

МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Ярославский государственный университет им.П.Г.Демидова»

УЧЕБНО-МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ

МЕТОДЫ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ В ИСКУССТВЕННОМ
ИНТЕЛЛЕКТЕ - 1

Направление подготовки (специальность):

02.04.02 Фундаментальная информатика и информационные технологии

Образовательная программа

Искусственный интеллект и компьютерные науки

очная форма обучения

Составитель:

ПАРАМОНОВ ИЛЬЯ ВЯЧЕСЛАВОВИЧ,
К.Ф.-М.Н., ДОЦЕНТ Ф-ТА ИВТ
ЯРГУ ИМ. П.Г. ДЕМИДОВА

г. Череповец - 2021

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

Основная литература:

- 1 Боярский, К. К. Введение в компьютерную лингвистику : Учебное пособие / К. К. Боярский. – Санкт-Петербург : Университет ИТМО, 2013. – 73 с. (ЭБС IPR BOOKS)

Дополнительная литература:

- 1 Проблемы компьютерной лингвистики и типологии : Сборник научных трудов / «ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ФАКУЛЬТЕТ РОМАНО-GERMANСКОЙ ФИЛОЛОГИИ НАУЧНО-МЕТОДИЧЕСКИЙ ЦЕНТР КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ. – Воронеж : Издательский дом ВГУ, 2017. – 246 с. URL: https://www.elibrary.ru/download/elibrary_35022773_12374093.pdf
- 2 Мамаев, И. Д. Русско-английский словарь основных терминов компьютерной лингвистики / И. Д. Мамаев // Лексикографическая копилка : Сборник научных статей / Под научной редакцией В.В. Гончаровой. – Санкт-Петербург : Санкт-Петербургский государственный экономический университет, 2020. – С. 83-92. URL: https://www.elibrary.ru/download/elibrary_43856875_79359109.pdf

Учебно-методическое обеспечение самостоятельной работы обучающихся по дисциплине (модулю)

- 1 Электронный университет Moodle ЯрГУ URL: <https://moodle.uniyar.ac.ru/>
- 2 Единое окно доступа к образовательным ресурсам URL: <http://window.edu.ru/>

Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля), включая перечень информационных справочных систем (при необходимости)

- 1 Пакет анализа естественного языка Stanza: официальный сайт. URL: <https://stanfordnlp.github.io/stanza/>
- 2 Пакет анализа естественного языка SpaCy: официальный сайт. URL: <https://spacy.io/>

Перечень информационных технологий, используемых при изучении дисциплины, включая программное обеспечение

1. Интерпретатор Python 3 (свободно распространяемое ПО)
Среда PyCharm Community Edition (свободно распространяемое ПО)

**Учебно-методические указания и рекомендации
к изучению тем лекционных и практических занятий, самостоятельной
работе студентов**

Содержание дисциплины

Наименование раздела дисциплины	Название темы с кратким содержанием
1	Лингвистика и информация. Язык. Предметы лингвистики и информационных технологий. Информация, свойства информации. Понятие формальности. Язык. Словарь. Естественные и формальные языки. Избыточность как свойство естественных языков. Знак. Слово. Лексема. Словоформа. Словоупотребление. Фонетика, морфология, синтаксис, семантика.
2	Лингвистическое моделирование. Модели языка как фундамент для построения систем автоматической обработки текста. N-граммы. Векторное представление слов.
3	Лингвистический процессор. Алгоритмы лингвистического разбора и анализа текста. Методы морфологического анализа, используемые в лингвистических процессорах. Морфологические словари. Алгоритмы синтаксического и семантического анализа для автоматических систем обработки текстов. Парсеры ЕЯ.
4	Формальные методы исследования структуры ЕЯ текста. Статистические методы анализа структур ЕЯ текста на морфологическом, синтаксическом, семантическом уровнях. Понятие о стилометрии.
5	Понятие тезауруса и онтологии естественного языка. Виды связей между словами и понятиями. Способы применения тезаурусов и онтологий при решении задач компьютерной лингвистики.

Примеры типовых индивидуальных заданий

№ п.п.	Тема
1	Уровни представления языковых выражений.
2	Морфологический анализ и синтез текстов. Поверхностная и глубинная морфология. Анализ композитов.
3	Основные типы представления синтаксической структуры предложения. Зависимости и составляющие. Дерево зависимостей.
4	Понятие синтаксического правила (синтагмы).

5	Грамматика и словарь.
6	Синтаксические признаки слова.
7	Валентностная структура предиката. Синтаксические и семантические валентности. Модель управления слова.
8	Основные типы компьютерных синтаксических ресурсов. Словари и корпуса текстов.
9	Глубокий анализ лингвистических данных: постановка задачи, основные методы и подходы.

Задание выполняется письменно в виде реферата.

Требования к выполнению индивидуального задания:

1. Наличие авторской позиции, самостоятельность суждений.
2. Степень раскрытия сущности проблемы:
 - соответствие плана теме реферата;
 - соответствие содержания теме и плану реферата;
 - владение терминологией и понятийным аппаратом проблемы;
 - полнота и глубина раскрытия основных понятий проблемы;
 - обоснованность способов и методов работы с материалом;
 - умение работать с литературой, систематизировать и структурировать материал;
 - умение обобщать, сопоставлять различные точки зрения по рассматриваемому вопросу, аргументировать основные положения и выводы
3. Уровень выбора источников
 - обоснованность,
 - полнота использования,
 - современность,
 - научность;
4. Соблюдение требований к оформлению, грамотность:
 - правильное оформление ссылок на используемую литературу;
 - соблюдение требований к объему реферата (от 10 до 15 страниц);
 - культура оформления: выделение абзацев
 - грамотность и культура изложения;
 - отсутствие орфографических и синтаксических ошибок,
 - отсутствие стилистических погрешностей;
 - отсутствие опечаток, сокращений слов, кроме общепринятых;
 - научный литературный стиль.

Работа, в которой дословно переписаны текст учебника, пособия или аналогичная работа, защищенная ранее другим студентом, не оценивается, а тема заменяется на новую.

Критерии оценки

«Отлично» — Выполнены все требования к заданию. Обозначена проблема и обоснована ее актуальность, сделан анализ различных точек зрения на рассматриваемую проблему и логично изложена собственная позиция; рассмотрены примеры использования разбираемых методов и подходов к решению задачи, сформулированы выводы, тема раскрыта полностью.

«Хорошо» – Выполнены все требования к заданию. Обозначена проблема и обоснована ее актуальность, сделан анализ хотя бы одной точки зрения на рассматриваемую проблему, изложена собственная позиция, но возможно нарушение логики изложения; рассмотрены

примеры использования разбираемых методов и подходов к решению задачи, но неполно, сформулированы выводы, тема раскрыта, но не полностью

«Удовлетворительно» – Выполнена большая часть требований. Проблема обозначена без обоснования ее актуальности, факты, методы, алгоритмы изложены частично и без анализа; рассмотрен пример использования разбираемых методов и подходов к решению задачи, но неполно или с ошибками

«Неудовлетворительно» – Не выполнены требования. Тема задания не раскрыта, обнаруживается существенное непонимание проблемы или задание не выполнено

Примеры вопросов к зачету

1. Предметы лингвистики и информационных технологий. Информация, свойства информации.
2. Естественные и формальные языки. Избыточность как свойство естественных языков. Знак. Слово. Лексема. Словоформа. Словоупотребление.
3. Фонетика, морфология, синтаксис, семантика.
4. Лингвистическое моделирование. Действующие модели языка .
5. Задачи компьютерной лингвистики в изучении естественного языка (ЕЯ).
6. Алгоритмы лингвистического разбора и анализа текста.
7. Методы морфологического анализа, используемые в лингвистических процессорах. Морфологические словари.
8. Алгоритмы синтаксического и семантического анализа для автоматических систем обработки текстов.
9. Формальные методы исследования структуры ЕЯ текста.
10. Статистические методы анализа структур ЕЯ текста на морфологическом, синтаксическом, семантическом уровнях.
11. Морфологический компонент системы автоматической обработки текстов.
12. Морфологическая структура слова и предложения.
13. Алгоритм синтаксического анализа.
14. Синтаксические отношения. Синтагмы. Синтаксическая структура предложения.

Критерии оценки

«Отлично» – ответ на вопросы показывает всестороннее знание темы, изученной литературы, изложен логично, аргументировано и в полном объеме. Основные понятия, выводы и обобщения сформулированы убедительно и доказательно. Продемонстрированы полные и глубокие навыки практического применения методов компьютерной лингвистики.

«Хорошо» – ответ на вопросы основан на твердом знании темы. Возможны недостатки в систематизации или в обобщении материала, неточности в выводах. Продемонстрированы хорошие навыки практического применения методов компьютерной лингвистики.

«Удовлетворительно» – ответ на вопросы базируется на знании основ предмета, но имеются значительные пробелы в изложении материала, затруднения в его изложении и систематизации, выводы слабо аргументированы, в содержании допущены теоретические ошибки. Продемонстрированы элементарные навыки практического применения методов компьютерной лингвистики.

«Неудовлетворительно» – оценивается ответ на вопросы, в котором обнаружено неверное изложение темы, систематизации знаний, обобщений и выводов нет. Навыки практического применения методов компьютерной лингвистики слабые и отрывочные или отсутствуют.