## Сжатый текст:

Естественный язык – знаковая система передачи и хранения информации, а также средство создания текста. Знаки, будь то устные или письменные, передают знания. Разные уровни языка (звуки, слова, фразы) выполняют разные функции, формируя структуру языка.

Автоматическая обработка текстов на естественных языках сталкивается с вызовами в выделении значений, тональности и прагматических отношений. Основная задача – выявление намерений, тональности и связей.

Компьютерная лингвистика включает гипергенеративные технологии, моделирование понимания текстов, системы поиска и реферирование текстов.

Развитие включает моделирование понимания смысла, морфологические анализаторы, поисковые системы и системы обработки текстов. Системы распознавания речи упрощают общение и поиск информации, используются в телефонии и для инвалидов.

Алгоритмическая обработка речи включает распознавание, понимание и синтез. Синтез речи охватывает процесс генерации голоса из текста с естественным звучанием и тембром. Различные методы включают директный и аллофонный подходы. Выбор подхода зависит от задачи, типа голосового аппарата и языковой структуры. Технологии компьютеров успешно моделируют и создают устройства для автоматического синтеза речи.

Автоматическая обработка письменного текста включает автоиндексирование и морфологический анализ словоформ. Задачи включают выделение и распознавание слов, а также их нормализацию с использованием автоматических словарей и информационно-поисковых тезаурусов. Синтаксис анализирует структуру предложений и применяется в автоматическом реферировании, включая квазиреферирование.

Позиционные методы генерации краткого содержания требуют учета контекста и выделения тривиальных фрагментов. Квазиреферирование просто и использует синтаксическое дерево. Шаблонные технологии широко используются в промышленности, представляя исходный текст в виде фрагментов действия. Нейросети, корпусная лингвистика и статистические методы являются основными принципами автоматического реферирования.

Электронный архив обеспечивает обработку и интерпретацию данных, а диахронический подход включает различные подпроекты, такие как разметка текстов на разные темы в национальном корпусе русского языка. Основные подходы к разметке включают актуации и абстракции, используя разные формализмы для представления данных.

Модель в компьютерной нейролингвистике (ЛП) описывает статические свойства языка и строится на теориях, разработанных лингвистами на основе текстов и интроспекции. Для обработки текста используется разделение процесса на уровни языка: графематический, морфологический, синтаксический, семантический и прагматический анализ. Эти этапы объединяются при анализе текста. При вводе текста модули ЛП учитывают особенности каждого уровня.

Решение прикладных задач иногда позволяет обойтись без развернутых уровней, например, в ранних экспериментальных программах. Структурно-статистические модели могут унифицировать информацию на разных уровнях, как в случае статистической модели частотности символов. В морфологическом синтезе используются лексема и морфологические характеристики словоформы.

Для моделирования синтаксиса предложены различные методы, включая генеративный подход Хомски, модели смысла и текстов, теорию синтаксических групп и математические языковые модели. В предложенном подходе используется дерево зависимостей для представления синтаксической структуры предложения, где узлы представляют слова, а дуги - синтаксические связи.

Деревья составляющих подходят для описания языков с простыми словородными словами. В рамках теории преобразователя, выполняющего переработку смыслов, предложены ориентация на синтез текста, многоуровневая интегральная модель, специальные средства и четырехпортовый словарь.

Модель Толково-комбинаторного словаря И.А. Мельника является основным инструментом исследования и описания языка, а язык рассматривается как преобразователь смыслов в тексты и наоборот. Модель должна быть действующей, лингвистической или алгоритмической, и осуществимой на вычислительной машине.

Слов: 461