## Round 2 results — Run TMACC_SeTA_sent2vec submitted from TMACC_SeTA

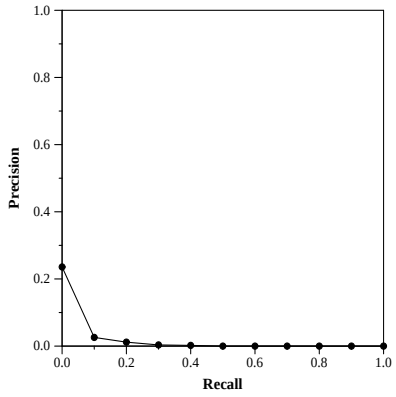
### Run Description

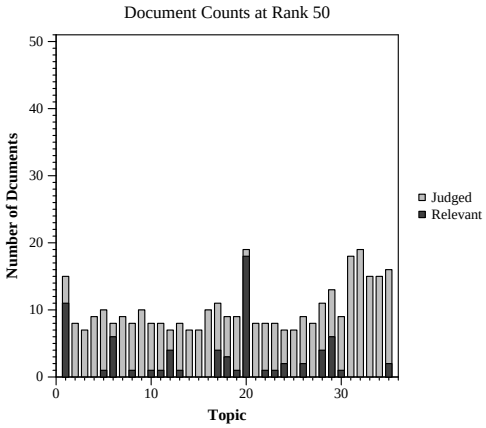
The submission was fully automatic. Topic <query> and <question> elements were Stop Word and POS tag filtered (using nltk) to produce a keyword list, that was expanded by querying a Word2Vec model (gensim) of the CORD-19 dataset for 10 most related terms, filtered by tfidf scores on the same corpus. Expanded keyword list was encoded as a vector in a doc2vec model, where each sentence is treated as document, while keeping reference to the id the corpus article it belongs to. 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 by aggregating and summing up the relevance scores of the top sentences returned by sent2vec.

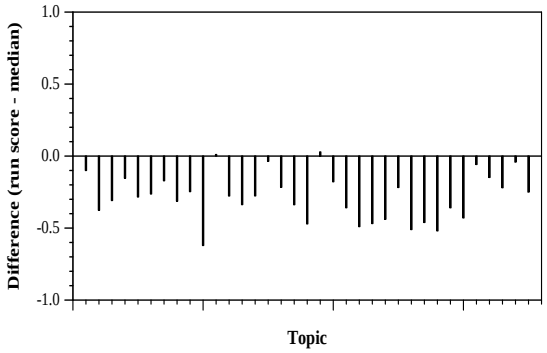
| Summary Statistics | |
|---|---|
| Run ID | TMACC_SeTA_sent2vec |
| Topic type | automatic |
| Contributed to judgment sets? | yes |

| Overall measures | |
|---|---|
| Number of topics | 35 |
| Total number retrieved | 35000 |
| Total relevant | 3002 |
| Total relevant retrieved | 352 |
| MAP | 0.0114 |
| Mean Bpref | 0.0877 |
| Mean NDCG@10 | 0.0702 |
| Mean RBP(p=0.5) | 0.0892 +0.0065 |

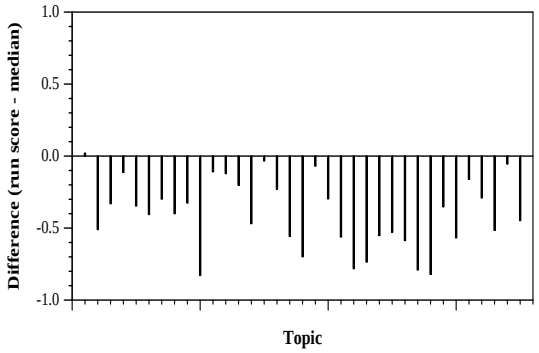
| Document Level Averages | |
|---|---|
| | Precision |
| At 5 docs | 0.1371 |
| At 10 docs | 0.1000 |
| At 15 docs | 0.0724 |
| At 20 docs | 0.0643 |
| At 30 docs | 0.0562 |
| R-Precision | |
| Exact | 0.0295 |



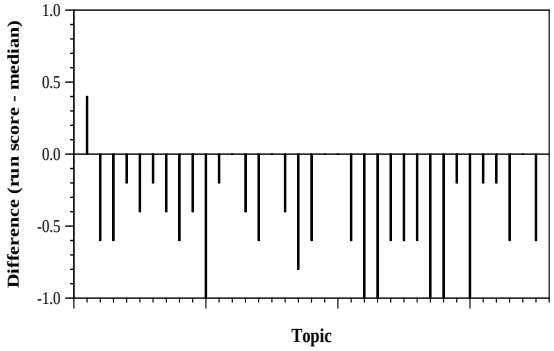




Per-topic difference from median bpref for all Round 2 runs

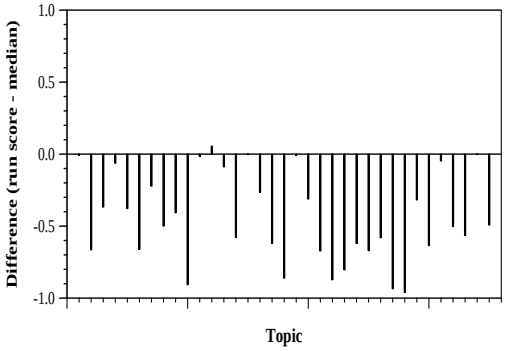



Per-topic difference from median NDCG@10 for all Round 2 runs

## Round 2 results — Run TMACC_SeTA_sent2vec 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