

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учебно-методическое объединение по образованию
в области информатики и радиоэлектроники

УТВЕРЖДАЮ

Первый заместитель Министра образования
Республики Беларусь

_____ А.Г.Баханович

Регистрационный № _____

**ЕСТЕСТВЕННО-ЯЗЫКОВОЙ ИНТЕРФЕЙС ИНТЕЛЛЕКТУАЛЬНЫХ
СИСТЕМ**

**Примерная учебная программа по учебной дисциплине
для специальности
6-05-0611-03 Искусственный интеллект**

СОГЛАСОВАНО

Председатель учебно-методического
объединения по образованию в
области информатики и
радиоэлектроники

_____ В.А.Богущ

СОГЛАСОВАНО

Начальник Главного Управления
профессионального образования
Министерства образования
Республики Беларусь

_____ С.Н.Пищов

СОГЛАСОВАНО

Проректор по научно-методической
работе Государственного учреждения
образования “Республиканский
институт высшей школы”

_____ И.В.Титович

Эксперт-нормоконтролер

Минск 2024

СОСТАВИТЕЛЬ:

Ю.Б.Крапивин, доцент кафедры интеллектуальных информационных технологий учреждения образования “Белорусский государственный университет информатики и радиоэлектроники”, кандидат технических наук, доцент.

РЕЦЕНЗЕНТЫ:

Кафедра интеллектуальных систем Белорусского государственного университета (протокол № 11 от 17.01.2024);

И.М.Бойко, ведущий научный сотрудник отдела информационных систем унитарного предприятия “Геоинформационные системы” Национальной академии наук Беларуси”, кандидат технических наук

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ В КАЧЕСТВЕ ПРИМЕРНОЙ:

Кафедрой интеллектуальных информационных технологий учреждения образования “Белорусский государственный университет информатики и радиоэлектроники” (протокол № 20 от 08.01.2024);

Научно-методическим советом учреждения образования “Белорусский государственный университет информатики и радиоэлектроники” (протокол № ____ от ____);

Научно-методическим советом по разработке программного обеспечения и информационно-коммуникационным технологиям Учебно-методического объединения по образованию в области информатики и радиоэлектроники (протокол № 7 от 18.03.2024)

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

ХАРАКТЕРИСТИКА УЧЕБНОЙ ДИСЦИПЛИНЫ

Примерная учебная программа по учебной дисциплине “Естественно-языковой интерфейс интеллектуальных систем” разработана для студентов учреждений высшего образования, обучающихся по специальности 6-05-0611-03 “Искусственный интеллект” в соответствии с требованиями образовательного стандарта общего высшего образования и примерного учебного плана вышеуказанной специальности.

Учебная дисциплина “Естественно-языковой интерфейс интеллектуальных систем” входит в число дисциплин государственного компонента модуля “Интерфейсы интеллектуальных систем” в учебном плане специальности 6-05-0611-03 “Искусственный интеллект”. Она предназначена для формирования представлений о способах организации диалога человек-компьютер, для обучения студентов подходам, которые используются при проектировании и реализации подсистем естественно-языкового и речевого интерфейсов интеллектуальных систем, а также для решения прикладных задач автоматической обработки текста. Программа имеет практическую направленность.

Воспитательное значение учебной дисциплины “Естественно-языковой интерфейс интеллектуальных систем” заключается в формировании у обучающихся технической культуры и научного мировоззрения; развитии исследовательских умений, аналитических способностей, креативности, необходимых для решения научных и практических задач; развитии познавательных способностей и активности: творческой инициативы, самостоятельности, ответственности и организованности; формировании способностей к саморазвитию, самосовершенствованию и самореализации.

Изучение данной учебной дисциплины способствует созданию условий для формирования интеллектуально развитой личности обучающегося, которой присущи стремление к профессиональному совершенствованию, активному участию в экономической и социально-культурной жизни страны, гражданская ответственность и патриотизм.

ЦЕЛИ, ЗАДАЧИ УЧЕБНОЙ ДИСЦИПЛИНЫ

Цели учебной дисциплины: ознакомление с основами компьютерной лингвистики; освоение навыков создания и использования лингвистических моделей, применяемых для реализации как естественно-языкового, так и естественно-речевого интерфейсов в интеллектуальных системах; изучение методов проектирования и построения систем анализа и синтеза речи, а также решения прикладных задач, связанных с автоматическим анализом и синтезом текстов на естественном языке.

Задачи учебной дисциплины:

изучение принципов организации диалога человек-компьютер;

изучение основ компьютерной лингвистики;

изучение лингвистических основ речевого интерфейса;

освоение представления о методах анализа и синтеза речевого сигнала и навыков практического использования систем автоматического распознавания и синтеза речи;

приобретение навыков практической реализации в виде компьютерных программ основных этапов анализа и синтеза естественно-языковых текстов и решений прикладных задач, связанных с автоматическим анализом и синтезом текстов на естественном языке.

Базовыми учебными дисциплинами по курсу “Естественно-языковой интерфейс интеллектуальных систем” являются “Общая теория интеллектуальных систем”, “Проектирование баз знаний”. В свою очередь учебная дисциплина “Естественно-языковой интерфейс интеллектуальных систем” является базой для учебной дисциплины “Технологии и инструментальные средства проектирования интеллектуальных систем” (учебная дисциплина компонента учреждения образования).

ТРЕБОВАНИЯ К УРОВНЮ ОСВОЕНИЯ СОДЕРЖАНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ

В результате изучения учебной дисциплины “Естественно-языковой интерфейс интеллектуальных систем” формируется следующая базовая профессиональная компетенция: проектировать естественно-языковые интерфейсы интеллектуальных систем.

В результате изучения учебной дисциплины студент должен:

знать:

основные принципы организации диалога человек-компьютер;

основные понятия и направления компьютерной лингвистики;

лингвистические модели, применяемые для реализации естественно-языкового интерфейса;

структуру и назначение естественно-языковой системы;

преимущества и основные характеристики речевого интерфейса;

структуру речевого сигнала, методы анализа и синтеза речевого сигнала;

методы синтеза речи по тексту, общую структуру и принципы работы синтезатора речи;

методы автоматического распознавания речи, принципы работы современных компьютерных систем распознавания речи;

подходы к решению прикладных задач автоматической обработки текста (информационного поиска, машинного перевода, автоматического реферирования, распознавания языка текста);

уметь:

формулировать цели и задачи разработки естественно-языкового интерфейса в интеллектуальных системах;

создавать и применять морфологические, синтаксические и семантические модели естественного языка для решения прикладных задач автоматической обработки текста и построения естественно-языкового интерфейса;

выявлять и формализовывать лингвистические знания;

выявлять фонетические закономерности и определять спектрально-временные характеристики речевых сигналов;

реализовывать в виде компьютерных программ основные этапы (морфологический, синтаксический, семантический) анализа естественно-языковых текстов;

владеть:

навыками проектирования и реализации систем анализа и синтеза речи,

навыками решения прикладных задач автоматического анализа и синтеза текстов на естественном языке;

навыками применения инструментов и технологий обработки естественного языка и речи.

Примерная учебная программа рассчитана на 320 учебных часов, из них – 140 аудиторных. Примерное распределение аудиторных часов по видам занятий: лекции – 68 часов, лабораторные занятия – 72 часа.

ПРИМЕРНЫЙ ТЕМАТИЧЕСКИЙ ПЛАН

Наименование раздела, темы	Всего аудиторных часов	Лекции	Лабораторные занятия
Введение	2	2	-
Раздел 1. Общение. Диалог. Язык	12	12	-
Тема 1. Компьютерная лингвистика и естественно-языковой интерфейс	2	2	-
Тема 2. Общение. Идеология построения модели общения	2	2	-
Тема 3. Проблема системы языка в современных условиях	4	4	-
Тема 4. Язык как система	4	4	-
Раздел 2. Основы компьютерной лингвистики	34	14	20
Тема 5. Введение в лингвистику	2	2	-
Тема 6. Морфология	6	2	4
Тема 7. Корпусы текстов естественного языка	6	2	4
Тема 8. Формальная модель синтаксиса	8	4	4
Тема 9. Семантика. Семантический и семантико-синтаксический анализ	12	4	8
Раздел 3. Естественно-языковой интерфейс	12	4	8
Тема 10. Знание и понимание как системы	2	2	-
Тема 11. Интерфейс интеллектуальной системы	8	2	8
Раздел 4. Прикладные задачи автоматической обработки текста	44	20	24
Тема 12. Информационный поиск	14	8	6
Тема 13. Автоматическое распознавание языка текста	10	4	6
Тема 14. Автоматическое реферирование документов	10	4	6
Тема 15. Машинный перевод текстов естественного языка	10	4	6
Раздел 5. Речевой интерфейс	8	4	4
Тема 16. Лингвистические основы речевого интерфейса	8	4	4
Раздел 6. Структура речевого сигнала. Анализ и синтез	14	6	8
Тема 17. Основы теории речеобразования	6	2	4
Тема 18. Методы анализа и синтеза речевого сигнала	8	4	4
Раздел 7. Синтез речи по тексту	8	4	4
Тема 19. Синтезатор речи	8	4	4
Раздел 8. Распознавание речи	6	2	4
Тема 20. Методы распознавания речи	6	2	4
Итого:	140	68	72

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

ВВЕДЕНИЕ

Основные положения компьютерной лингвистики. Компьютерная лингвистика как метод развития систем естественно-языкового общения. Модель общения. Естественный язык как система. Проблема формализации естественного языка.

Раздел 1. ОБЩЕНИЕ. ДИАЛОГ. ЯЗЫК

Тема 1. КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА И ЕСТЕСТВЕННО-ЯЗЫКОВОЙ ИНТЕРФЕЙС

Понятие компьютерной лингвистики (КЛ). Направления КЛ. Связь с другими науками и направлениями информатики в частности, понятие эргономики и т.п.

Тема 2. ОБЩЕНИЕ. ИДЕОЛОГИЯ ПОСТРОЕНИЯ МОДЕЛИ ОБЩЕНИЯ

Определения понятия “общение”. Правила взаимодействия между людьми. Способы общения. Основные особенности модели общения. Компьютерная лингвистика как метод развития систем естественно-языкового общения. Лингвистические модели. Основы модели общения. Понятие о человеко-машинных системах и требования к их разработке.

Тема 3. ПРОБЛЕМА СИСТЕМЫ ЯЗЫКА В СОВРЕМЕННЫХ УСЛОВИЯХ

Естественный язык как средство общения между людьми. “Машинный” язык как посредник в общении человека и компьютера. Математический аппарат представления языковых знаний в компьютере. Диалог как средство реализации общения. Типы диалога. Моделирование диалога. Формализация схемы и представления диалога в интеллектуальных системах. Диалоговые системы.

Тема 4. ЯЗЫК КАК СИСТЕМА

Элементы общей теории систем, помогающие формализовать языковой материал. Основные группы отношений, используемые для представления лингвистической информации. Язык и речь. Определения языка. Естественный язык (ЕЯ) как средство развития мышления. Функции языка: информационная, коммуникативная, воспитательная, развивающая, эстетическая. Понятия языкознания: слово, словоформа, лексема. Язык как семиотическая система. Основные понятия семиотики. Омонимия, типы омонимии. Синонимия. Автонимия.

Раздел 2. ОСНОВЫ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ

Тема 5. ВВЕДЕНИЕ В ЛИНГВИСТИКУ

Основные понятия и определения: лексика, морфология, синтаксис, семантика, грамматика. Словари: типология словарей, роль словаря в

реализации ЕЯ-системы. Уровни изучения текста. Общий алгоритм анализа и синтеза естественно-языковых текстов. Назначение и структура лингвистического процессора (ЛП). Этапы анализа и синтеза текста естественного языка (ТЕЯ) лингвистическим процессором. Представление лингвистических знаний, в том числе с использованием унифицированного способа абстрактного кодирования семантических сетей (SC-код).

Тема 6. МОРФОЛОГИЯ

Понятие морфологии. Классификация слов естественного языка. Слово и его состав. Морфологические категории слов. Декларативные и процедурные морфологические знания о языке. Морфологические словари. Морфологический анализ и синтез. Словообразование. Алгоритмы словообразовательного процесса. Методы обнаружения и исправления ошибок в тексте. Представление морфологических знаний в лингвистических базах знаний.

Тема 7. КОРПУСЫ ТЕКСТОВ ЕСТЕСТВЕННОГО ЯЗЫКА

Корпусы естественно-языковых текстов как основа знаний о языке. Виды корпусов текстов. Инструменты работы с корпусами текстов естественного языка.

Тема 8. ФОРМАЛЬНАЯ МОДЕЛЬ СИНТАКСИСА

Понятие предложения. Типология предложений. Поверхностно-синтаксический анализ. Первичный синтаксический разбор предложения. Словосочетание и члены предложения. Классификация словосочетаний. Типы сочетаемости слов. Декларативные синтаксические знания о языке. Этапы синтаксического анализа и синтеза. Формализация синтаксической модели русского языка.

Тема 9. СЕМАНТИКА. СЕМАНТИЧЕСКИЙ И СЕМАНТИКО-СИНТАКСИЧЕСКИЙ АНАЛИЗ

Семантический анализ ТЕЯ. Семантические модели представления ТЕЯ: компонентного подхода, модель Филмора, семантические сети, фреймы, Смысл-Текст, универсальный семантический код. Уровни представления ТЕЯ. Глубинные падежи и семантические валентности. Лексические функции и толкование слов.

Раздел 3. ЕСТЕСТВЕННО-ЯЗЫКОВОЙ ИНТЕРФЕЙС

Тема 10. ЗНАНИЕ И ПОНИМАНИЕ КАК СИСТЕМЫ

Типология знаний. Системное представление знаний. Структура знаний естественно-языкового интерфейса: лингвистические знания, знания о процессе общения и др. Проблемы грамматического и семантического кодирования. Использование языков представления знаний для формального описания знаний о языке в ЕЯ-системе.

Тема 11. ИНТЕРФЕЙС ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ

Структура интерфейса “человек-компьютер”. Интегральное использование лингвистических моделей для построения естественно-языкового интерфейса прикладной интеллектуальной системы.

Раздел 4. ПРИКЛАДНЫЕ ЗАДАЧИ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ТЕКСТА

Тема 12. ИНФОРМАЦИОННЫЙ ПОИСК

Задача информационного поиска. Структурно-функциональная схема системы информационного поиска. Метапоисковые средства: особенности построения и функционирования. Стратегии поиска документов. Статистические закономерности в ТЕЯ. Основы поисковой оптимизации: типы, структура поисковых запросов, факторы ранжирования результатов поиска. Инструменты поисковой оптимизации. Оценки эффективности систем классификации и информационного поиска.

Тема 13. АВТОМАТИЧЕСКОЕ РАСПОЗНАВАНИЕ ЯЗЫКА ТЕКСТА

Общая постановка задачи автоматического распознавания языка текста. Методы решения задачи автоматического распознавания языка текста: машинного обучения, лингво-статистические и др.

Тема 14. АВТОМАТИЧЕСКОЕ РЕФЕРИРОВАНИЕ ДОКУМЕНТОВ

Задача автоматического реферирования документов, подходы к решению. Виды рефератов, основные алгоритмы и методы, решения задачи автоматического реферирования текста.

Тема 15. МАШИННЫЙ ПЕРЕВОД ТЕКСТОВ ЕСТЕСТВЕННОГО ЯЗЫКА

Общая постановка задачи машинного перевода текстов. Классификация систем машинного перевода текстов. Уровни обеспечения систем машинного перевода текстов. Системы непрямого машинного перевода. Лингвистическое обеспечение систем непрямого машинного перевода: словари, грамматики и алгоритмы. Проблемы машинного перевода.

Раздел 5. РЕЧЕВОЙ ИНТЕРФЕЙС

Тема 16. ЛИНГВИСТИЧЕСКИЕ ОСНОВЫ РЕЧЕВОГО ИНТЕРФЕЙСА

Речевой интерфейс (РИ): общая структура речевого интерфейса, компьютерные системы синтеза, распознавания и понимания речи. Разделы лингвистики, которые используются для реализации РИ. Речевой сигнал и фонетика речи. Определение фонетики, просодики, фонемы, транскрипции. Звуки речи: фонемы и аллофоны. Классификация фонем русской речи. Слог и слоговая структура. Понятие аллофона. Типология аллофонов. Разбиение речевого потока на основные единицы. Просодика речи: ударение и интонация. Мелодика, ритмика и энергетика речи. Интонация и ее функции в речи.

Основные интонационные типы. Интонационные типы синтагм. Ударение и его функции в речи. Типология ударений.

Раздел 6. СТРУКТУРА РЕЧЕВОГО СИГНАЛА. АНАЛИЗ И СИНТЕЗ

Тема 17. ОСНОВЫ ТЕОРИИ РЕЧЕОБРАЗОВАНИЯ

Спектрально-временные характеристики речевого сигнала (РС). Информационная и модуляционная структура речи. Виды информации, заключенной в РС.

Тема 18. МЕТОДЫ АНАЛИЗА И СИНТЕЗА РЕЧЕВОГО СИГНАЛА

Определение спектра, формант и их характеристик. Формантная модель акустики речевого тракта. Использование формант для синтеза и распознавания речи. Распределение формантных частот гласных звуков русской речи. Разделение формант согласных звуков русской речи по месту и способу образования, временные диаграммы изменения амплитудных параметров частот согласных. Разнообразие методов синтеза речи: параметрический, компиляционный.

Раздел 7. СИНТЕЗ РЕЧИ ПО ТЕКСТУ

Тема 19. СИНТЕЗАТОР РЕЧИ

Общая структура синтезатора речи по тексту. Лингвистический (текстовый) процессор. Преобразование “буква-фонема”. Правила преобразования орфографического текста в фонемный. Синтагматическое членение фраз. Примеры преобразования текста в размеченный фонемный текст. Фонетический процессор. Типология фонетических процессоров. Артикуляторно-командная модель речевого тракта. Артикуляторно-формантный фонетический процессор. Формантный фонетический процессор.

Раздел 8. РАСПОЗНАВАНИЕ РЕЧИ

Тема 20. МЕТОДЫ РАСПОЗНАВАНИЯ РЕЧИ

Проблема распознавания речевых образов. Типология задач распознавания речевых образов. Виды характеристик и ограничений, накладываемых на систему автоматического распознавания речи (САРР). Распознавание раздельно произносимых слов из малого словаря. Задачи распознавания слитно произносимой речи. Общая схема распознавания речевых образов. Структура САРР. Методы автоматического распознавания речи. Системы распознавания речи: основные характеристики, оценка качества, практическое применение.

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

ЛИТЕРАТУРА

ОСНОВНАЯ

1. Автоматическая обработка текстов на естественном языке и анализ данных : учебное пособие / Е. И. Большакова [и др.]. – Москва : НИУ ВШЭ, 2017. – 269 с.
2. Гаврилова, Т. А. Инженерия знаний : модели и методы : учебник / Т. А. Гаврилова, Д. В. Кудрявцев, Д. И. Муромцев. – 2-е изд., стер. – Санкт-Петербург : Лань, 2018. – 324 с.
3. Головки, В. А. Основы компьютерных технологий : учебно-методическое пособие. – 2-е изд., испр. и доп. / В. А. Головки. А. А. Дудкин, Л. П. Матюшков. – Брест : БрГТУ, 2015. – 196 с.
4. Тампель, И. Б. Автоматическое распознавание речи : учебное пособие / И. Б. Тампель, А. А. Карпов. – Санкт-Петербург : ИТМО, 2016. – 138 с.
5. Хобсон, Л. Обработка естественного языка в действии / Л. Хобсон, Х. Ханнес, Х. Коул. – Санкт-Петербург : Питер, 2020. – 576 с.

ДОПОЛНИТЕЛЬНАЯ

6. Рассел, С. Искусственный интеллект : современный подход / С. Рассел, П. Норвиг ; пер. с англ. – 2-е изд. – Москва : Вильямс, 2006. – 1408 с.
7. Елисеева, О. Е. Естественно-языковой интерфейс интеллектуальных систем : учебное пособие / О. Е. Елисеева ; под науч. ред. проф. В. В. Голенкова. – Минск : БГУИР, 2009. – 151 с.
8. Маннинг, Д. Введение в информационный поиск / Д. Маннинг, П. Рагхаван, Х. Шютце. – Москва : Вильямс, 2011. – 528 с.
9. Лобанов, Б. М. Компьютерный синтез и клонирование речи / Б. М. Лобанов, Л. И. Цирульник. – Минск : Белорусская наука, 2008. – 342 с.

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ОРГАНИЗАЦИИ И ВЫПОЛНЕНИЮ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

При изучении учебной дисциплины рекомендуется использовать следующие формы самостоятельной работы:

- подготовка к лекциям и лабораторным занятиям;
- работа с учебно-методическими пособиями;
- чтение рекомендуемой литературы;
- выполнение лабораторных работ;

самостоятельное изучение тем и проблем, не выносимых на лекции и лабораторные занятия.

ПЕРЕЧЕНЬ РЕКОМЕНДУЕМЫХ СРЕДСТВ ДИАГНОСТИКИ КОМПЕТЕНЦИЙ СТУДЕНТА

Примерным учебным планом специальности 6-05-0611-03 “Искусственный интеллект” в качестве формы промежуточной аттестации по учебной дисциплине “Естественно-языковой интерфейс интеллектуальных систем” рекомендуются зачет и экзамен. Оценка учебных достижений студента производится по системе “зачтено/не зачтено” и десятибалльной шкале.

Для текущего контроля по учебной дисциплине и диагностики компетенций студентов могут использоваться следующие формы:

- устный опрос;
- защита лабораторных работ.

РЕКОМЕНДУЕМЫЕ МЕТОДЫ (ТЕХНОЛОГИИ) ОБУЧЕНИЯ

Основные рекомендуемые методы (технологии) обучения, отвечающие целям и задачам учебной дисциплины:

1. Теоретико-информационные:
 - устное логически-целостное изложение учебного материала (лекции);
 - объяснение;
 - консультирование.
2. Практико-операционные:
 - упражнения;
 - решение задач;
 - лабораторные занятия.

ПРИМЕРНЫЙ ПЕРЕЧЕНЬ ТЕМ ЛАБОРАТОРНЫХ ЗАНЯТИЙ

1. Лексико-грамматический анализ текстов естественного языка.
2. Построение и использование корпусов текстов естественного языка.
3. Синтаксический анализ текстов естественного языка.
4. Семантико-синтаксический анализ текстов естественного языка.
5. Построение и использование диалоговых систем с поддержкой естественного языка.
6. Информационный поиск.
7. Автоматическое распознавание языка текста.
8. Автоматическое реферирование документов.
9. Автоматический машинный перевод текстов.
10. Фонетические аспекты речевого сигнала.
11. Просодические аспекты речевого сигнала.
12. Методы синтеза речевого сигнала.
13. Автоматизированная система синтеза речи.
14. Автоматизированная система распознавания речи.

ПРИМЕРНЫЙ ПЕРЕЧЕНЬ КОМПЬЮТЕРНЫХ ПРОГРАММ
(необходимого оборудования, наглядных пособий и т. п.)

1. Программный пакет MS Office или аналогичный.
2. Средство разработки графических диаграмм MS Visio или аналогичное.
3. Открытая семантическая технология компонентного проектирования интеллектуальных систем.
4. Персональный компьютер, оборудованный звуковой картой и гарнитурой (наушники, микрофон).
5. Инструментальная среда записи и обработки звука.
6. Инструментальная среда проектирования программных приложений с использованием языков программирования высокого уровня.