



Министерство науки и высшего образования
Российской Федерации

Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Московский государственный технический университет
имени Н. Э. Баумана (национальный
исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

Факультет «Информатика и системы управления»

ОТЧЕТ по лабораторной работе №4

по дисциплине

«Информационный поиск и извлечение информации из
текстов»

Студент группы ИУ9-21М

_____ С.С. Погосян
(подпись, дата)

Руководитель

_____ Н.В. Лукашевич
(подпись, дата)

1. Постановка задачи

Задача-1

- Эксперт нашел 20 релевантных документов. Система нашла 4 документа в следующей последовательности релевантных и нерелевантных документов:
- RNRNNRRNNNN
- Какова средняя точность поиска – Average Precision

Задача-2

- При разметке релевантных документов эксперты использовали шкалу от 0 (нерелевантные документы) до 4 баллов.
- При тестировании систем выяснилось, что системы выдали следующие результаты поиска ответов на один и тот же запрос:
- **Система 1:** 4, 2, 3, 1, 2, 0 (и далее 0)
- **Система 2:** 3, 2, 4, 4, 4., далее 0.
-
- Какая система ищет лучше по мере NDCG.

2. Решение задачи 1

1. R $\Rightarrow Pr_1 = \frac{1}{1}$

2. N $\Rightarrow Pr_2 = \frac{1}{2}$

3. R $\Rightarrow Pr_3 = \frac{2}{3}$

4. N $\Rightarrow Pr_4 = \frac{2}{4}$

- 5. $N \Rightarrow Pr_5 = \frac{2}{5}$
- 6. $R \Rightarrow Pr_6 = \frac{3}{6}$
- 7. $R \Rightarrow Pr_7 = \frac{4}{7}$
- 8. $N \Rightarrow Pr_8 = \frac{4}{8}$
- 9. $N \Rightarrow Pr_9 = \frac{4}{9}$
- 10. $N \Rightarrow Pr_{10} = \frac{4}{10}$
- 11. $N \Rightarrow Pr_{11} = \frac{4}{11}$

$$AP = \frac{1 + \frac{2}{3} + \frac{1}{2} + \frac{4}{7}}{20} = 0.136 \quad - \text{средняя точность}$$

3. Решение задачи 2

$$DCG = \sum_{i=1}^N \frac{rel_i}{\log_2(i+1)}$$

$$DCG_1 = \sum_{i=1}^5 \frac{rel_i}{\log_2(i+1)} = \frac{4}{\log_2 2} + \frac{2}{\log_2 3} + \frac{3}{\log_2 4} + \frac{1}{\log_2 5} + \frac{2}{\log_2 6} = 7.96$$

$$DCG_2 = \frac{3}{\log_2 2} + \frac{2}{\log_2 3} + \frac{4}{\log_2 4} + \frac{4}{\log_2 5} + \frac{4}{\log_2 6} = 9.53$$

$$idealDCG_1 = \frac{4}{\log_2 2} + \frac{3}{\log_2 3} + \frac{2}{\log_2 4} + \frac{2}{\log_2 5} + \frac{1}{\log_2 6} = 8.14$$

$$idealDCG_2 = \frac{4}{\log_2 2} + \frac{4}{\log_2 3} + \frac{4}{\log_2 4} + \frac{3}{\log_2 5} + \frac{2}{\log_2 6} = 10.58$$

$$NDCG_1 = \frac{DCG_1}{idealDCG_1} = \frac{7.96}{8.14} = 0.977$$

$$NDCG_2 = \frac{9.53}{10.58} = 0.900$$

$$NDCG_1 > NDCG_2 \Rightarrow \text{первая система лучше}$$