

- 1) Распознавание именованных сущностей. Области применения распознавания именованных сущностей. Машинное обучение с учителем. Логистическая регрессия. Метод опорных векторов. Оценка качества распознавания именованных сущностей.
- 2) Распознавание именованных сущностей. Машинное обучение с учителем. Нейронные сети прямого распространения. Алгоритм обратного распространения ошибки.
- 3) Распознавание именованных сущностей. Разметка последовательности. Условные случайные поля. Алгоритм Витерби.
- 4) Распознавание именованных сущностей. Разметка последовательности. Рекуррентные нейронные сети.
- 5) Векторные представления слов. Дистрибутивная гипотеза. Локальные модели векторов слов: модели skip-gram и continuous bag of words.
- 6) Векторные представления слов. Дистрибутивная гипотеза. Вычислительная сложность softmax. Иерархический softmax и negative sampling.
- 7) Векторные представления слов. Дистрибутивная гипотеза. Матрица совместной встречаемости слов. Глобальные модели векторов слов: модель GloVe.
- 8) Векторные представления слов. Проблема редких слов в моделях векторных представлений слов. Мера близости слов. Способы оценки качества векторов слов.
- 9) Необходимость применения символьного состава слов в методах обработки текстов. Мешок символьных последовательностей. Парное кодирование байтов. Модель fasttext.
- 10) Символьные представления слов на базе рекуррентных нейронных сетей. Символьные представления слов на базе сверточных нейронных сетей. «Расширяющиеся» сверточные нейронные сети.
- 11) Сегментация текста: задачи токенизации и определения границ предложений.
- 12) Задача определения языка текста. Профили языка. Оценка качества классификации. Проблема коротких и мультязычных текстов.
- 13) Задача определения языка текста. Naïve Bayes классификатор. Оценка качества классификации. Проблема коротких и мультязычных текстов.
- 14) Задача определения частей речи и морфологического анализа. Разметка последовательности. Multi-label классификация.
- 15) Задача нормализации слов. Ripple Down Rules и LemmaGen. Грамматическая омонимия и способы борьбы с ней.
- 16) Задача нормализации слов. Нормализация слов как задача классификации. Грамматическая омонимия и способы борьбы с ней.
- 17) Синтаксический анализ. Дерево составляющих. Понятие формальной грамматики, иерархия Хомского. Генерация текста по формальной грамматике.
- 18) Синтаксический анализ. Дерево составляющих. Разбор предложения по грамматике составляющих. Алгоритм Кока-Янгера-Касами. Неоднозначность разбора, выбор лучшего разбора.
- 19) Синтаксический анализ. Дерево зависимостей. Разбор предложения на основе переходов: Arc-standard, Arc-eager.

- 20) Синтаксический анализ. Дерево зависимостей. Проективность деревьев разбора. Методы разбора для непроективных деревьев: Attardi's system, Online geordering.
- 21) Синтаксический анализ. Дерево зависимостей. Stack LSTM. Синтаксический разбор на основе Stack LSTM.
- 22) Синтаксический анализ. Дерево составляющих и дерево зависимостей. Оценка качества построенных деревьев разбора.
- 23) Задача машинного перевода. Модель зашумленного канала. Оценка качества машинного перевода.
- 24) Статистический машинный перевод. Языковая модель.
- 25) Статистический машинный перевод. Модель перевода. Выравнивания. IBM Model 1.
- 26) Статистический машинный перевод. Декодирование.
- 27) Нейросетевой машинный перевод. Архитектура кодировщик-декодировщик. Методы кодирования.
- 28) Нейросетевой машинный перевод. Архитектура кодировщик-декодировщик. Методы декодирования.
- 29) Кореферентность и анафоричность. Постановка задачи разрешения кореферентности. Классификация подходов.
- 30) Разрешение кореферентности. Модель "пара упоминаний". Отбор пар упоминаний. Классификатор пар упоминаний.
- 31) Разрешение кореферентности. Модель "пара упоминаний". Объединение пар упоминаний в кореферентные цепочки.
- 32) Разрешение кореферентности. Модель "сущность - упоминание".
- 33) Разрешение кореферентности. Модели на основе ранжирования. Модель турнира. RankSVM.
- 34) Разрешение кореферентности. Поиск упоминаний в тексте. Оценка качества.
- 35) Разрешение кореферентности. Оценка качества.
- 36) Варианты постановки задачи извлечения мнений: анализ тональности текста, извлечение мнений. Методы оценки качества.
- 37) Задача автоматического реферирования. Варианты постановок задач.
- 38) Задача автоматического реферирования. Статистические методы решения задачи. Графовые методы решения задачи.
- 39) Задача автоматического реферирования. Оценка качества построенных рефератов. Ручные, автоматизированные и автоматические методы оценки качества.
- 40) Диалоговые системы. Диалоги с определенной целью. Общая схема решения.
- 41) Диалоговые системы. Диалоги без цели. Идеи подходов к решению задачи.
- 42) Варианты постановок задачи извлечения отношений. Области применения задачи.
- 43) Базы знаний. Примеры. Постановка задачи привязки к базе знаний.