

Раздел "Компьютерная лингвистика"

1 Задание

- Основные этапы развития компьютерной лингвистики:
Первый этап - 1954-1966 год (Джорджтаунский эксперимент: перевод предложений с русского языка на английский и разработка систем перевода текста).

Второй этап - 1966-1969 гг (Формирование компьютерной лингвистики как науки).

Третий этап - 1970 - конец 80-ых годов (Распространение и популяризация компьютерной лингвистики в связи с увеличением числа компьютеров)

Четвертый этап - середина нулевых годов XXI века (Проникновение компьютерной лингвистики во все сферы, которые ранее не были связаны напрямую с обработкой текста и речи)

- Возможные пути развития компьютерной лингвистики: Дальнейшее развитие языковых моделей искусственного интеллекта, мультимодальная языковая обработка, межкузыковая и многоязычная языковая обработка.

2 Задание.

Майкрософт: В видео можно наблюдать сбой программы обработки речи, прямо в самом начале, когда система не смогла разобрать слова спикера и перевела "back to the topic" как "back to shoot". Далее неоднократно в видео происходят такие казусы, как на 0:56, а программа не смогла обработать слово "words" и вывела на экран пробелы.

Это связано с двумя вещами: во-первых на тот момент технологии распознавания разговорной речи были лишь в зачаточном состоянии и имели большую погрешность. А во-вторых - из-за того, что спикер немного проглатывает окончания система не могла корректно определить слово.

Шотландцы в лифте: система лифта не смогла правильно проанализировать речь героев из-за трудно понимаемого диалекта

шотландцев, который не всегда могут понять даже живущие рядом с ними англоговорящие нации.

4 Задание.

5 Задание:

Что такое квантитативная лингвистика?

Квантитативная лингвистика – это направление компьютерной лингвистики (которая в свою очередь является одним из разделов общей лингвистики), в рамках которого изучение языка основывается на применении статистических методов и теории вероятности.

Чем занимается компьютерная лингвистика?

- Создание инструментов для машинного перевода и автоматической генерации текста на иностранном языке.
- Делает информацию доступной для людей, которые не могут читать или писать.
- Обрабатывает большие объемы текстов, которые появляются в Интернете каждый день. Определяет важные темы и извлекать из них полезную информацию.
- Разрабатывает системы искусственного интеллекта, способные взаимодействовать с людьми.
- Развивает новые технологии, такие как голосовые помощники и системы автоматической классификации и обработки текста.

В чем отличие компьютерной лексикографии от традиционной?

В отличие от традиционной, машинная (компьютерная) лексикография – относительно молодая наука, реализующая традиционные наработки в технических средах и создающая разнообразные электронные словари. К основным задачам компьютерной лексикографии относятся также задачи разработки технологий составления электронных словарей и управления терминологией (Terminology Management).

В чем заключаются современные методы анализа текста?

Распознавание речи. Компьютер может переводить голосовую речь в текст. Это требуется для любого приложения, которое выполняет голосовые команды или общается с человеком в чате. Например, так работают умные колонки с голосовым помощником Алисой.

Генерация естественного языка. Перевод структурированных, то есть табличных, данных, в текст на естественном языке. Можно сказать, что эта задача противоположна распознаванию речи.

Определение смысла слова. Компьютер может точно определить значение слова после семантического анализа предложения. Например, слово «замок» может иметь разные значения: «механическое устройство для запирания дверей» или «здание с фортификационными сооружениями». Задача NLP — определить, какой смысл имеет это слово в тексте.

Анализ эмоциональной окраски текста. Алгоритмы обработки естественного языка могут получать из текста его субъективные характеристики, например эмоции.

Определение перекрестных ссылок. Во время анализа текста он разбивается на токены — небольшие фрагменты, например отдельные слова. При дальнейшем анализе требуется сохранить и учесть их взаимосвязь.

Распознавание именованных сущностей. В текстах часто встречаются имена собственные: имя человека, название города, валюты и так далее. Задача NLP — правильно их выявить, чтобы корректно использовать при обработке текста и генерации ответа.

Задание 6.



Задание 7.

8 Задание

Звуковые корпуса текста могут быть использованы в различных сферах, включая:

Звуковые корпуса текста используются для изучения звуковых систем языков, их вариаций и изменений. Они также могут быть использованы для обучения студентов фонетике иностранных языков.

Звуковые корпуса текста используются для разработки и тестирования алгоритмов распознавания речи и синтеза речи. Они также могут быть использованы для обучения и тестирования систем автоматического перевода.

Исследования в области искусственного интеллекта: Звуковые корпуса текста могут быть использованы для обучения и тестирования систем искусственного интеллекта, которые могут понимать и генерировать звуковую информацию.

Э