

Round 2 results — Run TMACC_SeTA_tfidf submitted from TMACC_SeTA

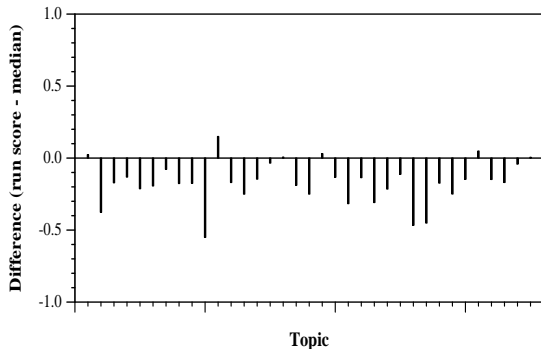
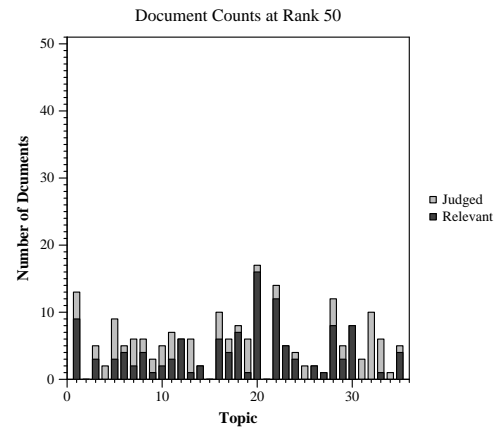
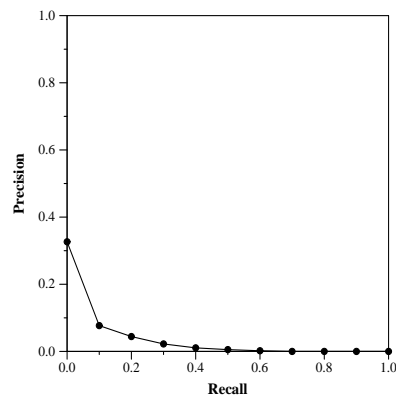
Run Description

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 model was then filtered by using tfidf scores. The 1000 most related doc were returned.

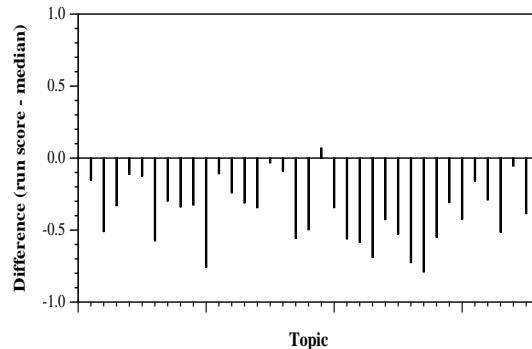
Summary Statistics	
Run ID	TMACC_SeTA_tfidf
Topic type	automatic
Contributed to judgment sets?	no

Overall measures	
Number of topics	35
Total number retrieved	35000
Total relevant	3002
Total relevant retrieved	732
MAP	0.0243
Mean Bpref	0.2018
Mean NDCG@10	0.1011
Mean RBP(p=0.5)	0.1286 +0.8144

Document Level Averages	
	Precision
At 5 docs	0.1314
At 10 docs	0.1143
At 15 docs	0.0952
At 20 docs	0.0900
At 30 docs	0.0867
R-Precision	
Exact	0.0614

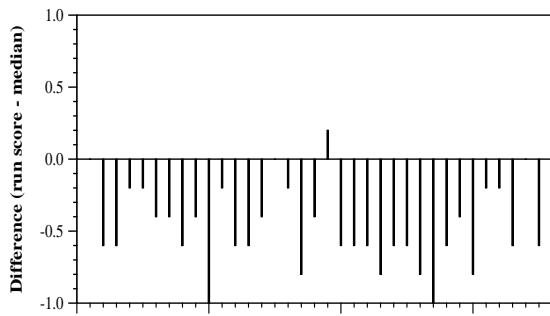


Per-topic difference from median bpref for all Round 2 runs

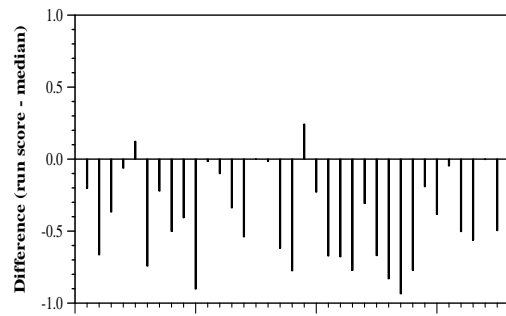


Per-topic difference from median NDCG@10 for all Round 2 runs

Round 2 results — Run TMACC_SeTA_tfidf 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