Дорожкин Денис, группа 425

Задача

Везде округляем до 2х знаков после запятой.

1) представляем запрос и документы в виде векторов

Q = (1,0,1,0); D1 = (27,3,0,14); D2 = (4,33,33,0); D3 = (24, 0, 29, 17)

2) покажем tf для всех слов для каждого документа

tf\_car\_D1 = 27; tf\_car\_D2 = 4; tf\_car\_D3 = 24  
tf\_auto\_D1 = 3; tf\_auto\_D2 = 33; tf\_auto\_D3 = 0  
tf\_insurance\_D1 = 0; tf\_insurance\_D2 = 33; tf\_insurance\_D3 = 29  
tf\_best\_D1 = 14; tf\_best\_D2 = 0; tf\_bext\_D3 = 17

tf\_D1 = (27, 3, 0, 14); tf\_D2 = (4, 33, 33, 0) tf\_D3 = (24, 0, 29, 17)

3) tf-idf вес всех слов всех документов

tf-idf = tf \* idf (для векторов перемножаем поэлементно)

tf-idf\_D1 = (27, 3, 0, 14) \* (1.65, 2.08, 1.62, 1.5) = (44.55, 6.24, 0, 21)  
tf-idf\_D2 = (4, 33, 33, 0) \* (1.65, 2.08, 1.62, 1.5) = (6.6, 68.64, 53.46, 0)  
tf-idf\_D3 = (24, 0, 29, 17) \* (1.65, 2.08, 1.62, 1.5) = (39.6, 0, 46.98, 25.5)

4) отнормируем веса

tf-idf\_D1 = (0.9, 0.13, 0, 0.4)  
tf-idf\_D2 = (0.08, 0.79, 0.6, 0)  
tf-idf\_D3 = (0.6, 0, 0.71, 0.38)

5) Посчитаем cos углов между векторами

cos(Q, D1) = 0.9 / (sqrt(2) \* sqrt(0.9^2 + 0.13^2 + 0.4^2)) = 0.64  
cos(Q, D2) = (0.08 + 0.6) / (sqrt(2) \* sqrt(0.08^2 + 0.79^2 + 0.6^2)) = 0.48  
cos(Q, D3) = (0.6 + 0.71) / (sqrt(2) \* sqrt(0.6^2 + 0.71^2 + 0.38^2)) = 0.92

Ответ

Чем больше cos, тем меньше угол между вектором запроса и документа.   
Документы выстроятся по порядку D3 -> D1 -> D2