

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 02.06.2025 12:22:08
Уникальный программный ключ:
sa953a01204891083f939673078ef1a989dae18a

Федеральное государственное автономное образовательное учреждение высшего образования
«Российский университет дружбы народов имени Патриса Лумумбы»
Факультет искусственного интеллекта
(наименование основного учебного подразделения (ОУП)-разработчика ОП ВО)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

ЛИНГВИСТИЧЕСКИЕ ОСНОВЫ АНАЛИЗА ЕСТЕСТВЕННОГО ЯЗЫКА

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

**02.03.02 ФУНДАМЕНТАЛЬНАЯ ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ,
09.03.03 ПРИКЛАДНАЯ ИНФОРМАТИКА**

(код и наименование направления подготовки/специальности)

Освоение дисциплины ведется в рамках реализации основной профессиональной образовательной программы высшего образования (ОП ВО):

**ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: РАЗРАБОТКА И ОБУЧЕНИЕ
ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ**

(наименование (профиль/специализация) ОП ВО)

2025 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Лингвистические основы анализа естественного языка» входит в программу бакалавриата «Искусственный интеллект: разработка и обучение интеллектуальных систем» по направлению 02.03.02 «Фундаментальная информатика и информационные технологии» и изучается в 4 семестре 2 курса. Дисциплину реализует Кафедра прикладного искусственного интеллекта. Дисциплина состоит из 3 разделов и 26 тем и направлена на изучение формирования у студентов фундаментальных знаний в области лингвистики, необходимых для эффективного анализа, интерпретации и предобработки текстовых данных. Курс раскрывает основные понятия структуры языка, лингвистических уровней (фонетика, морфология, синтаксис, семантика), знакомит с методами разметки, препроцессинга и построения языковых моделей. Эти знания необходимы как для реализации прикладных NLP-задач, так и для правильной постановки и интерпретации результатов в области искусственного интеллекта.

Целью освоения дисциплины является дать студентам системное представление о структуре, правилах и моделях естественных языков, научить анализировать тексты на различных лингвистических уровнях, понимать специфику текстовых данных для последующего использования инструментов NLP и машинного обучения, избежать типичных ошибок в постановке и реализации задач.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Лингвистические основы анализа естественного языка» направлено на формирование у обучающихся следующих компетенций (части компетенций):

Таблица 2.1. Перечень компетенций, формируемых у обучающихся при освоении дисциплины (результаты освоения дисциплины)

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
УК-4	Способен к коммуникации в межличностном и межкультурном взаимодействии на русском как иностранном и иностранном(ых) языке(ах) на основе владения взаимосвязанными и взаимозависимыми видами репродуктивной и продуктивной иноязычной речевой деятельности, такими как аудирование, говорение, чтение, письмо и перевод в повседневной, социокультурной, учебно-профессиональной, официально-деловой и научной сферах общения	УК-4.1 Знает принципы построения устного и письменного высказывания на государственном и иностранном языках; требования к деловой устной и письменной коммуникации; УК-4.2 Умеет применять на практике устную и письменную деловую коммуникацию; УК-4.3 Владеет методикой составления суждения в межличностном деловом общении на государственном и иностранном языках, с применением адекватных языковых форм и средств;
ПК-2	Способен эффективно работать с большими объемами данных, включая их предварительную обработку, анализ и визуализацию, с целью извлечения полезной информации для обучения	ПК-2.2 Демонстрирует навыки анализа данных с использованием статистических методов и инструментов; ПК-2.3 Владеет методами работы с различными алгоритмами машинного обучения и глубокого обучения для решения различных задач;

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
	моделей искусственного интеллекта	

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Лингвистические основы анализа естественного языка» относится к блоку по выбору блока образовательной программы высшего образования.

В рамках образовательной программы высшего образования обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Лингвистические основы анализа естественного языка».

Таблица 3.1. Перечень компонентов ОП ВО, способствующих достижению запланированных результатов освоения дисциплины

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
УК-4	Способен к коммуникации в межличностном и межкультурном взаимодействии на русском как иностранном и иностранном(ых) языке(ах) на основе владения взаимосвязанными и взаимозависимыми видами репродуктивной и продуктивной иноязычной речевой деятельности, такими как аудирование, говорение, чтение, письмо и перевод в повседневной, социокультурной, учебно-профессиональной, официально-деловой и научной сферах общения	Русский язык и культура речи; <i>Иностранный язык**</i> ; <i>Русский язык (как иностранный)**</i> ; Русский язык для иностранных студентов; Второй иностранный язык (практический курс);	<i>Иностранный язык в профессиональной деятельности**</i> ; <i>Русский язык в профессиональной деятельности**</i> ; Русский язык для иностранных студентов; Преддипломная практика;
ПК-2	Способен эффективно работать с большими объемами данных, включая их предварительную обработку, анализ и визуализацию, с целью извлечения полезной информации для обучения моделей искусственного интеллекта	Технологическая (проектно-технологическая) практика (учебная); Статистические методы и первичный анализ данных; Введение в базы данных; Программирование на языке Python;	<i>Цифровые двойники**</i> ; <i>Информационный поиск**</i> ; <i>Анализ временных рядов**</i> ; <i>Основы больших языковых моделей**</i> ; <i>Нейронные сети</i> ; <i>Nadoor, SPARK</i> ; <i>Анализ естественного языка с помощью методов искусственного интеллекта</i> ; <i>Оптимизация моделей машинного обучения</i> ; <i>Практикум по обработке естественного языка (NLP)</i> ; <i>Основы глубокого обучения</i> ; <i>Обработка и анализ изображений и видео с</i>

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
			помощью методов искусственного интеллекта; Проектирование и разработка систем компьютерного зрения; Программирование на языке C++; Основы робототехники**; Преддипломная практика; Технологическая (проектно-технологическая) практика (производственная); Эксплуатационная практика (производственная);

* - заполняется в соответствии с матрицей компетенций и СУП ОП ВО

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Лингвистические основы анализа естественного языка» составляет «3» зачетные единицы.

Таблица 4.1. Виды учебной работы по периодам освоения образовательной программы высшего образования для очной формы обучения.

Вид учебной работы	ВСЕГО, ак.ч.		Семестр(-ы)
			4
Контактная работа, ак.ч.	51		51
Лекции (ЛК)	17		17
Лабораторные работы (ЛР)	0		0
Практические/семинарские занятия (СЗ)	34		34
Самостоятельная работа обучающихся, ак.ч.	57		57
Контроль (экзамен/зачет с оценкой), ак.ч.	0		0
Общая трудоемкость дисциплины	ак.ч.	108	108
	зач.ед.	3	3

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Таблица 5.1. Содержание дисциплины (модуля) по видам учебной работы

Номер раздела	Наименование раздела дисциплины	Содержание раздела (темы)		Вид учебной работы*
Раздел 1	Структура и уровни естественного языка	1.1	Язык как система: структура, функции, уровни анализа	ЛК
		1.2	Фонетика и фонология — звуковая сторона языка; письменные и устные тексты	ЛК
		1.3	Морфология — строение слова, морфемы, части речи; основы лемматизации и стемминга	ЛК
		1.4	Анализ текстов с точки зрения звучания, разбор фонем	СЗ
		1.5	Построение морфологической разметки, определение морфем	СЗ
		1.6	Групповое обсуждение задач морфологического анализа (русский и английский язык)	СЗ
		1.7	Практика лемматизации и стемминга на простых корпусах текстов	СЗ
		1.8	Типовые ошибки морфологической обработки в NLP	СЗ
Раздел 2	Синтаксис и основы структурной разметки текста	2.1	Синтаксис: понятие синтаксической структуры, координаты и связи в предложении	ЛК
		2.2	Методы синтаксического разбора (парсинга): деревья зависимостей и составляющие	ЛК
		2.3	Синтаксическая неоднозначность и способы её разрешения	ЛК
		2.4	Практика ручного синтаксического разбора предложения	СЗ
		2.5	Составление деревьев синтаксических связей для коротких фраз	СЗ
		2.6	Разработка правил идентификации основных синтаксических структур	СЗ
		2.7	Обзор существующих языковых ресурсов (UD, SynTagRus, Penn Treebank)	СЗ
		2.8	Практика работы с корпусами синтаксически размеченных текстов	СЗ
		2.9	Групповой разбор случая синтаксической неоднозначности и её устранения	СЗ
Раздел 3	Семантика, прагматика и языковые ресурсы для NLP	3.1	Семантика: значения слов и отношений; синонимия, омонимия, полисемия	ЛК
		3.2	Основы прагматики: контекст, импликации, речевые акты	ЛК
		3.3	Корпусная лингвистика: ресурсы и их применение в анализе текстов (WordNet, Национальный корпус русского языка, Wiki-разметки)	ЛК
		3.4	Анализ лексической неоднозначности и определение смысловых оттенков	СЗ
		3.5	Практика построения простых семантических сетей	СЗ
		3.6	Разбор связи текста и контекста, прагматический анализ микротекстов	СЗ
		3.7	Знакомство с электронными корпусами и их структурой	СЗ
		3.8	Групповой проект: построение собственного мини-корпуса и аннотирование	СЗ
		3.9	Итоговая дискуссия: современные направления применения лингвистических основ в ИИ	СЗ

* - заполняется только по **ОЧНОЙ** форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – практические/семинарские занятия.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Таблица 6.1. Материально-техническое обеспечение дисциплины

Тип аудитории	Оснащение аудитории	Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости)
Лекционная	Аудитория для проведения занятий лекционного типа, оснащенная комплектом специализированной мебели; доской (экраном) и техническими средствами мультимедиа презентаций.	
Семинарская	Аудитория для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная комплектом специализированной мебели и техническими средствами мультимедиа презентаций.	
Для самостоятельной работы	Аудитория для самостоятельной работы обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом специализированной мебели и компьютерами с доступом в ЭИОС.	

* - аудитория для самостоятельной работы обучающихся указывается **ОБЯЗАТЕЛЬНО!**

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

1. Гречко, В.А. Теория языкознания: Учеб. пособие/В.А. Гречко. - М.:Высш шк., 2003.— 375 с: ил. ISBN 5-06-004292-8

2. Лаппо, М. А. Самоидентификация: семантика, прагматика, языковые ресурсы: монография / М. А. Лаппо; Мин-во образования и науки РФ, Новосиб. гос. пед. ун-т. – Новосибирск: Изд-во НГПУ, 2013. – 180 с. ISBN 978-5-85921-977-3

Дополнительная литература:

1. Солнышкина Марина Ивановна, Макнамара Даниэль С., Замалетдинов Радиф Рифкатович ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА И ИЗУЧЕНИЕ СЛОЖНОСТИ ДИСКУРСА // Russian Journal of Linguistics. 2022. Выпуск номер 2 том 26, С.317-341

Ресурсы информационно-телекоммуникационной сети «Интернет»:

1. ЭБС РУДН и сторонние ЭБС, к которым студенты университета имеют доступ на основании заключенных договоров

- Электронно-библиотечная система РУДН – ЭБС РУДН

<https://mega.rudn.ru/MegaPro/Web>

- ЭБС «Университетская библиотека онлайн» <http://www.biblioclub.ru>

- ЭБС «Юрайт» <http://www.biblio-online.ru>

- ЭБС «Консультант студента» www.studentlibrary.ru

- ЭБС «Знаниум» <https://znanium.ru/>

2. Базы данных и поисковые системы

- Sage <https://journals.sagepub.com/>

- Springer Nature Link <https://link.springer.com/>

- Wiley Journal Database <https://onlinelibrary.wiley.com/>

- Наукометрическая база данных Lens.org <https://www.lens.org>

Учебно-методические материалы для самостоятельной работы обучающихся при освоении дисциплины/модуля:*

1. Курс лекций по дисциплине «Лингвистические основы анализа естественного языка».

* - все учебно-методические материалы для самостоятельной работы обучающихся размещаются в соответствии с действующим порядком на странице дисциплины **в ТУИС!**

РАЗРАБОТЧИК:

Заведующий кафедрой
прикладного искусственного
интеллекта

Должность, БУП

Подпись

Подолько Павел
Михайлович

Фамилия И.О.

РУКОВОДИТЕЛЬ БУП:

Заведующий кафедрой
прикладного искусственного
интеллекта

Должность, БУП

Подпись

Подолько Павел
Михайлович

Фамилия И.О.

РУКОВОДИТЕЛЬ ОП ВО:

Заведующий кафедрой
прикладного искусственного
интеллекта

Должность, БУП

Подпись

Подолько Павел
Михайлович

Фамилия И.О.