

## Round 1 results — Run ir\_covid19\_cle\_lmd submitted from IR\_COVID19\_CLE

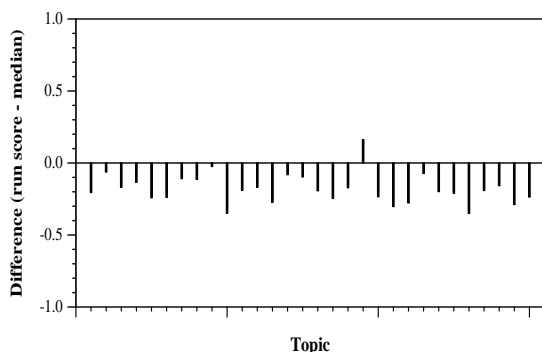
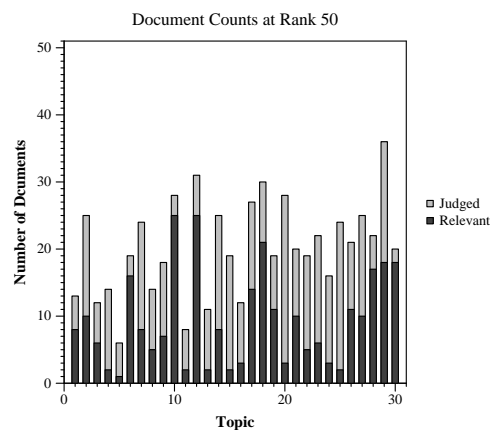
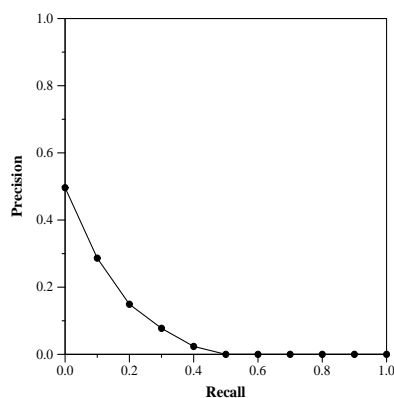
### Run Description

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 LM Dirichlet similarity which are language model-based similarities that rely on relative frequency of occurrence of words in documents. In LMD model is the best out of box model for short queries. For our approach we have used a default value for  $k = 2000$  Reference Paper: Zhai, C., & Lafferty, J. (2001). A Study of Smoothing Methods for Language Models. Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '01). ACM, .

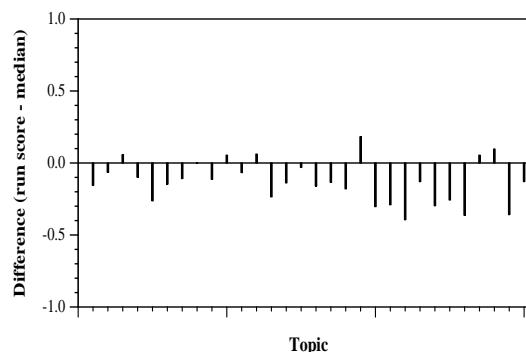
| Summary Statistics            |                    |
|-------------------------------|--------------------|
| Run ID                        | ir_covid19_cle_lmd |
| Topic type                    | automatic          |
| Contributed to judgment sets? | no                 |

| Overall measures         |        |
|--------------------------|--------|
| Number of topics         | 30     |
| Total number retrieved   | 2601   |
| Total relevant           | 2352   |
| Total relevant retrieved | 349    |
| MAP                      | 0.0703 |
| Mean Bpref               | 0.1454 |
| Mean NDCG@10             | 0.2286 |

| Document Level Averages |           |
|-------------------------|-----------|
|                         | Precision |
| At 5 docs               | 0.3267    |
| At 10 docs              | 0.3067    |
| At 15 docs              | 0.2822    |
| At 20 docs              | 0.2833    |
| At 30 docs              | 0.2400    |
| R-Precision             |           |
| Exact                   | 0.1513    |



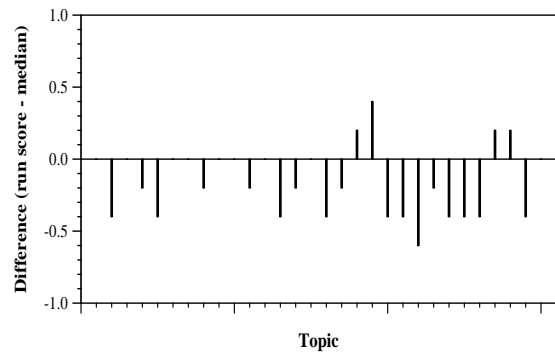
Per-topic difference from median bpref for all Round 1 runs



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

Round 1 results — Run ir\_covid19\_cle\_lmd submitted from IR\_COVID19\_CLE

---



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