ILS-Z534 - Assignment 2: Retrieval Algorithm and Evaluation

Prahasan Gadugu

Summary of the findings:

- A closer look at the table shows that the recall increases gradually from @5 to @100. Recall should 1 when all documents are retrieved without taking relevant score into consideration.
- On the contrary, the precision values drop gradually.
- We know that Precision = True Positives/(True Positives + False Positives)
- Therefore, as we retrieve more documents, precision value should drop since it will penalize misclassified documents with each result.
- Mean Average Precision (MAP) calculates the performances of relevancy tests over several queries. In this case, MAP has been calculated for 50 queries. Precision with just one query is equivalent to MAP for just 1 query.
- NDCG calculates the ratio of Discounted Cumulative Gain to the Ideal Discounted Cumulative Gain.
- In this case, NDCG is also calculated for 50 queries. Therefore, if the ideal ranks match with our model ranks, the NDCG will be 1.
- My model shows that NDCG is between range 0.1-0.2 meaning that our model is at an offset of 0.2 from the ideal scenario.
- My algorithm implements Classic Similarity model using Standard Analyzer to remove stop words affecting the initial search (easySearch.java). Given a string input it would give the relevance score.
- The evaluation tables for different algorithms for short and Long query are given below,

Short Query Evaluation Table for different algorithms:

Evaluation Metric	My Algorithm (TF-IDF)	Vector Space Model (Classic Similarity)	BM25	Language Model with Dirichlet Smoothing	Language Model with Jelinek Mercer Smoothing
P@5	0.148	0.296	0.308	0.348	0.288
P@10	0.146	0.302	0.3	0.326	0.28
P@20	0.135	0.26	0.271	0.289	0.243
P@100	0.098	0.1648	0.1678	0.169	0.1602
Recall@5	0.0234	0.0539	0.0488	0.062	0.0534
Recall@10	0.0539	0.096	0.0879	0.0995	0.0911
Recall@20	0.0839	0.1416	0.1378	0.1467	0.1302
Recall@100	0.2196	0.3578	0.3572	0.3468	0.332
MAP	0.105	0.1976	0.2005	0.2059	0.1932
MRR	0.2863	0.4696	0.4772	0.4785	0.4637
NDCG@5	0.1627	0.3116	0.3246	0.3517	0.3063
NDCG@10	0.1621	0.3183	0.3199	0.342	0.3011
NDCG@20	0.1619	0.3043	0.3124	0.3326	0.2888
NDCG@100	0.1919	0.3214	0.3259	0.3311	0.3121

Long Query Evaluation Table for different algorithms:

Evaluation Metric	My Algorithm (TF-IDF)	Vector Space Model (Classic Similarity)	BM25	Language Model with Dirichlet Smoothing	Language Model with Jelinek Mercer Smoothing
P@5	0.128	0.256	0.284	0.256	0.232
P@10	0.124	0.244	0.244	0.242	0.214
P@20	0.114	0.22	0.234	0.234	0.212
P@100	0.0742	0.1406	0.1488	0.146	0.1378
Recall@5	0.0187	0.0349	0.0402	0.0403	0.0406
Recall@10	0.0368	0.0621	0.0703	0.0708	0.0658
Recall@20	0.0595	0.1051	0.1159	0.127	0.1136
Recall@100	0.1743	0.2929	0.3167	0.334	0.2901
MAP	0.0664	0.153	0.1676	0.1586	0.1514
MRR	0.2563	0.4523	0.4597	0.3475	0.364
NDCG@5	0.1388	0.2819	0.3029	0.2499	0.2348
NDCG@10	0.1341	0.2682	0.2729	0.2473	0.2294
NDCG@20	0.131	0.2577	0.2718	0.259	0.241
NDCG@100	0.1466	0.2694	0.2872	0.2753	0.2609