

1. Основные задачи NLP.
2. Представления текстовых данных.
3. Предобработка текста, лемматизация, стемминг.
4. Методы машинного обучения для классификации текстовых документов на основе частотных мер (TF-IDF).
5. Деревья решений
6. наивный байесовский классификатор
7. логистическая регрессия в задаче классификации текстов.
8. Языковые модели. Word embeddings.
9. Нейросетевые модели языка: word2vec, fasttext. Мера семантической близости.
10. Классификация текстов на основе нейросетевых моделей языка.
11. Кластеризация текстовых документов.
12. Тематическое моделирование Методы LSA, pLSA.
13. Аддитивная регуляризация тематических моделей в BigARTM
14. Классификация текстов с помощью глубоких нейронных сетей: CNN, LSTM.
15. Задачи обработки последовательностей: машинный перевод, автоматическое реферирование (summarization), вопросно-ответные системы.
16. Механизм внимания (attention).
17. Архитектуры encoder-decoder-attention.
18. Transfer learning в задачах анализа текстов. Self-Attention.
19. Архитектуры трансформеров: BERT, GPT в задачах классификации текстов, предсказания пропущенных слов, генерации текстов. Fine-tuning трансформеров.
20. Построение диалоговых систем. Архитектура диалоговых систем.
21. Модули понимания естественного языка (NLU) и диалоговый менеджер (DM).
22. Сложности построения диалоговых систем.
23. Проектирование UX/UI диалоговых ассистентов в чатах и голосе.
24. Обзор современных фреймворков для построения диалоговых систем: DeepPavlov, Rasa, Just AI Conversational Platform