

Учреждение образования
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ
Кафедра интеллектуальных информационных технологий

Отчёт
по курсу «Естественно-языковой интерфейс интеллектуальных
систем»

Лабораторная работа №4
«Семантико-синтаксический анализ текстов естественного языка»

Выполнили студенты группы 121701:	Липский Р. В. Жолнерчик И. А. Стронгин А. В.
Проверил:	Крапивин Ю.Б.

Минск 2024

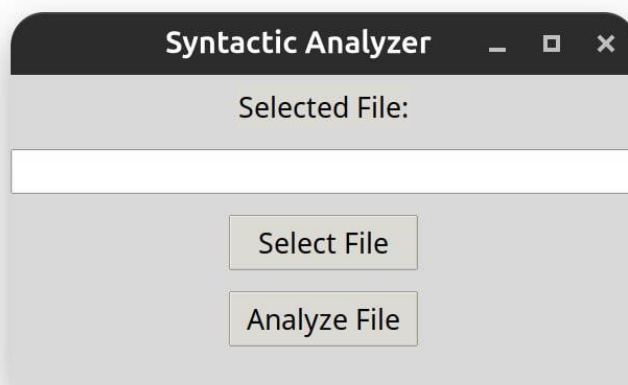
Цель работы:

освоить принципы разработки прикладных сервисных программ для решения задачи автоматического семантико-синтаксического анализа ТЕЯ.

Задание:

познакомиться со значением, структурой и функциональностью, предоставляемой базовым для решения задачи автоматического семантико-синтаксического анализа ТЕЯ. Закрепить навыки программирования при решении задач автоматической обработки ТЕЯ.

Интерфейс:



Стартовое окно выбора файла

Окно предоставляет возможность выбора файла для проведения анализа с помощью кнопки “Select File” и подтверждение выбора с помощью кнопки “Analyze File”, которая запускает процесс анализа файла.

Save

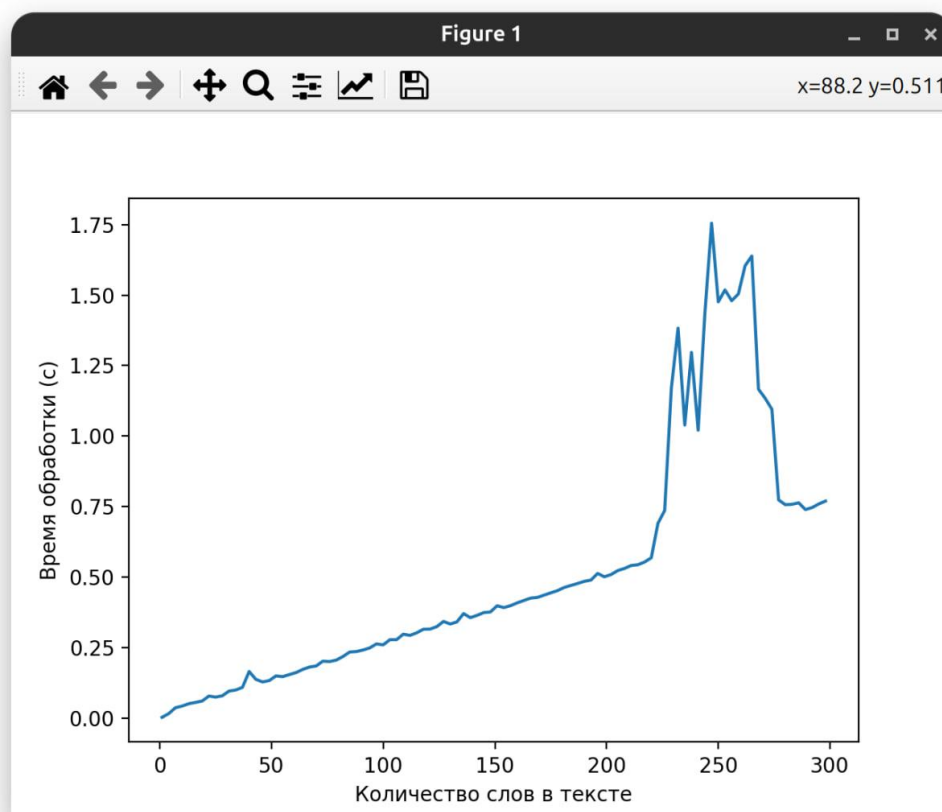
Analysis Report

Artificial intelligence AI touted as the harbinger of technological revolution has proven time and again to be nothing more than an overhyped and underwhelming endeavor While some perceive AI as groundbreaking advancement the reality paints far bleaker picture Numerous AI companies have emerged claiming to possess the panacea to all our problems but their contributions have been disappointingly lackluster Let's start with AlphaCorp once prominent AI company that promised to revolutionize industries with its advanced algorithms and machine learning techniques However their products have consistently fallen short of expectations Their flagship AI system touted as game changer in healthcare has been riddled with errors and inaccuracies putting patients lives at risk It's clear that AlphaCorp's AI is nothing more than glorified marketing gimmick incapable of delivering on its lofty promises Another infamous player in the AI arena is BetaTech They claim to have developed revolutionary AI platform that can automate complex business processes but the reality is far from their claims Countless businesses that eagerly adopted BetaTech's AI solution soon discovered its inherent flaws The system was plagued with inefficiencies resulting in costly mistakes and widespread frustration It became painfully evident that BetaTech's AI was nothing more than facade poorly disguised attempt to capitalize on the AI hype And let's not forget about GammaSoft company that boasted about its AI's ability to analyze and predict stock market trends Many investors lured by the promise of unprecedented profits eagerly invested their hard earned money into GammaSoft's AI powered trading platform However the results were catastrophic The AI consistently made incorrect predictions leading to devastating financial losses for those who trusted in its capabilities GammaSoft's AI turned out to be nothing more than an elaborate deception preying on the gullibility of investors These examples are just the tip of the iceberg when it comes to the failures of AI companies The truth is the field of AI has been marred by overpromises and underdeliveries Companies that claim to possess cutting edge technologies often turn out to be mere charlatans exploiting the naivety of consumers and businesses alike In conclusion AI has proven to be deeply flawed and disappointing endeavor The grandiose claims made by AI companies have repeatedly been exposed as hollow and unreliable From AlphaCorp's disastrous healthcare AI to BetaTech's inefficient business automation and GammaSoft's disastrous stock market predictions the landscape of AI is littered with broken promises and shattered dreams Rather than being the panacea to our problems AI has become cautionary tale of hubris and deception

Окно вывода результата синтаксического разбора

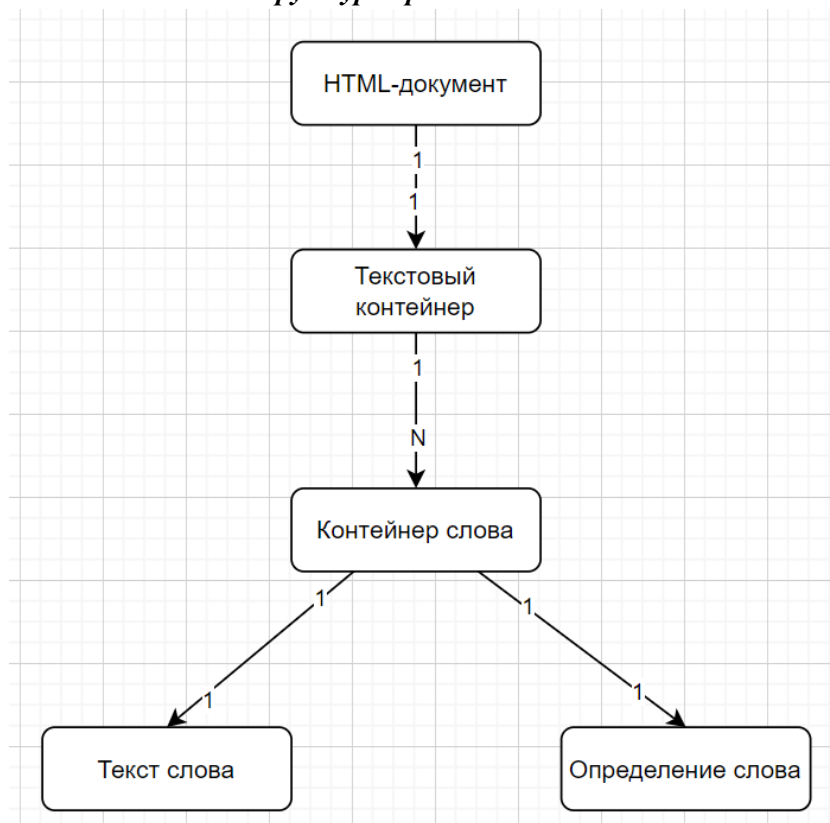
После успешного выбора файла и триггера кнопки активации процесса анализа пользователь получает результат в виде HTML-файла. На рисунке изображён пример результата анализа предложений. При наведении на слово пользователь получает определение слова. Также при нажатии на слово его определение можно корректировать.

Тест производительности:



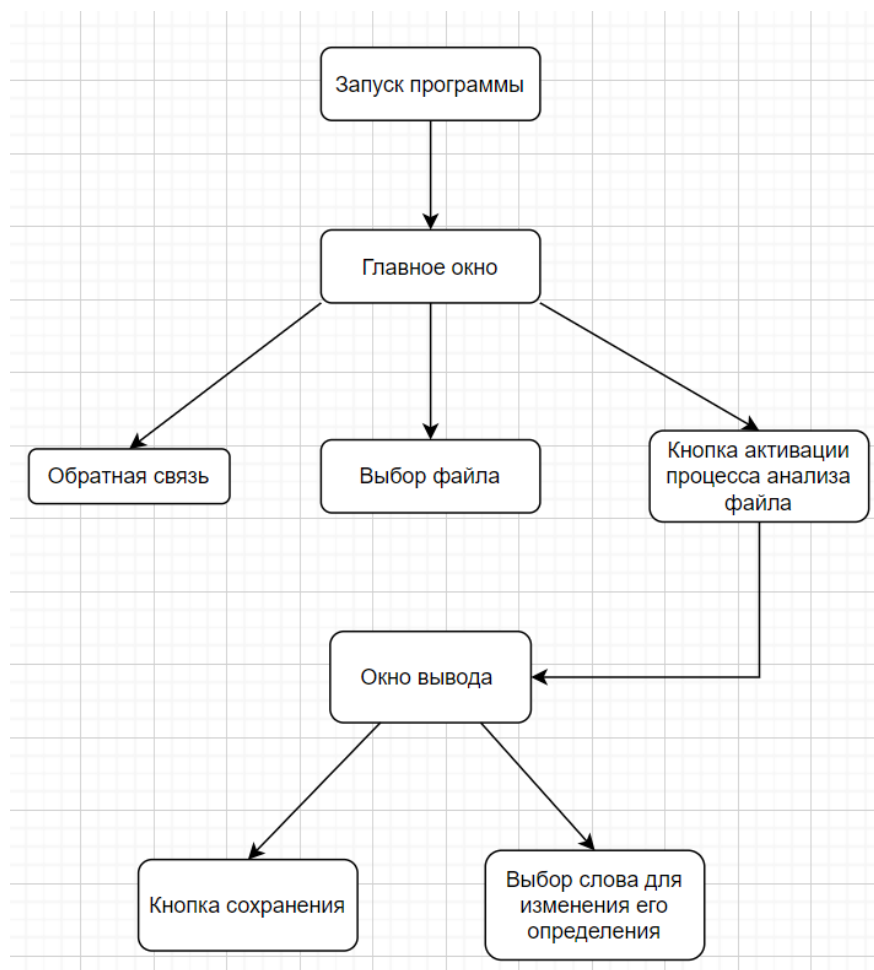
На данном графике изображена зависимость времени проведения синтаксического анализа (в секундах) от количества слов в анализируемом тексте. Исходя из графика очевидно, что зависимость имеет линейный характер. Резкий скачок во времени обработки текстов более 250 слов объясняется переходом машины, на которой исполнялось приложение, в энергосберегающий режим.

Структура хранения данных



Результаты синтаксического анализа представлены в виде редактируемого отчёта в формате HTML-документа. Один документ содержит в себе один текстовый контейнер, который в свою очередь содержит N контейнеров слов (где N – количество слов в тексте). Каждый контейнер слова содержит в себе информацию о тексте слова (отображается всегда) и определение слова (отображается при наведении курсора).

Структурная схема приложения



После запуска программы пользователя встречает главное окно, содержащее три элемента: кнопку для обратной связи, кнопку для выбора файла и кнопку для активации процесса анализа файла. Кнопка для активации процесса анализа файла запускает процесс синтаксического анализа файла и отображает пользователю редактируемый интерактивный отчёт. Внутри отчёта пользователь имеет возможность просматривать и изменять определения слов текста, а также сохранить актуальную версию отчёта.

Описание алгоритмов:

Разбиение текста на словоформы и получение их определений:

Начало – Получаем файл формата TXT/RTF – Считываем текст файла – Получаем список словоформ – Получаем определения каждой – Генерируем размеченный HTML-текст – Отображаем HTML-файл – Конец

Генерация размеченного HTML-текста: Начало – Получаем результат синтаксического анализа – Загружаем шаблон HTML-отчёта – Вставляем слова и семантическую информацию – Возвращаем сгенерированный HTML-текст – Конец.

Алгоритм работы интерактивного HTML-отчёта: Начало – Ожидаем действий пользователя (1)
– Если пользователь закрыл отчёт, то переходим к шагу (2) – Если пользователь не изменял определение слова, возвращаемся к шагу (1) – Сохраняем новое определение – Переходим к шагу (1) – Конец (2).

Вывод:

В результате выполнения данной лабораторной работы мы изучили возможности библиотеки spaCy и nltk в проведении семантического анализа текстов, разработали алгоритмы для семантического анализа текстов, а также дружелюбный пользовательский интерфейс. Наработки, созданные в рамках данной лабораторной работы, могут быть полезны для последующих исследований и разработок в сфере анализа естественного языка, машинного обучения и лингвистического анализа.