Equação 1

$$sim(d_j, q) = \frac{\vec{d_j} \cdot \vec{q}}{|\vec{d_j}| \times |\vec{q}|} = \frac{\sum_{i=1}^t w_{i,j} \cdot w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} \times \sqrt{\sum_{i=1}^t w_{i,q}^2}}$$
(1)

A fim de diminuir o peso de termos altamente frequentes, usa-se o fator IDF (Inverted Document Frequency), que é a frequencia inversa de documentos que contem um termo, dada por  $IDF(k_i) = \log \frac{N}{n_i}$ 

$$IDF(k_i) = log \frac{N}{n_i} \tag{2}$$

$$w_{i,j} = freq_{i,j} \cdot IDFk_i \tag{3}$$