Учреждение образования

«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ»

Кафедра интеллектуальных информационных технологий

**Отчет по лабораторной работе №3**

**по курсу**

**«Естественно-языковой интерфейс интеллектуальных систем»**

# на тему: «Семантико-синтаксический анализ текстов естественного языка»

Выполнили студенты группы 921701: Соловьёв А.М.

Шило М.Ю.

Проверил ассистент кафедры ИИТ: Крапивин Ю.Б.

**Минск, 2022**

**Цель работы:**

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

**Задачи лабораторной работы:**

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.

**Методические указания:**

Требуется спроектировать и реализовать программно 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 семантико-синтаксического анализа предложений входного текста согласно варианта задания;
* взаимодействие с пользователем посредствам графического интерфейса (интерфейс должен быть интуитивно-понятным и дружественным 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).

Рекомендуется использовать функциональность стандартной, а также специализированных библиотек языка программирования Python для обработки естественного языка, например, nltk, WordNet, EuroWordnet, ConceptNet, FrameNet.

Вариант задания выбирается студентом самостоятельно и согласовывается с преподавателем. Средства разработки выбираются студентом самостоятельно. Защита лабораторной работы предполагает демонстрацию работоспособности всех реализованных функций в соответствии с требованиями.

**Вариант задания: 8**

Язык текста - английский, формат входного документа - RTF

**Ход работы**

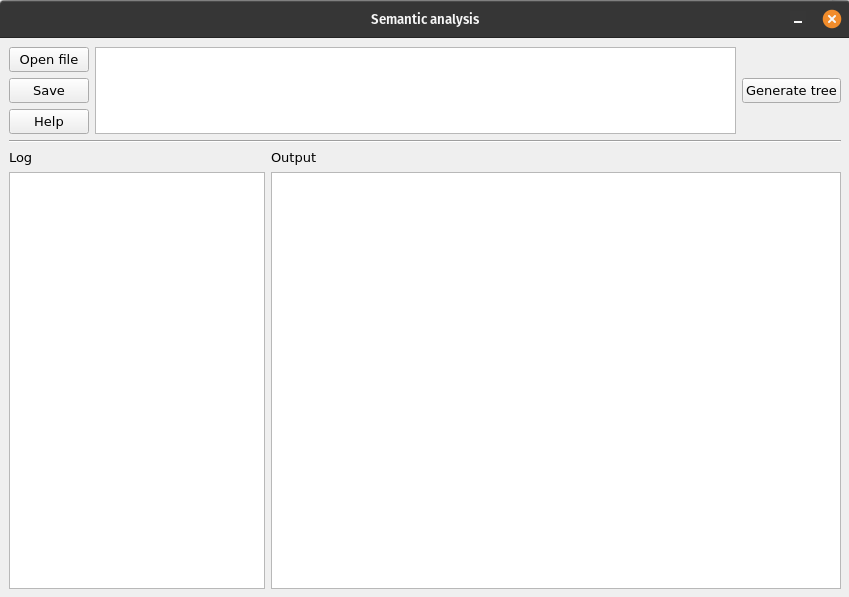
Используемая библиотека **-** [**nltk**](https://www.nltk.org/)

**Наличие системы средств помощи пользователю** реализуется в виде

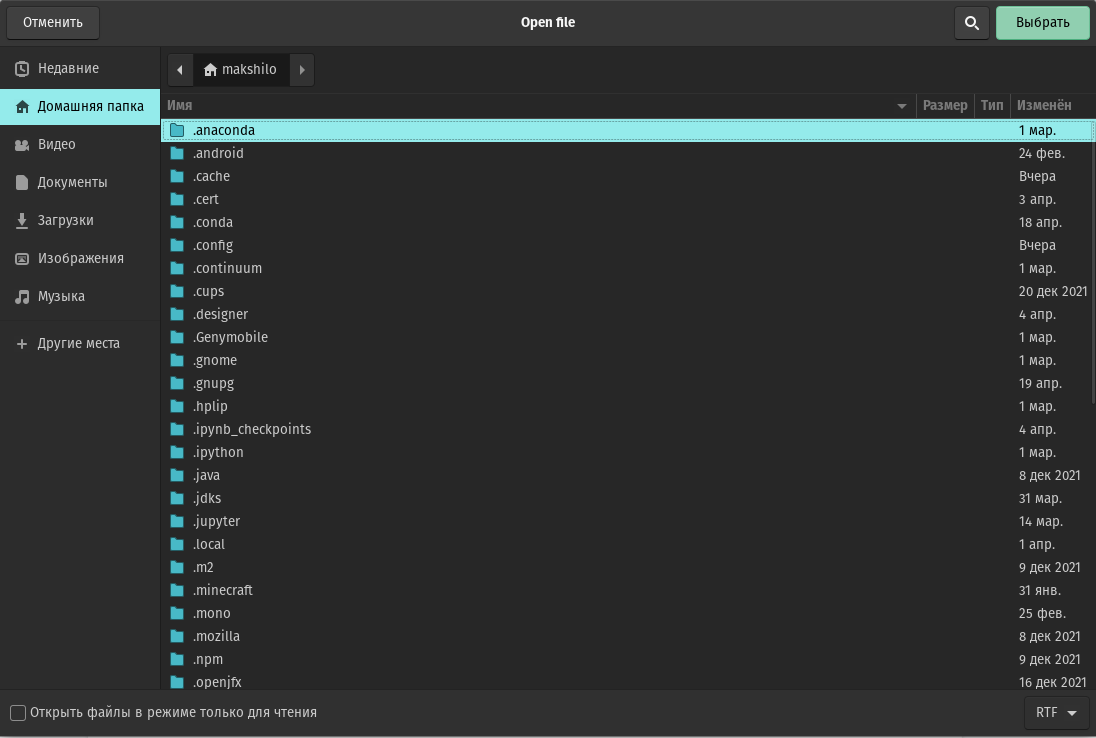
диалогового окна при нажатии кнопки *help* в углу основного окна интерфейса. Сами компоненты при этом подписаны и интуитивно понятны.

**Интерфейс программы**

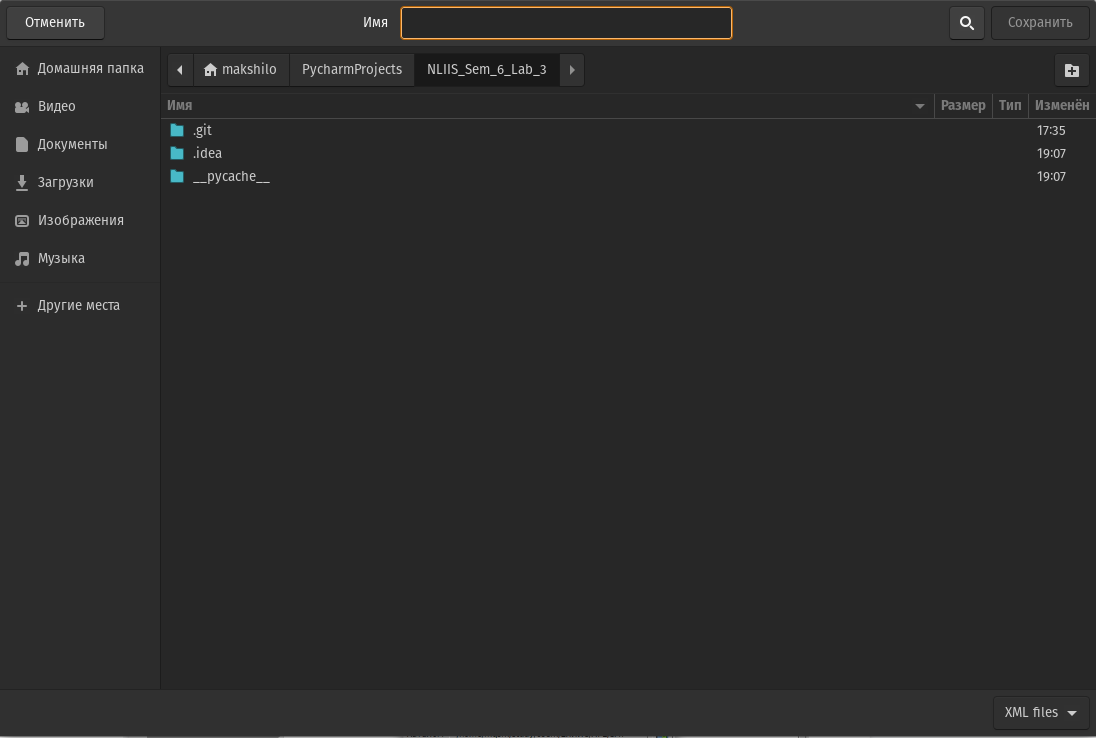
Главное окно выглядит следующим образом



При нажатии на кнопку открытия файла, открывается диалоговое окно с выбором файла в формате rtf. Для удобства текст можно изменять прямо в поле ввода.



Сохранение в xml документ по аналогии



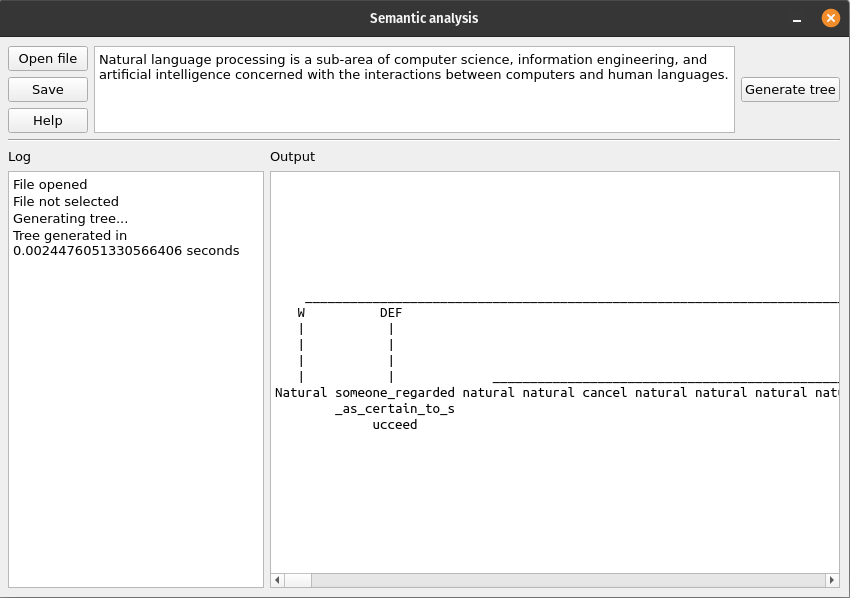
На панели слева ведется логирование действий.

**Примеры входных и выходных данных**

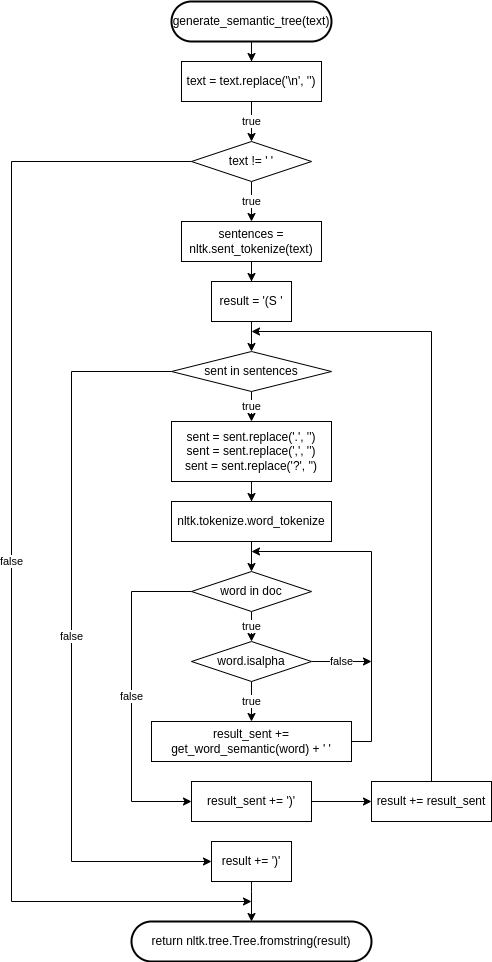
Вход:

Файл в формате rtf с текстом

Natural language processing is a sub-area of computer science, information engineering, and artificial intelligence concerned with the interactions between computers and human languages.

Выход

Для обработки и хранения данных используются встроенные в библиотеку структуры данных, а также встроенные средства языка python.

Блок-схема основной функции генерации семантического дерева:

Оценка быстродействия: Дерево для среднего по размерам предложения генерируется до двух секунд.

**Работа с графическим интерфейсом**

Для создания графического пользовательского интерфейса использовалась библиотека [PyQt5](https://pypi.org/project/PyQt5/). Основными компонентами интерфейса являются текстовое окно, отображающее поля ввода, вывода и логирования, панель с кнопками для управления и индивидуальное диалоговое окно. При проектировании интерфейса возникла сложность с выводом семантического дерева в поле вывода ввиду отсутствия должным образом описанной документации по поводу параметров поля вывода.

**Перспективы использования приложения**

Обработка естественного языка - одно из центральных направлений искусственного интеллекта. Распознавание естественного языка и синтез речи как никогда актуальны в информационном обществе. Однако даже при наличии унифицированного языка представления знаний довольно сложно транслировать его структуры на естественный язык. Семантический анализатор текстов ЕЯ может помочь подойти к решению этой задачи. Тем не менее его недостаточно для синтезирования естественного текста, нужны ещё правила согласования слов в словосочетаниях и словосочетаний в предложениях, при этом не теряя смысл. Семантический анализатор текстов ЕЯ позволяет получать смысловую информацию слова.

**Вывод**

В ходе работы были закреплены навыки работы со стандартными структурами данных и алгоритмами на python, работа с библиотекой синтаксического анализа на примере английского языка, работа с документами, а также приобретены навыки создания пользовательского интерфейса на примере библиотеки PyQt5. Изучен на практике автоматический семантический анализ текста естественного языка.