Vipul Munot
(vipmunot)

Long query

Evaluation metric	Your algorithm	Vector Space Model	BM25	Language Model with Dirichlet Smoothing	Language Model with Jelinek Mercer Smoothing
P@5	0.0	0.2	0.6	0.0	0.2
P@10	0.0	0.3	0.3	0.2	0.3
P@20	0.15	0.25	0.3	0.25	0.25
P@100	0.04	0.1	0.1	0.1	0.09
Recall@5	0.0	0.0323	0.0968	0.0	0.0323
Recall@10	0	0.0968	0.0968	0.0645	0.0968
Recall@20	0.968	0.1613	0.1935	0.1613	0.1613
Recall@100	0.1290	0.3226	0.3226	0.3226	0.2903
MAP	0.0210	0.0966	0.1333	0.0693	0.0853
MRR	0.0833	0.5	1.0	0.125	0.2
NDCG@5	0	0.214	0.6399	0.0	0.1312
NDCG@10	0	0.2785	0.4153	0.1357	0.2369
NDCG@20	0.1060	0.2545	0.3732	0.1924	0.2259
NDCG@100	0.0982	0.2881	0.3519	0.2399	0.2478

Short Query

Evaluation metric	Your algorithm	Vector Space Model	BM25	Language Model with Dirichlet Smoothing	Language Model with Jelinek Mercer Smoothing
P@5	0.2	0.4	0.6	0.6	0.4
P@10	0.1	0.5	0.5	0.5	0.5
P@20	0.1	0.4	0.3	0.35	0.25
P@100	0.06	0.09	0.10	0.09	0.1
Recall@5	0.0323	0.06	0.0968	0.0968	0.0645
Recall@10	0.0323	0.16	0.1613	0.1613	0.1613
Recall@20	0	0.25	0.1935	0.2258	0.1613
Recall@100	0.1935	0.29	0.3226	0.2903	0.3226
MAP	0.0631	0.1833	0.1894	0.1404	0.1462
MRR	1.0	1.00	1.0	0.5	1.0
NDCG@5	0.3392	0.5531	0.7227	0.4913	0.5531
NDCG@10	0.2201	0.5801	0.6208	0.4666	0.5704

NDCG@20	0.1804	0.4786	0.4341	0.3704	0.3681
NDCG@100	0.2112	0.3804	0.4036	0.3180	0.3726

Summary:

Here we use the BM25 Algorithm with default k1 = 1.2, b = 0.75 values. The BM25 has considered term frequency and scaling by document length as observed. BM25 has almost similar values for long queries and short queries. The classic similarity (Vector Space Model) saves only one byte of data hence, its accuracy decreases. My algorithm and classic should have same values but it won't be possible because compare file uses TopDocs function. The term smoothing describes techniques for adjusting the maximum likelihood estimate of probabilities to produce more accurate probabilities. The name smoothing comes from the fact that these techniques tend to make distributions more uniform, by adjusting low probabilities such as zero probabilities upward, and high probabilities downward. Whenever a probability is estimated from few counts, smoothing has the potential to significantly improve estimation.

For Information Retrieval, we need higher precision, hence, we can use Vector Space Model and BM25 Models with comparison to other models.