

**ФУНКЦИОНАЛЬНОСТЬ ЛИНГВИСТИЧЕСКОГО ПРОЦЕССОРА
В ЗАДАЧАХ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ТЕКСТА
НА ПРИМЕРЕ CORENLP**

Обработка естественного языка (NLP – Natural language processing) – это общее направление технологий искусственного интеллекта и компьютерной лингвистики, которое изучает проблемы компьютерного анализа и синтеза текстов на естественных языках (ЕЯ).

Используя инструменты NLP, существует возможность находить быстрые и качественные решения для таких прикладных задач, как машинный перевод, информационный поиск, автоматическое аннотирование и реферирование, распознавание и синтез речи и т.д.

Лингвистический процессор (ЛП) – это базовая компонента любой современной системы автоматической обработки ЕЯ. Лингвистический процессор выполняет лингвистическую обработку текста в следующей последовательности:

1. Токенизация: текст разбивается на предложения, в каждом предложении выделяются токены (слова, знаки препинания и другие элементы текста – числа, формулы и пр).

2. Морфологический анализ: для каждой выделенной в процессе токенизации словоформы определяются грамматические признаки и лексико-грамматический класс.

3. Синтаксический анализ: для каждого предложения текста строится синтаксическая структура в виде дерева.

4. Семантический анализ заключается в выделении семантических отношений и формировании семантического представления предложения.

Для поддержки задач автоматической обработки текста ЕЯ существует множество доступных инструментальных средств. Мы сосредоточим внимание на использовании языка программирования Python и библиотеки Stanford CoreNLP. Stanford CoreNLP предлагает следующий набор модулей для решения ряда задач автоматической обработки текста:

- частеречная разметка;
- распознавание именованных сущностей;
- синтаксический анализ на основе грамматики зависимостей;
- разрешение кореферентности.

На основании вышеизложенного можно сделать вывод, что библиотека Stanford Core NLP – это универсальный инструмент для работы с ЕЯ. Широкий функционал CoreNLP позволяет просто и эффективно решать прикладные задачи в рамках научного направления обработка естественного языка которые ставит перед нами Natural language processing.