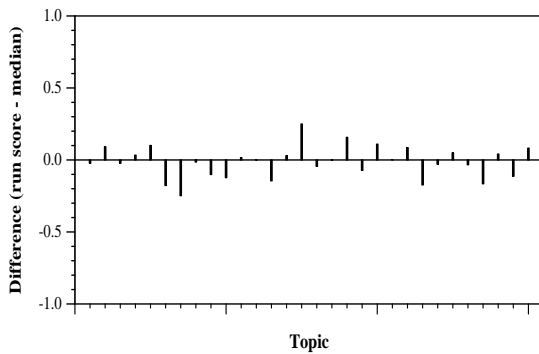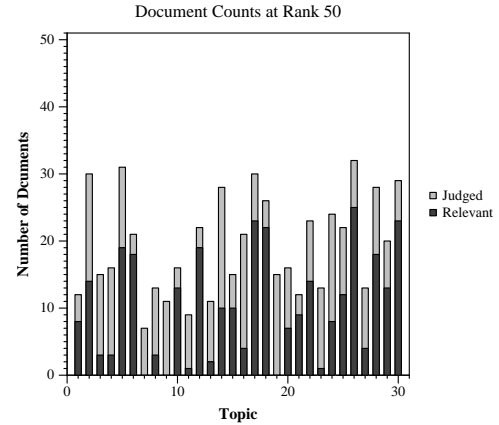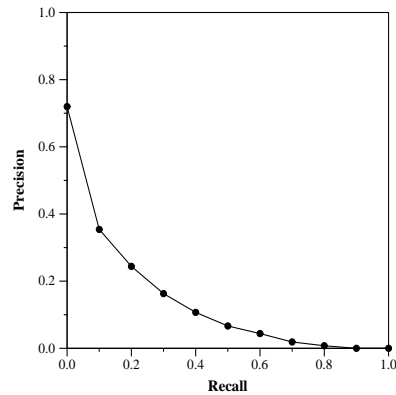## Run Description

44 participants undertook a manual task using used our experimental search system (with results returned from the Bing Search API). Upon completion of the task, participants selected two queries that they felt returned good results. This was repeated over three topics. Post experiment, rankings for each query were obtained, with the Borda method run over the rankings to obtain scores to highlight the queries that were broadly accepted by participants over each topic.
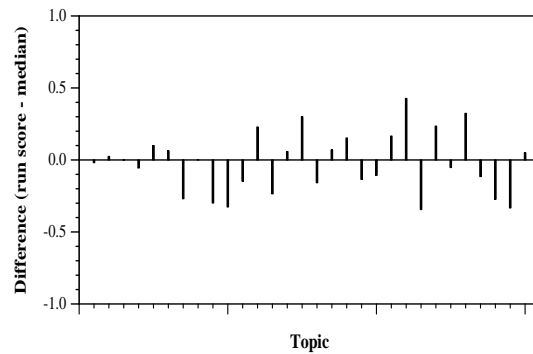
Summary Statistics	
Run ID	wistud_bing
Topic type	manual
Contributed to judgment sets?	yes

Overall measures	
Number of topics	30
Total number retrieved	16223
Total relevant	2352
Total relevant retrieved	970
MAP	0.1269
Mean Bpref	0.3110
Mean NDCG@10	0.3362

Document Level Averages	
	Precision
At 5 docs	0.4467
At 10 docs	0.3733
At 15 docs	0.3289
At 20 docs	0.3117
At 30 docs	0.2522
R-Precision	
Exact	0.1722

Document Counts at Rank 50
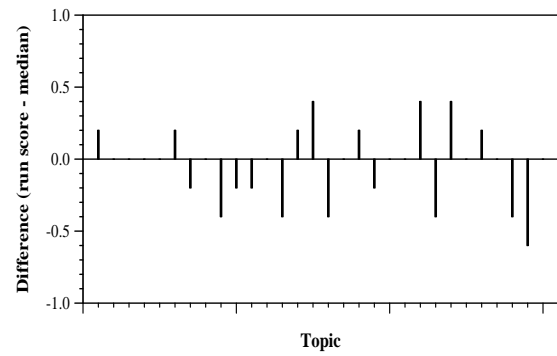

Per-topic difference from median bpref for all Round 1 runs

Per-topic difference from median NDCG@10 for all Round 1 runs

Per-topic difference from median P@5 for all Round 1 runs