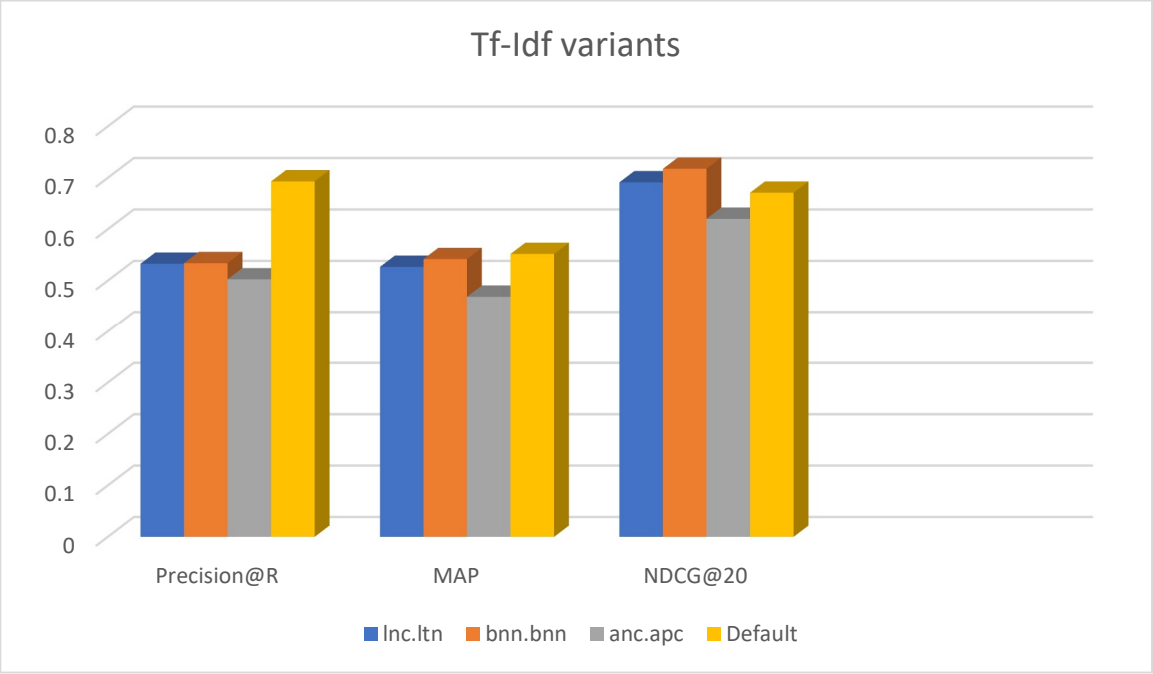
CS 753/853: Topics / Information Retrieval Fall 2019
Programming Assignment 3: TF-IDF

1. Implement Tf-Idf

Solution: The table lists metrics for Precision at R (RPrec), mean-average precision (MAP) and NDCG@20 (ng_cut_20).



Measures	lnc.ltn	bnn.bnn	anc.apc	Default
Precision@R	0.5328	0.5364	0.5026	0.693
MAP	0.5268	0.542	0.4686	0.552
NDCG@20	0.6912	0.7175	0.6202	0.671

Which of these variants perform best?
Of the various tf-idf methods we tried, the bnn.bnn variant seemed to perform better than the both methods. it seems to be it is closest to the Lucene default ranking model.

Do they perform better or worse than Lucene's default ranking model?
None of them are better than lucene default ranking model.

Using the standard error method, analyze if the the difference is significant.
From the graph above, we can see that there is no significant difference as the errors overlap.

2. Spearman's Rank Correlation Coefficient.

lnc.ltn 0.461285680325717
bnn.bnn 0.050986974031029794
anc.apc 0.4648760696699176

anc.apc is the best of all the three and is the closet to Lucene's standard model.
As we can see from the above results anc.apc is the closest to the lucene's default model while the second closest is lnc.ltc and bnn.bnn is the third closest with the maximum difference in ranking

3. Graduate students (voluntary for undergrads): TF-IDF

for Section Queries.
we have retrieved ranking for each section in each page in the file of the test200 benchmark. but we couldn't figure out how to use the combination given to evaluate using hierarchical qrels fie;.
we have evaluated the values and find the precision@R i.e 0.693.