

Onde t é o número de termos da consulta presentes no documento i ; N é igual à quantidade de documentos que a base possui e n_i é o número de documentos em que o termo t aparece.

2.4.2 BM25

O modelo BM25 (*Best Match 25*) é resultado de experimentos sobre o modelo probabilístico clássico (MPC) de *Robertson e Jones* (1976) utilizando três princípios: frequência inversa de documentos, frequência dos termos e normalização pelo tamanho dos documentos (BAEZA-YATES; RIBEIRO-NETO, 2013). A Equação 2 é a equação de similaridade do BM25.

$$sim_{BM25}(d_j, BUSCA) = \sum_{k_1[q, d_j]} B_{i,j} \times \log\left(\frac{N - n_i + 0,5}{n_i + 0,5}\right) \quad (2)$$

onde

$$B_{i,j} = \frac{(K_1 + 1)f_{i,j}}{K_1 \left[(1 - b) + b \frac{len(d_j)}{avg_doclen} \right] + f_{i,j}} \quad (3)$$

sendo:

- K_1 e b : Constantes empíricas;
- N : Número de documentos da base;
- n_i : Número de documentos em que o termo t aparece;
- $f_{i,j}$: Frequência do termo k_1 no documento d_j ;
- $len(d_j)$: Tamanho em caracteres do documento d_j ;
- avg_doclen : Tamanho médio dos documentos da base.

Baeza-Yates e Ribeiro-Neto (2013) indicam que as constantes K_1 e b podem receber os valores 1 e 0,75, respectivamente, para coleções genéricas.

2.4.3 Métrica MRR

Em sistemas de buscas e respostas que tem por objetivo recuperar uma resposta em vez de produzir uma lista de documentos é desejável uma métrica que favoreça resultados onde a primeira resposta correta esteja no topo do *ranking* (BAEZA-YATES; RIBEIRO-NETO, 2013). Uma métrica que apresenta boas