

### Задание 1:

Запрос: a b.

Коллекция:

1. a b c d
2. a a a
3. b b c
4. a b b c

Других документов нет. Применить языковую модель к коллекции. Сравнить  $l=0.5$ ,  $l=0.9$ . ( $l$  - лямбда).

Вероятности термов по всей коллекции:

P(a)	P(b)	P(c)	P(d)
5/14	5/14	3/14	1/14

Вероятности термов по документам:

d/P(t Md)	P(a Md)	P(b Md)	P(c Md)	P(d Md)
d1	1/4	1/4	1/4	1/4
d2	3/3	0/3	0/3	0/3
d3	0/3	2/3	1/3	0/3
d4	1/4	2/4	1/4	0/4

$l = 0.5$ :

1.  $P(q|d1) = 0.09215$
2.  $P(q|d2) = 0.12117$
3.  $P(q|d3) = 0.09141$
4.  $P(q|d4) = 0.13010$

Ранжирование:  $d4 > d2 > d1 > d3$

$l = 0.9$ :

1.  $P(q|d1) = 0.06797$
2.  $P(q|d2) = 0.03341$
3.  $P(q|d3) = 0.02270$
4.  $P(q|d4) = 0.12663$

Ранжирование:  $d4 > d1 > d2 > d3$

При  $l = 0.9$  множитель, соответствующий терму b в d2, становится слишком маленьким (т.к. b отсутствует в d2) и d2 начинает уступать d1. Считаю, что вариант с  $l = 0.9$  более правильный, т.к. он отдал предпочтение документу, в котором присутствуют все термы запроса, а не один терм, но много раз.