Title: Laboratory 3
Date: 09/11/21
Student names: Vincent Achukwu, Alan Hian Wu

3. Investigating BM25 Parameter Settings

3.2 Basic Experimental Procedure

For each run, we used BM25.

TREC6	MAP	GMAP
k = 1.2, b = 0.75	0.2153	0.0949
k = 1.2, b = 0	0.2233	0.1035
k = 1.2, b = 1	0.2184	0.0916
k = 1, b = 0	0.2229	0.1035

TREC7	MAP	GMAP
k = 1.2, b = 0.75	0.1806	0.0946
k = 1.2, b = 0	0.1281	0.0390
k = 1.2, b = 1	0.1840	0.0894
k = 1, b = 0	0.1624	0.0709

| TREC8	MAP	GMAP
k = 1.2, b = 0.75	0.1806	0.0946
k = 1.2, b = 0	0.1281	0.0390
k = 1.2, b = 1	0.1840	0.0894
k = 1, b = 0	0.2445	0.1366

Conclusion:
We know the k parameter value affects how often terms appear in a document, whereas the b parameter value affects the document length and ranges between 0 and 1. The default value for k is 1.2 and for b it's 0.75. We've altered the values for b to 0.75, 0, and 1, and for k we used 1.2 and 1. It seems like the **map** value increases for various values of k, hence this

would require more testing and many values of k to see this pattern. Varying the b parameter changes the length of the document normalisation rate.

3.3 Optimisation of Parameter Values

Running the test.sh file generates the CSV file by also specifying the output file name and the trec dataset passed as arguments in the command line. The graphs we generated are shown below.
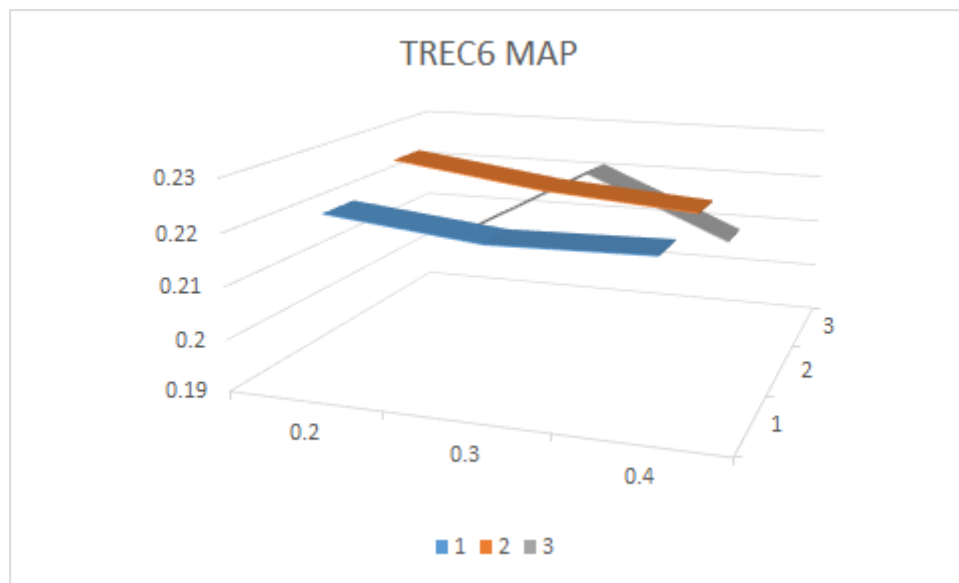
We obtained these results by having the k values `for k in 1.0 2.0 3.0` and the b values `for b in 0.2 0.3 0.4`
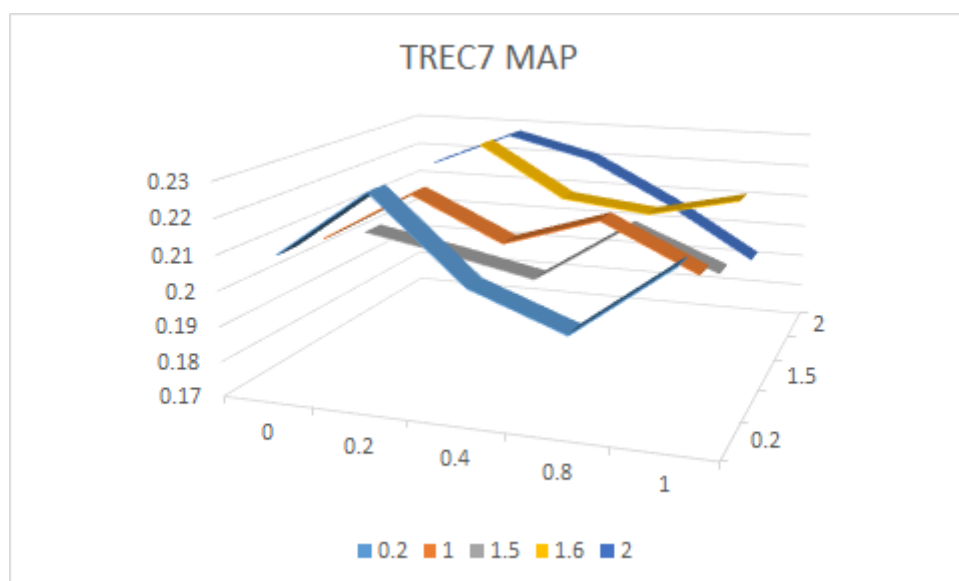
We obtained these results by having the k values `for k in 0.2 1.0 1.5 1.6 2.0` and the b values `for b in 0.0 0.2 0.4 0.8 1.0`

TREC8 MAP
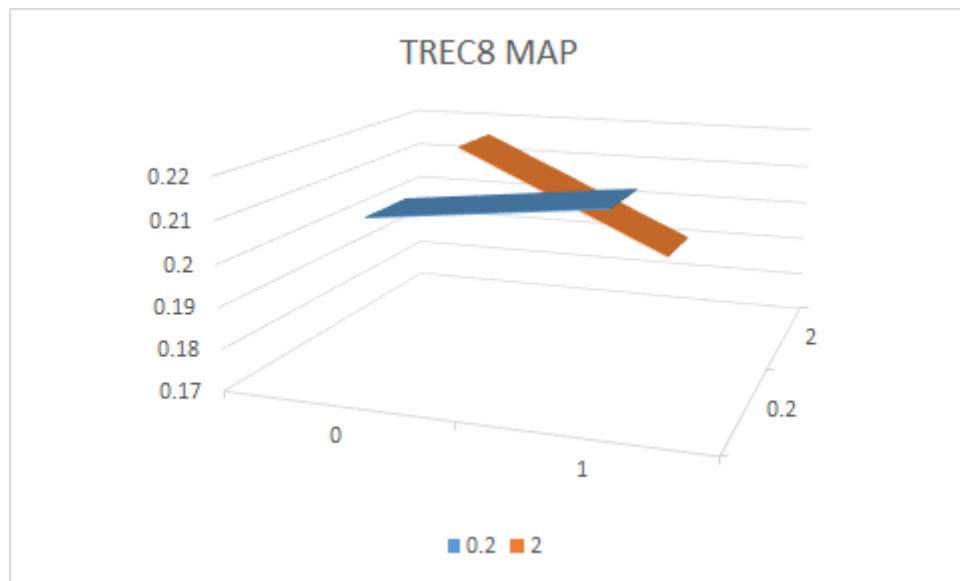
We obtained these results by having the k values `for k in 0.2 2.0` and the b values `for b in 0.0 1.0`