



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 №3

по дисциплине

«Информационный поиск и извлечение информации из  
текстов»

Студент группы ИУ9-21М

\_\_\_\_\_  
(подпись, дата) С.С. Погосян

Руководитель

\_\_\_\_\_  
(подпись, дата) Н.В. Лукашевич

## 1. Постановка задачи

## 2. Решение

За основу взяты следующие факты:

- К большому террору в революционной Франции привели восстания федералистов
- Ножницы превратили прозу Гюстава Леруша в стихи Блеза Сандра.
- В годы правления Иди Амина из Уганды были изгнаны десятки тысяч азиатов

В приведенной в приложении программе было реализовано следующее:

1. Класс `PrepareFile` выполняет считывание текста из файла (в котором находятся все статьи, соответствующие своим фактам), токенизацию, текста и составление матрицы лемматизированных документов (предложений).
2. Класс `WeightMatrix` вычисляет матрицу терм-документ и нормализует ее
3. Класс `VectorOfRequest` вычисляет вектор запроса, на основе пространства состояний (набора лемм) и нормализует его
4. Класс `RankingDocument` рассчитывает косинусную близость между нормализованным запросом и нормализованной матрицей терм-документ и выдает ранжированный список вес-документ
5. Функция `main()` выполняет запрос

**Примечание:** в расчетах матрицы терм-документ использовалась следующая формула:

$$W_{td} = tf \cdot idf$$

Также использовались стоп слова при формировании запроса Код доступен по ссылке :