

Round 1 results — Run wistud_indri submitted from wistud

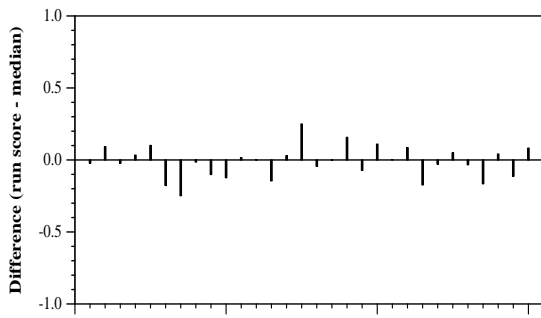
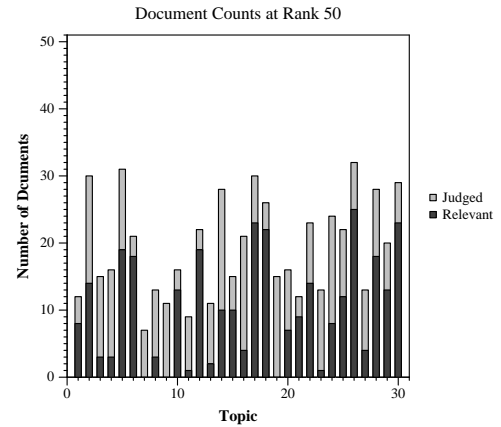
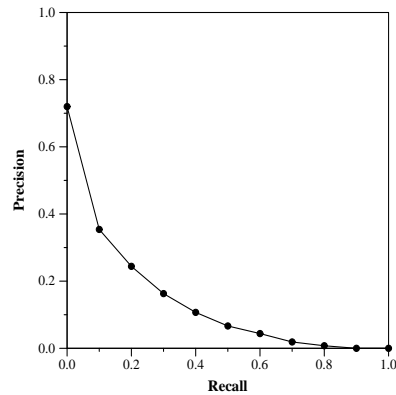
Run Description

45 participants undertook a manual task using our experimental search system (using the collection indexed with Indri, using QL with $\mu=2500$). 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.

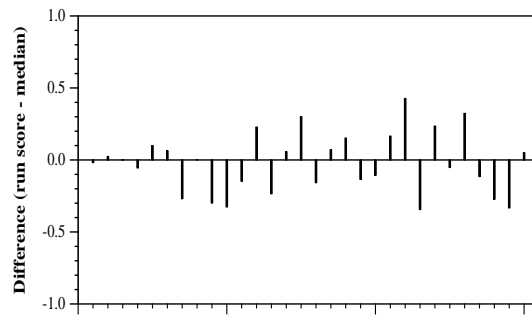
Summary Statistics	
Run ID	wistud_indri
Topic type	manual
Contributed to judgment sets?	no

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

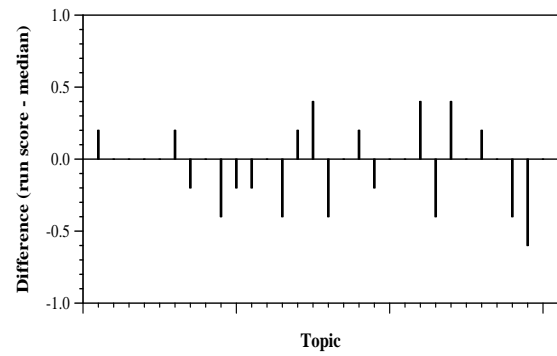


Per-topic difference from median bpref for all Round 1 runs



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

Round 1 results — Run wistud_indri submitted from wistud



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