Programming Assignment 3

(Team 3)

Team Members:

Name	UNH ID	Webcat Username	Assignment #
Amith Ramanagar Chandrashekar	928101607	ar1184	3
Vaughan Coder	966427992	vwc1010	3
Sepideh Koohfar	928334858	sk1015	3
Pooja Himanshu Oza	900539783	pho1003	3

1. **Scores Table**

Search	**lnc.ltn**	**bnn.bnn**	**anc.apc**	**LuceneDefault**
map	0.5607	0.5419	0.5382	0.6016
Rprec	0.5689	0.5358	0.5517	0.5966
ndcg_cut_20	0.728	0.7168	0.7047	0.7696

As is shown, the lnc.ltn ndcg20 performed the highest. With LNC.LTN also having the highest non-default score. BNN performed better than anc.apc in the instance of Map but not on Rprecision.

S.E.	LNC	BNN	ANC
Map	0.017501	0.017837	0.017145
Rprec	0.016351	0.016931	0.015958
NDCG20	0.018124	0.016334	0.015494

These answers are statistically significant based on the standard error calculations performed. With that said, in all cases these methods underperformed the Lucene default.

2. SpearMan coefficient (LNC) =0.8365662285737226

   SpearMan coefficient (BNN) =0.05846498804551581

   SpearMan coefficient (ANC) =0.5868918476402237

   LNC appears to be the most strongly correlated.

SectionPath

Search	lnc.ltn	bnn.bnn	anc.apc	LuceneDefault
map	0.5375	0.474	0.5096	0.5669

Rprec	**0.5471**	**0.4876**	**0.5216**	**0.5634**
ndcg_cut_20	0.6949	0.6589	0.6714	0.7057

The approach followed for section queries was to process into our individual query, each of the values. We removed all special characters from the sections and concatenated them. The idea behind this being we would improve our TF-IDF scoring by increasing our knowledge of what items occurred frequently in the query.