

Учреждение образования
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ»

Кафедра интеллектуальных информационных технологий

Отчёт по лабораторной работе №1 по курсу «ЕЯзИИС»

Выполнили студенты
группы 021703:
Озерец Д.А
Анищенко Е.С
Бадей С.П

Проверил:
Крапивин Ю.Б.

МИНСК 2023

1. Цель:

Освоить принципы разработки прикладных сервисных программ для решения задачи автоматического лексического и лексико-грамматического анализа текста естественного языка.

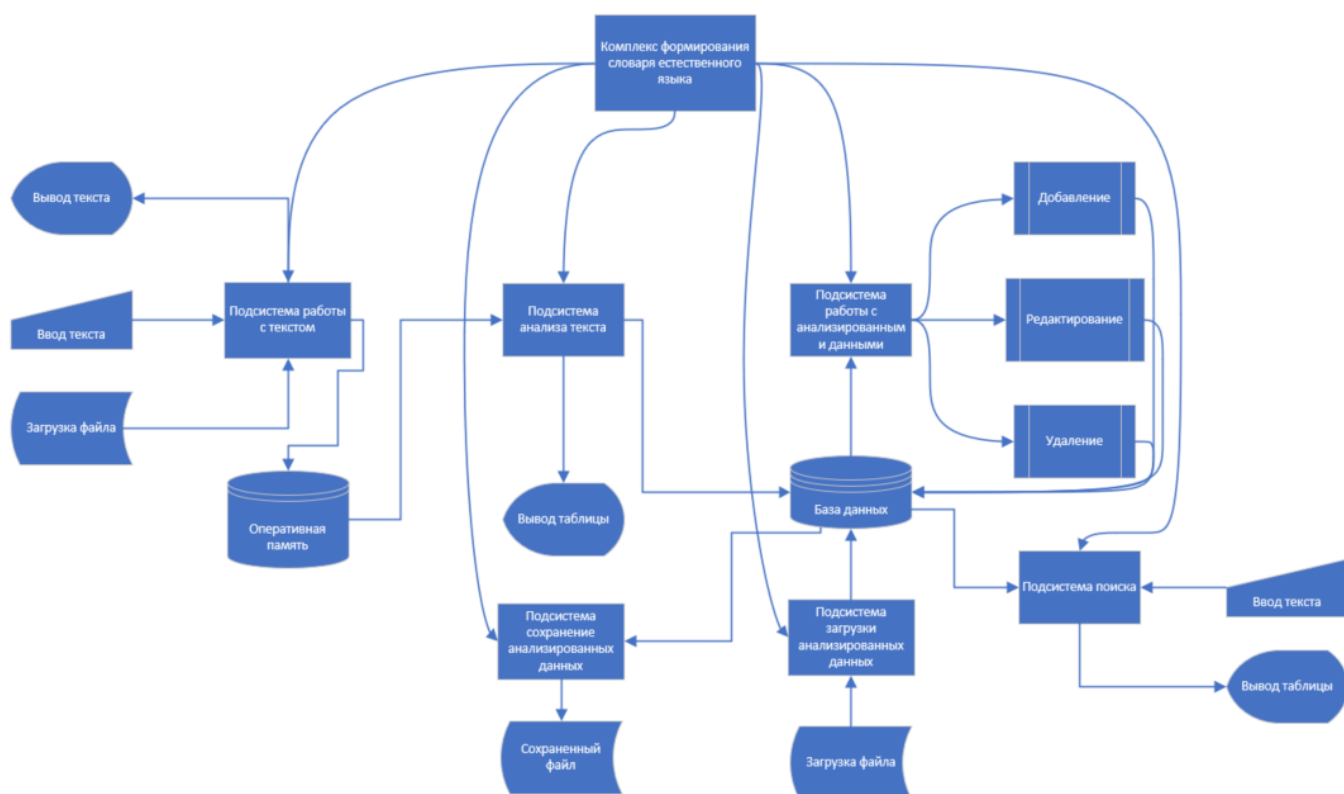
2. Постановка задачи:

1. Познакомиться с назначением, структурой и функциональностью, предоставляемой базовым ЛП для решения задачи автоматического лексического и лексико-грамматического анализа ТЕЯ.

2. Закрепить навыки программирования при решении задач автоматической обработки ТЕЯ.

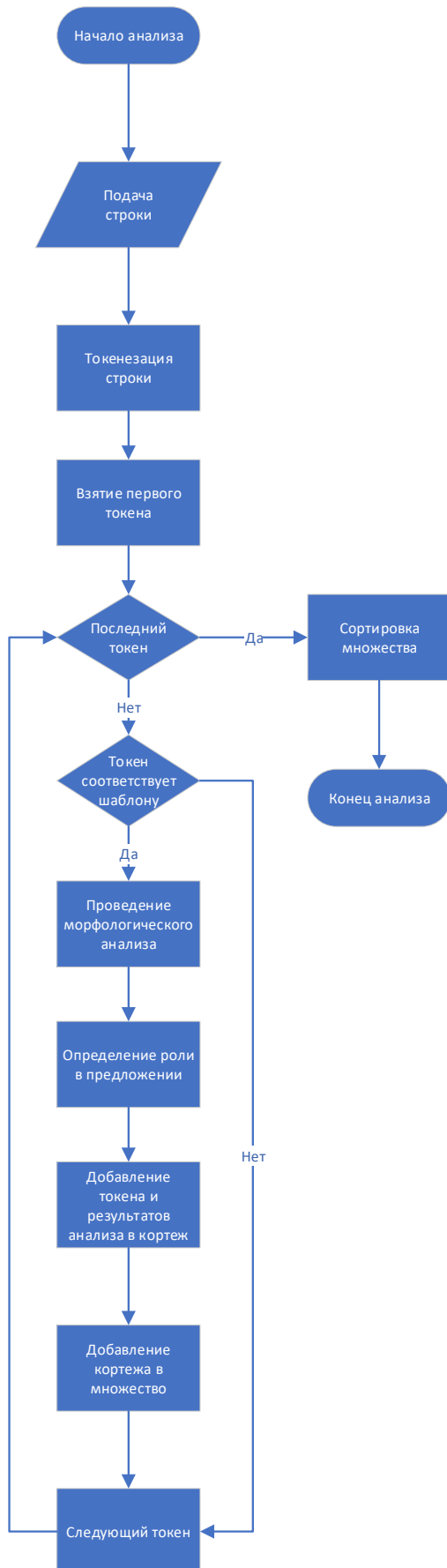
Задание 4. Список слов, упорядоченный по алфавиту и включающий только лексемы с дополнительно оформленными записями о месте и роли данного слова в составе предложения. К такой информации относится описание того, каким членом предложения может быть данное слово и в какой форме (падеж, число, время и т.п.). Например, если это существительное в именительном падеже, то оно может выступать в роли подлежащего; если это существительное в родительном падеже, то оно может быть дополнением; если это прилагательное, то оно может быть определением и т.п.

3. Структурно-функциональная схема разработанного приложения:



4. Описание структур хранения данных, алгоритмов их обработки, необходимых для реализации базовых требований к разработанной программе.

Для хранения данных использовался стандартный тип set, а также файлы форматов TXT и RTF. Алгоритм продемонстрирован ниже:



Для реализации указанного приложения использовались стандартные средства Python, а также следующие библиотеки:

- Django – для графического интерфейса;
- pymorphy3 – для морфологического анализа;
- spacy – для токенизации;
- striprt – для парсинга RTF.

Графический интерфейс разработанного приложения:

The screenshot displays the application's graphical user interface. On the left, there is a text input area with a sample text about Mr. and Mrs. Dursley. Below the input are two buttons: 'Загрузка файла' (File Upload) and 'Выгрузить текст' (Export Text). On the right, there is a table showing the morphological analysis of the text. The table has three columns: 'Лексема' (Lexeme), 'Морф разбор' (Morphological Analysis), and 'Роль в предложении' (Role in Sentence). Below the table are buttons for 'Поиск' (Search), 'Анализировать' (Analyze), and a search icon. There are also icons for adding, deleting, and downloading data.

Лексема	Морф разбор	Роль в предложении
а	Не определен	СОЮЗ
абсолютно	Обстоятельство	Н
автомобиль	Подлежащее	СУЩ, неод, мр, ед, им
автомобильную	Определение	ПРИЛ жр, ед, вн

В данном интерфейсе реализовано:

1. Загрузка файла
2. Выгрузка текста
3. Анализ текста
4. Удаление, добавление, редактирование проанализированных слов
5. Поиск словам и результатам анализа
6. Сохранение полученного анализа
7. Загрузка проанализированных данных
8. Кнопка помощи

По времени дольше всего выполняется анализ текста и загрузка результата в БД:

```
--- 1.1392643451690674 seconds ---
```

5. Выводы.

В результате выполнения лабораторной работы были закреплены навыки работы с языковым материалом. Были освоены принципы разработки прикладных сервисных программ для решения задачи автоматического лексического и лексико-грамматического анализа текста естественного языка. Получены навыки работы с Python, Django, pymorph3, spacy, БД.