

Учреждение образования
“БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ ”

Кафедра интеллектуальных информационных технологий

Отчет по лабораторной работе №3
по курсу: “Естественно-языковой интерфейс интеллектуальных систем”

Выполнили студенты группы 121701:

Пашин Н.А.
Маевский В.Г.

Проверил:

Крапивин Ю.Б.

Минск

2022

Цель работы: освоить на практике основные принципы автоматического реферирования документов с учетом функционала, предоставляемого технологией OSTIS.

Вариант 13:

Язык текста: русский, немецкий.

Предметная область: Научные статьи по медицине, Критика предметов изобразительного искусства.

Требования к разрабатываемой системе:

- на входе – текстовые документы одинакового размера (например, 10 страниц формата А4), содержащие тексты из предметных областей на естественных языках согласно варианту подлежащие процедуре автоматического реферирования;
- на выходе – активная ссылка на исходный документ и построенный реферат документа, состоящий из 2-х разделов:
 - классического реферата;
 - реферата в виде списка ключевых слов;
- наличие средств сохранения в файл и распечатки полученной на выходе информации;
- интерфейс системы должен быть предельно простым и доступным для пользователей любого уровня, содержать понятный набор инструментов и средств, а также help-средства.

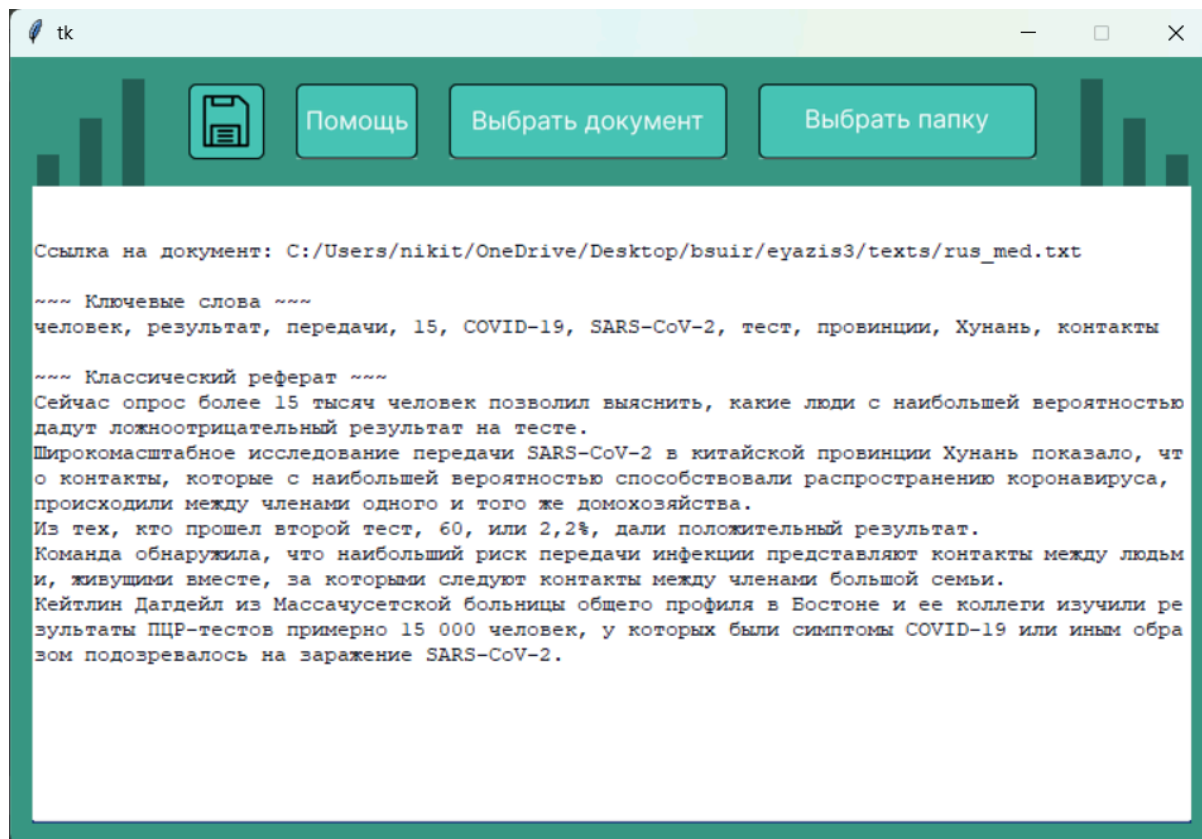
Структура разработанной системы

Система реализована на языке Python версия 3.10 . Графический интерфейс системы создан с помощью библиотеки tkinter. Система может принимать тексты путем ввода абсолютного пути файла. Выполняет реферирование введенного текста. Все результаты сохраняются в текстовый файл.

Классический реферат вычисляется при помощи библиотеки sklearn и ее

методов TfidfVectorizer и cosine_similarity. Реферат коротких слов вычисляется при помощи библиотеки nltk и ее метода FreqDist.

Тестирование:



Описание и особенности применения готовых к использованию компонент

Tkinter– это набор библиотек и инструментов для разработки графических пользовательских интерфейсов (GUI) на языке программирования Python.

NLTK — это библиотека для обработки естественного языка, которая предоставляет инструменты и ресурсы для работы с текстами.

Scikit-learn – библиотека для Python, предназначенная для выполнения научных и инженерных вычислений. Она строится поверх более базовой библиотеки NumPy и предоставляет множество дополнительных функций и инструментов, которые облегчают выполнение различных вычислительных задач. В данном случае для преобразования текстовых данных в числовые векторы.

Вывод:

В результате выполнения лабораторной работы была разработана система автоматического реферирования, с возможностью создания реферата в виде списка ключевых слов, классического реферата. Автоматическое реферирование производится для текстов на русском и немецком языках.