Jack Davey
2/15/21
COEN 169

1. RawTF Evaluation
    a. The RawTF algorithm with stopwords removed and stemming implemented performed better than the one without, with a ~17% increase in average precision. Both retrieved a total of 3000 documents, however the RawTF algorithm with stemming and no stopwords retrieved 196 of the 442 relevant documents, 3 more than the non-stemmed version's 193. According to the "eval_data_stemmed_nostopw_index" file, the stemmed query utilized the arabic stemmer. After running a third query instead with the porter stemmer, average precision improved more than 36%, from 0.1174 to 0.0859.
    b. Without stemming implemented, stopword removal significantly harmed average precision. Stopword removal appears to make no difference when stemmed using the arabic stemmer, and should not be implemented together to save computing resources. Stopword removal significantly improved average precision when using the porter stemmer however, so its usage should be dependent on stemmer choice.
2. RawTF, RawTFIDF, LogTFIDF, and Okapi Analysis
    a. In testing the RawTF, RawTFIDF, LogTFIDF, and Okapi retrieval algorithms, no noticeable patterns arose when testing combinations of stopword removal and stemming with each, however each algorithm (except RawTFIDF) performed best with no stopw and stemming implemented. (in all experiments with stemming, the porter stemmer was used). Overall, the Okapi algorithm with stopwords removed and stemming achieved the highest average precision, 0.3584, and performed considerably better with no stopw & no stemming, as well as stopw & stemming, than any other algorithm. Okapi suffers considerably when stopw aren't removed and stemming isn't implemented, achieving half the average precision with no stopw and stemming. The second highest score was achieved using LogTFIDF, with no stopw and stemming, which retrieved 255 relevant documents, 31 less than Okapi's 286. As mentioned earlier, RawTFIDF performed slightly better without stemming with or without stopwords, and should not be used with the porter stemmer. RawTF was by far the worst performer, with the lowest average precision in every instance, but benefitted from stemming slightly, particularly with the porter stemmer.

	Remove Stopwords & Stemming	Remove Stopwords & No Stemming	No Removing Stopwords & Stemming	No Removing Stopwords & No Stemming
RawTF	0.1174P, 0.0859A	0.0449	0.0859	0.0712
RawTFIDF	0.2137	0.2170	0.1260	0.1861
LogTFIDF	0.3241	0.1369	0.1657	0.2687
Okapi	0.3584	0.3126	0.3004	0.1727