

Improved VSM with Term Frequency Weighting

$$q = (x_1, \dots, x_N)$$

x_i = count of word W_i in query

$$d = (y_1, \dots, y_N)$$

y_i = count of word W_i in doc

$$\text{Sim}(q, d) = q \cdot d = x_1 y_1 + \dots + x_N y_N = \sum_{i=1}^N x_i y_i$$

What does this ranking function intuitively capture?

Does it fix the problems of the simplest VSM?