

Round 2 results — Run TMACC_SeTA_tfidf_q submitted from TMACC_SeTA

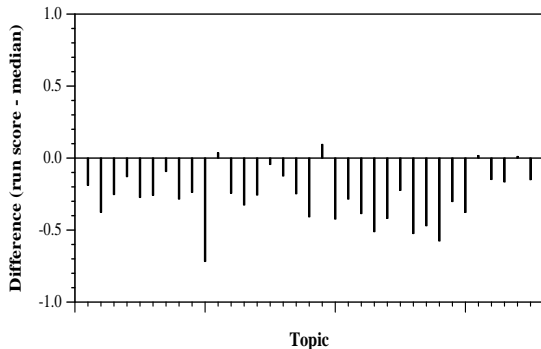
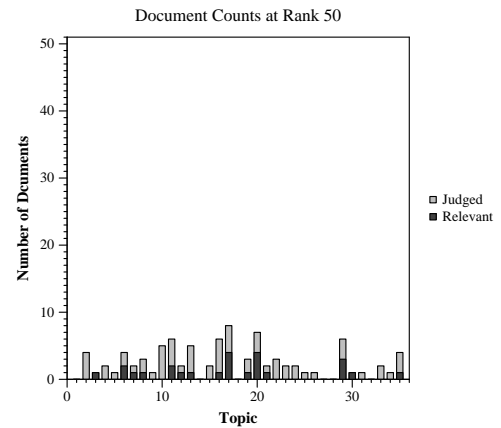
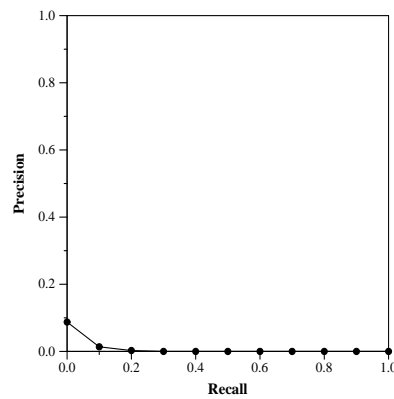
Run Description

The submission was fully automatic. Topic <query> and <question> elements were Stop Word and POS tag filtered (using nltk) to produce a keyword list, that was expanded by querying a Word2Vec model (gensim) of the CORP-19 dataset for 10 most related terms, filtered by tfidf scores on the same corpus. Expanded keyword list was encoded as a vector in a doc2vec model. This model was trained using gensim lib after pre-processing for phrase extraction with spacy, textacy and Abner NER for multi-word named entity. The 1000 most related doc were returned.

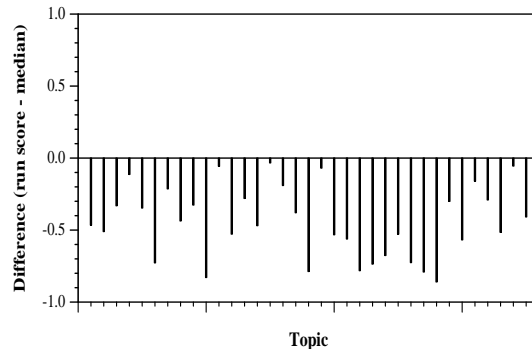
Summary Statistics	
Run ID	TMACC_SeTA_tfidf_q
Topic type	automatic
Contributed to judgment sets?	no

Overall measures	
Number of topics	35
Total number retrieved	35000
Total relevant	3002
Total relevant retrieved	376
MAP	0.0039
Mean Bpref	0.1060
Mean NDCG@10	0.0270
Mean RBP(p=0.5)	0.0278 +0.8940

Document Level Averages	
	Precision
At 5 docs	0.0286
At 10 docs	0.0314
At 15 docs	0.0286
At 20 docs	0.0214
At 30 docs	0.0190
R-Precision	
Exact	0.0127

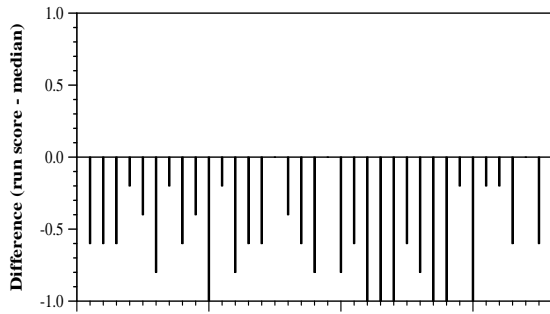


Per-topic difference from median bpref for all Round 2 runs

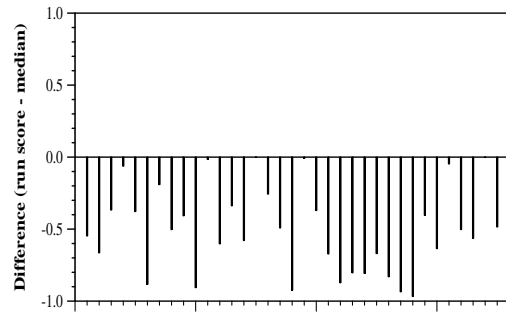


Per-topic difference from median NDCG@10 for all Round 2 runs

Round 2 results — Run TMAcc_SeTA_tfidf_q submitted from TMAcc_SeTA



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