Автоматическая обработка текста

26.12.2022. Начало в 15-30 (для онлайн-студентов), 16-20 (для оффлайн-студентов).

Устный экзамен, 2 вопроса, 10 минут на подготовку.

Темы:

1. **Предобработка текста. Токенизация, лемматизация, стемминг, стоп-слова.**
2. **Векторное представление текста. Мешок слов, TF-IDF. Сходство текстов в векторном пространстве.**
3. **Эмбеддинги. Символьные эмбеддинги, эмбеддинги слов, предложений, документов. Контекстуализированные и неконтекстуализированные эмбеддинги.**
4. **Языковая модель: определение, оценка качества. Классические языковые модели: примеры.**
5. **Языковая модель: определение, оценка качества. Нейросетевые языковые модели: примеры.**
6. **Моделирование последовательностей: типы, примеры задач, примеры подходов.**
7. **Классификация текстов. Постановка задачи, примеры подходов.**
8. **Seq2seq модели. Постановка и задачи, примеры подходов.**
9. **Рекуррентные нейронные сети: RNN, GRU, LSTM. Основные принципы, примеры применения для анализа текстов.**
10. **Сверточные нейронные сети. Основные принципы, примеры применения для анализа текстов.**
11. **Архитектура «трансформер». Основные принципы, примеры применения для анализа текстов.**
12. **BERT и модели семейства BERT. Основные принципы. Обучение и дообучение с помощью BERT.**
13. **Question Answering. Типы задач, наборы данных, примеры подходов.**
14. **Суммаризация текстов. Экстрактивная и абстрактивная суммаризация. Наборы данных, примеры подходов.**
15. **Симплификация текстов. Наборы данных, примеры подходов.**
16. **Лингвистические семантические ресурсы. Определение лингвистического ресурса, примеры ресурсов.**
17. **Синтаксический анализ текста. Грамматики зависимостей и составляющих. Трибанки. Основные подходы к синтаксическому парсингу. Оценка качества парсинга.**
18. **Дискурсивный анализ текста. Теория риторических структур (RST). Дискурсивное дерево и дискурсивный граф. Примеры применения.**