HomeWork-1 Report Template

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

|  |  |
| --- | --- |
| Model | Score |
| ES (built-in) | 0.3017 |
| Okapi TF | 0.2439 |
| TF-IDF | 0.3050 |
| Okapi BM-25 | 0.1817 |
| Unigram LM with Laplace smoothing | 0.2651 |
| Unigram LM with Jelinek-Mercer smoothing | 0.2481 |

**Inference on the above results**

TF-IDF and ES (built-in) are the top performers, suggesting that classic methods of scoring document relevance remain highly effective for the task at hand. Their success implies that distinguishing relevant documents based on the frequency of terms is crucial for this dataset.

The lower performance of Okapi BM-25 is unexpected given its reputation and suggests a possible mismatch with the dataset or task characteristics. It might be worth investigating the dataset and task setup to understand why BM-25 underperforms in this context.

Language model approaches (with both types of smoothing) show moderate performance, indicating their utility but also highlighting that more sophisticated or task-specific tuning might be necessary to outperform more straightforward statistical methods like TF-IDF.

# Retrieval Model Performance

# [ Highlight the scores more than 0.28]

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Average Precision | Precision at 10 | Precision at 30 |
| ES (built-in) | 0.2739 | 0.3560 | 0.3120 |
| Okapi TF | 0.2149 | 0.3600 | 0.3220 |
| TF-IDF | 0.2745 | 0.3960 | 0.3480 |
| Okapi BM-25 | 0.1472 | 0.2720 | 0.2440 |
| Unigram LM with Laplace smoothing | 0.2135 | 0.3600 | 0.2907 |
| Unigram LM with Jelinek-Mercer smoothing | 0.2021 | 0.3080 | 0.2573 |

**Inference on above retrieval model results**

TF-IDF's superior performance suggests its effectiveness in weighting term importance across the dataset, making it highly capable of identifying relevant documents both in early retrieval and on average.

Variability in model performance across different metrics underscores the importance of selecting the right model based on the specific needs of a retrieval task. For instance, if the goal is to maximize the relevance of the top 10 results, models like Okapi TF and the Unigram LM with Laplace Smoothing perform comparably to more complex models.

The generally lower performance of Okapi BM-25 may reflect the model's sensitivity to the specific dataset or task parameters, such as the document length distribution or the use of query-specific tuning parameters.

The effectiveness of language models (with smoothing) in certain contexts suggests that incorporating understanding of language structure and term distribution can be valuable, though they may require careful tuning and adaptation to match the performance of more straightforward statistical methods like TF-IDF.

# 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) | 0.2674 | 0.3720 | 0.3093 |
| Okapi TF | 0.1282 | 0.2600 | 0.1960 |
| TF-IDF | 0.2544 | 0.3800 | 0.3280 |
| Okapi BM-25 | 0.1156 | 0.2440 | 0.2027 |
| Unigram LM with Laplace smoothing | 0.2190 | 0.2720 | 0.2080 |
| Unigram LM with Jelinek-Mercer smoothing | 0.2131 | 0.2720 | 0.2027 |

1. 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) | 0.2760 | 0.3080 | 0.2867 |
| Okapi TF | 0.2172 | 0.3480 | 0.2933 |
| TF-IDF | 0.2509 | 0.3120 | 0.3053 |
| Okapi BM-25 | 0.1094 | 0.1560 | 0.1133 |
| Unigram LM with Laplace smoothing | 0.2104 | 0.3400 | 0.2893 |
| Unigram LM with Jelinek-Mercer smoothing | 0.2013 | 0.2400 | 0.1787 |

**Inference on the above pseudo-relevance results**

**Original Performance**

TF-IDF stands out as the top performer in the original setup, particularly in Precision at 10 and 30, highlighting its robustness.

ES (built-in) also shows strong performance, especially in Precision at 10, indicating its effectiveness in surfacing relevant documents at the top of the results.

Okapi BM-25 and Unigram LM with Jelinek-Mercer smoothing lag behind, with BM-25 showing notably lower performance across all metrics.

**After Adding Top 5 Distinctive Terms to Each Query**

TF-IDF experiences a slight decline in Average Precision but maintains strong performance in Precision at 10. This suggests that while the added terms may slightly dilute overall relevance, they still help surface relevant documents at the top.

ES (built-in) sees a minor drop in Average Precision but improves in Precision at 10, indicating that the distinctive terms likely helped refine the top results even if they slightly impacted overall precision.

Okapi TF and Okapi BM-25 suffer significant drops in performance, particularly in Average Precision and Precision at 30, suggesting that the added terms might have introduced noise or irrelevant information for these models.

Unigram LMs show mixed results; while Laplace smoothing slightly improves in Average Precision, Jelinek-Mercer smoothing sees improvements across the board, suggesting some compatibility between the added terms and the smoothing techniques.

**After Adding Top 5 Significant Terms from Elasticsearch Aggs to Each Query**

ES (built-in)'s performance slightly improves in Average Precision, but Precision at 10 and 30 decreases. This could indicate that while the significant terms generally enhance query relevance, they might not always align with the top documents' relevance criteria.

Okapi TF shows improvement across all metrics compared to the addition of distinctive terms, suggesting a better alignment with the significant terms from Elasticsearch aggs.

TF-IDF sees a decrease in performance across all metrics, indicating that the added terms might not always be beneficial, possibly diluting the effectiveness of its relevance scoring.

Okapi BM-25 continues to struggle, especially in Precision at 10 and 30, highlighting that the modifications further reduce its effectiveness.

Unigram LM with Laplace smoothing and Jelinek-Mercer smoothing both show a decrease in performance compared to the original but are relatively stable compared to the second modification, suggesting some resilience to the addition of significant terms.

**Table showing the Query used for Evaluation**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Query number | 95 | 68 | 57 | 97 | 98 |
| Original Query | describe a computer application to crime solving | 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. | discuss how MCI has been doing since the Bell System breakup. | identify instances of fiber optics technology actually in use. | identify individuals or organizations which produce fiber optics equipment. |
| Processed Query | ['comput', 'applic', 'crime', 'solv'] | ['actual', 'studi', 'unsubstanti', 'concern', 'safeti', 'manufactur', 'employe', 'instal', 'worker', 'finediamet', 'fiber', 'insul', 'product'] | ['mci', 'bell', 'system', 'breakup'] | ['instanc', 'fiber', 'optic', 'technolog', 'actual'] | ['individu', 'organ', 'produc', 'fiber', 'optic', 'equip'] |
| Processed Query - Pseudo RF **( Only MS students)** | ['comput', 'applic', 'crime', 'solv', ‘supercomput’,’faster’,’model’,’,machin’,’million’] | ['actual', 'studi', 'unsubstanti', 'concern', 'safeti', 'manufactur', 'employe', 'instal', 'worker', 'finediamet', 'fiber', 'insul', 'product', ‘fiber’,’reinforc’,’technolog’,’compani’, ‘lead’] | ['mci', 'bell', 'system', 'breakup', ‘local’, ‘charge’, ‘longdist’, ‘breakup’, ‘reduct’] | ['instanc', 'fiber', 'optic', 'technolog', 'actual', ‘glass’, ‘work’, ‘manufactur’,’swirbul’, ‘technolog’] | ['individu', 'organ', 'produc', 'fiber', 'optic', 'equip', ‘satellit’,’technolog’, ‘manufactur’, ‘opticalwaveguid’,’compani’] |