

3. Investigating BM25 Parameter Settings

3.2 Basic Experimental Procedure

The **k1** factor determines the impact of term frequency in a document. A typical value for **k1** would be **1.5** The **b** value determines the degree of document length normalisation. The value of **b** can vary in the range **0 to 1**

| TREC 6 | MAP | gm_map |
|-------------------|--------|--------|
| k = 1,2, b = 0.75 | 0.2153 | 0.0949 |
| k = 1.2, b = 1 | 0.2037 | 0.0755 |
| k = 1, b = 1 | 0.2064 | 0.0775 |
| k = 0.5 , b = 1 | 0.2142 | 0.0832 |
| k = 0, b =0 | 0.1532 | 0.0363 |
| k = 1.5, b = 0 | 0.2040 | 0.0777 |

| TREC 7 | MAP | gm_map |
|--------------------|--------|--------|
| k = 1.2 , b = 0.75 | 0.1806 | 0.0946 |
| k = 1.2, b = 1 | 0.1642 | 0.0787 |
| k = 1, b = 1 | 0.1698 | 0.0804 |
| k = 0.5 , b = 1 | 0.1760 | 0.0821 |
| k = 0, b = 0 | 0.1146 | 0.0302 |
| k= 1.5, b = 0 | 0.1699 | 0.0844 |

| TREC 8 | MAP | gm_map |
|-------------------|--------|--------|
| k = 1.2, b = 0.75 | 0.2357 | 0.1343 |
| k = 1.2, b = 1 | 0.2199 | 0.1193 |
| k = 1, b = 1 | 0.2267 | 0.1234 |
| k = 0.5 , b = 1 | 0.2376 | 0.1305 |
| k = 0, b =0 | 0.1541 | 0.0568 |
| k = 1.5, b = 0 | 0.2282 | 0.1292 |

From running multiple tests using different **b** and **k** values in different topic sets. We noticed that **MAP** value was the highest for **k = 1.2** and **b = 0.75** in **TREC6** and **7**. In **TREC8** the highest map value was returned for **k = 0.5** and **b = 1**. Overall the **MAP** value seems to increase within certain range of **k**.

3.3 Optimisation of Parameter Values