

Задание 7 от 28.09.20 (Relevance feedback)

Валентин Александров, 620 группа

Постановка задачи

- Запрос: "отбор кандидатов" (**q**)
- Пользователь отметил релевантными два документа
 - "Кандидат отобрать претендент" (**d1**)
 - "Отбор выбрать претендент" (**d2**)
- Объем коллекции (**N**) – 1 млн. документов
- Df:
 - отбор 70000
 - кандидат – 70000,
 - претендент - 30000
 - отобрать – 50000
 - выбрать - 70000

Как изменится запрос ($q_{modified}$), если

- $\alpha=0.7$ (коэффициент учета запроса)
- $\beta=0.3$ (коэффициент учета релевантных документов)
- Запрос представляется как вектор частот
- Документ представляется как нормализованный вектор tf.idf

Решение

$$q_{modified} = \alpha q + \beta \frac{d_1 + d_2}{2}$$

term	df	idf	d1	d2	q	$q_{modified}$
отбор	70 000	1.1549	0	0.5171	1	0,777565
кандидат	70 000	1.1549	0.4995	0	1	0,774925
претендент	30 000	1.5228	0.6586	0.6819	0	0,201075
отобрать	50 000	1.301	0.5627	0	0	0,084405

term	df	idf	d1	d2	q	<i>q_{modified}</i>
<i>выбрать</i>	70 000	1.1549	0	0.5171	0	0,077565