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 №3-4

“Семантико-синтаксический анализ текстов естественного языка”

Выполнил студент группы 221702: Скоробогатый А.Л

Проверил: Крапивин Ю.Б.

Минск 2025

**Цель работы:**

Освоить принципы разработки прикладных сервисных программ для решения задачи автоматического семантико-синтаксического анализа текста естественного языка.

**Задачи лабораторной работ**

**Задание:**

1. Познакомиться с назначением, структурой и функциональностью, предоставляемой базовым ЛП для решения задачи автоматического семантико-синтаксического анализа ТЕЯ.

­2.Закрепить навыки программирования при решении задач

автоматической обработки ТЕЯ.

**Вариант 6**

**Язык текста:**

Английский

**Формат входного документа:**

HTML

**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

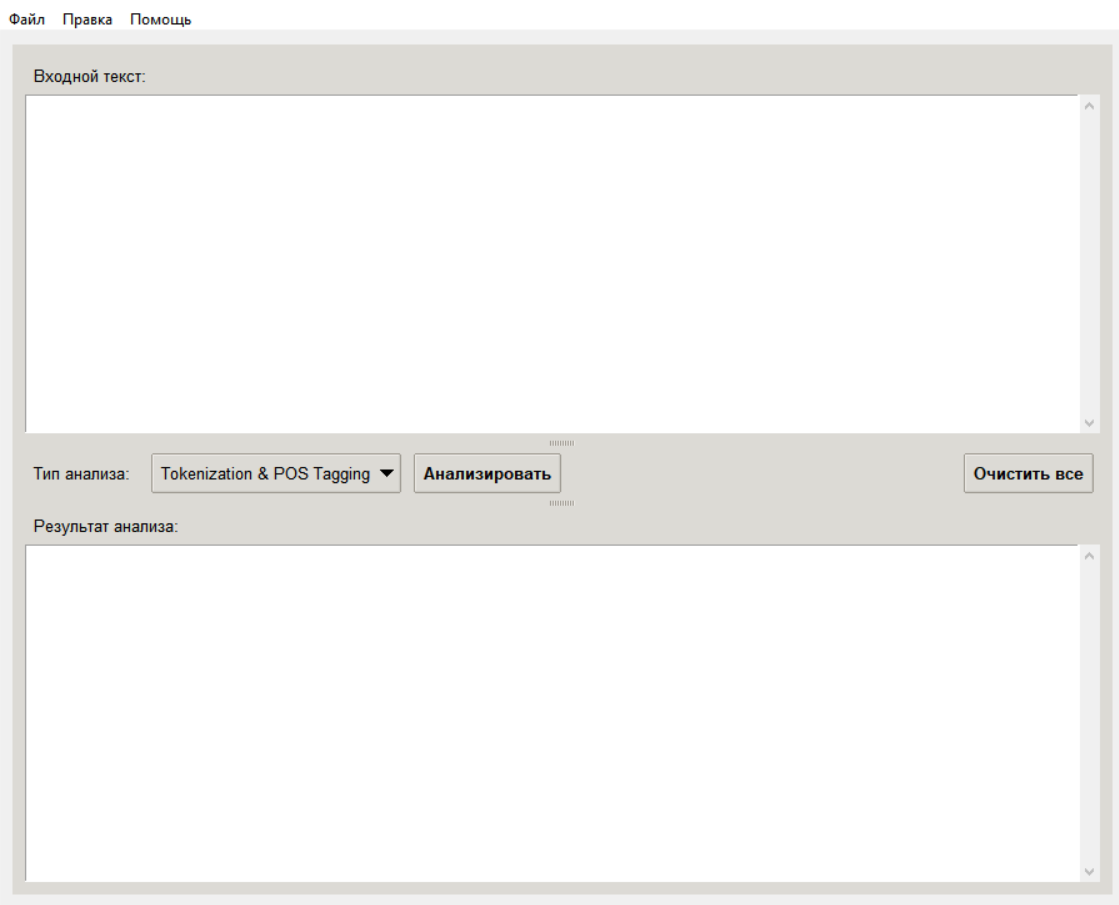
**spacy** — для синтаксического анализа зависимостей (Dependency Parsing).

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

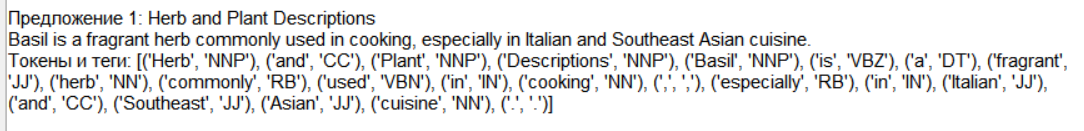
**json**: Встроенная в Python библиотека для работы с форматом JSON

**Скриншоты разработанной системы:**

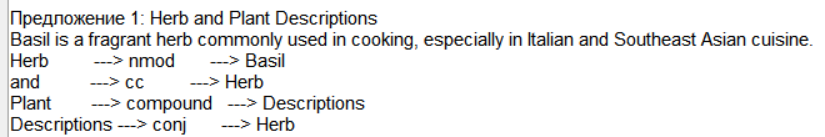
**Главное меню:**



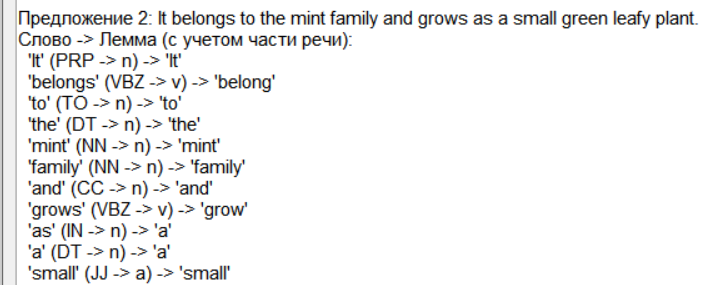
**Токенизация:**



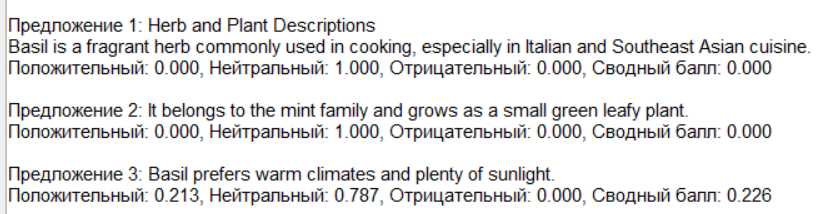
**Дерево зависимостей:**



**Лемматизация:**



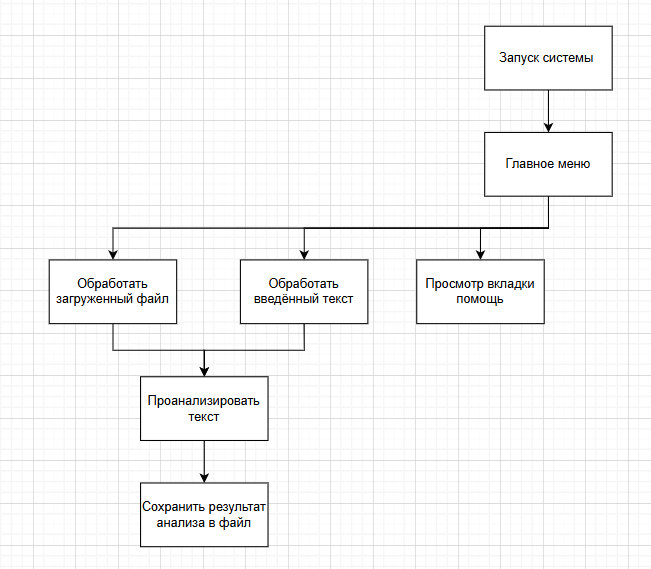
**Тональность:**



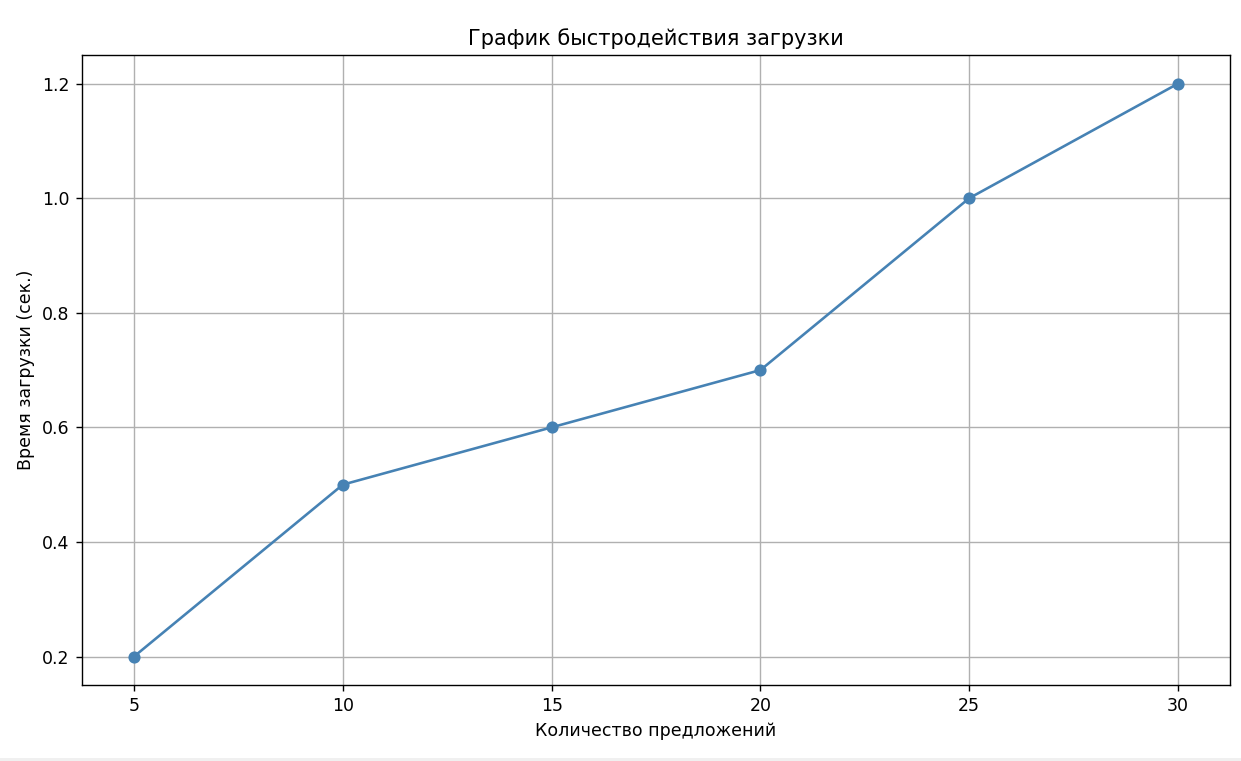
**Вкладка помощь:**



**Структурно-функциональная схема разрабатываемого приложения:**



**График оценки быстродействия приложения:**



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

В ходе выполнения лабораторной работы была разработана модульная система для автоматического семантико‑синтаксического анализа текстов на русском языке. На вход программа принимает документы в самых распространённых форматах, автоматически выполняет токенизацию, лемматизацию и определение частей речи с помощью библиотеки NLTK, строит синтаксические деревья зависимостей с использованием SpaCy и потом сопоставляет семантические роли субъекта, предиката и объекта, обращаясь к лексическим базам WordNet и ConceptNet. Пользователь может загружать текст, запускать анализ, просматривать результаты в виде графа связей прямо в интерфейсе и в любой момент редактировать или сохранять полученные структуры. Интуитивно понятный дизайн, всплывающие подсказки и встроенная справочная система делают процесс работы комфортным даже для тех, кто впервые сталкивается с подобными инструментами.

Перспективы развития системы заключаются прежде всего в расширении языковой базы и обогащении семантической составляющей. Включение поддержки европейских и азиатских языков, подключение специализированных онтологий и использование фреймов из FrameNet дадут возможность переходить от простого разбора «субъект–предикат–объект» к многослойным семантическим сценариям. Кроме того, оптимизация для обработки больших потоков текстовых данных с применением асинхронных очередей и распределённых вычислений позволит развёртывать сервис в облачных инфраструктурах и корпоративных системах, где требуется автоматическое извлечение фактов, семантический поиск и автоматическое резюмирование документов.