## Run Description

We have used the data set with all the documents from corpus. Using the metadata available, we added cord_id. We used around 48000 files for indexing. We used "Paper_id", "Title Id" and "Abstract" and "body_text" 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 LM Jelinek-Mercer similarity. Reference Paper: Zhai , C., & Lafferty, J. (2001). A study of smoothing methods for language models applied to Ad Hoc information retrieval. In Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '01).

Summary Statistics	
Run ID	ir_covid19_cle_LMJ
Topic type	feedback
Contributed to judgment sets?	no

Overall measures	
Number of topics	35
Total number retrieved	30954
Total relevant	3002
Total relevant retrieved	1436
MAP	0.1090
Mean Bpref	0.3042
Mean NDCG@10	0.3046
Mean RBP(p=0.5)	0.3378 +0.2634

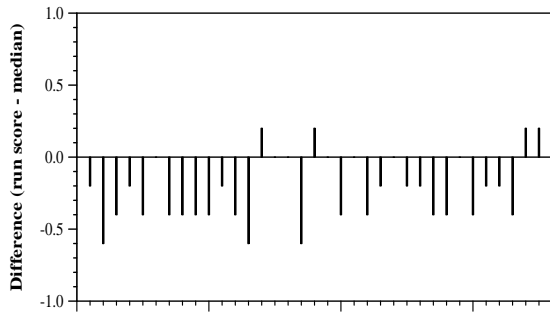| Document Level Averages	
	Precision
At 5 docs	0.3771
At 10 docs	0.3400
At 15 docs	0.2933
At 20 docs	0.2757
At 30 docs	0.2390
R-Precision	
Exact	0.1673

Document Counts at Rank 50
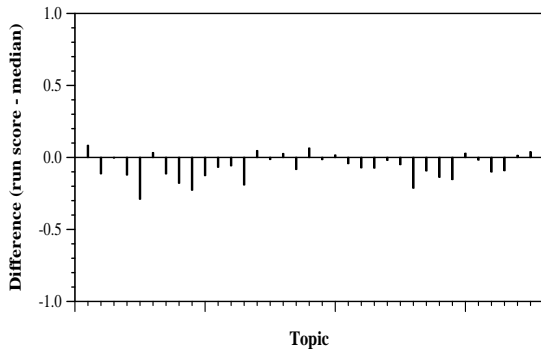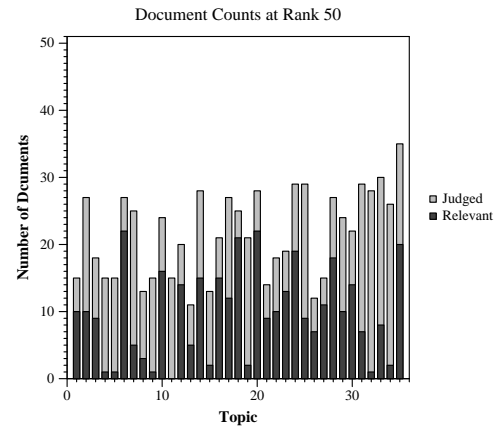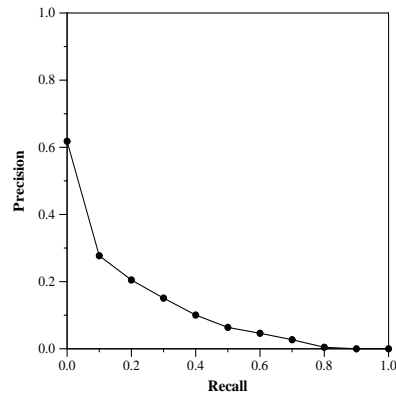
Per-topic difference from median bpref for all Round 2 runs
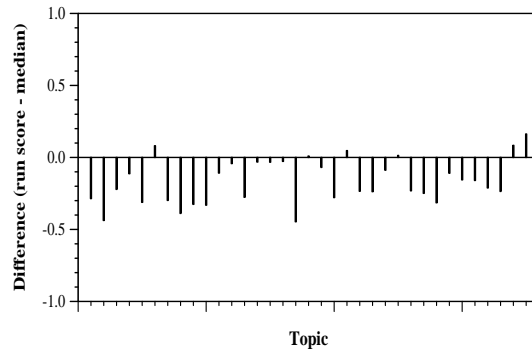
Per-topic difference from median NDCG@10 for all Round 2 runs

Per-topic difference from median P@5 for all Round 2 runs
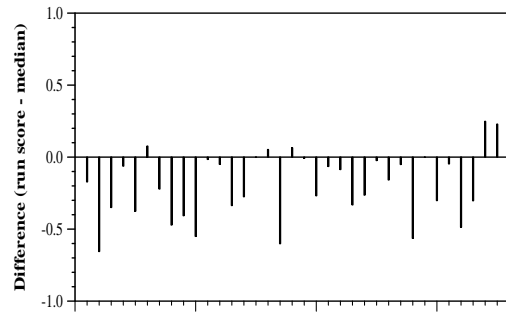
Per-topic difference from median RBP(p=0.5) for all Round 2 runs