## Весь необработанный текст из конспектов лекций по Естественному языку:

1 лекция

ИлЕстественный язык21.09.20231 я ЛекцияМетоды работы с естественнымязыкомИнформационная технология проблемаавто-обработки текстовой информациипредставленной на естеств. языкахЯзык = знаковая системапередача 1 храненияинформациипередача информациисоздания / и приём 1 текстаОсновной элемент языковой системы-Знак

Знания передана знанийустная или письменная 1()Могут не отличатьсяЧеловек рень 3 импораздельнаяФонеми - состав элементы речиСлоги, морфемы сюва,Слово т.п единица,обозн. понятияЗадачей току. анализа являетсявиделения из текста напленований,тонетий, а также определениясигм. и прагм. отн. между ними

Задача т разработка машинногоперевода, и поиск информации-диатого- вопросные системыЕдиницы разного уровне:люрфемысловасловосочетанияФразысерх дразовке единстваМерархическая П системаСлово 2 сновосочетание иУстойчивыеКомпьютерная лингвистикаавтомати. обработка и информации

Основные направления:Компьютерная лингвистикагипергенстовые технологии предст. текстаИнформау. - поисто вне системыМаши иный переводКомпьютерное под елизования()общение+Вопрос ответ+ Диалоги+ связные текстыРаспозн. речиРасп. сим КоловИзвлеч. данныхАвто -рферирование тексто в

Современные разработки:Проблема: Мод ерирования процессапонимания смысла текстоЛингв. средстваделпаративные процедурныеМорфологии разборСловари унна аслови речиИнформау.- поисковые системы1) Документальные хранен текстыдокументов)2) Файоографические описаниярайтов)3) Смешные

Инф. - юистовые системыИнф.- юисто вой ции -это формальныйязык - описание содерж. документовИнделирования - описание документана инф. поис Сто вом язываПоисковый т образ документатематика документа опис. совокупн.дейсторовКудто снова и терминыАлгоритм. обработка звукащей речи1)Распозн. речи2)понимание речи3) Синте, речи

Системы распозн. речи5 По назнач.по потребительным Качествампо теханиз мат функусионированияПрименение рече технологийУпр. БДРгеуп- технии системдинсковияэлектровейшиработытелефонияСоздания инвалидов

Системы распозн. речи:1) Дкуст. людель2) ритви Ст. люденьподходи построения:2)Изобретатель свий2) БионическийСистема распозн речисемант.синтакс.ленсии.торфолронет.Роколог

Уровни линк модели.1)Фонет. уровень2)Фоколо и и. уровня3) Лексии. уровень4) Семантикатестирре гистр 41)допусаГолосовое управление

2 лекция

я2 Лекция05 10.2023Автоматический синтез речиЭто процесс генерации человеческогоголоса, из текстаПлавное звучание свяВерная интонацияестеств. голос, и тембрречиРасст. ударенийРасши Фровка+ аббревеатур+ чисел+ спец. знаков+ сокращений

от середины одной фонеми,1)Диронный подходдо середины другой2)Аллофонный подход фонема» окружениеОслева и справабересной - выбор 8 визн. элементов3)4 п.из реч. базыСуществуют недостатки подходов.1)Нест. тембр2) Слишком работизированный3) Дойты чёт недостатковНекоторые слова произносетсяс искажения ми (вилоть довыпадение звуков из слова)

Ограничение на выборметода синтезаречи:1)Задача (её тип)2) Голосовой аппарат не ро вкаРечевая волна, качест его,сигнал путь и истои нимувозбуждает действие органов,фильеры времени, Артитуляциониеорганы, гладное сцепление.3) Структу ра ЯзыкаРед возможностей языковойструктуры - объеснение и описаниеязыка, входная рече вае волна, мо-рфологи нес Сте и синтакси ческиеструктуры

4) ТехнологииВозможности успешно моделироватьсоздавать устройства,иРеневая наука техноло гин уифровыекомпьютеры. Мощные, компавание, инедорогие устройства, работающиереальном времени.

Следующая тема:А втоматическая обработкаписьменного текстаАвто- индексирование9 документа рА вто матичевоеРучноеиндексирование 1инденирование ТКачественностьно Минусов нетНизная производит.

Анализ структур словоформ.Задачи морф. анализа:Выделение словоформ из текстаРаспозн. словИх нормализацияАвтоматические словари1)СловариРаспозн. сюворформ, и их нормализ.2) Информационно- поисковыетезаурусыСодержат информацию ототношениях между словарями

3 лекция

Синта Ксинес Сий я анал из текста19.10. г023ЛекцияАвтомат ичестоеРеферированиеЭто составление токспертов№ Направление/ кратное суюи1 документовквазирефорию вание

СтатическиеПозиционныеИндинаториюНе годыПозиционные ме годы2. Негод заглавил) не год комализациии ключе вое предложениесвязь с левы м и правыВыделение трифтов

Квазирефериро ваниеПрислущество - в простоте.Метод синтоксического наборатекстаНетод понимания встественногоязываСуществуют общедосту пиияпрограмми квази реферито вато,например, в состав сер веиеныхвозможностей системыМ.слодовы был в входит режимАвторедерат

Шаблонные технологииОтносительно прости и надёжны.Находят шировое промышленное применение.Главная особенность этих технологийсостоит в том что содержано изеготекста предста влено в Иля видеФрагментов действа51 МР таблицыМетод фокусировки вниманиеПрактика 21.10. 2023Голосо вой полнотник АлисаСтатист. методы3 рефериро ванияНейр. Сети

4 лекция

дляЛекция11. 202302.

Корпуская лингвистика3 принципа.1-авто -извленение сведентов, и не2 -обработва и ИФ.3- про вриди интерпр.электронный архивэл. бит мотемагор пус те всто всут орус+ синхроние Илит+ прниторныйдиагронический ний

по индейсацииЭтапи разметкиНТМ -разметкатхтРазличные подпро виточе, иполу а в о на- ]сохране не

О раз метвеКорпусние ме нед жерыНауиональный торпусРусстого языкаПредставление лиг венсти ческихданныхЛинавист. информ ацияФормализми,ограниченеоктологи

Подходы и предста вре нию данных2. Осно в а иные на разметие- Основанные на актютацияхС.Ош. на абстракции1.Ом на формализмах

5 лекция

16.11.2023Мод ели ро в а ниевКомпьютернойлинг вистикеРазработка лингви стическогопро цессора(ЛП), предполагаето писание ст ическихсвойств, обрабатываемого языкаЕЯ и это описание организуетсянаи модель языка

Используемые в кл поделиязыка, обычно строятся наоснове теорий, создаваемых лингвистами,путем изучение различных текстов,и на основе с воейлинг выстинес интуации(инфроснойции)111 Сложность ЕЯ его описанияи обработ ни, ведёт и разбиениюэтого отдельного процесса наотдельные эта пы, соответствующиеуровнем языкаПри а нал изе текста отдельныелюдули № ви полне ют.

2) Графематиче Лита на из2) Морфологичес анализ3) Сиком селовей анализ4) Семантичес ний, и прагматческийанализтакимобразом, АП можнорасмат ривать наи личогоэта шныйэта и про бразо ванияпри водятьв случае анализа текста каждоеего предложение, во внутренияпредставление его смысла инаоборот,в случае синтезаСоответствующая ходель язываможет называться структура

Хоте полные подели КЛ требуютуче та всех о что в них уровит я, нваи наличе соответ ствующих модутей,при решение шекоторые прикладшихзадач можно обовлась без5 предста влп отдельных уровийк примеру в раннихэксперементаль ны+ программахкл, обрабатываемые текстыотно сились и очень узнимпро блемным о областем.Если одним примером редуциюной модели, нике достато ино частовопрочает ой е вляется язытовая модельчастотности символов и и + состант

Такая статистическая модельотображает лингвистиче скуюинформацию на уровне сим волов(бучев) текста и ее достаточнонапример, для вые влениев тексте, и ли для распознаванияего языковой принадлежностиОтметим, и то возможныструктурно- ота тистическиелюдели, в которых, при представленииотдель ни + уровней ЕЕ унитываетсята или иная статисти на - словиту.

При морфологическом синтезе,и сход ними данными являются:левсема, и конкретные мордологическиехарактеристики запрашиваемой сковоформывДля моделирования синтаксисав раммах предложено было большоешисю разных идей, и методов,от личающихся способом они сашяа синтаксиса языка.Подходы и созданию моделей.1) Генеративный подход Хомский2) модель Смысл 2) текст Мельную3) теория синтами насти групп Гладший4) Математическая языковая модельСтузов

В рамках данного подхода,для предста вление синтаксине стойструктуры предложения используетсяболее нагледный и распространённыйдеревья зависимостейспосо бВузлах дерева расположеныслова предложе ние (в торне, обычно)глагол сказуемое, а каждая дугадерева, связывающая пару узков,инте протируется как синтайшнескаяподчинные льная связь между ними.

Деревья составляющих большеподходят для описания языков,с жест Силь породном слов,предста в ление с их помощьюразор ванных и непро ективныхконструкций, требует расширениеиспользуемого грамматическогоформоанализа.Валентность- с пособность слова,присоединеть другие словаАнтант- слова- связка

В соотваствии с этой теорией, Еврассма три вается, как особого родапреобраз ователь выполняю щийперера бот из заданных смыслов всоответствующие им смысли1)Ориентация на синтез текста2) Многоуровневый, модульный3) Интегральная модель4) Специаль ные средстваписа име синта ктиги5) 4 пор на словарьЭТАП11

6 лекция

Модель Толково-комбинаторного словаряИ.А. МельнукаОсновным орудием исследо вание иописание языка является И.А. Мельнуюпризнаёт действующие людели, типасмысл (=) текстПодобная модель некоторогоконкретного языка - Это сложноорганизо ванная совокупность пра вил,чисто механичес кое применениякоторых должно обеспе или вать:- Переход от данного текста, нарассматриваемом языке, и формальо писанию смысла этого текстат.е.:

к его семантическому представлению(= смысловой записи) Разные, ноинтунтивно юнетике (синокимичные)тексты, должны получать одинако вые.или ходе-бы сисанимыми дививалетныепредставление. А темест, имеющийболее одного мысла,И.А. Мельную исходит изследующего тезиса:стест венный язывеиЭто особого род апреобразо ватель, в и по-пнающий переработвузаданных с мы с ро в, в

соответ ствующие и м тексты,люви б аданных тексто всоответствующие им смысли11Ауф.Другими словами, языкрассмат ри вается как определённоесоответст вид между смыслали итекстами Плюс, некоторый механизм,реализующий это соответствие, ввиде ной кретной процедуры, т.евыполняющий переход от смыслови певестам, и обратно.

4Смысок 1=) текстМодель9также должка быть действующейРмоделью.иОна представляет устройство,и Фимеющее вход и выход, иперерабатывающее то, итоюдаетсяему на вход (знание), в нечтодругое, что выдаётся на выходе(результат)с +Модели бываютЛинг влоти чисваяАгори минесвая

Модель (мысолет тевстдолжна быть б. адана совршенноформально- посредством однознанных,и рогичес Сли последовательныхформулизован, не требующих привлечениякакой либо добавочной информацииВ качестве понтрольногопритерия, вид вигается принципальнаяосуществи мость модели, илилюбогоее фрагмента, на вычислительноймашине