

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Дисциплина: «Анализ данных»

Домашнее задание на тему:
«Лабораторная работа №16»

Выполнил: Осипов Лев,
студент группы 301ПИ (1).

СОДЕРЖАНИЕ

Теоретическая часть.....	3
Задание 1	3
Задание 2	3
Задание 3	3
Задание 4	3
Список литературы	4

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

Задание 1

Человеку легче среди двух документов выбрать более релевантный, банально сравнивая их. Работает так называемый «принцип контраста», основанный на относительности. Взять и сравнить между собой два документа представляется человеку (ассесору) в разы проще, чем оценивать релевантность численно или же сортировать по релевантности кучу документов (здесь может быть много неоднозначностей).

Задание 2

Ответ на это вопрос опять же заключается в самих ассесорах, то есть людях. Выбрать среди многочисленных пунктов зачастую бывает очень сложно. Приведу пример: даже в различных тестах и опросах применяется модель «Да – скорее да – нейтрально – скорее нет – нет», вместо, предположим, десятибалльной шкалы (хотя есть и такие). Чем меньше вариантов, тем легче ассесорам. К тому же, большая вариативность может породить неоднозначность в оценке, что будет только усложнять обучение модели, несмотря на точные алгоритмы.

Задание 3

В случае второго пользователя ответ очевиден: первый документ и второй. Как мы говорили на лекции, пользователь вряд ли откроет второй документ, не посмотрев на первый. Очевидно, что эта пара будет самая информативная.

Что касается другого пользователя, так как он кликнул по первому же документу в списке, мы не можем утверждать, смотрел ли он на остальные. Поэтому, не обладая этими данными, мы не можем составить информативную пару.

Задание 4

Первая мысль, которая приходит в голову, – это брать элементы с начала и конца списка, то есть наиболее разные по релевантности.

Но документы с такой разницей в релевантности могут не очень хорошо обучить алгоритм, поэтому можно сделать немного по-другому – разделить список на несколько блоков, и попарно брать элементы из этих блоков.

СПИСОК ЛИТЕРАТУРЫ

- 1) **Анализ данных (Программная инженерия) –**
[http://wiki.cs.hse.ru/Анализ_данных_\(Программная_инженерия\)](http://wiki.cs.hse.ru/Анализ_данных_(Программная_инженерия))