**Министерство образования Республики Беларусь**

**Учреждение образования**

**«Белорусский государственный университет**

**информатики и радиоэлектроники»**

ФАКУЛЬТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И УПРАВЛЕНИЯ

Кафедра интеллектуальных информационных технологий

**Отчёт по лабораторной работе №3**

**по курсу «ЕЯзИИС» на тему:**«Разработка системы автоматического реферирования документов»

|  |  |
| --- | --- |
| Выполнили студенты группы 921703: | Кравцов Михаил Сергеевич  Новицкий Сергей Владимирович  Семякин Артем Вадимович  Губаревич Иван Андреевич |
| Проверил: | Крапивин Юрий Борисович |

**Минск 2022**

Содержание

[1. Цель работы и вариант 3](#__RefHeading___Toc150_3878622861)

[2. Информация о текстовой коллекции документов 4](#__RefHeading___Toc152_3878622861)

[3. Описание системы, данных и алгоритмов 5](#__RefHeading___Toc154_3878622861)

[3.1. Описание структуры системы 5](#__RefHeading___Toc156_3878622861)

[3.2. Описание типов данных 5](#__RefHeading___Toc158_3878622861)

[3.3. Описание алгоритмов 6](#__RefHeading___Toc160_3878622861)

[3.4. Результат тестирования системы 7](#__RefHeading___Toc164_3878622861)

[4. Использование библиотек 9](#__RefHeading___Toc299_1650505274)

[5. Вывод 10](#__RefHeading___Toc242_1095348702)

1. Цель работы и вариант

Цель работы - освоить на практике основные принципы автоматического реферирования документов.

Вариант 4:



1. Информация о текстовой коллекции документов

Были использованы тексты из газет и классических произведений на соответствующем языке. В проекте можно найти следующие файлы:

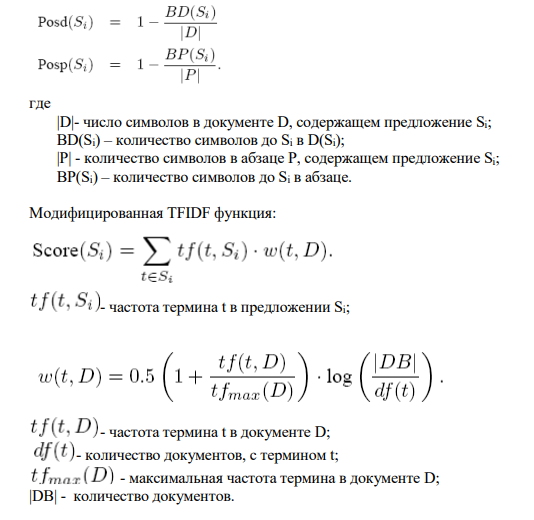
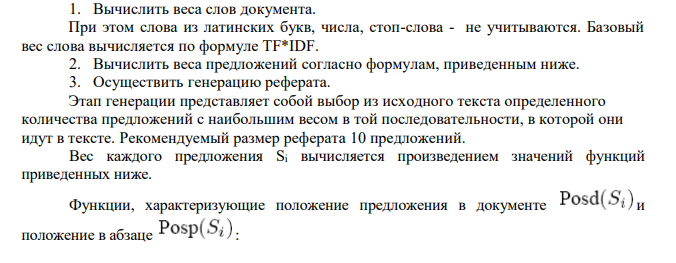
* computerScience.txt — содержит тренировочный текст на немецком языке;
* essayLiterature.txt — содержит тренировочный текст на русском языке;

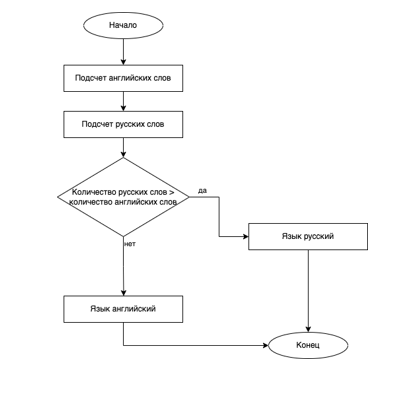
1. Описание системы, данных и алгоритмов
   1. Описание структуры системы

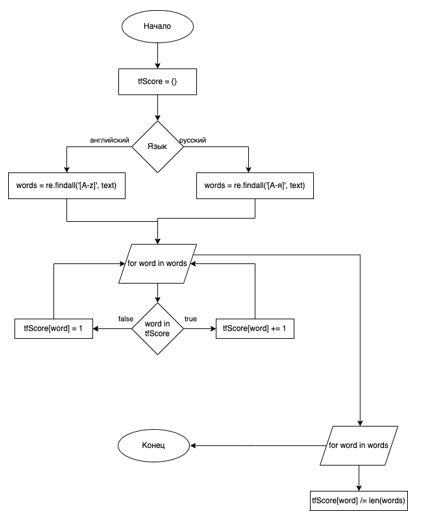
Система представляет собой веб-приложение, написанное с использованием Flask, присутствуют классы, ответственные за интерфейс и логику за ним. Разбор текста происходит с применением метода Sentence extraction и библиотеки nltk и встроенным ML. Система реализована на языке python.

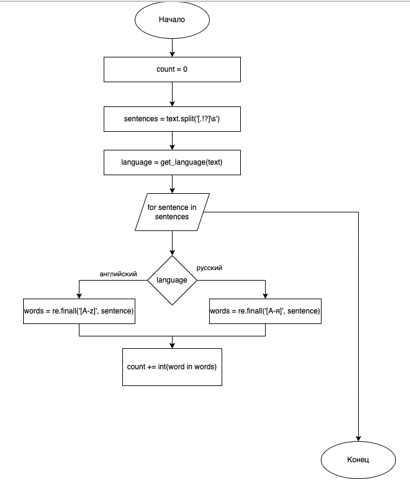
* 1. Описание типов данных

В данной системе мы использовали такие типы данных как строка, массив, ассоциативный массив.

* 1. Описание алгоритмов

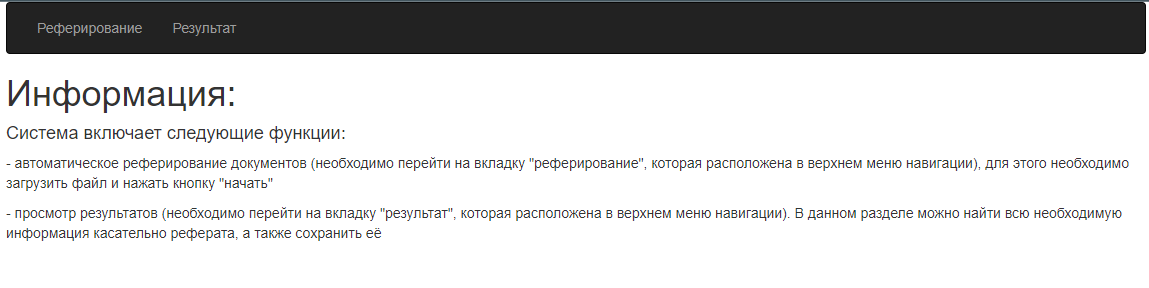


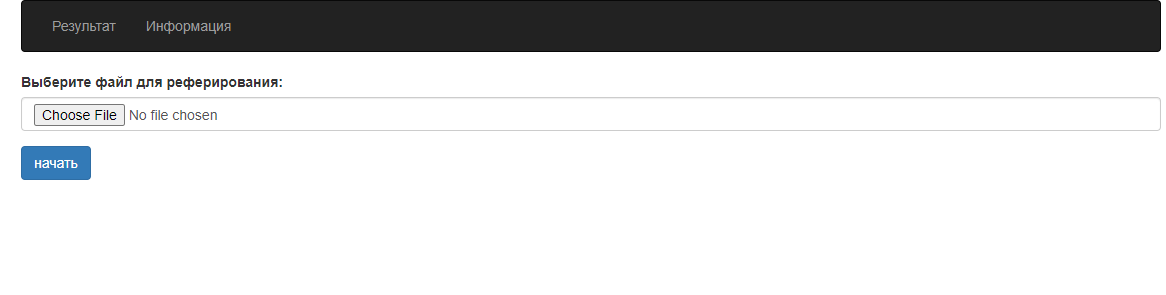


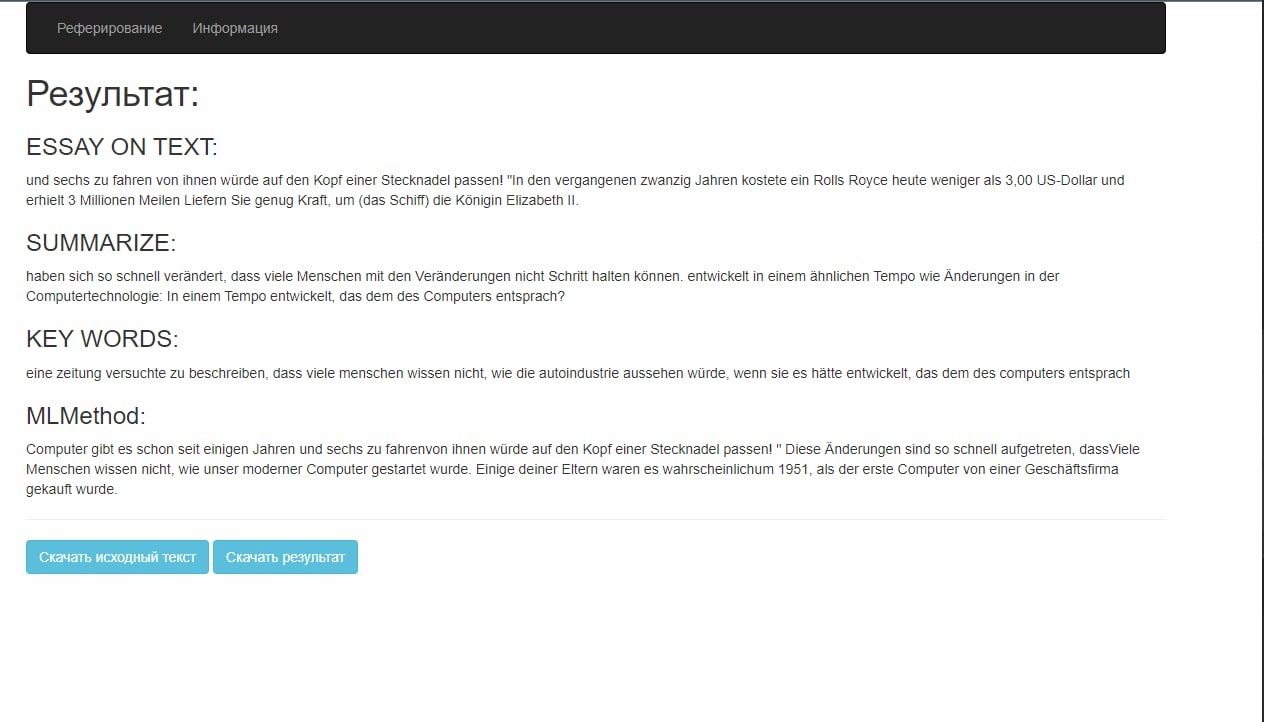


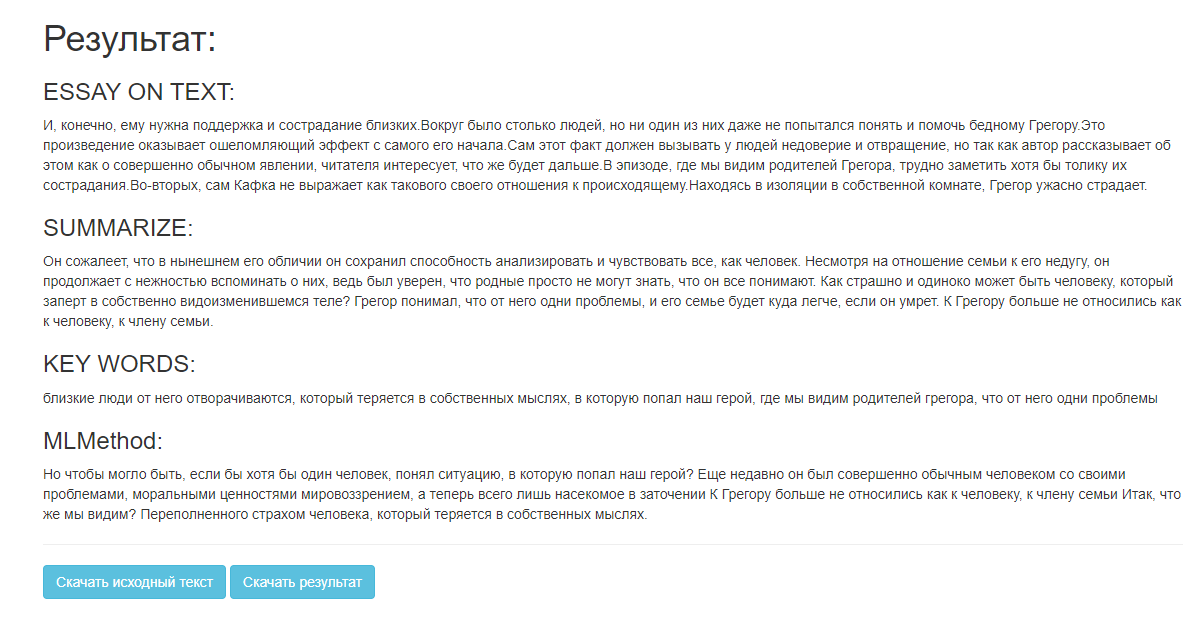
* 1. Результат тестирования системы

Пример взаимодействия с программой выглядит следующим образом









1. Использование библиотек

Для реализации системы использовалось библиотека nltk со встроенным Machine Learning, которая позволяет работать с естественным языков (в данном случае с русским и немецким).

Для выполнения лабораторной работы использовались следующие библиотеки:

1. nltk
2. nltk.corpus
3. math
4. gensim.summarization
5. rake\_nltk
6. string
7. nltk.corpus
8. nltk.cluster.util
9. numpy
10. networkx
11. sys
12. Вывод

В данной лабораторной работе была реализована система для автоматического реферирования текстов в виде веб приложения с удобным пользовательским интерфейсом, поддерживаются два языка: немецкий и русский.