Министерство образования Республики Беларусь

Учреждение образования “Белорусский государственный университет

информатики и радиоэлектроники”

Факультет информационных технологий и управления

Кафедра интеллектуальных информационных технологий

Отчёт по дисциплине

“Естественно-языковой интерфейс интеллектуальных систем”

Лабораторная работа №2

“Построение и использование корпусов текстов естественного языка”

Выполнил студент группы 221702: Скоробогатый А.Л

Проверил: Крапивин Ю.Б.

Минск 2025

Лабораторная работа №2

**Цель работы:**

1. Изучить принципы построения корпусов текстов, виды разметки и способы аннотирования, инструменты работы с корпусами текстов.
2. Построить корпус текстов и разработать корпусный менеджер.

**Задание:**

1. Сформировать электронный корпус текстов по выбранной предметной

области.

2. Используя результаты лабораторной работы №1 (возможность получения

лингвистических сведений для произвольной лексемы естественного

языка) разработать корпусный менеджер, обеспечивающий базовую

функциональность работы с созданным корпусом текстов.

**Вариант 15:**

Растения

**Язык текста:**

Английский

**Формат входного документа:**

TXT

**Используемые средства разработки:**

**Tkinter** — стандартная библиотека Python для создания графических пользовательских интерфейсов (GUI)

**NLTK (Natural Language Toolkit)** — библиотека для обработки естественного языка

**re** — стандартный модуль Python для регулярных выражений, используется для поиска слов в контексте (конкорданс).

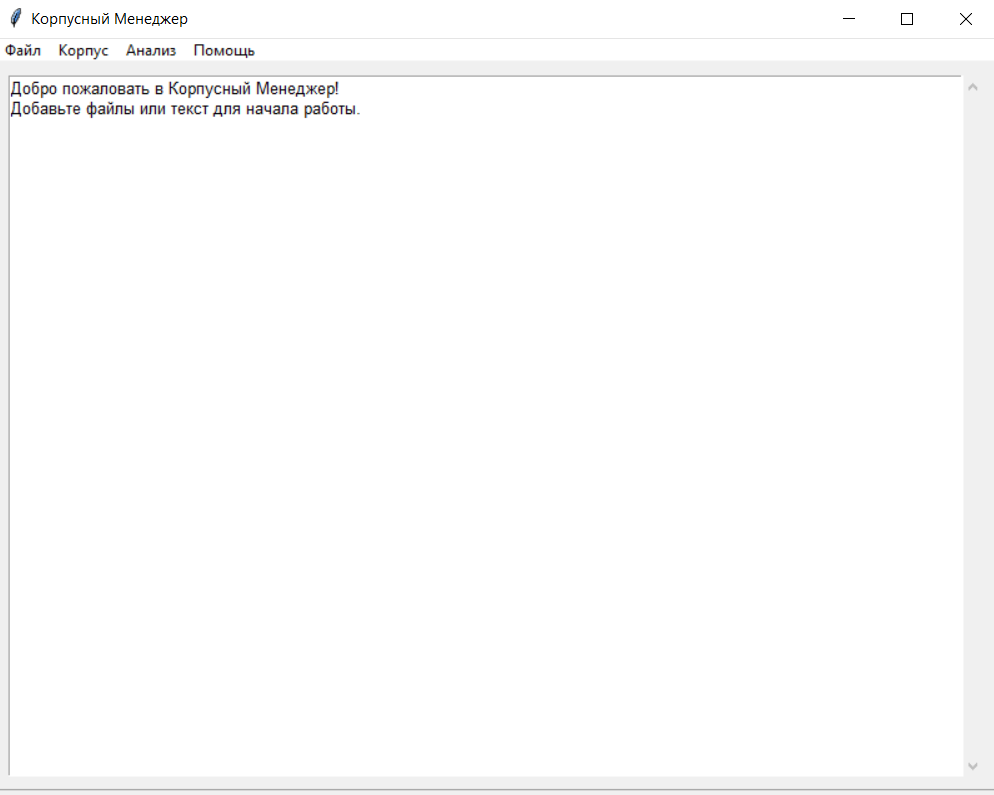
**collections.Counter** — встроенный Python-модуль для подсчёта частоты слов

**pickle** — для сериализации и сохранения/загрузки корпуса в файл.

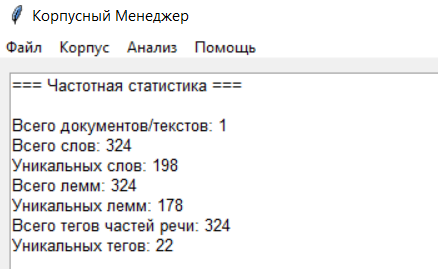
**os** — используется для работы с именами файлов и путями.

**Скриншоты разработанной системы:**

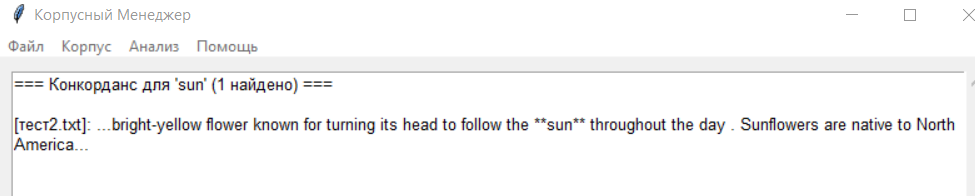
**Главное меню:**



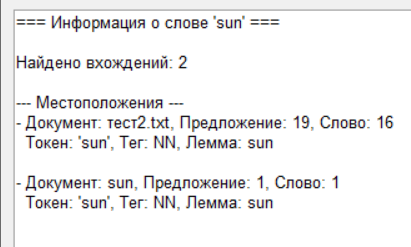
**Статистика корпуса:**



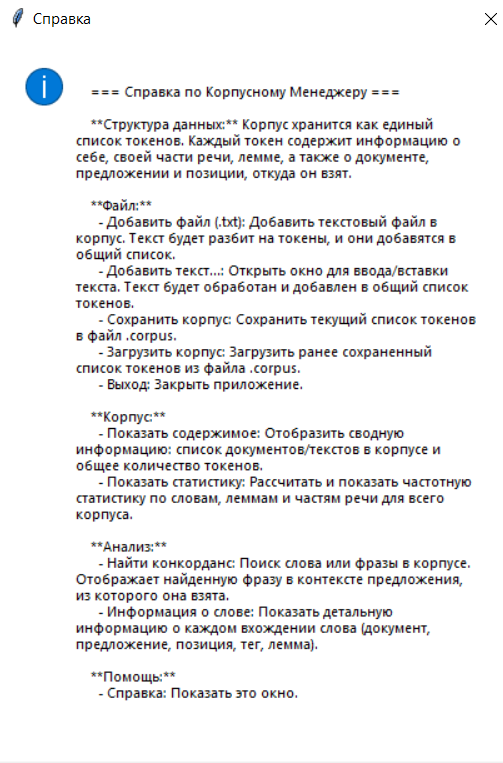
**Найти конкорданс:**



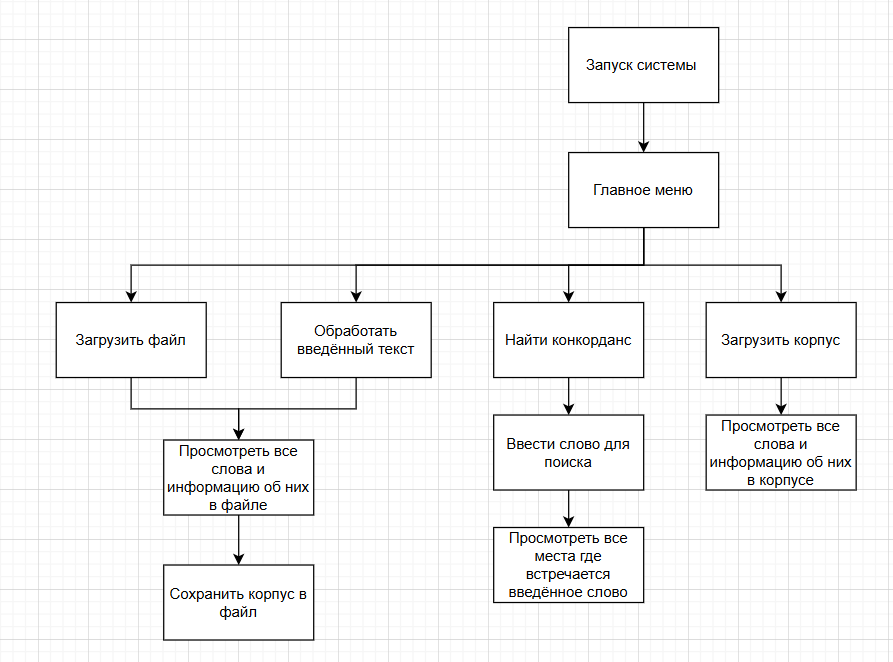
**Информация о слове:**



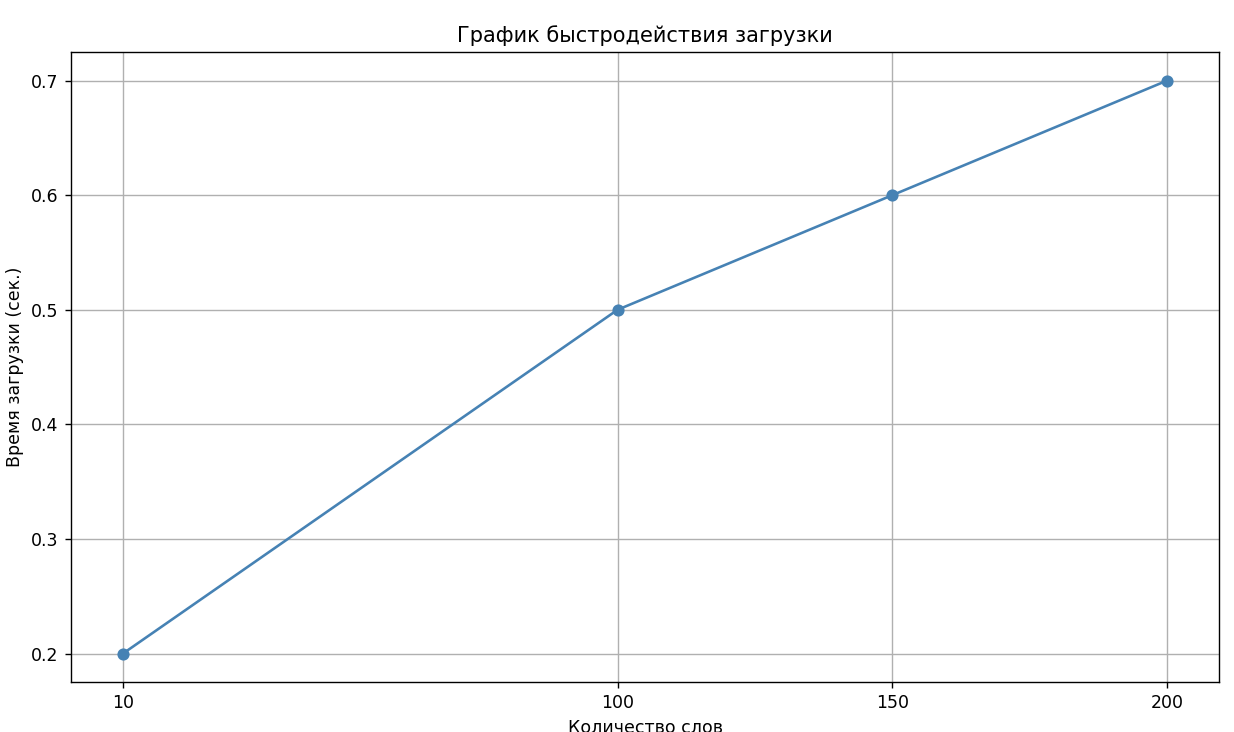
**Вкладка помощь:**



**Структурно-функциональная схема разрабатываемого приложения:**



**График оценки быстродействия приложения:**



**Выводы по работе и по перспективам развития приложения:**

Разработанное приложение представляет собой инструмент для построения и анализа текстового корпуса с удобным графическим интерфейсом. В процессе работы была реализована возможность добавления текстов как из файлов, так и вручную, с последующей автоматической лингвистической обработкой — включая токенизацию, определение частей речи и лемматизацию. Программа также предоставляет функции получения статистики по корпусу, поиска слова в конкордансе, а также отображения лемм и тегов для заданного слова.

Программа уже может быть полезна в учебных и исследовательских целях, однако существует потенциал для дальнейшего развития. В перспективе возможно улучшение визуализации статистики, добавление экспорта данных в другие форматы, а также реализация поддержки других языков. Кроме того, можно расширить функциональность поиска, добавив морфологическую нормализацию запросов. Перевод приложения в веб-формат также может существенно расширить круг пользователей и сделать работу с корпусом более гибкой и доступной.