

Домашнее задание №4

Асланов. Андрей, ИУГ-НИИ

Формулировка языковой модели для информационного поиска:

$$P(q|d) = \prod_{t \in q} ((1-\lambda)p(t) + \lambda P(t|M_d))$$

Запрос к поисковой системе состоит из двух слов: а, в.

В коллекции следующие документы:

Документы		В конкретном док-те	
		"а" встретится	"в" встретится
doc1	а, в, с, d	1 раз	1 раз
doc2	а, а, а	3 раза	0 раз
doc3	в, в, с	0 раз	2 раза
doc4	а, в, в, с	1 раз	2 раза
$\Sigma 14$ док-тов		$\Sigma 5$ раз	$\Sigma 5$ раз

Применить языковую модель.

Сравнить результаты упорядочения выдачи при $\lambda=0,5$ и $\lambda=0,9$.

Для $\lambda=0,5$:

$$P(q|doc_1) = \left[(1-0,5) \cdot \left(\frac{5}{14} \right) + 0,5 \cdot \left(\frac{1}{4} \right) \right] \left[0,5 \cdot \frac{5}{14} + 0,5 \cdot \frac{1}{4} \right] = \left(\frac{1}{2} \cdot \frac{5}{14} + \frac{1}{2} \cdot \frac{1}{4} \right) \cdot \frac{1}{2} \left(\frac{5}{14} + \frac{1}{4} \right) \approx 0,921$$

$$P(q|doc_2) = \left[(1-0,5) \cdot \frac{5}{14} + 0,5 \cdot \frac{3}{3} \right] \left[0,5 \cdot \frac{5}{14} + 0,5 \cdot \frac{0}{3} \right] = \left(\frac{1}{2} \cdot \frac{5}{14} + \frac{1}{2} \cdot 1 \right) \cdot \frac{1}{2} \cdot \frac{5}{14} = \frac{95}{784} \approx 0,121$$

$$P(q|doc_3) = \left[(1-0,5) \cdot \frac{5}{14} + 0,5 \cdot \frac{0}{3} \right] \left[0,5 \cdot \frac{5}{14} + 0,5 \cdot \frac{2}{3} \right] = \left(\frac{1}{2} \cdot \frac{5}{14} \right) \cdot \left(\frac{1}{2} \cdot \frac{5}{14} + \frac{1}{2} \cdot \frac{2}{3} \right) \approx 0,0914$$

$$P(q|doc_4) = \left[(1-0,5) \cdot \frac{5}{14} + 0,5 \cdot \frac{1}{4} \right] \left[0,5 \cdot \frac{5}{14} + 0,5 \cdot \frac{2}{4} \right] \approx 0,1301$$

Ранжирование: doc4 \rightarrow doc2 \rightarrow doc1 \rightarrow doc3.

Для $\lambda=0,9$:

$$P(q|doc_1) = \left[(1-0,9) \cdot \frac{5}{14} + 0,9 \cdot \frac{1}{4} \right] \left[(1-0,9) \cdot \frac{5}{14} + 0,9 \cdot \frac{1}{4} \right] = \frac{73}{280} \cdot \frac{73}{280} \approx 0,0679$$

$$P(q|doc_2) = \left[(1-0,9) \cdot \frac{5}{14} + 0,9 \cdot \frac{3}{3} \right] \left[(1-0,9) \cdot \frac{5}{14} + 0,9 \cdot \frac{0}{3} \right] \approx 0,0334$$

$$P(q|doc_3) = \left[(1-0,9) \cdot \frac{5}{14} + 0,9 \cdot \frac{0}{3} \right] \left[(1-0,9) \cdot \frac{5}{14} + 0,9 \cdot \frac{2}{3} \right] = \frac{5}{140} \cdot \left(\frac{5}{140} + \frac{9}{10} \cdot \frac{2}{3} \right) \approx 0,0227$$

$$P(q|doc_4) = \left[(1-0,9) \cdot \frac{5}{14} + 0,9 \cdot \frac{1}{4} \right] \left[(1-0,9) \cdot \frac{5}{14} + 0,9 \cdot \frac{2}{4} \right] \approx 0,12994$$

Ранжирование: doc4 \rightarrow doc1 \rightarrow doc2 \rightarrow doc3