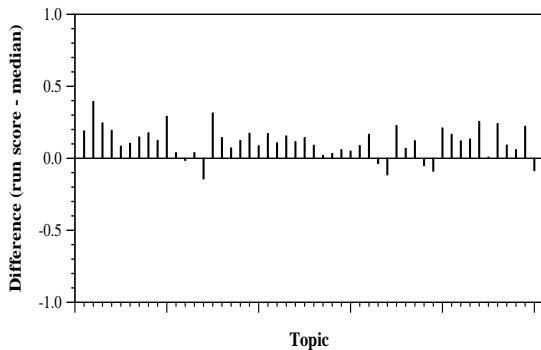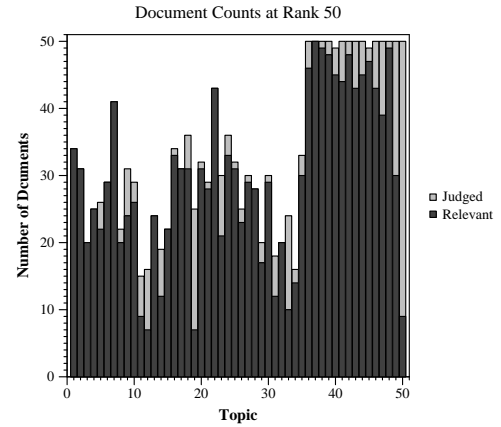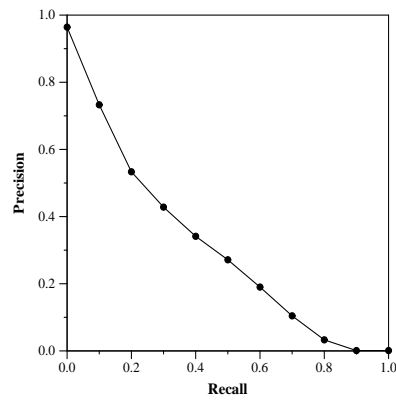d LM Dirichlet) and deep neural language models (BERT, RoBERTa and XLNet) extracted from metadata and full text indices.

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

Overall measures	
Number of topics	50
Total number retrieved	49922
Total relevant	10910
Total relevant retrieved	6834
MAP	0.3068
Mean Bpref	0.5759
Mean NDCG@20	0.7312
Mean RBP(p=0.5)	0.8302 +0.0052

Document Level Averages	
	Precision
At 5 docs	0.9080
At 10 docs	0.8400
At 15 docs	0.7813
At 20 docs	0.7460
At 30 docs	0.6740
R-Precision	
Exact	0.3546

Document Counts at Rank 50

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

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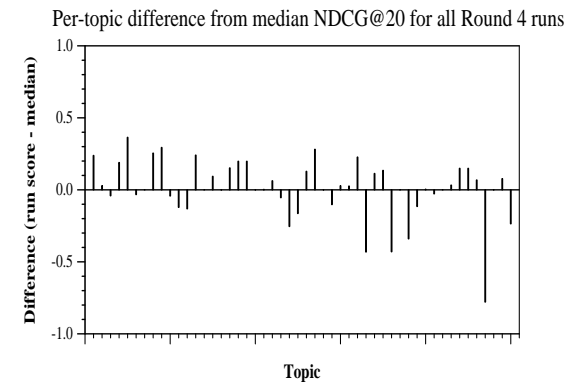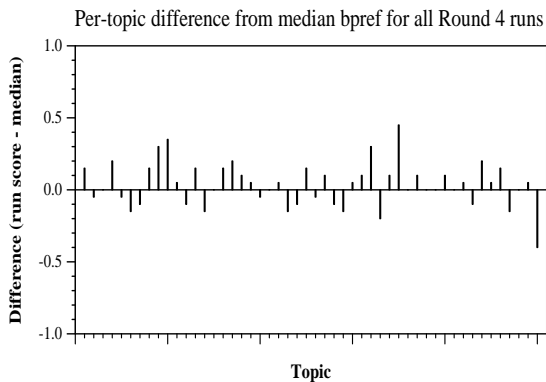
Per-topic difference from median P@20 for all Round 4 runs

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