

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ ИМЕНИ Н. Э. БАУМАНА
Факультет информатики и систем управления
Кафедра теоретической информатики и компьютерных технологий

Лабораторная работа №2
по курсу «Информационный поиск»

«TF IDF.
Ранжирование документов по запросу»

Выполнил:
студент группы ИУ9-21М
Беляев А. В.

Проверила:
Лукашевич Н. В.

Москва 2019

Term	df	idf	d1	d2	d3
car	18165	1.65	27	4	24
auto	6723	2.08	3	33	0
Insu- rance	19241	1.62	0	33	29
best	25235	1.5	14	0	17

Рис. 1:

1 Цель работы

Для запроса *car insurance* и документов, представленных в таблице на Рис. 1, вычислить веса документов по отношению к запросу и упорядочить документы.

2 Ход работы

Каждый документ представлен своим вектором, составленным из частот слов:

- $d1 = (27, 3, 0, 14)$
- $d2 = (4, 33, 33, 0)$
- $d3 = (24, 0, 29, 17)$

Перемножаем вектора с соответствующими значениями *idf*:

- $d1 = (27 * 1.65, 3 * 2.08, 0, 14 * 1.5) = (44.55, 6.24, 0, 21)$
- $d2 = (4 * 1.65, 33 * 2.08, 33 * 1.62, 0) = (6.6, 68.64, 53.46, 0)$
- $d3 = (24 * 1.65, 0, 29 * 1.62, 17 * 1.5) = (39.6, 0, 46.98, 25.5)$

Далее необходимо нормализовать полученные вектора:

- $d1 = (0.9, 0.13, 0, 0.42)$
- $d2 = (0.08, 0.8, 0.6, 0)$
- $d3 = (0.6, 0, 0.71, 0.4)$

Теперь необходимо векторизовать запрос *car insurance*. $q = (1, 0, 1, 0)$

Вычисляем итоговый вес документа по отношению к запросу через скалярное произведение векторов документа и запроса:

- $W_{d1} = d1 * q = 0.9$

- $W_{d2} = d_2 * q = 0.68$
- $W_{d3} = d_3 * q = 1.31$

Таким образом, документы будут отранжированы таким образом: **D3, D1, D2**.

3 Выводы

В ходе работы был изучен способ ранжирования документов по степени соответствия запросу *car insurance*. Были вычислены веса документов по отношению к запросу.