

Модуль I. Лекция 4.
Области применения информационных
технологий в лингвистике



План

1. Автоматический анализ и синтез звучащей речи
2. Автоматическое распознавание текста

Автоматический анализ и синтез звучащей речи

Обзор методов анализа и синтеза звучащей речи с использованием информационных технологий.



Анализ речи

Использование алгоритмов для анализа голосовых сигналов и их преобразование в текст.



Синтез речи

Генерация звучащей речи на основе текстовых данных с использованием синтезаторов.



Автоматическое распознавание текста

Рассмотрим методы и алгоритмы для автоматического распознавания текста с помощью информационных технологий.

Оптическое распознавание символов (OCR)

Технология, позволяющая преобразовывать отсканированный текст в электронный формат для дальнейшей обработки.

Распознавание рукописного текста

Алгоритмы, позволяющие распознавать рукописный текст и преобразовывать его в печатный формат.

Автоматическое извлечение информации

Методы для автоматического извлечения структурированных данных из текстовых источников.

В связи с принципиальными различиями в способах цифровизации и обработки звучащей речи и печатных текстов в нашей работе эти явления рассматриваются в разных параграфах. Первый параграф посвящен вопросам автоматической обработки и синтеза звучащей речи, а во всех последующих рассматриваются автоматические операции, производимые над печатными текстами.

При автоматическом анализе звучащей речи она преобразуется в печатный текст, над которым можно производить дальнейшие операции. Автоматический синтез звучащей речи представляет собой обратный процесс преобразования печатного текста, существующего в цифровой форме, в звучащий текст на естественном человеческом языке.

Процесс автоматического анализа речи включает следующие этапы:

- 1) ввод звучащей речи в компьютер с помощью микрофона.
- 2) выделение компьютерной программой в звуковом потоке отдельных знаков.
- 3) идентификация выделенных знаков звучащей речи со знаками языка.

Современные методы анализа текста и речи

Изучим современные методы и подходы к анализу текста и речи с использованием информационных технологий.

1 Анализ тональности текста

Определение настроения текста на основе его смыслового содержания и лексических характеристик.

2 Распознавание именованных сущностей

Выделение и классификация именованных сущностей в тексте, таких как имена людей, организации, места и т.д.

3 Машинный перевод

Использование алгоритмов и моделей для автоматического перевода текста с одного языка на другой.

Автоматическое распознавание речи

Рассмотрим технологии и методы автоматического распознавания произнесенной речи.



Машинное обучение в лингвистике

Исследуем применение методов машинного обучения в лингвистических исследованиях.

Классификация текстов

Обучение моделей для автоматической классификации текстов по определенным категориям.

Автоматическая аннотация текстов

Методы для создания автоматических аннотаций и категоризации текстовых данных.

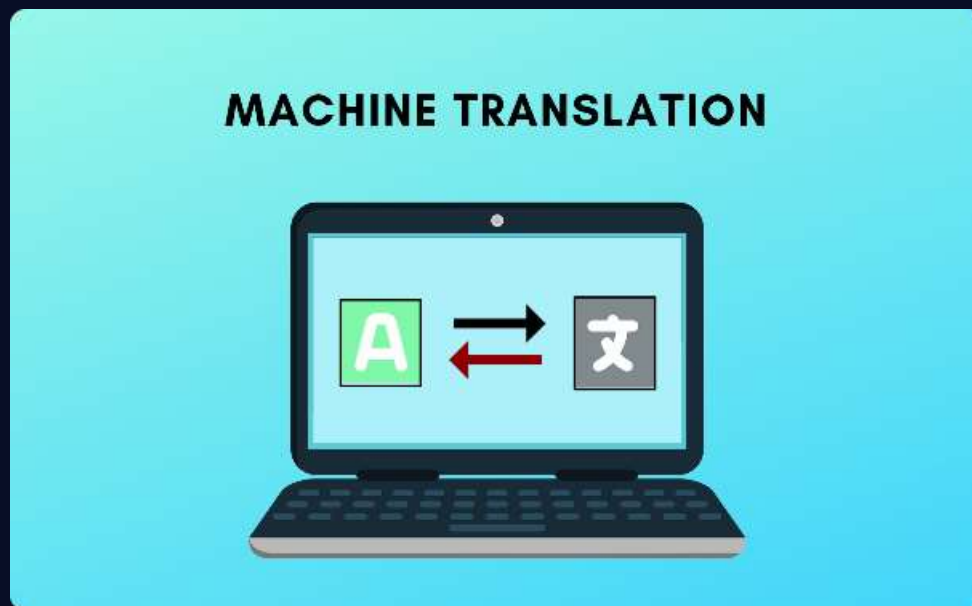
Языковые модели

Разработка и использование моделей для предсказания следующего слова в тексте.



Автоматический перевод и машинный переводчик

Изучим технологии автоматического перевода и разработку машинных переводчиков.




Методы машинного перевода

Обзор различных методов и алгоритмов машинного перевода с использованием информационных технологий.



Разработка машинных переводчиков

Исследование разработки машинных переводчиков для реального времени и их применение в различных сферах.



Компьютерные программы для изучения иностранных языков

Рассмотрим программы и ресурсы для изучения иностранных языков с использованием информационных технологий.

1 Интерактивные упражнения и игры

Использование компьютерных программ для обучения иностранным языкам через интерактивные упражнения и игры.

2 Автоматическая оценка произношения

Программы для оценки и коррекции произношения иногда бывают очень точными и полезными.

3 Электронные словари и глоссарии

Разработка и использование электронных словарей и глоссариев для обучения иностранным языкам.

Вопросы для обсуждения

1. Что такое автоматический анализ речи?
2. Из каких этапов состоит процесс автоматического анализа речи?
3. Как проводится анализ по фонемного распознавания звуков?
4. Что такое спектограмма? Назовите виды спектограмм
5. Когда и в каких случаях применяется автоматический анализ речи?
6. Назовите компьютерные программы используемые для распознавания автоматического анализа речи?

Спасибо

за

ВНИМАНИЕ!!!