

## CS980 Assignment 1

Ajesh Vijayaragavan

MS Computer Science

### Search and Indexing:

#### BM25

- Primary method implemented is BM25 similarity which is a ranking function that ranks documents according to relevance. Lucene index for paragraphCorpus collection from trec-carv2.0 using TREC-CAR dataset (Cited below).
- Search retrieval model and run file formation.
- Used trec-car tool bindings as maven dependency provided in pom.xml.
- Baseline: BM25 section path query.
- Ground Truth: benchmarkY1train article.qrels (For Pages)
  - benchmarkY1train hierarchical.qrels (For Sections).

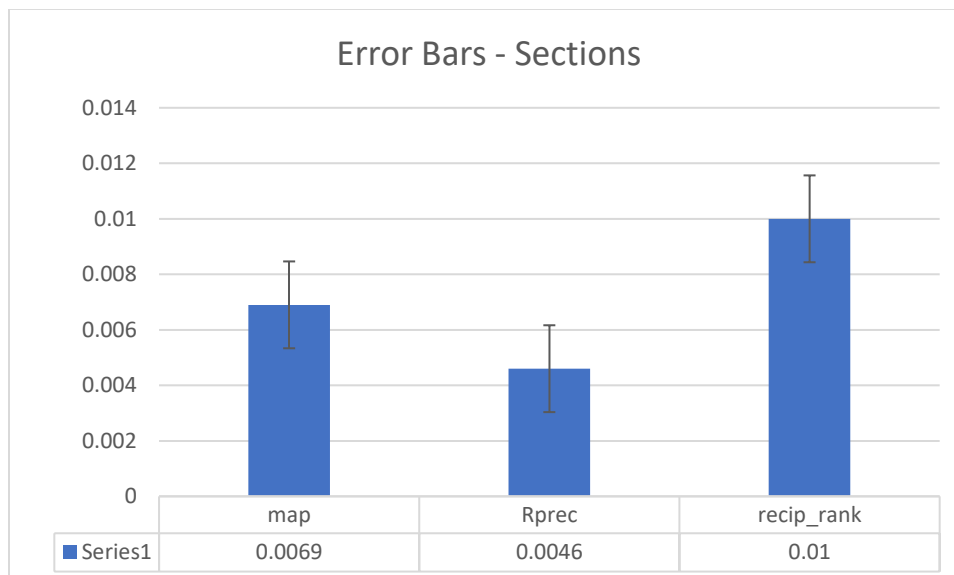
#### Evaluation Results

1. Gives the evaluation results with standard error bars in a table (one row per method, one column per measure)

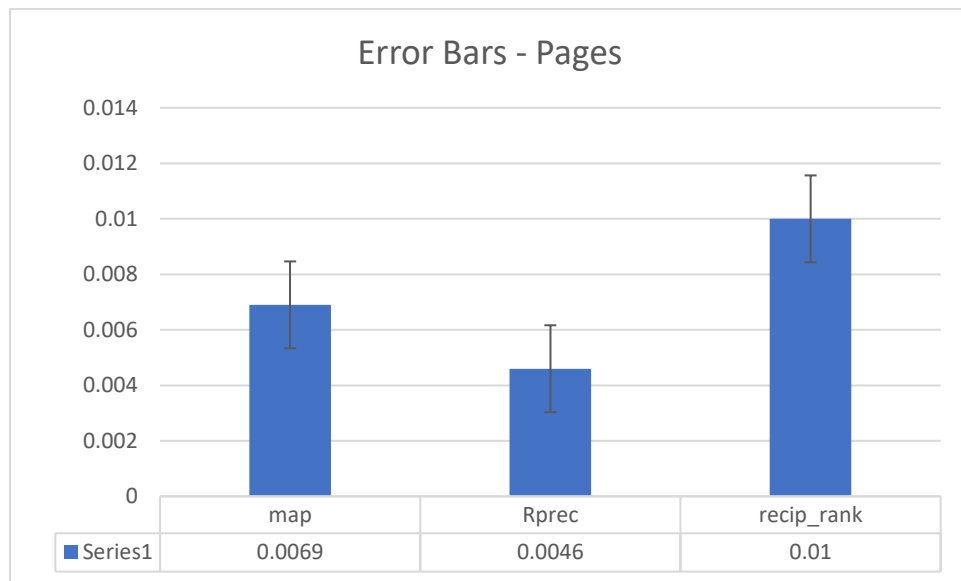
- Row : method, Column: Evaluation Measures.

Method	Mean Average Precision (MAP)	Mean Reciprocal Rank (MRR)	R-Precision (Precision@R)
BM25-Sections-all	0.0810	0.1386	0.4035
BM25-Pages-all	0.0069	0.0046	0.0010

- Evaluation methods : Mean Average Precision (MAP), Mean Reciprocal Rank (MRR), R-Precision (Precision@R).
- Error bar for sections using hierarchical.qrels



- Error bar for pages using article.qrels



**Github Link:** <https://github.com/Ajeshv/Lucene-search-and-indexing>

**Dataset Citation:**

Laura Dietz, Ben Gamari.

"TREC CAR: A Data Set for Complex Answer Retrieval".

Version 1.5, 2017.

<http://trec-car.cs.unh.edu>