



Информационные ресурсы в финансовом мониторинге

НИЯУ МИФИ, Кафедра финансового мониторинга.

В.Ю. Радыгин, Г.К. Крючков

Курсовая работа

Источники данных

1. Уголовный кодекс РФ:
https://www.consultant.ru/document/cons_doc_LAW_10699/
2. Налоговый кодекс РФ. Часть первая:
https://www.consultant.ru/document/cons_doc_LAW_19671/
3. Федеральный закон "О контрактной системе в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд": https://www.consultant.ru/document/cons_doc_LAW_144624/
4. Уголовно-процессуальный кодекс РФ:
https://www.consultant.ru/document/cons_doc_LAW_34481/
5. Бюджетный кодекс РФ:
https://www.consultant.ru/document/cons_doc_LAW_19702/

Задание курсовой работы

Соответствие вариантов и источников:

Вариант	1	2	3	4	5	6	7	8	9	10
Источник	1	5	2	4	3	3	4	2	5	1

Варианты 1,3,5,7,9

Для указанного в вашем варианте источника данных реализуйте следующий функционал:

1. Загрузку документа с сайта;
2. Выборку текстовой части, без лишних вводных (заголовочных) и заключительных блоков (без всего, что не вошло в основные разделы);
3. Удаление названий частей, глав, разделов, статей и другой информации, относящейся к структуре текста, а не содержанию;
4. Выполните подготовку текста к анализу (лемматизацию, удаление знаков препинания и стоп-слов, нормализацию и т.д.);
5. Составьте предметный указатель для 100 основных слов, с указанием номера строки и символа, где слово встречается впервые в исходном (до всех обработок) тексте.
6. Дополнительное задание (для тех, кто пропустил 2 и более лекции):
 - i. Возьмите произвольный отрывок рассматриваемого документа (например, главу)

- ii. Сгенерируйте в любой нейронной сети «документ» на эту тему (приведите запрос и результат, а также укажите нейронную сеть)
- iii. Сравните зависимость количества уникальных слов в тексте от количества слов в тексте (нарастающим итогом в виде графиков)
- iv. Сделайте вывод о различии документа, написанного человеком и нейронной сетью.

Варианты 2,4,6,8,10

Для указанного в вашем варианте источника данных реализуйте следующий функционал:

1. Загрузку документа с сайта;
2. Выборку текстовой части, без лишних вводных (заголовочных) и заключительных блоков (без всего, что не вошло в основные разделы);
3. Удаление названий частей, глав, разделов, статей и другой информации, относящейся к структуре текста, а не содержанию;
4. Выполните подготовку текста к анализу (лемматизацию, удаление знаков препинания и стоп-слов, нормализацию и т.д.);
5. Определите примерные позиции в тексте, в которых происходит ступенчко-образное приращение числа слов словаря документа.
6. Дополнительное задание (для тех, кто пропустил 2 и более лекций):
 - i. Возьмите произвольный отрывок рассматриваемого документа (например, главу) достаточного для дальнейшей работы размера
 - ii. Сгенерируйте в любой нейронной сети «документ» на эту тему (приведите запрос и результат, а также укажите нейронную сеть)
 - iii. Сравните зависимость количества уникальных слов в тексте от количества слов в тексте (нарастающим итогом в виде графиков)
 - iv. Сделайте вывод о различии документа, написанного человеком и нейронной сетью.