



## Модуль 2. Обработка естественного языка

### Тема 5. Введение в обработку естественных языковых текстов

Лингвистика как наука о языке. Представление об уровнях представления языка – фонетика, морфология, синтаксис, семантика. Лингвистика и прагматика. Лингвистическое моделирование. Действующие модели языка. Теория «Смысл – Текст» как фундамент для построения систем автоматической обработки текста.

### Тема 6. Методы обработки естественных языков

Анализ и синтез текста. Морфологический и синтаксический анализ. Парсинг. Различные подходы к синтаксическому анализу: анализ «сверху вниз» и «снизу вверх». Языковая неоднозначность как принципиальное свойство языка и методы ее разрешения при автоматической обработке текста. Интерактивное разрешение лексической и синтаксической неоднозначности. Правилые и статистические подходы к автоматической обработке текста. Алгоритм синтаксического анализа. Синтаксические отношения. Синтагмы. Синтаксическая структура предложения.

### Тема 7. Вопросно-ответные системы

Вопросно-ответные системы: основы вопросно-ответной системы, архитектуры вопросно-ответной системы, установление смысла вопроса и порождение ответов. Распознавание имён людей, географических названий и других сущностей, различные подходы к распознаванию именованных сущностей

### Тема 8. Программирование и проектирование систем обработки естественных языков

Задачи морфологического анализа, морфологический разбор, стемминг, лемматизация. Понятия лексемы, словоформы, леммы, морфемы, псевдо-основы и псевдо-окончания. Грамматические категории. Словоизменительная парадигма. Морфотактика. Структура данных морфологического словаря, лексикона. Грамматические модели русского языка в контексте автоматической обработки. Минимальное расстояние редактирования. Алгоритм подсчета расстояния Левенштейна. Практика по подсчету минимального расстояния Левенштейна. Понятие статистической языковой модели. Области применения. N-граммы.

## **5. Дополнительная полезная информация**

Дисциплина предназначена для формирования элементов следующих компетенций образовательной программы:

ПК-1. Способен адаптировать и применять методы и алгоритмы машинного обучения для решения прикладных задач в различных предметных областях.

Форма промежуточной аттестации: дифференцированный зачёт.

Наименование оценочного средства: практические работы № 1-6 (собеседование по результатам выполнения практических работ); индивидуальное проектное задание; контрольные работы.