HomeWork-1 Report

Score of Top Relevant File of a Sample Query for each Retrieval Model

Model	Score
ES (built-in)	11.46703
Okapi TF	1.9071363603383173
TF-IDF	4.493555637486965
Okapi BM-25	11.456762917351623
Unigram LM with Laplace smoothing	967.8645382931072
Unigram LM with Jelinek-Mercer smoothing	986.9008829501246

Inference on the above results

Sample query considered: Query number 85

ES (built-in): This model takes into consideration term frequency, inverse document frequency, and field-length normalization. Good score here indicates that the retrieved document has high term frequency for important query terms and low document frequency across the indexed documents. This model uses a different version of BM.

Okapi TF: Overall Okapi-TF gives a lower score compared to other models. Term frequencies are not that high for query terms but this model does not consider IDF or normalize document length.

TF-IDF: Retrieves a document with relatively high term frequency for important query terms while considering their rarity across the corpus.

Okapi BM-25: Retrieves documents with optimal balance between term frequency, inverse document frequency, and document length normalization.

Unigram LM with Laplace smoothing: The significantly higher score suggests the document contains all query terms with high frequency, effectively penalizing unseen terms.

Unigram LM with Jelinek-Mercer smoothing: Retrieves documents with balanced term frequency and document frequency distribution, reflecting interpolation between document-specific and corpus-wide probabilities.

Retrieval Model Performance

[Highlight the scores more than 0.28]

Model	Average Precision	Precision at 10	Precision at 30
ES (built-in)	0.3085	0.4600	0.3853
Okapi TF	0.2718	0.4400	0.3760
TF-IDF	0.2955	0.4320	0.3813
Okapi BM-25	0.3033	0.4520	0.3787
Unigram LM with Laplace smoothing	0.3211	0.5200	0.3920
Unigram LM with Jelinek-Mercer smoothing	0.2800	0.4160	0.3760

Inference on above retrieval model results

ES (built-in): Performs well overall, with high Average Precision and effective retrieval within the top 10 and top 30 results.

Okapi TF and TF-IDF: Achieve decent precision within the top 10 and top 30 results, but slightly lower Average Precision compared to ES.

Okapi BM-25: Shows better overall performance than Okapi TF and TF-IDF, with higher Average Precision and effective retrieval within the top 10 and top 30 results.

Unigram LM with Laplace smoothing: Shows a higher performance with the highest Average Precision and excellent precision within the top 10 and top 30 results. The Unigram LM with Laplace smoothing model excels in precision due to its ability to handle out-of-vocabulary terms, reduce overfitting, and mitigate the impact of sparsity.

Unigram LM with Jelinek-Mercer smoothing: While achieving reasonable performance, precision within the top 10 and top 30 results is slightly lower compared to Laplace smoothing.

Pseudo-relevance Feedback Improvements[ONLY MS STUDENTS]

[The highlighted scores that indicate an improvement in the average precision score of the model]

1. Result after adding the top 5 distinctive terms to each query.

Model	Average Precision	Precision at 10	Precision at 30
ES (built-in)	0.3129	0.4720	0.3853
Okapi TF	0.2744	0.4400	0.3720
TF-IDF	0.2999	0.4320	0.3813
Okapi BM-25	0.3078	0.4640	0.3787
Unigram LM with Laplace smoothing	0.3135	0.5040	0.3907
Unigram LM with Jelinek-Mercer smoothing	0.2823	0.4160	0.3773

2. Results after adding top 5 significant terms from Elasticsearch aggs to each query.

Model	Average Precision	Precision at 10	Precision at 30
ES (built-in)	0.3349	0.4760	0.4080
Okapi TF	0.2988	0.4640	0.3973
TF-IDF	0.3237	0.4640	0.4027
Okapi BM-25	0.3307	0.4800	0.4013
Unigram LM with Laplace smoothing	0.3357	0.5240	0.4080
Unigram LM with Jelinek-Mercer smoothing	0.3191	0.4600	0.4027

Inference on the above pseudo-relevance results

The results demonstrate that incorporating pseudo-relevance feedback, particularly by adding top distinctive terms or significant terms from Elasticsearch aggregations to the queries, leads to improvements in average precision scores across all retrieval models. This enhancement is evident in both the Precision at 10 and Precision at 30 metrics as well. Pseudo relevance feedback has overall helped to improve the performance of the models.

Table showing the Query used for Evaluation – Top 5 distinctive terms

Query number	95	87	97	68	77
Original Query	Document must describe a computer application to crime solving.	Document will report on current criminal actions against officers of a failed U.S. financial institution.	Document must identify instances of fiber optics technology actually in use.	Document will report actual studies, or even unsubstantiated concerns about the safety to manufacturing employees and installation workers of fine-diameter fibers used in insulation and other products.	Document will report a poaching method used against a certain type of wildlife.

Processed Query	comput crime	offic institut	fiber optic technolog	safeti worker diamet fiber	poach wildlif
Processed Query - Pseudo RF **(Only MS students)**	comput crime feder fbi viru ncic hacker	offic institut feder billion	fiber optic technolog satellit	safeti worker diamet fiber asbesto	poach wildlif wa

Table showing the Query used for Evaluation – Top 5 significant terms from ES aggs

Query number	59	64	77	87	97
Original Query	Document will report a type of weather event which has directly caused at least one fatality in some location.	Document will report an event or result of politically motivated hostage-taking.	Document will report a poaching method used against a certain type of wildlife.	Document will report on current criminal actions against officers of a failed U.S. financial institution..	Document must identify instances of fiber optics technology actually in use.
Processed Query	weather least locat	coup attempt	poach wildlif	offic institut	fiber optic technolog
Processed Query - Pseudo RF **(Only MS students)**	weather least locat forecast temperatur rain wind snow	hostag hostages lebanon proirania n shiit kidnap	poach wildlif poacher tusk poaching antipoach poachers	offic institut research institute thrift institutions deposit	fiber optic technolog hairthin space mad resolidifi fibers telescope

References:

https://www.elastic.co/guide/en/elasticsearch/reference/current/docs-termvectors.html

Used ChatGpt for debugging some part of the code