

Round 2 results — Run allFiltering submitted from IRLabKU

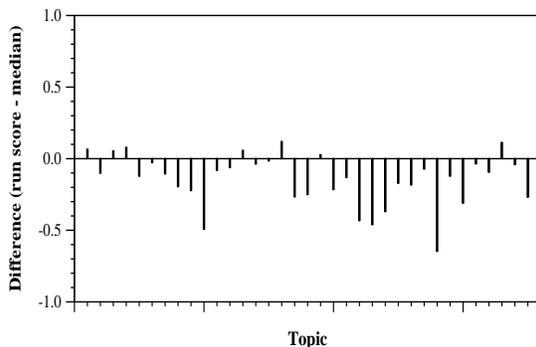
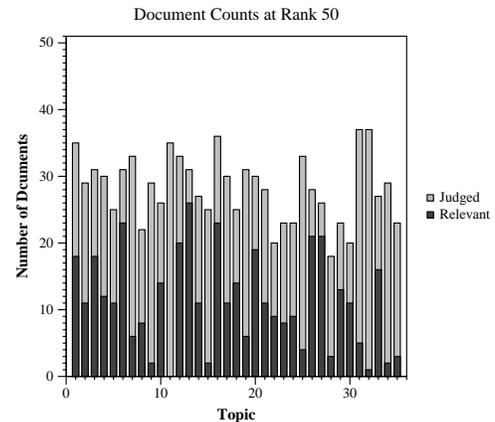
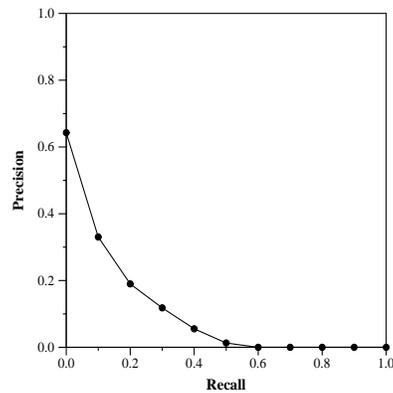
Run Description

–We create 3 indexes: 1) Based on title+abstract, 2) Based on full-text, and 3) Based on title+abstract+paragraph (a document is split into para1, para2, ..., paraK, and we create K+1 documents of the form title+abstract, title+abstract+para1, title+abstract+para2. –We create 4 different types of queries: 1) Query + named_entities(question), 2) Query + named_entities(question) + named_entities(narrative), 3) Question + named_entities(query), and 4) Question + named_entities(query) + named_entities(narrative) –For each query type we create 4 runs (default BM25 + default RM3): 1) We search in index of title+abstract, 2) We search in index of full-text, 3) We search in index of title+abstract+paragraph (where we only keep the first occurrence of a document and delete the following duplicates), and 4) We search in index of title+abstract+paragraph (where we don't remove the duplicates) Finally, based on these 4 runs, we use reciprocal rank fusion to get 1 run. –Now, we have 4 fusion run, which we use to create a final reciprocal rank fusion (Fusion of Fusions). We also create a fusion of the 16 runs without an intermediate fusion (Fusion of Runs). –Finally, we use reciprocal rank fusion between the 27 automatic runs, that did not contribute to the pooling, from round 1, as well as our own 2 fusion runs.

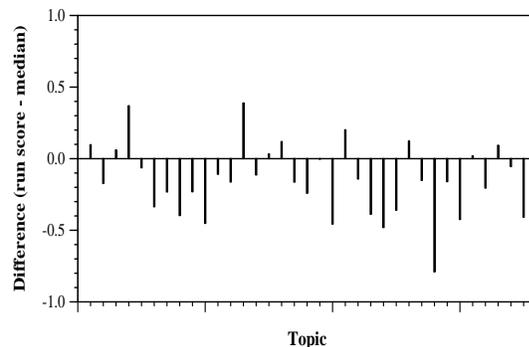
Summary Statistics	
Run ID	allFiltering
Topic type	automatic
Contributed to judgment sets?	yes

Overall measures	
Number of topics	35
Total number retrieved	35000
Total relevant	3002
Total relevant retrieved	1224
MAP	0.0981
Mean Bpref	0.2267
Mean NDCG@10	0.3237
Mean RBP(p=0.5)	0.3488 +0.0019

Document Level Averages	
	Precision
At 5 docs	0.4343
At 10 docs	0.4086
At 15 docs	0.3505
At 20 docs	0.3286
At 30 docs	0.2876
R-Precision	
Exact	0.1733

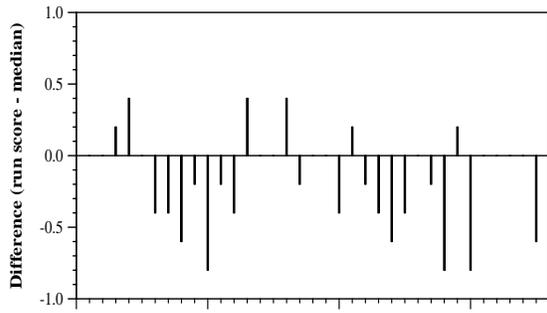


Per-topic difference from median bpref for all Round 2 runs

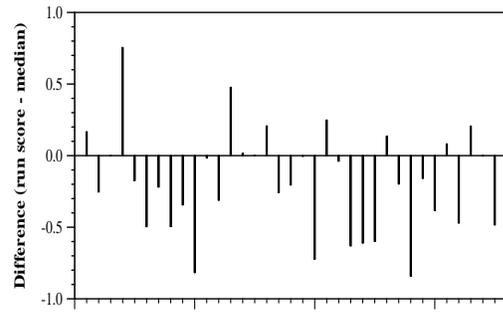


Per-topic difference from median NDCG@10 for all Round 2 runs

Round 2 results — Run allFiltering submitted from IRLabKU



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