

Round 1 results — Run ir_covid19_cle_ib submitted from IR_COVID19_CLE

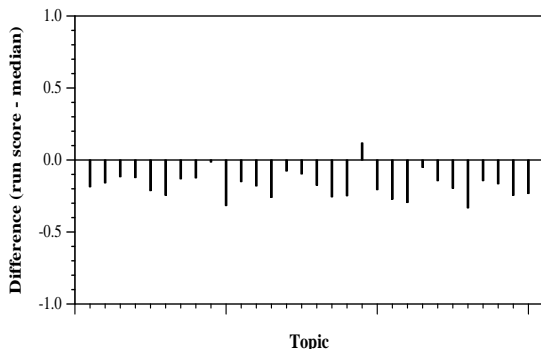
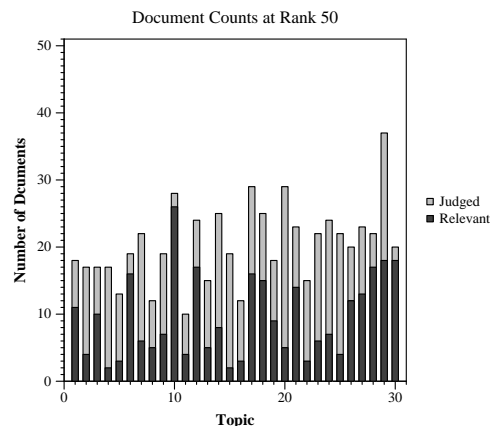
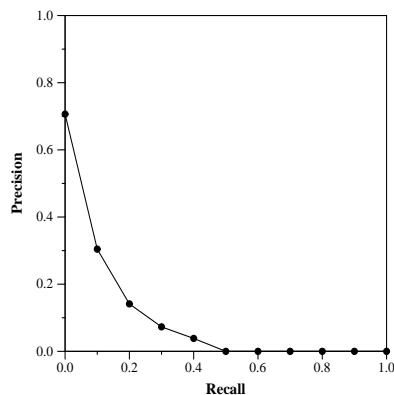
Run Description

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 Information based (IB Similarity) which models rely on normalized values of occurrence of a word in documents. Information Models are characterized by three elements that are normalization function, probability distribution and retrieval function. Reference Paper: Clinchant , S., & Gaussier, E. (2010). Information-based models for ad hoc IR. In Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval (SIGIR '10).

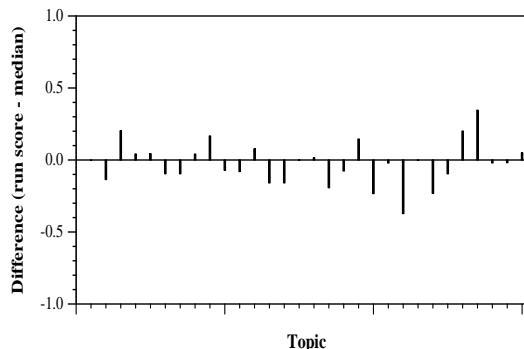
Summary Statistics	
Run ID	ir_covid19_cle_ib
Topic type	automatic
Contributed to judgment sets?	no

Overall measures	
Number of topics	30
Total number retrieved	2559
Total relevant	2352
Total relevant retrieved	361
MAP	0.0814
Mean Bpref	0.1517
Mean NDCG@10	0.3342

Document Level Averages	
	Precision
At 5 docs	0.4867
At 10 docs	0.3867
At 15 docs	0.3533
At 20 docs	0.3083
At 30 docs	0.2600
R-Precision	
Exact	0.1560

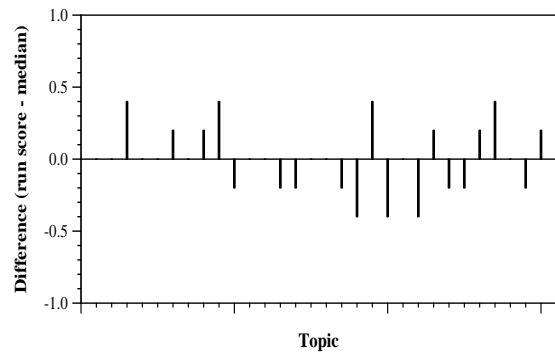


Per-topic difference from median bpref for all Round 1 runs



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

Round 1 results — Run ir_covid19_cle_ib submitted from IR_COVID19_CLE



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