

Round 2 results — Run bm25\_safir\_nvsm\_rrf submitted from ims\_unipd

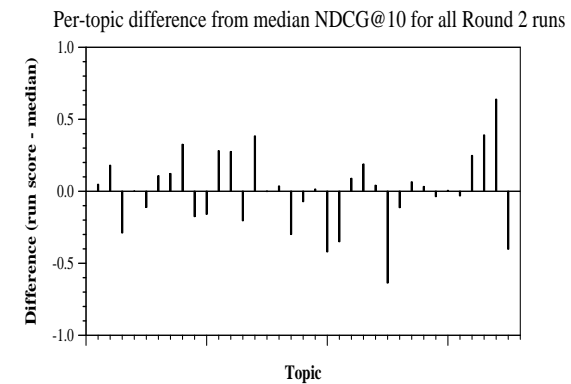
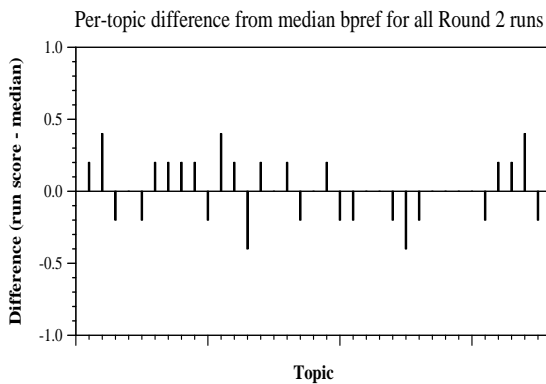
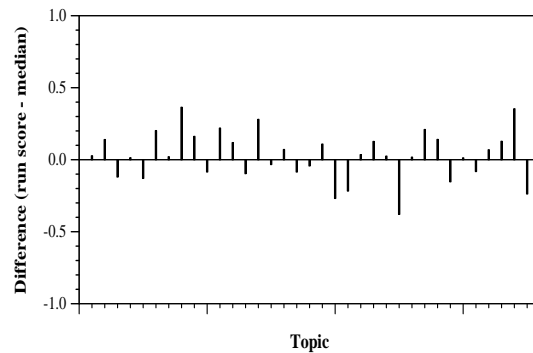
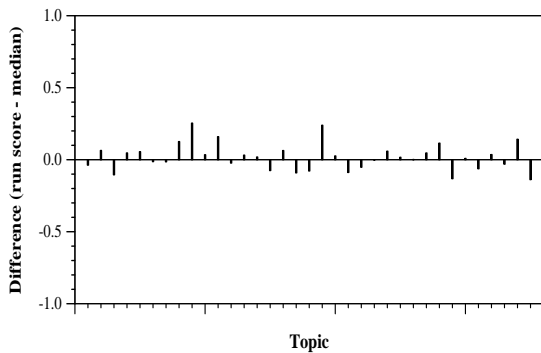
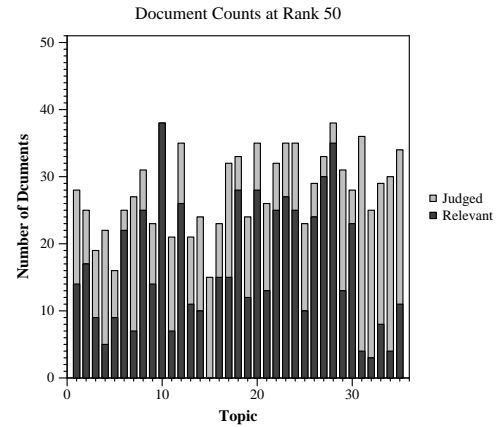
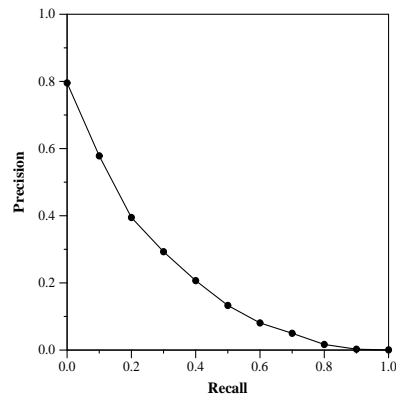
Run Description

We adopt SAFIR (three models) + RM3, NVSM + RM3, and BM25 + RM3. Then we perform RRF to combine the five runs.

Summary Statistics	
Run ID	bm25_safir_nvsm_rrf
Topic type	automatic
Contributed to judgment sets?	no

Overall measures	
Number of topics	35
Total number retrieved	32036
Total relevant	3002
Total relevant retrieved	1611
MAP	0.2068
Mean Bpref	0.3875
Mean NDCG@10	0.4981
Mean RBP(p=0.5)	0.5379 +0.0140

Document Level Averages	
	Precision
At 5 docs	0.6171
At 10 docs	0.5657
At 15 docs	0.4971
At 20 docs	0.4629
At 30 docs	0.3990
R-Precision	
Exact	0.2728



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