

Министерство образования Республики Беларусь

**Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»**

ФАКУЛЬТЕТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И УПРАВЛЕНИЯ

Кафедра интеллектуальных информационных технологий

**Отчёт по лабораторной работе №2
по курсу «ЕЯзИИС» на тему:
«Синтаксический анализ текстов естественного языка»**

Выполнил студент группы 921703:	Кравцов Михаил Сергеевич
Проверил:	Крапивин Юрий Борисович

Минск 24.04.2022

1. Условие	3
1.1. Цель	3
1.2. Постановка задачи	3
1.3. Задание	3
2. Схемы программы	4
3. Оценка быстродействия	6
4. Используемые библиотеки и разработанные классы	7
Выводы	8

1. Условие

1.1. Цель

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

1.2. Постановка задачи

1. Познакомиться с назначением, структурой и функциональностью, предоставляемой базовым ЛП для решения задачи автоматического синтаксического анализа ТЕЯ.
2. Закрепить навыки программирования при решении задач автоматической обработки ТЕЯ.

1.3. Задание

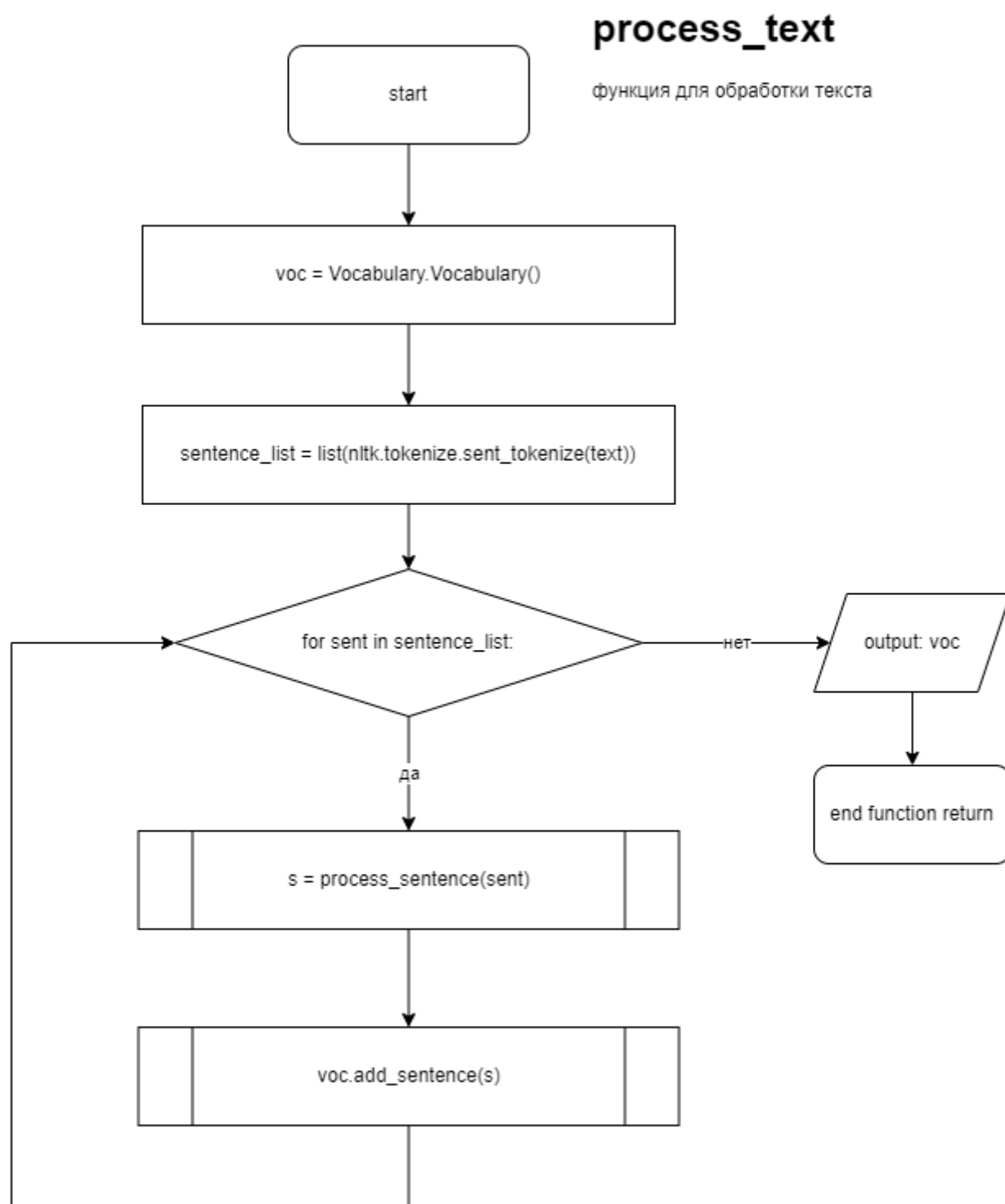
Требуется спроектировать и программно реализовать структуры хранения данных, алгоритмы их обработки, необходимые в рамках следующих базовых требований к разрабатываемому приложению:

- входные данные – текст заданного естественного языка;
- выходные данные – структуры, полученные при проведении автоматического синтаксического анализа предложений входного текста согласно варианта задания;
- взаимодействие с пользователем посредством графического интерфейса (интерфейс должен быть интуитивно-понятным и дружелюбным пользователю);
- наличие системы средств помощи пользователю;
- обеспечение возможности построения, сохранения, просмотра, редактирования, документирования автоматически получаемого результата либо заданной его части;
- поддержка различных форматов представления входных данных (TXT, RTF, PDF, HTML, DOC, DOCX).

Вариант:

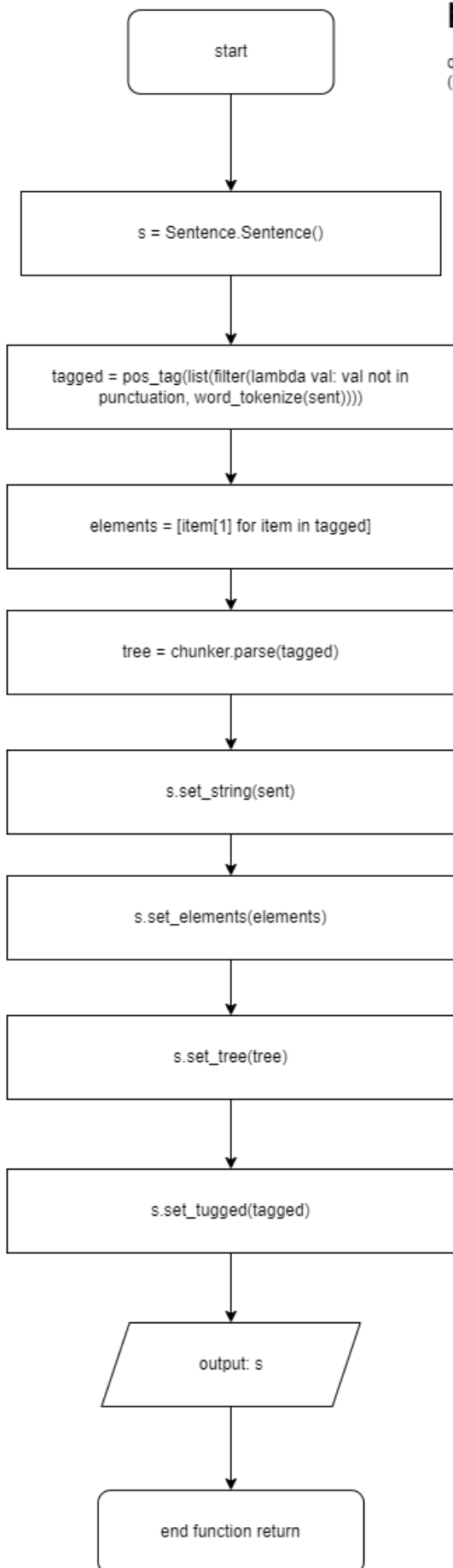
11	Английский	DOCX
----	------------	------

2. Схемы программы



process_sentence

функция для обработки предложения
(используется в process_text)



3. Оценка быстродействия

Быстродействие программы, как известно, зависит от сложности применённого алгоритма. Поэтому подсчитаем сложность алгоритма используя нотацию $O()$.

Создаем объект типа `Vocabulary`, в котором будет храниться результат. Преобразуем входной текст в предложения и над каждым предложением проделываем следующие операции: производим синтаксический анализ и сохраняем результат (объект типа `Sentence`) в `Vocabulary`.

Синтаксически анализ проводится следующим образом: помечаем все слова, предварительно их отфильтровав от знаков препинания функцией `pos_tag()` библиотеки `nltk`. После чего сохраняем все элементы предложения и составляем дерево с использованием функции `chunker.parse()`. Результаты сохраняем в объекте типа `Sentence`.

Сложность составляет $O(n*m*r*q)$, где n – кол-во полученных предложений, m – кол-во полученных слов, r – кол-во элементов предложения, q – расходы на вызов функций библиотеки `nltk`.

4. Используемые библиотеки и разработанные классы

При выполнении поставленных задач были использованы следующие библиотеки:

- 1) Tkinter
- 2) nltk
- 3) docx
- 4) pickle
- 5) re
- 6) string

Были разработаны классы Sentence, Vocabulary, App, Controller, Table, View, также функции для обработки engine.

Код можно посмотреть по ссылке - https://github.com/mixa1334/laba2_eyazis_6sem

Выводы

В данной лабораторной работе было создано и протестировано приложение для проведения автоматического синтаксического анализа предложений текста естественного языка (английский). Приложение получилось простое в использовании и довольно функциональным, что дает ему преимущества.