

1- Практическое занятие: Введение в алгоритмы NLP

Цели занятия:

- Понять основные понятия и терминологию NLP.
- Изучить эволюцию алгоритмов обработки естественного языка.
- Применить базовые алгоритмы для решения практических задач.

Задания:

Задание 1: Основные понятия NLP

- **Вопрос:** Определите и кратко опишите следующие термины:
 1. Токенизация
 2. Стемминг
 3. Лемматизация
 4. Нормализация
 5. Часть речи (POS tagging)

Задание 2: Эволюция алгоритмов NLP

- **Вопрос:** Расположите в хронологическом порядке следующие подходы к NLP и кратко описывайте каждый из них:
 1. Правила на основе шаблонов
 2. Статистические модели (например, n-граммы)
 3. Машинное обучение (например, наивный байесовский классификатор)
 4. Глубокое обучение (например, сети LSTM)
 5. Трансформеры (например, BERT, GPT)

Задание 3: Применение базовых алгоритмов

- **Задача:** Используя Python и библиотеку nltk, напишите код для выполнения следующих задач:
 1. Токенизация текста на предложения и слова.
 2. Применение стемминга и лемматизации на примере введенного текста.
 3. Выполнение POS tagging на текстовом наборе.

Задание 4: Исследование современных алгоритмов

- **Вопрос:** Выберите одну современную модель NLP (например, BERT или GPT) и подготовьте краткую презентацию о ее архитектуре, области применения и преимуществах по сравнению с предыдущими подходами.

Задание 5: Обсуждение

- Проведите обсуждение в группе о том, как изменения в алгоритмах NLP повлияли на различные области, такие как машинный перевод, анализ чувств и чат-боты.
- Запишите несколько примеров успешного применения NLP в реальном мире.