

Homework-1 Report

Score of Top Relevant File of a Sample Query for each Retrieval Model – Sample Query Selected - 85

Model	Score
ES (built-in)	10.106513
Okapi TF	3.232
TF-IDF	2.593
Okapi BM-25	6.712
Unigram LM with Laplace smoothing	-26.993
Unigram LM with Jelinek-Mercer smoothing	-8.244

Inference on the above results

The built-in Elasticsearch model rates the document as quite relevant with a score of 10.10, while Okapi TF and TF-IDF models assign lower scores of 3.232 and 2.593, respectively, indicating a slightly lower perceived relevance. The Okapi BM-25 model shows a score of 6.712, reflecting its effectiveness in modern search engines. Language models with Laplace and Jelinek-Mercer smoothing show negative scores (-51.094 and -28.412), which is typical due to their probabilistic nature and log probability outputs, indicating comparatively lower relevance within their scoring system but not necessarily implying absolute irrelevance.

Retrieval Model Performance

[Highlight the scores more than 0.28]

Model	Average Precision	Precision at 10	Precision at 30
ES (built-in)	0.1925	0.3087	0.2725
Okapi TF	0.2448	0.4080	0.3280
TF-IDF	0.1696	0.2800	0.2507
Okapi BM-25	0.2517	0.4280	0.3227
Unigram LM with Laplace smoothing	0.2279	0.4400	0.3253
Unigram LM with Jelinek-Mercer smoothing	0.2672	0.4240	0.3333

Inference on above retrieval model results

The average precision scores from the given data suggest that the Okapi BM-25 model is the most effective, with the highest score, indicating its strength in balancing term frequency with document length normalization for this specific corpus. The Unigram Language Model with Jelinek-Mercer smoothing follows closely, hinting at its proficiency in handling the probabilistic nature of language within this dataset. Notably, the traditional TF-IDF model lags behind, reflecting its limitations in capturing term significance across documents. The ES built-in model shows reasonable performance, but it does not lead in this metric, which could be due to the nuances of this test collection or the nature of the queries not aligning perfectly with the algorithms ES employs. The lower performance of the Unigram Language Model with Laplace smoothing suggests that its smoothing technique might not be as suited to the dataset as Jelinek-Mercer's approach.

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	0.1883	0.2840	0.2693
Okapi TF	0.2458	0.4040	0.3080
TF-IDF	0.1693	0.2840	0.2533
Okapi BM-25	0.2517	0.4280	0.3227
Unigram LM with Laplace smoothing	0.2239	0.4200	0.3133
Unigram LM with Jelinek-Mercer smoothing	0.2672	0.4240	0.3333

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

Model	Average Precision	Precision at 10	Precision at 30
Okapi TF	0.1812	0.3120	0.2533
TF-IDF	0.1478	0.2240	0.2307
Okapi BM-25	0.2225	0.3680	0.3053
Unigram LM with Laplace smoothing	0.1599	0.3160	0.2773
Unigram LM with Jelinek-Mercer smoothing	0.2458	0.3600	0.3147

Inference on the above pseudo-relevance results

The implementation of pseudo-relevance feedback via query expansion with the top 5 distinctive terms from Elasticsearch aggs has produced a varied impact across different information retrieval models. Notably, the Unigram Language Model with Jelinek-Mercer smoothing demonstrated an increase in average precision, suggesting that the additional terms were beneficial in enhancing the model's retrieval effectiveness. In contrast, models such as Okapi TF, TF-IDF, and Okapi BM-25 experienced a decline in average precision, implying that the inserted terms might have introduced noise, thereby diluting the relevance of retrieved documents. The precision at 10 and precision at 30 metrics further reflect these mixed results, indicating that the effectiveness of pseudo-relevance feedback is not uniform across models and heavily relies on the relevance of the added terms to the user's search intent.

Table showing the Query used for Evaluation

Query number	85	59	56	71	64
Original Query	Document will discuss allegations, or measures being taken against, corrupt public officials of any governmental jurisdiction worldwide	Document will report a type of weather event which has directly caused at least one fatality in some location.	Document will include a prediction about the prime lending rate, or will report an actual prime rate move.	Document will report incursions by land, air, or water into the border area of one country by military forces of a second country or a guerrilla group based in a second country	Document will report an event or result of politically motivated hostage-taking.
Processed Query	['alleg', 'measur', 'corrupt', 'public', 'offici', 'ani', 'government', 'jurisdict', 'worldwid']	['weather', 'directli', 'caus', 'least', 'fatal', 'locat']	['prime', 'lend', 'rate', 'prime', 'rate']	['border', 'area', 'countri', 'militari', 'forc', 'second', 'countri', 'guerrilla', 'base', 'second', 'countri']	['result', 'polit', 'motiv', 'hostagetak']
Processed Query - Pseudo RF (Only MS students)	['prison', 'polic', 'prosector', 'parti']	['tornado', 'thunderstorm', 'offic']	['bank', 'lower']	['rubber', 'lebanon', 'govern']	['north', 'novemb', 'speaker']