**1. Предмет компьютерной лингвистики. Понятие. Задачи.**

Компьютерная лингвистика – это синтез лингвистических и информационных наук с целью разработки систем, способных обрабатывать естественный язык. Ее область включает в себя разработку алгоритмов для анализа, понимания и генерации текста на естественных языках. Основные направления включают обработку текста (включая сегментацию и анализ), машинный перевод (основанный на статистике, глубоком обучении и правилах), диалоговые системы (включая виртуальных ассистентов) и информационный поиск и извлечение. Целью компьютерной лингвистики является автоматизация задач обработки текста и коммуникации на естественном языке.

**2. Прикладные задачи компьютерной лингвистики.**

Прикладные задачи компьютерной лингвистики охватывают широкий спектр областей. Это включает разработку систем машинного перевода, таких как онлайн-переводчики и системы для автоматического перевода документов. Далее, диалоговые системы, используемые в чат-ботах, виртуальных ассистентах и системах клиентского обслуживания, анализ настроений и сентиментов для определения эмоциональной окраски текста и мониторинга общественного мнения. Информационный поиск и извлечение методов для эффективного поиска информации в текстовых источниках. Это может включать создание поисковых систем, способных анализировать и индексировать большие объемы текста для точного поиска. Распознавание речи также важная область, включая разработку систем для распознавания и транскрибирования речи в текст, а также системы распознавания голосовых команд и диктовки. Наконец, генерация текста, где разрабатываются алгоритмы и модели для автоматической генерации текста на естественном языке, например, новостных статей или кратких описаний. Эти задачи играют ключевую роль в информационных технологиях, медиа, медицине, образовании и других отраслях.

**3. Основные задачи лингвистического обеспечения процессов сбора, накопления, обработки и поиска информации и подходы к их решению.**

Основные задачи лингвистического обеспечения процессов сбора, накопления, обработки и поиска информации включают морфологический анализ, где определяются грамматические характеристики слов для автоматического определения начальных форм и грамматических характеристик. Синтаксический анализ строит структуру предложений, определяя синтаксические связи между словами, как подлежащее, сказуемое и дополнение. Семантический анализ определяет значения слов и выражений, а также распознает семантические отношения, что позволяет понимать смысл текста и выявлять связи между концептами. Просодический анализ анализирует интонацию, акцент и другие аспекты звучания речи, что полезно при распознавании речи и анализе эмоциональной окраски текста. Разметка и аннотирование текстов с помощью тегов или меток описывают различные лингвистические и семантические характеристики текста, что помогает создавать размеченные корпуса для обучения и оценки алгоритмов обработки текста. Подходы к решению этих задач включают использование лингвистических правил, машинное обучение на размеченных данных, статистические методы анализа текста и комбинацию различных подходов для достижения наилучших результатов, с учетом специфики задачи и ее контекста.

**4. Лексика. Лексическое и грамматическое значение слова, многозначность, полисемия, омонимия, синонимия.**

Лексика представляет собой совокупность слов, их форм и выражений в определенном языке. Лексическое значение слова определяет его основной смысл и конкретный образ или концепт, связанный с ним. Например, "солнце" имеет значение "светило, освещающее Землю". Грамматическое значение слова определяет его роль в предложении и его грамматические характеристики, такие как часть речи и временная форма. Например, в "Солнце светит ярко", "светит" - глагол настоящего времени, третьего лица единственного числа. Многозначность слова возникает, когда оно имеет несколько лексических значений, вызванных контекстом или разным употреблением. Полисемия – это многозначность, где различные значения слова связаны общим смысловым ядром, как у "лука" - овоща и оружия. Омонимия – это многозначность, когда разные значения слова не связаны между собой, как "лук" (овощ) и "лук" (оружие). Синонимия – это сходство значений различных слов, таких как "большой" и "великий", с различиями в оттенках и стилистических характеристиках.

**5. Морфология. Морфемы, их классификация, роль в составе слова.**

Морфология – это раздел лингвистики, изучающий структуру слова и его морфемные составляющие. Морфема – это минимальная значащая единица языка, несущая значение или грамматическую функцию. Морфемы классифицируются по типу значения: семантические морфемы несут смысл, например, "дом" в "домик", а грамматические определяют грамматические характеристики, как "-ов" в "домов". По месту в слове морфемы могут быть приставками (например, "под-" в "подход"), суффиксами (как "-ость" в "красота"), корнями (например, "книг-" в "книга") или окончаниями (как "-ам" в "книгам"). По типу функции морфемы делятся на флексии, меняющие грамматические характеристики (например, падежные окончания), и деривации, формирующие новые слова (например, суффикс "-ниц-" в "ученица"). Морфемы играют важную роль в формировании структуры слова, определении его значения и функции в предложении, а также указании на грамматические характеристики и семантические оттенки. Их сочетание образует разнообразные слова и словоформы, обеспечивая выражение широкого спектра концепций и идей в языке.

**6. Формальная модель синтаксиса. Модели синтаксической структуры предложения. 7. Автоматический синтаксический анализ ТЕЯ. Системы составляющих.**

Формальная модель синтаксиса представляет собой абстрактную систему правил, которые определяют правильные комбинации слов и фраз в предложениях в соответствии с грамматикой языка. Основные модели синтаксической структуры предложения включают в себя модель структурной грамматики, описывающую иерархические отношения между словами и фразами с помощью контекстно-свободных грамматик, и модель зависимостей, представляющую слова в графе с зависимостями между ними. Автоматический синтаксический анализ (АСА) включает в себя методы составляющих, которые разбивают предложение на составляющие, и методы зависимостей, которые определяют зависимости между словами. Примеры методов включают алгоритмы CYK и методы, основанные на рекурсивном спуске для составляющих, а также алгоритмы Arc-Standard и Arc-Eager для зависимостей. Каждый метод имеет свои преимущества и недостатки, и выбор зависит от конкретных задач и требований.

**7. Автоматический синтаксический анализ ТЕЯ. Системы составляющих.**

Автоматический синтаксический анализ (АСА) методом ТЕЯ (трансформационно-экстрактивный анализ) использует техники трансформации и экстракции для построения синтаксической структуры предложений, часто представленной в виде синтаксических деревьев или графов зависимостей. Системы составляющих, как подход к АСА, разбивают предложение на составляющие (или конституенты), такие как фразы или подфразы, и затем объединяют их в иерархическую структуру для описания синтаксических отношений. Примеры таких систем включают Stanford Parser, основанный на статистических методах и машинном обучении, Berkeley Parser, известный своей точностью и производительностью, и Google SyntaxNet, использующий нейронные сети для синтаксического анализа. Эти системы предоставляют различные методы и подходы к автоматическому синтаксическому анализу и находят применение в различных областях, включая машинный перевод и обработку естественного языка.

**8. Автоматический синтаксический анализ ТЕЯ. Деревья подчинения.**

Автоматический синтаксический анализ методом ТЕЯ (трансформационно-экстрактивный анализ) использует деревья подчинения для описания синтаксической структуры предложений. Деревья подчинения – это структуры данных, где каждое слово предложения представлено узлом, а связи между словами - ребрами. Например, для предложения "The cat chased the mouse" дерево подчинения показывает, что "chased" является корнем, а "cat" и "mouse" подчинены ему. Деревья подчинения используются для описания структуры предложений и выделения синтаксических отношений между словами. Они являются основным инструментом в автоматическом синтаксическом анализе и используются в различных областях обработки естественного языка, таких как машинный перевод и анализ текста.

**9. Семантика. Виды семантических отношений между словами естественного языка. Формальные модели семантики.**

Семантика, важный раздел лингвистики, изучает значения слов и их взаимосвязи в контексте естественного языка. Различные виды семантических отношений между словами включают синонимию, антонимию, гипонимию и гиперонимию, а также гибридные формы семантики. 1). Синонимия отражает отношение между словами схожего или идентичного значения, как "автомобиль" и "машина". 2). Антонимия выражает противоположные значения, как "большой" и "маленький". 3). Гипонимия и гиперонимия обозначают отношение между словами, где одно слово является частью или подтипом другого, например, "яблоко" к "фрукт". 4). Гибридная семантика объединяет различные семантические отношения, как в слове "автомобильный", где сочетаются синонимия и гипонимия. В формальных моделях семантики используются различные подходы: 1). Модель дистрибутивной семантики анализирует контекст встречаемости слов и определяет их семантические свойства на основе соседних слов в тексте. 2). Модель векторного пространства представляет слова в виде векторов в n-мерном пространстве, где семантически близкие слова имеют близкие векторы. 3). Модель представления знаний использует формальные структуры, такие как онтологии и сети знаний, для описания семантических отношений между словами. Эти модели семантики помогают компьютерным системам понимать и работать с естественным языком, что полезно для задач машинного перевода, анализа текста и ответов на вопросы.

**10. Семантика. Модели представления знаний: семантические сети, фреймы, лексико-семантические комплексы.**

Модели представления знаний играют ключевую роль в семантике естественного языка и являются основой для множества компьютерных приложений в области лингвистики. Вот основные модели: 1). Семантические сети: Эти сети представляют знания в виде графа, где узлы представляют концепты, а ребра - отношения между ними. Эффективны для моделирования лексических отношений, таких как синонимия и антонимия. 2). Фреймы: Фреймы описывают типичные сценарии или ситуации, включая слоты для характеристик и действий. Используются для моделирования знаний о предметах и сценариях. 3). Лексико-семантические комплексы: Эта модель объединяет лексическое значение слова с его семантическим контекстом и окружением. Позволяет учитывать многозначность слова и его семантическое окружение при анализе текста. Эти модели помогают компьютерным системам понимать смысл текста, что важно для систем вопросно-ответной обработки, машинного перевода, анализа текста и других задач обработки естественного языка. Они обеспечивают основу для создания интеллектуальных систем, способных обрабатывать и интерпретировать естественный язык с высокой точностью и эффективностью.

**11. Семантика. Модель представления знаний «Смысл - текст».**

Модель "Смысл - текст" утверждает, что текст несет определенный смысл, формируемый из значений слов, фраз и предложений, их взаимосвязей и контекста. Принципы модели: 1). Связь смысла и текста: Текст передает смысл, который формируется из его элементов. 2). Контекстуальная зависимость: Смысл зависит от контекста, включая социокультурный и языковой контекст. 3). Интерпретация и понимание: Читатель интерпретирует смысл текста на основе своих знаний и контекста. 4). Многоуровневость: Текст имеет разные уровни смысла, включая лексический, синтаксический, семантический и прагматический. 5). Контекстуальная адаптация: Смысл текста может меняться в зависимости от контекста использования и аудитории. Модель "Смысл - текст" объясняет передачу информации в текстах и их восприятие читателем или слушателем. Она важна для разработки систем обработки естественного языка и понимания текстовых данных компьютерами.

**12. Методы обнаружения и исправления ошибок. Задача о редакционном расстоянии и метод ее решения.**

Методы обнаружения и исправления ошибок в тексте включают алгоритмы, направленные на определение и исправление опечаток, грамматических ошибок и стилистических несоответствий. Одним из ключевых методов является задача о редакционном расстоянии и его решение. Задача о редакционном расстоянии заключается в определении минимального числа операций (вставок, удалений, замен символов), необходимых для преобразования одного слова в другое. Это позволяет определить сходство между словами и их значимость. Методы решения задачи о редакционном расстоянии включают динамическое программирование и алгоритмы, такие как алгоритмы Левенштейна и Дамерау-Левенштейна. 1). Алгоритм Левенштейна основан на динамическом программировании и находит минимальное количество редакторских операций для преобразования одной строки в другую. Это эффективный метод, но не всегда дает оптимальное решение. 2). Алгоритм Дамерау-Левенштейна включает операцию транспозиции (перестановки символов), что позволяет учесть опечатки, вызванные перестановкой символов. Он подходит для более точного определения расстояния между словами. Эти методы применяются для исправления опечаток и грамматических ошибок, а также для поиска похожих слов и выявления ошибок ввода пользователя. Они используются в автоматической проверке правописания и других приложениях обработки текста.

**13. Методы обнаружения и исправления ошибок. Задача об операционновзвешенном редакционном расстоянии и метод ее решения. Алфавитновзвешенное редакционное расстояние. Решения строковых задач в терминах взвешенного редакционного графа.**

Методы обнаружения и исправления ошибок в тексте включают использование редакционного расстояния для измерения различий между строками. Операционно-взвешенное и алфавитно-взвешенное редакционные расстояния расширяют классическую задачу о редакционном расстоянии, учитывая стоимость операций и символов. 1). Операционно-взвешенное редакционное расстояние учитывает стоимость каждой редакторской операции (вставка, удаление, замена, транспозиция) в зависимости от их значимости. 2). Алфавитно-взвешенное редакционное расстояние учитывает разные стоимости операций в зависимости от символов, которые они затрагивают. Решение задачи о редакционном расстоянии в терминах взвешенного редакционного графа представляет задачу в виде графа, где вершины - префиксы строк, а ребра - редакторские операции с соответствующей стоимостью. Для нахождения минимального расстояния применяются алгоритмы поиска кратчайшего пути, такие как алгоритм Дейкстры или алгоритм А\*. Эти методы эффективно исправляют опечатки и другие ошибки в тексте, учитывая их разнообразие и контекст. Они применяются в автоматической проверке правописания, машинном переводе и других задачах обработки текста.

**14. Семантический анализ ТЕЯ. Методы и подходы.**

Семантический анализ текстов на естественном языке (ТЕЯ) включает использование компьютерных методов и алгоритмов для извлечения смысла из текстовых данных. Некоторые методы и подходы включают: 1). Методы машинного обучения: Используют алгоритмы, такие как нейронные сети, SVM и случайные леса, для анализа семантики текста, обучая модели на размеченных данных. 2). Алгоритмы обработки естественного языка (NLP): Включают токенизацию, лемматизацию, выделение именованных сущностей и синтаксический анализ для анализа текста. 3). Векторные представления слов и эмбеддинги: Используют методы, такие как Word2Vec и GloVe, для представления слов в векторном пространстве и решения задач, таких как поиск похожих слов или классификация текста. 4). Семантические графы и базы знаний: Построение графов или баз знаний, отражающих семантические отношения в тексте на основе онтологий или баз данных, таких как WordNet или DBpedia. 5). Глубокие нейронные сети и трансформеры: Используют сети, такие как BERT и GPT, для решения сложных задач анализа текста, включая понимание контекста и генерацию текста.

**15. Автоматизация процесса построения размеченных корпусов текстов.**

1). Сбор и парсинг данных: Включает веб-скрапинг и использование API для сбора текстовых данных из различных источников, а также их парсинг для извлечения нужной информации. 2). Автоматическая разметка: Применяются алгоритмы машинного обучения для автоматической разметки текстов, включая части речи, именованные сущности и синтаксические структуры. 3). Семантическая аналитика: Используются алгоритмы анализа текста для извлечения семантической информации, такой как ключевые слова и тематическое моделирование. 4). Интеграция с инструментами аннотирования: Автоматизированный процесс интегрируется с инструментами аннотирования для быстрой и эффективной разметки текстов. ). Обратная связь и корректировка: После автоматической разметки текстов может потребоваться обратная связь и корректировка результатов экспертами или обучение моделей на размеченных данных. 6). Масштабирование: Автоматизированный процесс обеспечивает масштабирование создания корпусов текстов для разметки больших объемов данных.

**16. Технология построения корпусов текстов.**

Технология построения корпусов текстов обычно включает в себя восемь основных этапов. На первом этапе определяются цели и параметры корпуса, включая выбор жанра текстов, их объем, язык, структуру и другие характеристики. Затем следует этап сбора текстовых данных, который может осуществляться различными способами, такими как веб-скрапинг, использование открытых баз данных или запросы к API. После сбора текстов идет этап предобработки, где тексты подвергаются различным операциям, таким как удаление лишних символов, токенизация, удаление стоп-слов и другие. Затем происходит разметка текстов, которая может включать в себя разметку частей речи, именованных сущностей и других лингвистических атрибутов. После этого размеченные тексты сохраняются в базе данных для последующего доступа и анализа. Важным этапом является проверка качества данных в корпусе, включая анализ точности разметки и ее полноты. После построения корпуса необходимо обеспечить доступ к нему для исследователей и разработчиков через специальные интерфейсы или API. Наконец, корпус текстов требует постоянного обновления и поддержки, включая добавление новых текстов, улучшение разметки и расширение или оптимизацию базы данных.

**17. Синтаксический анализ ТЕЯ. Методы и подходы.**

Синтаксический анализ текста на естественном языке (ТЕЯ) представляет собой процесс анализа структуры предложений с целью выявления синтаксических отношений между словами. Методы и подходы к такому анализу могут быть разнообразными и включать: 1). Метод рекурсивного спуска, который базируется на идее рекурсивного спуска по грамматике языка, где грамматика разбивается на набор правил, реализуемых через рекурсивные функции. 2). Метод анализа сверху-вниз, где анализ начинается с начального символа грамматики и движется вниз по дереву разбора до листовых узлов, соответствующих терминалам. 3). Метод анализа снизу-вверх, где анализ начинается с терминалов и постепенно комбинирует их для получения более сложных структур, пока не будет получено начальное правило грамматики. 4). Методы, основанные на стохастических моделях, использующие вероятностные модели для принятия решений о структуре предложения, обучаемые на размеченных корпусах текстов. 4). Глубокое обучение, включающее в себя использование глубоких нейронных сетей, таких как рекуррентные и сверточные нейронные сети, для автоматического извлечения синтаксических структур из текста, обучаемых на больших объемах данных. 5). Комбинированные методы, которые объединяют элементы различных подходов, например, рекурсивный спуск с нейронными сетями, для достижения лучшей производительности или решения специфических задач.

**18. Размеченные корпуса текстов.**

Размеченные корпуса текстов - это коллекции текстов, к которым добавлена разметка с информацией о лингвистических характеристиках, таких как части речи и синтаксическая структура. Их создание может быть ручным, когда аннотаторы добавляют разметку непосредственно на текст, или автоматическим, с применением алгоритмов машинного обучения. Методы включают и полуавтоматическую разметку, где используется комбинация ручной и автоматической разметки. Размеченные корпуса текстов необходимы для обучения и оценки моделей в области обработки естественного языка, таких как различные алгоритмы синтаксического и семантического анализа. Они используются для обучения моделей машинного обучения и оценки их качества. Некоторые известные размеченные корпуса включают Penn Treebank для синтаксического анализа и WordNet для семантического анализа.

**19. Автоматический морфологический анализ ТЕЯ. Методы и подходы.**

Автоматический морфологический анализ текстов в обработке естественного языка (NLP) определяет морфологические характеристики слов, включая часть речи, число и падеж. Различные методы и подходы включают: 1). Статистические методы, основанные на анализе статистических характеристик текста и словоформ, таких как скрытые марковские модели (HMM), методы машинного обучения (например, метод опорных векторов) и нейронные сети. 2). Лингвистические правила, использующие морфологические словари и правила для определения части речи и других характеристик слова. 3). Сложение морфем, анализирующее словоформы как сумму морфем, что позволяет определить их морфологические характеристики. 4). Комбинированные подходы, объединяющие различные методы для улучшения точности и производительности, например, сочетание статистических методов с лингвистическими правилами или использование ансамблей моделей машинного обучения. Каждый метод имеет свои преимущества и недостатки и может применяться в зависимости от требований конкретной задачи.

**20. Понятие корпуса текстов. Основные характеристики и назначение.**

Корпус текстов - это собрание большого количества документов, используемых в лингвистических исследованиях и обработке естественного языка (NLP). Основные характеристики и назначение корпусов включают: 1). Объем и разнообразие: Корпусы содержат тысячи или миллионы документов различных тематик, стилей и жанров, позволяя анализировать языковые явления в различных контекстах. 2). Разметка и аннотация: Некоторые корпуса содержат разметку с информацией о частях речи, синтаксической структуре и семантике, что делает их ценным ресурсом для обучения и оценки алгоритмов в NLP. 3). Исследования и разработка: Корпусы используются для лингвистических исследований и разработки методов и моделей в обработке естественного языка. 4). Обучение и оценка моделей: Они служат основой для обучения и оценки моделей в NLP, таких как модели языка и машинного перевода. 5). Ресурс для приложений: Корпусы также используются для создания и улучшения приложений, таких как системы автоматического перевода и чат-боты. Таким образом, корпусы текстов играют важную роль в исследованиях и приложениях в области NLP, предоставляя доступ к большому объему данных и разнообразию языковых явлений.

**21. Лингвистический процессор. Структура и состав ЛП.**

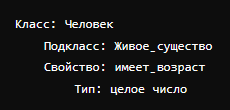
Лингвистический процессор (ЛП) - это программное обеспечение для обработки текста на естественном языке. Он состоит из следующих компонентов: 1). Токенизатор разбивает текст на отдельные токены, например, слова или знаки пунктуации. 2). Морфологический анализ определяет грамматические характеристики каждого слова, такие как часть речи и падеж. 3). Синтаксический анализ анализирует структуру предложений и определяет синтаксические отношения между словами. 4). Семантический анализ анализирует смысловое содержание текста и интерпретирует значения слов в контексте. 5). Дискурсивный анализ исследует связь между предложениями и организацию текста. 6). Прагматический анализ учитывает контекст общения и цели говорящего для понимания текста. 7). Генерация и анализ речи включает в себя создание и понимание речевых данных.Структура ЛП представляет собой пайплайн, где каждый компонент обрабатывает текст последовательно, передавая результаты следующему для дальнейшей обработки. Это обеспечивает гибкость и масштабируемость системы для обработки текста на естественном языке.

**22. WordNet: структура и основные характеристики. Применение в задачах автоматической обработки ТЕЯ.**

WordNet - это лексическая база данных, представляющая собой семантический тезаурус, организованный в виде сети синонимов и связанных понятий. Вот основные характеристики и структура WordNet: 1). Структура WordNet: 1.1). WordNet организован как сеть синонимов, где каждый узел представляет собой лексическое понятие (синонимическую группу) или семантическую категорию. 1.2). Синонимические группы связаны между собой семантическими отношениями, такими как гипонимия (отношение часть-целое), гиперонимия (отношение целое-часть), ассоциация и пр. 1.3). Каждое слово в WordNet связано с одним или несколькими синонимами и может иметь ассоциации с другими словами в базе данных. 2). Характеристики WordNet: 2.1). В WordNet содержится обширная лексическая информация для английского языка, охватывающая существительные, глаголы, прилагательные и наречия. 2.3). Каждый элемент в WordNet имеет уникальный идентификатор, который обеспечивает легкий доступ к семантическим связям и определениям. 2.4). WordNet предоставляет определения и примеры использования для множества слов, что делает его полезным ресурсом для различных приложений в области обработки естественного языка. 3). Применение в задачах автоматической обработки ТЕЯ: 3.1). WordNet широко используется в различных задачах обработки текста, таких как информационный поиск, распознавание именованных сущностей, машинный перевод, анализ тональности и другие. 3.2). В информационном поиске WordNet может быть использован для расширения запросов пользователя синонимами или связанными терминами, улучшая точность и полноту поисковых результатов. 3.3). В задачах анализа тональности WordNet может помочь в определении семантической окраски слов и выражений, помогая выявить положительные, отрицательные или нейтральные отзывы. 3.4). WordNet также может быть использован в семантическом анализе текста для выявления связей между словами и понимания их значения в контексте.

**23. Онтология. Структура и ее формальное описание.**

Онтология - это формализованная модель предметной области, включающая следующие основные элементы структуры и формальное описание: 1). Элементы онтологии: 1.1). Понятия (термины): Описывают объекты, сущности или концепции в предметной области. 1.2). Отношения: Устанавливают связи и взаимодействия между понятиями. 1.3). Свойства: Характеристики объектов или концепций, определяющие их состояние или атрибуты. 1.4). Аксиомы: Логические утверждения, описывающие ограничения и правила в предметной области. 2). Формальное описание: 2.1). Язык описания: Обычно онтологии формализуются с использованием языков, таких как RDF, OWL или языков программирования. 2.2). Описание понятий и их связей: Понятия и отношения могут быть описаны с использованием концептов и свойств. 2.3). Аксиомы и ограничения: Онтологии могут включать логические утверждения, определяющие правила взаимодействия и ограничения на использование понятий. 3). Пример формального описания (на языке OWL):



Онтологии находят широкое применение в искусственном интеллекте, семантической веб-технологии, базах данных и других областях для формализации знаний и обеспечения их семантического понимания компьютерными системами.

**24. Применение онтологий в задачах автоматической обработки ТЕЯ.**

Онтологии важны для обработки текстов и естественного языка, формализуя знания о предметной области. Например: 1). Семантический поиск: Улучшают точность поиска, уточняя запросы и соотнося результаты, что полезно для медицинских запросов. 2). Категоризация текстов: Помогают автоматически классифицировать новости по событиям или темам. 3). Разрешение неоднозначности: Определяют семантические связи между словами, что помогает при переводах. 4). Извлечение информации: Помогают извлекать информацию о заболеваниях и лечении из медицинских текстов. 5). Семантическая аннотация: Создают структурированные и семантически богатые корпуса текстов. 6). Автоматический семантический анализ: Распознают смысл предложений и выявляют ключевые концепции и отношения.

**25. Задача общения пользователей с автоматизированными интеллектуальными информационными системами на естественном языке. Подходы к решению, существующие проблемы реализации.**

Natural Language Understanding (NLU) - ключевая задача в обработке естественного языка, существующие подходы к которой включают: 1). Правила и шаблоны: Используют предопределенные правила и шаблоны для обработки запросов пользователей, что хорошо подходит для систем с ограниченным функционалом. 2). Статистические методы и машинное обучение: Применяют методы машинного обучения для извлечения семантической информации из текста и гибкой обработки запросов. 3). Гибридные подходы: Комбинируют правила и статистические методы для повышения точности и универсальности обработки запросов. Однако существуют проблемы, такие как: 1). Полисемия и омонимия: Сложность определения значений слов в разных контекстах. 2). Сложные запросы: Неоднозначные или грамматически сложные запросы могут затруднить понимание системой. 3). Недостаток данных: Для обучения моделей NLU требуются большие объемы размеченных данных. 4). Проблемы с речью: Распознавание и преобразование речи в текст могут быть сложными. 5). Интерпретация намерений: Точное определение намерений пользователя может быть сложным, особенно для многофункциональных систем. Для решения этих проблем необходимо совершенствовать методы обработки естественного языка и иметь больше данных для обучения.

**26. Понятия информации, данных, знаний, предметной области, автоматизированной информационной системы.**

В контексте компьютерной науки и информационных технологий: 1). Данные - необработанные факты или результаты измерений. 2). Информация - организованные данные, имеющие смысл и пригодные для принятия решений или обмена знаниями. 3). Знания - умения, опыт и понимание, извлеченные из информации и используемые для решения задач. 4). Предметная область - область знаний, объектов и явлений, изучаемая в определенном контексте. 5). Автоматизированная информационная система (АИС) - набор программ и оборудования для сбора, обработки и передачи информации в определенной предметной области.

**27. База знаний (БЗ), источники и методы пополнения БЗ, место БЗ в интеллектуальной системе.**

База знаний (БЗ) - это хранилище информации о предметной области, которое обеспечивает компьютеру доступ к необходимым знаниям для решения задач. Ее обновление и пополнение происходят из разных источников: 1). Экспертные знания от специалистов. 2). Новые данные и результаты исследований. 3). Литература и научные публикации. Опыт использования и обратная связь пользователей. БЗ могут пополняться различными методами: 1). Ручное добавление экспертами. 2). Автоматическое извлечение с использованием NLP. 3). Машинное обучение для анализа данных. Место БЗ в интеллектуальной системе зависит от ее конкретного назначения, но она часто выступает важным компонентом, предоставляющим знания и данные для принятия решений и выполнения функций системы.

**28. Интеллектуальная информационная система (ИИС). Отличительные особенности ИИС.**

Интеллектуальная информационная система (ИИС) представляет собой комплекс программ и оборудования, применяющий методы искусственного интеллекта для сбора, хранения, обработки, анализа и использования информации, обеспечивая автоматизированную поддержку принятия решений. Основные особенности ИИС: 1). Использование методов искусственного интеллекта, таких как машинное обучение и обработка естественного языка. 2). Автоматизация решения сложных задач путем анализа данных и принятия решений на их основе. 3). Обработка неструктурированной информации, включая тексты, изображения и аудиофайлы. 4). Адаптивность и возможность самообучения системы для улучшения производительности и качества решений. 5). Решение сложных и нестандартных задач, для которых традиционные методы могут быть неэффективны. 6). Интеграция с существующими системами и данными для совместного использования и анализа информации из различных источников.