

Компьютерная лингвистика
Лекция № 1.

Компьютерная ЛИНГВИСТИКА как наука

Содержание лекции

1. Определения компьютерной лингвистики
2. Компьютерная лингвистика и искусственный интеллект
3. Разделы компьютерной лингвистики
4. Связь компьютерной лингвистики с другими областями знаний

Определения компьютерной ЛИНГВИСТИКИ

1. Деятельность по формализации знаний о естественном языке на разных его уровнях с целью использования в компьютерных технологиях.
2. Область знаний, решающая проблемы общения человека и компьютера на естественном языке.
3. Широкая область использования компьютерных инструментов - программ, компьютерных технологий организации и обработки данных – для моделирования функционирования языка в тех или иных условиях, ситуациях, проблемных областях.

Связь компьютерной лингвистики с искусственным интеллектом

- ▣ Искусственный интеллект – направление информатики, целью которого является разработка компьютерных систем, способных выполнять функции, традиционно считающиеся интеллектуальными, - понимание естественного языка, логический вывод, использование накопленных знаний, обучение, планирование действий и др.

Области искусственного интеллекта, использующие данные компьютерной ЛИНГВИСТИКИ

- Представление знаний
- Разработка экспертных систем
- Обучение и выявление закономерностей
- Обработка естественно-языковых текстов
- Разработка диалоговых систем и естественно-языкового (в том числе устного) интерфейса

Разделы компьютерной ЛИНГВИСТИКИ

- 1. Прикладная лексикография – рассматривает возможные типы машинных словарей, способы кодирования лингвистической информации в машинных словарях, пути составления и пополнения таких словарей.

```
-NNAM=NNAM
абориген_О=тубылец_О
аборигена_Г=тубыльца_Г
аборигену_Д=тубыльцу_Д
аборигена_А=тубыльца_А
аборигеном_І=тубыльцам_І
аборигене_Р=тубыльцу_Р
аборигены_РО=тубыльцы_РО
аборигенов_РГ=тубыльцаў_РГ
аборигенам_РД=тубыльцам_РД
аборигенов_РА=тубыльцаў_РА
аборигенами_РІ=тубыльцамі_РІ
аборигенах_РР=тубыльцах_РР
+
-VP=VP
благоустроить_С=упарадкаваць_С
благоустрою_ІФ1=упарадкую_ІФ1
благоустроишь_ІФ2=упарадкуеш_ІФ2
благоустроит_ІФ3=упарадкуе_ІФ3
благоустроим_ІФ1Р=упарадкуем_ІФ1Р
благоустроите_ІФ2Р=упарадкуеце_ІФ2Р
благоустроят_ІФ3Р=упарадкуюць_ІФ3Р
благоустрой_М2=упарадкуй_М2
благоустройте_М2Р=упарадкуйце_М2Р
благоустроил_ІРМ=упарадкаваў_ІРМ
благоустроила_ІРФ=упарадкавала_ІРФ
благоустроило_ІРН=упарадкавала_ІРН
благоустроили_ІРР=упарадкавалі_ІРР
благоустроив_В=упарадкаваўшы_В
+
```

Разделы компьютерной лингвистики

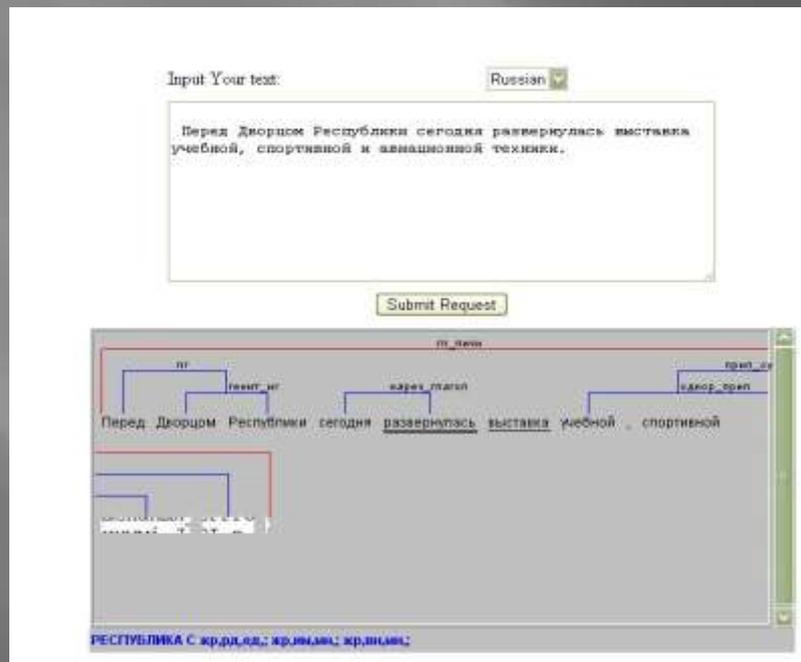
2. Составление машинных грамматик – формализация грамматических знаний о естественном языке.

Артыкул NNIMO 6 DC ..
Дзяржаўная JJFO ўлада NNIFO ў INR Рэспубліцы NNIFR Беларусь NPIFO
ажыццяўляецца VILIR3 на INR аснове NNIFR падзелу NNIMG яе PP3FG на INA
заканадаўчую JJFA ,, выканаўчую JJFA і CC судовую JJFA ..
Дзяржаўныя JJPO органы NNIMPO ў INR межах NNIFPR сваіх PSPG
паўнамоцтваў NNINPG самастойныя JJPO : : яны PP3PO
ўзаемадзейнічаюць VIIR3P паміж INI сабой PLI ,, стрымліваюць VIIR3P і CC
ўраўнаважваюць VIIR3P адзін DC1MA аднаго DC1MU ..
Артыкул NNIMO 7 DC ..
У INR Рэспубліцы NNIFR Беларусь NPIFO устанаўліваецца VILIR3
прынцып NNIMO вяршэнства NNING права NNING ..
Дзяржава NNIFO ,, ўсе PTPO яе PSPO органы NNIMPO і CC службовыя JJPO
асобы NNAFPO дзейнічаюць VIIR3P у INR межах NNIFPR Канстытуцыі NNIFG і CC
прынятых LPPPR у INR адпаведнасці NNIFR з INI ёю PP3FS актаў NNIMPG
заканадаўства NNING ..
Прававыя JJPO акты NNIMPO або CC іх PP3PO асобныя JJPO палажэнні NNINPO
,, прызнаныя LPPPO ва INR ўстаноўленым LPPMR законам NNIMI парадку NNIMR
супярэчнымі JJPI з INI палажэннямі NNINPI Канстытуцыі NNIFG ,, не XNOT
маюць VIIR3P юрыдычнай JJFG сілы NNIFG ..
Нарматыўныя JJPO акты NNIMPO дзяржаўных JJPG органаў NNIMPG
публікуюцца VILIR3P або CC даводзяцца VILIR3P да ING ўсеагульнага JJNG
ведама NNING іншым PTMI прадугледжаным LPPMI законам NNIMI спосабам NNIMI

Фрагмент корпуса текста с морфологическим аннотированием

Разделы компьютерной лингвистики

3. Построение алгоритмов лингвистического анализа – разработка алгоритмов анализа естественно-языковых текстов на всех языковых уровнях: фонемном, морфемном, морфологическом, синтаксическом, прагматическом.



Построение синтаксического дерева зависимостей (проект Автоматическая Обработка Текста)
<http://aot.ru/>

Разделы компьютерной лингвистики

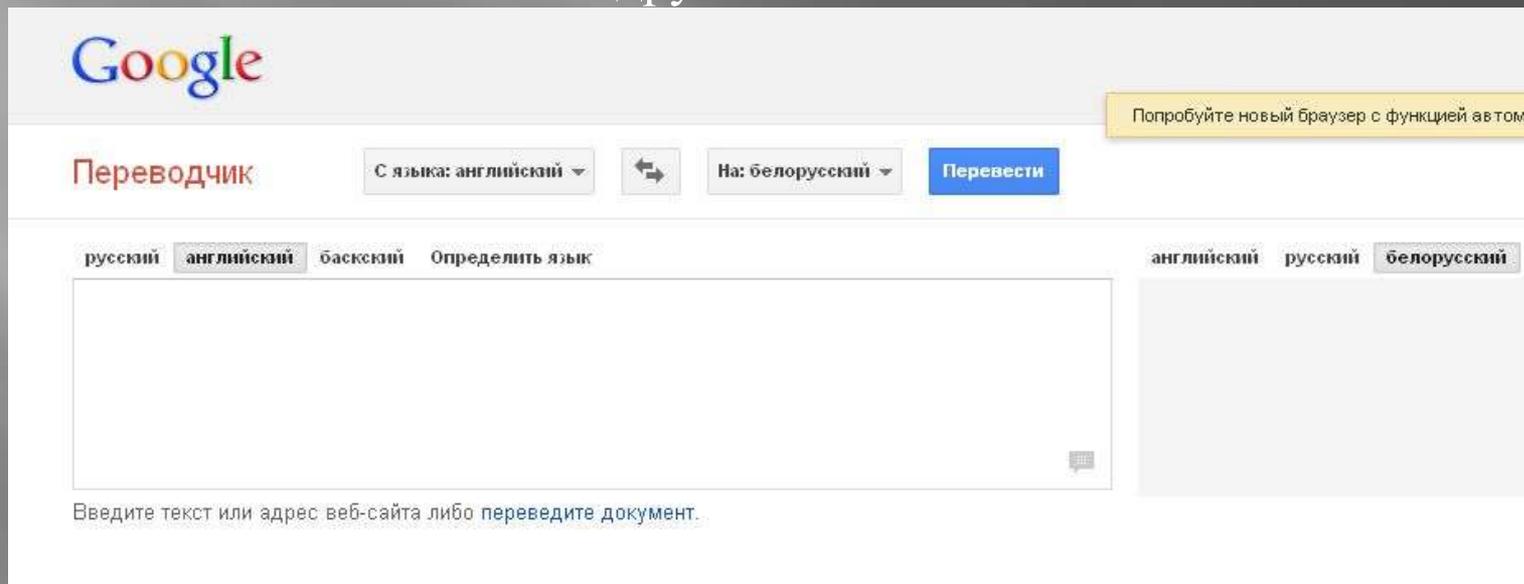
- 4. Корпусная лингвистика – занимается разработкой, созданием и использованием текстовых (лингвистических) корпусов.

The screenshot shows the website of the National Corpus of the Russian Language. The main header reads "Национальный корпус русского языка". The left sidebar contains navigation links: "главная", "решения", "вход в корпус", "что такое корпус?", "состав и структура", "статистика", "графы", "частоты", "морфология", "сборщики", "анализатор", "символы", "параметры текстов", "стабильность", "форум", "в проекте", "участники проекта", "публикации", "программные средства", "шаблоны в корпусе", "использование корпуса", "другие корпуса". The main content area features a "Новости проекта" section with several news items dated from 2012 to 2013, detailing updates to the corpus, including the addition of new text types and the expansion of the parallel corpus.

Главная страница сайта Национального корпуса русского языка
<http://ruscorpora.ru>

Разделы компьютерной лингвистики

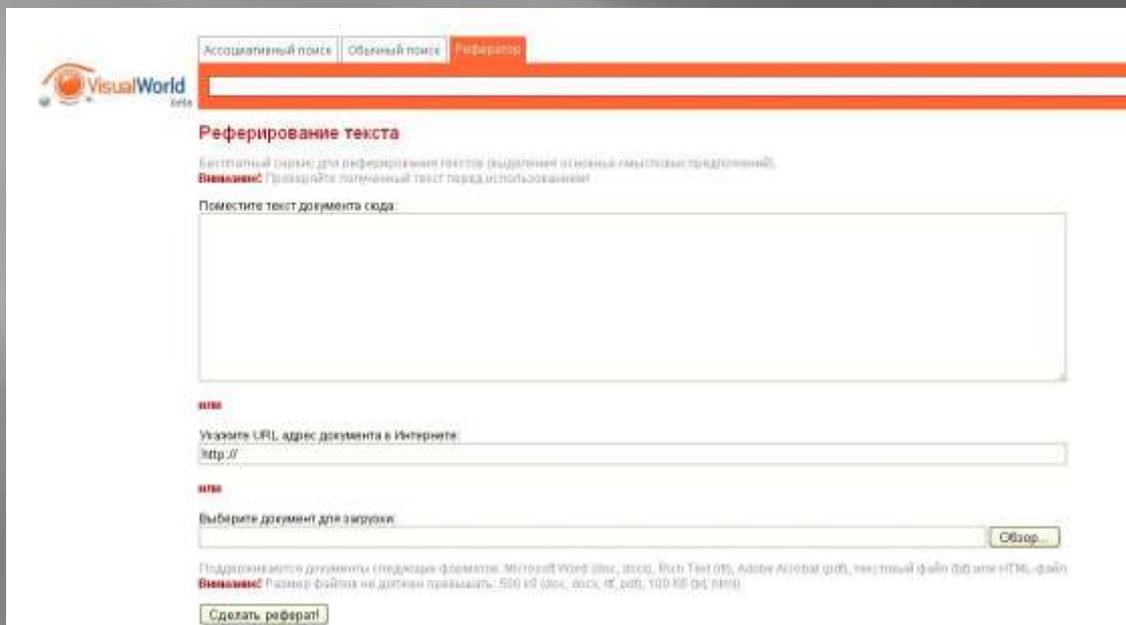
5. Машинный перевод - направление научных исследований, связанных с построением автоматических систем перевода текстов с одного естественного языка на другой .



Главная страница сайта Google-переводчик
<http://translate.google.com>

Разделы компьютерной ЛИНГВИСТИКИ

6. Автоматическое реферирование – создание программ сжатого представления естественно-языковых текстов.



The screenshot shows the 'Visual World' website interface for text summarization. At the top, there are three tabs: 'Ассоциативный поиск', 'Обычный поиск', and 'Рефератор', with 'Рефератор' being the active tab. Below the tabs is a search bar. The main heading is 'Реферирование текста'. A description states: 'Бесплатный сервис для реферирования текстов (выделения цитируемых фрагментов текста)'. A note says: 'Внимание! Проводится полный текст перед использованием!'. There is a large text input area with the prompt 'Поместите текст документа сюда:'. Below this is a section for 'ИЛИ' with a text input field for 'Укажите URL адрес документа в Интернете' and a 'http://' prefix. Another 'ИЛИ' section has a file selection input field and a 'Сбор...' button. At the bottom, there is a 'Сделать реферат!' button and a note: 'Поддерживаются документы следующих форматов: Microsoft Word (doc, docx), Rich Text (rtf), Adobe Acrobat (pdf), текстовый файл (txt) или HTML-файл'. A warning note says: 'Внимание! Размер файла не должен превышать: 500 кб (doc, docx, rtf, pdf), 100 кб (rtf, html)'.

Сервис построения реферата на сайте Визуальный мир
<http://visualworld.ru/referat.jsp>

Разделы компьютерной ЛИНГВИСТИКИ

7. Полнотекстовый информационный поиск – разработка информационно-поисковых систем с лингвистическим анализом текстов информационного массива и запроса пользователя.



Главная страница информационно-поисковой системы Яндекс
<http://ya.ru/>

Связь компьютерной лингвистики с другими науками

Компьютерная лингвистика

Общее языкознание

Общий объект изучения —
естественный язык

Связь компьютерной лингвистики с другими науками

Компьютерная лингвистика

Информатика

Компьютерная лингвистика изучает язык с позиции его использования в информационных системах

Связь компьютерной лингвистики с другими науками

Компьютерная
лингвистика

Математика

Главная задача компьютерной лингвистики —
выражение знаний о естественном языке в
формализованном виде, с помощью
математического аппарата

Связь компьютерной лингвистики с другими науками

Компьютерная лингвистика

Семиотика

Формализация – перекодирование информации, выраженной знаками одной семиотики средствами другой.

Связь компьютерной лингвистики с другими науками

Компьютерная лингвистика

Психология и психолингвистика

При создании искусственных систем анализа и синтеза естественно-языковых текстов необходимо понимать, как устроено языковое сознание и мышление человека.