Round 1 results — Run ir_covid19_cle_dfr submitted from IR_COVID19_CLE

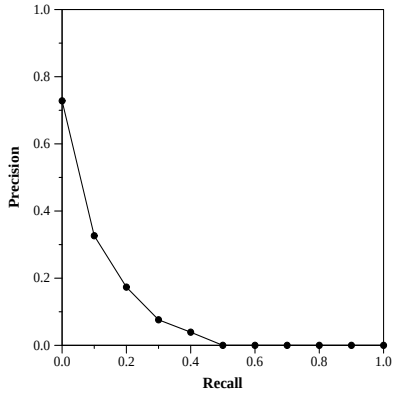
Run Description

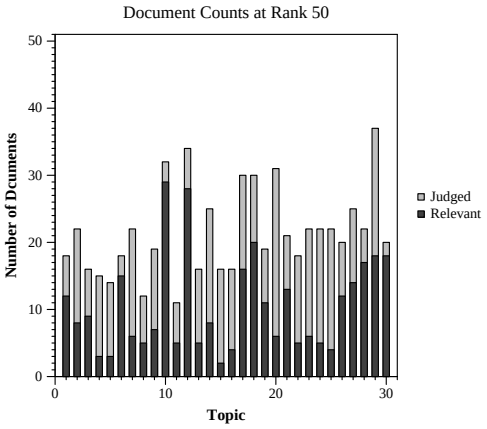
We have used the data set with all the documents from corpus Commercial use subset, Non-commercial use subset, Custom license subset and bioRxiv/medRxiv subsets. We used "Paper_id", "Title Id" and "Abstract" to index all the documents using Apache Lucene. We have indexed every document for all tokens present with in the document. However, in a collection of documents these tokens can be repeating in multiple documents as well. Here, we use inverted index to store tokens repeating in multiple indexes, so that when searched for a specific token, we can narrow down the search documents specifically all documents that token is present. We have used the query of the topic for querying the index. We parsed the query with English Analyzer and searched on the abstract text field of the index. For each query, We have retrieved the Top 100 documents and the relevance scores using Divergence from Randomness (DFR) similarity model, which is based on randomness model, first normalization and term frequency normalization. Reference Paper: RIJSBERGEN, C., & AMATI, G. (2002). Probabilistic models of information retrieval based on measuring the divergence from randomness. ACM Trans. Inf. Syst. 20, 4 (October 2002).

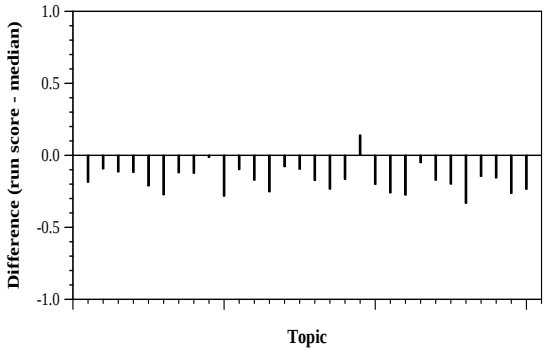
Summary Statistics	
Run ID	ir_covid19_cle_dfr
Topic type	automatic
Contributed to judgment sets?	yes

Overall measures	
Number of topics	30
Total number retrieved	2564
Total relevant	2352
Total relevant retrieved	385
MAP	0.0902
Mean Bpref	0.1601
Mean NDCG@10	0.3600

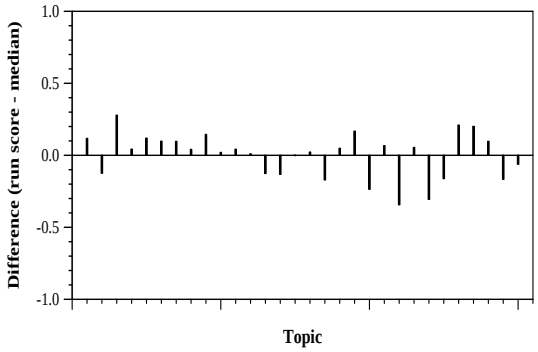
Document Level Averages	
	Precision
At 5 docs	0.5067
At 10 docs	0.4233
At 15 docs	0.3600
At 20 docs	0.3300
At 30 docs	0.2822
R-Precision	
Exact	0.1633





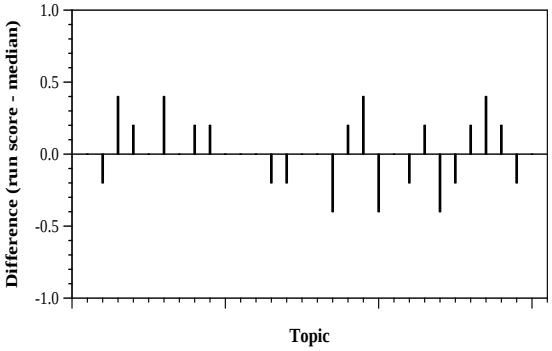


Per-topic difference from median bpref for all Round 1 runs



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

Round 1 results — Run ir_covid19_cle_dfr submitted from IR_COVID19_CLE



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