

Из учебной программы

ОСНОВЫ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ

учреждения образования
по учебной дисциплине для специальностей:

6-05-0231-01 Современные иностранные языки (английский, немецкий)
6-05-0231-01 Современные иностранные языки английский, китайский)
6-05-0231-01 Современные иностранные языки (немецкий, английский)

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Учебная программа «Основы компьютерной лингвистики» предназначена для студентов 3 курса дневной формы получения высшего образования, обучающихся по специальности 6-05-0231-01 «Современные иностранные языки (английский, немецкий / английский, китайский / немецкий, английский)».

Учебная дисциплина «Основы компьютерной лингвистики» входит в модуль дисциплин профилизации «Компьютерная лингвистика» вузовского компонента. Данная дисциплина направлена на овладение необходимым инструментарием деятельности эффективного специалиста.

Целью изучения дисциплины является формирование у студентов лингвистических и технологических компетенций в области компьютерной лингвистики. В рамках указанных целей выделяются следующие *задачи*:

- приобретение студентами компетенций в области лингвистической безопасности и её аспектов (коммуникативного, когнитивного, образовательного), навыков ответственной коммуникации и работы с языковыми данными в современных цифровых средах;
- приобретение студентами опыта в области постановки и решения профильных задач на основе использования современных систем обработки вербальных данных, навыков работы с современными интеллектуальными системами и сервисами;
- освоение студентами основных методов и средств компьютерной обработки языка и текста, получение навыков комплексного использования средств вычислительной техники в решении языковых задач;
- приобретение начального опыта образовательной коммуникации с интеллектуальными системами;
- стимулирование процессов самопознания и самосовершенствования;

– воспитание патриотизма и гражданственности, уважительного отношения к историческому наследию и культурным традициям белорусского народа в контексте цифровых трансформаций.

Достижение поставленных задач при изучении дисциплины «Основы компьютерной лингвистики» обеспечивается соответствующим уровнем систематизированных, осознанных и устойчивых знаний и умений студента.

В результате изучения учебной дисциплины студент должен *знать*:

– научные и общекультурные основы возникновения технологий компьютерной лингвистики;

– принципы и возможности компьютерной обработки языка и текста в коммуникативных системах;

– тенденции развития и основные направления современных методов компьютерной обработки языковых феноменов;

– понятие, задачи, технологии и правила использования современных информационных сервисов, средств и систем;

– требования цифровой культуры и информационной безопасности при организации и осуществлении образовательного процесса;

уметь:

– эффективно использовать аппаратное и программное обеспечение генерации, анализа, обработки и воспроизведения текстовой информации;

– анализировать и оценивать способы и средства моделирования текстовой информации;

– эффективно использовать программное обеспечение обработки компьютерных моделей связного текста;

– осуществлять лингвистическую, семантическую и гипертекстовую разметку текстовой информации,

– извлекать и обрабатывать метаязыковую информацию из текстов.

владеть:

– стратегиями ответственного коммуникативного поведения в цифровых образовательных средах;

– средствами и приемами эффективного применения цифровых сервисов и систем обработки текста и речи;

– правилами и приемами образовательного взаимодействия с системами искусственного интеллекта.

Процесс освоения учебной дисциплины «Основы компьютерной лингвистики» должен обеспечить у обучающихся формирование следующих компетенций:

УК. Решать стандартные задачи профессиональной деятельности на основе применения информационно-коммуникационных технологий.

СК. Решать прикладные лингвистические задачи на основе применения языковых компьютерных моделей.

Дисциплина «Основы компьютерной лингвистики» связана с учебными дисциплинами «Информационные технологии», «Искусственный интеллект», «Машинный перевод», «Компьютерной моделирование».

На занятиях рекомендуется использовать цифровой учебный инструментарий в сетевых образовательных средах, применять различные формы самостоятельной работы студентов и проектные задания. Изложение материала по дисциплине «Основы компьютерной лингвистики» предполагает обязательное использование интеллектуальных систем.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

5 семестр

Тема 1. Понятие и основания компьютерной лингвистики

1.1 Предмет, цели, методы и этапы развития компьютерной лингвистики. Формальные языки и грамматики. Статистические, вероятностные методы изучения и моделирования языка. Информационная безопасность. Лингвистическая и коммуникативная безопасность. Образовательная безопасность и защита данных.

1.2 Основания компьютерной лингвистики. Языковедческие основания компьютерной лингвистики. История лингвистики в контексте развития представления о языке и речевом субъекте. Логико-математические основания компьютерной лингвистики. Языковые парадоксы. Математика, логика и язык. Логика высказываний. Логика предикатов. Технологические основания компьютерной лингвистики.

Тема 2. Автоматическая обработка языковых данных

2.1 Данные, текст и язык. Данные. Структуры данных. Текстовые данные. Цифровое представление данных. Кодировки и форматы текстовых данных. Дефиниции текста в различных исследовательских контекстах. Письменный и устный текст. Мультимодальный текст.

2.2 Электронный текст. Текст как информационная система. Текст в электронной информационной среде.

2.3 Автоматическая обработка текста как технологическая основа компьютерной лингвистики. История автоматической обработки текста. Уровни автоматической обработки: Text Processing, Natural Language Processing.

Тема 3. Этапы обработки языковых данных

3.1 Препроцессинг языковых данных. Графематическая обработка текста. Нормализация и токенизация текста. Лемматизация и стемминг. Парсинг текста. Автоматическая разметка текста. Структурная разметка. Лингвистическая разметка (синтаксическая, семантическая просодическая, морфемная). Речевой препроцессинг. Цифровизация речевого сигнала. Нормализация и форматирование сигнала.

3.2 Компьютерный анализ языка и текста. Языковые средства и методы процессинга. Программные библиотеки процессинга в АОТ. Клеточные автоматы. Защита данных средствами NLP. Статистический анализ текстов. Статистические модели. Марковские цепи. Вероятностный анализ текстов.

Понятие вероятности и её основные категории. Теорема Байеса и апостериорные вероятности. Байесовский подход.

3.3 *Постпроцессинг языковых данных.* Форматирование языковых данных. Компьютерная корректура. Автореферирование. Улучшение читаемости. Адаптация результатов автоматической обработки под конкретную задачу.

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Основная литература:

1. Актуальные вопросы обеспечения информационной безопасности : пособие для педагогов учреждений образования, реализующих образовательные программы общего среднего образования / В. Ю. Арчаков, А. Л. Баньковский, О. С. Макаров [и др.]. – Минск : Народная асвета, 2021. – 168 с.

2. Гринчук, С. Н. Облачные технологии и сервисы Веб 2.0 в образовательной деятельности : учебно-методическое пособие / С. Н. Гринчук, И. А. Дзюба, Е. В. Шакель ; Мин. образования Республики Беларусь, ГУО «Республиканский институт высшей школы». – Минск : РИВШ, 2023. – 162 с.

3. Сизанов, А. Н. Интерактивные и компьютерные технологии в воспитательной работе: опыт России и Беларуси : учебно-методическое пособие / А. Н. Сизанов. – Минск : РИВШ, 2013. – 186 с.

Дополнительная литература:

4. Информационные технологии в образовании : учебник для вузов / Е. В. Баранова, М. И. Бочаров, С. С. Куликова, Т. Б. Павлова. – Санкт-Петербург : Лань, 2022. – 296 с.

5. Сысоев, П. В. Информационные и коммуникационные технологии в лингвистическом образовании : учебно-методическое пособие / П. В. Сысоев. – М. : ЛИБРОКОМ, 2015. – 264 с.

6. Старостина, С. А. Информационные технологии в филологии : учебное пособие для вузов / С. А. Старостина. – Тамбов : ТГУ им. Г.Р. Державина, 2022. – 118 с.

Перечень средств диагностики

Компьютерное тестирование при выполнении практических заданий.
Устный опрос во время занятий.
Оценка письменных отчетов о выполнении практических заданий.
Оценка контрольных тестовых заданий по разделам дисциплины.
Дифференцированный зачет.

Примерный список тем к дифференцированному зачету

1. Языковедческие основы компьютерной лингвистики.
2. Логико-математические основы компьютерной лингвистики.

3. Технологические основы компьютерной лингвистики. АОТ.
4. Логика высказываний.
5. Устная и письменная речь их этапы и уровни.
6. Лингвистическая безопасность.
7. Предмет, цели и этапы развития компьютерной лингвистики.
8. Препроцессинг, процессинг, построение.
9. Нормализация и токенизация текста.
10. Лемматизация и стемминг.
11. Парсинг текста.
12. Автоматическая разметка текста.
13. Лингвистическая разметка.
14. Социолингвистическая разметка.
15. Речевой препроцессинг.
16. Статистический анализ текстов.
17. Формальные языки и грамматики.
18. Статистические и вероятностные методы моделирования языка.
19. Стохастическая парадигма изучения языка.
20. Безопасность в образовательной коммуникации с LLM.
21. Прикладные направления компьютерной лингвистики.
22. Компьютерные представления языковых структур.
23. Лингвистический анализ в компьютерной лингвистике.
24. Обработка многоязычных текстов.
25. Основы создания диалоговых систем.
26. Компьютерная лексикография и терминоведение.
27. Компьютерный перевод
28. Определение, цели и задачи лингвистических корпусов.
29. Виды и назначение лингвистических корпусов.
30. Разметка лингвистических корпусов.
31. Проблемы качества корпусных данных.
32. Статистические подходы в корпусной лингвистике.
33. Многоязычные корпуса.
34. Корпуса устной речи.
35. Использование корпусов в образовании.
36. Корпусные грамматики.

Методические рекомендации по организации и выполнению самостоятельной работы студентов

Частью процесса обучения компьютерной лингвистике является организация самостоятельной работы студентов. Самостоятельная работа студентов предусматривает использование возможностей информационно-коммуникационных технологий, библиотечных ресурсов, самостоятельный образовательный диалог с системами генеративного искусственного интеллекта, чтение актуальных научно-технических публикаций на английском языке с целью формирования информационно-коммуникационных компетенций и удовлетворения познавательных интересов, а также выполнение упражнений и тестовых заданий.