

Текстовое ранжирование

Совпадение по слова

Просто проверяем вхождение слов запроса в документ

Запрос описывается вектором, где единицы стоят для встретившихся в нем слов

Документ аналогично

Ранжирование через расстояние хэмминга между вектором запроса и документа

Текстовое ранжирование

Модель векторного пространства

$D_i = \{w_{i1}, w_{i2}, w_{i3}, \dots, w_{it}\}$ - вектор документа
 t – размерность вектора. $|V|$ - размерность словаря
 w_{ij} - вес j -го термина в документе i

$s(Q, D_i)$ - Мера сходства документа и запроса

$$w_{ij} = tf_{ij \in Q}(t_j, d_i) \cdot idf_j(t_j, D)$$

$tf_{ij \in Q}$ - частота j слова в документе i

idf_j - инвертированная частота слова j

