

Министерство образования Республики Беларусь

Учреждение образования

“Белорусский государственный университет информатики и  
радиоэлектроники”

Кафедра информационных интеллектуальных технологий

**Лабораторная работа 4**

**Вариант 7**

**“Семантико-синтаксический анализ  
текстов естественного языка”**

Выполнил

гр.121701

Мулярчик Д.С.

Лемантович Д.К.

Проверил

Крапивин Ю. Б

Минск 2024

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

**Используемые инструменты:** Python с PyQt и Natural Language Toolkit.

**Структуры хранения:** TXT-файлы

### Структурно-функциональная схема

Структурно-функциональная схема приложения представлена черным ящиком:

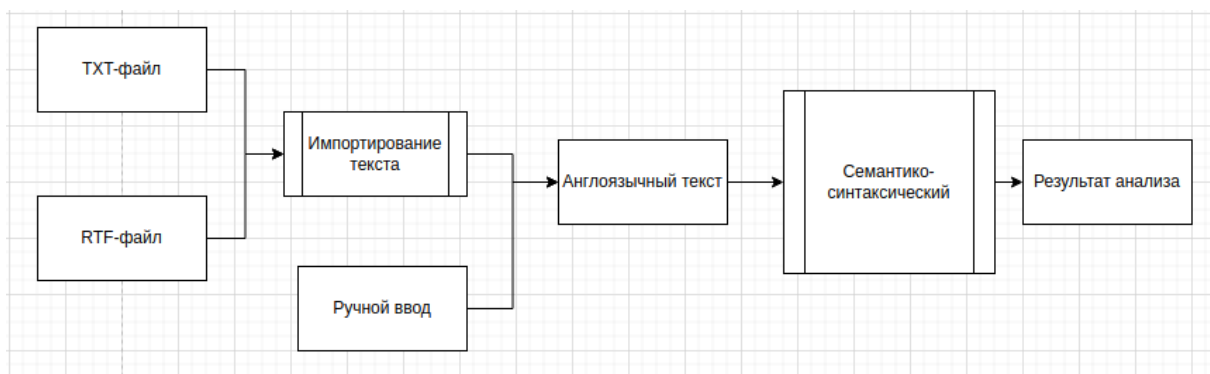


Рис.1 - Схема

## Структуры хранения

После ввода текста, он разбивается на предложения. Предложения хранятся в виде списка. Анализ хранится в виде словаря, ключ это предложение, а значение это словарь. В этом словаря ключами являются "Defenition", "Example", "Synonyms", "Antonyms", "Hyponyms", "Hypernyms".

## Алгоритм обработки

1. Вход: текст на естественном языке.
2. Вызов функции **semantic\_anallysis(text)**, где text - текст на естественном языке.
3. Разбиение текста на слова по пробелам.
4. Берем **i(i=0)** слово, для **i** слова достаём информацию из модуля **wordnet** информацию о нём.
5. Достаём информацию о определении **i** слова.
6. Достаём информацию о пример **i** слова.
7. Достаём информацию о синонимах **i** слова.
8. Достаём информацию о антонимах **i** слова.
9. Достаём информацию о гипонимах **i** слова.
10. Достаём информацию о гиперонимах **i** слова.
11. Если слов в предложении нет, переходим к шагу 13.
12. Достаём следующие слово увеличиваем **i** на 1 и переходим к шагу 4
13. Открываем окно с результатом.
14. Выход: словарь с результатом.

## Алгоритм фильтрации и поиска

1. Вход: данные - словарь, который получен после обработки текста, запрос - в случае фильтрации по слову или части речи - строка.
2. Фильтрация предложений по слову, результат список предложений.
3. Возвращаем пользователю

4. Выход: список предложений.

## Пример работы приложения

1. При запуске приложения нас встречает интерфейс:

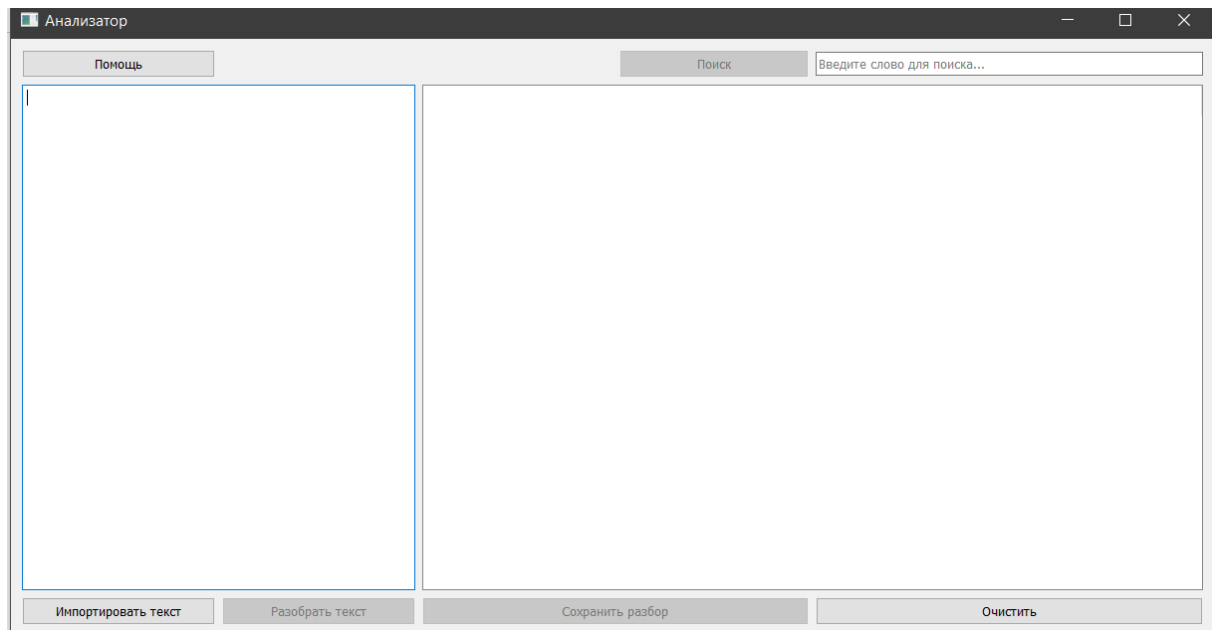


Рис.2 - Интерфейс

2. В него можно написать текст или импортировать текст из TXT- или RTF-файла, нажав на кнопку «Импортировать файл». Пример текста:

Have you ever thought about what your future life is going to be like? What are you going to do when you finish school? It is never too early or late to start thinking about your future career. Maybe you enjoy some of the subjects at school more than others. If you do, this is a good sign, because they will guide you to your future profession.

Рис.3 - Пример текста

3. Пример разбора данного текста после нажатия кнопки «Разобрать текст»:

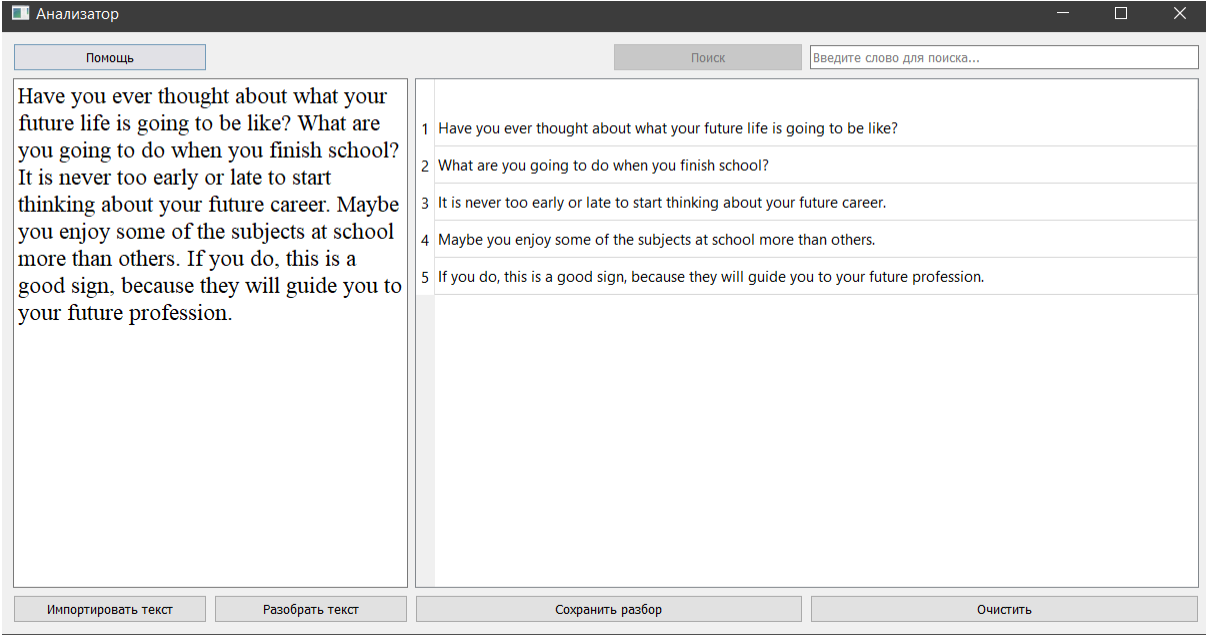


Рис.4 - Пример разбора

4. При двойном нажатии на ячейку таблицы, выводится результат семантического анализа.

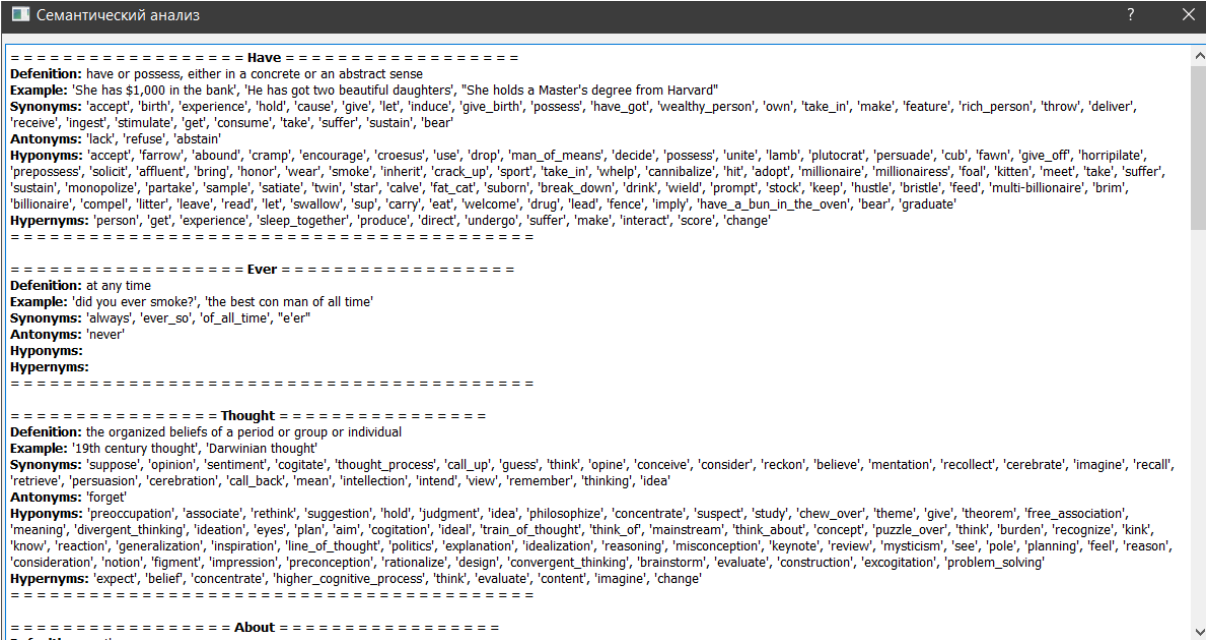


Рис. 6 - Пример анализа

**Вывод:** В результате выполнения лабораторной работы была разработана программа для семантико-синтаксического анализа текста на естественном языке. В будущем эта программа может применяться для разбора крупных текстов как в целях машинного обучения, так и в целях использования в учреждениях образования и у лингвистов.

Вопросы:

Как происходит Семантико-синтаксический анализ:

Токенизация - Текст разбивается на отдельные слова или токены. Это помогает выделить основные элементы для дальнейшего анализа.

Лемматизация - Слова приводятся к своей базовой форме (лемме).

Морфологический анализ - определение вероятных лемм для каждой словоформы.

Постморфологический анализ - устранение неоднозначности слов.

Синтаксический анализ - определение границ именных и числовых групп и вводных конструкций..

Семантический анализ - Анализируются значения слов и фраз в контексте. Это включает в себя определение синонимов, антонимов, многозначности слов и выявление смысловых связей между различными частями текста.

Машинный перевод: Автоматическое преобразование текста из одного языка на другой.

Разметка частей речи: Определение частей речи (существительные, глаголы, прилагательные и т. д.) в тексте.

Извлечение информации: Автоматическое извлечение структурированных данных из текста (например, извлечение имен, дат, сумм и т. д. из финансовых отчетов).

Анализ настроений: Определение тональности текста (положительная, отрицательная, нейтральная) с целью оценки эмоциональной окраски.

Распознавание именованных сущностей: Выделение именованных сущностей, таких как имена людей, местоположения, организации и другие важные элементы.

Генерация текста: Создание текста на основе шаблонов, обучения на больших объемах данных или использования моделей генерации текста, таких как GPT.

Классификация текста: Определение принадлежности текста к определенным категориям или классам (например, спам/не спам, категория новостной статьи и т. д.)

Анализ зависимостей: Построение графов зависимостей между словами в предложении для понимания их логической структуры и взаимосвязей.