CS 6200 - Information Retrieval

Homework - 1

1. **Retrieval Model Performance**

   [Highlighted the scores more than 0.28]

Model	Average Precision	Precision at 10	Precision at 30
ES (built in)	0.3019	0.4200	0.3680
Okapi TF	0.2510	0.4000	0.3293
TF-IDF	0.2959	0.4120	0.3573
Okapi BM-25	0.3020	0.4160	0.3653
Unigram LM with Laplace Smoothing	0.2420	0.4080	0.3107
Unigram LM with Jelinek Mercer Smoothing	0.3007	0.4720	0.3733

Able to meet average precision of 0.28 for TF-IDF and Okapi BM-24
Able to meet average precision of 0.25 for Unigram LM with JM smoothing

2. Pseudo-relevance Feedback Improvements

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

a. General Algorithm - Adding top 5 distinctive terms to the query

Model	Average Precision	Precision at 10	Precision at 30
ES (built in)	0.2839	0.3200	0.2987
Okapi TF	0.2897	0.3840	0.3373
TF-IDF	0.2710	0.3080	0.2867
Okapi BM-25	0.2781	0.3160	0.2973
Unigram LM with Laplace Smoothing	0.2417	0.3400	0.2880
Unigram LM with Jelinek Mercer Smoothing	0.2779	0.3480	0.3107

Precision increases for Okapi-TF when implementing a general algorithm for pseudo-relevance feedback.

b. Elasticsearch "Significant Terms" : Adding top 5 significant terms to the query

Model	Average Precision	Precision at 10	Precision at 30
ES (built in)	0.1568	0.1680	0.1760
Okapi TF	0.1741	0.2400	0.2133
TF-IDF	0.1555	0.1640	0.1707
Okapi BM-25	0.1559	0.1720	0.1720
Unigram LM with Laplace Smoothing	0.1594	0.2240	0.1960
Unigram LM with Jelinek Mercer Smoothing	0.1459	0.2120	0.1800

The Average precision for each of the models is relatively low.