

BM25 can also Discover Syntagmatic Relations

$$d1=(x_1, \dots, x_N) \quad \text{BM25}(w_i, d1) = \frac{(k+1)c(w_i, d1)}{c(w_i, d1) + k(1-b + b*|d1|/avdl)}$$

$$x_i = \frac{\text{BM25}(w_i, d1)}{\sum_{j=1}^N \text{BM25}(w_j, d1)}$$

$$b \in [0, 1]$$

$$k \in [0, +\infty)$$

IDF-weighted $d1=(x_1 * IDF(w_1), \dots, x_N * IDF(w_N))$

The highly weighted terms in the context vector of word w are likely syntagmatically related to w .