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

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

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

Model	Average Precision	Precision at 10	Precision at 30
ES (built in)	0.1416	0.1920	0.1867
Okapi TF	0.1423	0.2360	0.2053
TF-IDF	0.1398	0.1760	0.1800
Okapi BM-25	0.1406	0.1840	0.1867
Unigram LM with Laplace Smoothing	0.1376	0.2680	0.2040
Unigram LM with Jelinek Mercer Smoothing	0.1498	0.2360	0.2053