

Лабораторная работа №2. Синтаксический анализ текстов естественного языка

Цель работы:

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

Задачи лабораторной работы:

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.

Методические указания:

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).

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

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.

Требования к отчету:

В отчете представить, в том числе графически, используя такие программные средства, как Microsoft Visio или Draw.io:

- структурно-функциональную схему разработанного приложения;
- описание структур хранения данных, алгоритмов их обработки, необходимых для реализации базовых требований к разработанной программе;
- оценку быстродействия приложения;
- выводы по работе и по перспективам использования приложения.

Отчет предоставить для проверки в электронном виде.

Варианты заданий:

№	Язык текста	Формат входного документа
1	Русский	TXT,
2	Русский	RTF
3	Русский	PDF
4	Русский	DOC,
5	Русский	DOCX
6	Английский	HTML
7	Английский	TXT,
8	Английский	RTF
9	Английский	PDF
10	Английский	DOC,
11	Английский	DOCX
12	Английский	HTML