

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 29.05.2025 12:26:00
Уникальный программный ключ:
sa953a0120d891083f939673078ef1a989dae18a

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

МЕТОДЫ АНАЛИЗА И ХРАНЕНИЯ БОЛЬШИХ ДАННЫХ

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

02.04.02 ФУНДАМЕНТАЛЬНАЯ ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

(код и наименование направления подготовки/специальности)

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

УПРАВЛЕНИЕ ДАННЫМИ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

(наименование (профиль/специализация) ОП ВО)

2025 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Методы анализа и хранения больших данных» входит в программу магистратуры «Управление данными и искусственный интеллект» по направлению 02.04.02 «Фундаментальная информатика и информационные технологии» и изучается во 2 семестре 1 курса. Дисциплину реализует Кафедра прикладного искусственного интеллекта. Дисциплина состоит из 3 разделов и 26 тем и направлена на изучение формирования у студентов системных знаний и практических навыков работы с масштабируемыми инфраструктурами и методами хранения и анализа данных в условиях роста их объема, разнообразия, скорости поступления и требований к надежности и безопасности. Курс развивает компетентность в архитектуре современных систем Big Data, интеграции с распределенными и облачными решениями, автоматизации работы с данными, анализе потоковых, нереляционных и смешанных данных, а также внедрении эффективных методологий хранения, обработки и аналитики в промышленных и исследовательских задачах искусственного интеллекта.

Целью освоения дисциплины является научить студентов проектировать и реализовывать решения для хранения, обработки и анализа больших и сложных потоков данных, понимать производственные и исследовательские стандарты современных распределённых систем, грамотно выбирать технологию под задачу (Hadoop, Spark, Kafka, облачные хранилища и др.), обеспечивать масштабируемость, надежность, защиту и эффективность в рамках комплексных AI/Data Science-проектов.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Методы анализа и хранения больших данных» 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)
ПК-1	Способен разрабатывать и применять алгоритмы интеллектуальной обработки данных для решения задач профессиональной деятельности	ПК-1.1 Знает существующие системы хранения и анализа данных, алгоритмы интеллектуальной обработки данных; ПК-1.2 Умеет модифицировать алгоритмы интеллектуальной обработки данных; ПК-1.3 Имеет навыки использования и применения существующих и модифицированных систем хранения и анализа данных, алгоритмов интеллектуальной обработки данных для решения задач профессиональной деятельности;

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Методы анализа и хранения больших данных» относится к обязательной части блока 1 «Дисциплины (модули)» образовательной программы высшего образования.

В рамках образовательной программы высшего образования обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Методы анализа и хранения больших данных».

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

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
ПК-1	Способен разрабатывать и применять алгоритмы интеллектуальной обработки данных для решения задач профессиональной деятельности	<i>Обработка мультимодальных данных**;</i> <i>Современные устройства центров обработки больших данных**;</i>	Преддипломная практика; <i>Искусственный интеллект и интернет вещей**;</i> Генеративный искусственный интеллект; Большие языковые модели (на основании трансформеров); Методы оптимизации;

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

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Методы анализа и хранения больших данных» составляет «4» зачетные единицы.

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

Вид учебной работы	ВСЕГО, ак.ч.	Семестр(-ы)
		2
Контактная работа, ак.ч.	51	51
Лекции (ЛК)	17	17
Лабораторные работы (ЛР)	17	17
Практические/семинарские занятия (СЗ)	17	17
Самостоятельная работа обучающихся, ак.ч.	66	66
Контроль (экзамен/зачет с оценкой), ак.ч.	27	27
Общая трудоемкость дисциплины	ак.ч.	144
	зач.ед.	4

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	Архитектуры и инфраструктура анализа и хранения больших данных	1.1	Архитектурные основы Big Data: концепция, требования, типы данных, сценарии применения	ЛК
		1.2	Выбор архитектур: локальные, распределённые, облачные, гибридные решения хранения и обработки	ЛК
		1.3	Хранилища данных: Data Lake, Data Warehouse, Data Mesh, современные форматы хранения (Parquet, ORC и др.)	ЛК
		1.4	Развертывание и настройка мини-кластера Hadoop/Spark или облачного хранилища	ЛР
		1.5	Оценка и тестирование современных форматов хранения данных: сравнение по скорости и объёму	ЛР
		1.6	Обзор современных платформ и сервисов управления большими данными (выбор под задачу)	СЗ
		1.7	Анализ ландшафта индустриальных проектов Big Data: сравнение архитектур и инфраструктур	СЗ
		1.8	Практический разбор кейсов проектирования инфраструктуры — от стартапа до корпорации	СЗ
Раздел 2	Технологии хранения и обработки больших данных: распределённые и потоковые методы	2.1	Распределённые файловые системы и базы данных: HDFS, Cassandra, ClickHouse	ЛК
		2.2	Потоковые платформы и обработка событий: Apache Kafka, Spark Streaming, Flink	ЛК
		2.3	Методы обеспечения согласованности, доступности, надёжности и отказоустойчивости в Big Data	ЛК
		2.4	Настройка и выполнение запросов в распределённой базе данных (например, ClickHouse или Cassandra)	ЛР
		2.5	Организация обработки потоковых данных: создание пайплайна Kafka + Spark Streaming	ЛР
		2.6	Лабораторное задание по реализации процедур резервного копирования и восстановления в продвинутых хранилищах	ЛР
		2.7	Разбор CAP-теоремы, типов согласованности и требований к масштабированию	СЗ
		2.8	Обсуждение сценариев применения потоковых и распределённых хранилищ (индустриальные кейсы)	СЗ
		2.9	Практика мониторинга, поддержки и резервирования в распределённых инфраструктурах	
Раздел 3	Интеллектуальный анализ данных и интеграция с AI/Data Science	3.1	Методы интеграции Big Data с аналитическими, ML и AI-системами: вызовы оркестрации, пайплайны	ЛК
		3.2	Масштабируемая аналитика: совместное использование Spark ML, Dask, параллельных библиотек	ЛК
		3.3	Безопасность, приватность и этические аспекты анализа и хранения больших данных в проектах искусственного интеллекта	ЛК
		3.4	Интеграция распределённого хранения с ML-аналитикой (Spark ML, Dask, BigQuery ML)	ЛР
		3.5	Реализация сквозного аналитического	ЛР

Номер раздела	Наименование раздела дисциплины	Содержание раздела (темы)		Вид учебной работы*
			пайплайна: ingestion, storage, обработка, ML-вывод	
		3.6	Эксперимент по оценке и обеспечению безопасности/приватности в инфраструктуре Big Data	ЛР
		3.7	Разбор практик построения сквозных пайплайнов AI/Big Data и управления жизненным циклом данных	СЗ
		3.8	Критический анализ реальных проектов Data Science на больших данных: bottlenecks, риски, удачные решения	СЗ
		3.9	Дискуссия по вопросам приватности, этики, compliance и ответственности при эксплуатации Big Data в AI	

* - 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.	
Компьютерный класс	Компьютерный класс для проведения занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная персональными компьютерами (в количестве 25 шт.), 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. Параскевов, А. В. Большие данные: учебное пособие / А. В. Параскевов, А. Э. Сергеев. - Москва; Вологда: Инфра-Инженерия, 2024. - 148 с. - ISBN 978-5-9729-2120-1. - Текст: электронный. - URL: <https://znanium.ru/catalog/product/2169699>
2. Григорьев, А. А. Методы и алгоритмы обработки данных: учебное пособие / А.А. Григорьев, Е.А. Исаев. — 2-е изд., перераб. и доп. — Москва: ИНФРА-М, 2024. — 383 с. + Доп. материалы [Электронный ресурс]. — (Высшее образование: Бакалавриат). — DOI 10.12737/1032305. - ISBN 978-5-16-015581-4. - Текст: электронный. - URL: <https://znanium.ru/catalog/product/2084190>
3. Анализ больших данных: учебное пособие / И. Б. Тесленко, В. Е. Крылов, А. М. Губернаторов [и др.]. — Москва: КноРус, 2025. — 295 с. — ISBN 978-5-406-14006-2. — URL: <https://book.ru/book/955989>

Дополнительная литература:

1. Дадян, Э. Г. Методы, модели, средства хранения и обработки данных: учебник / Э.Г. Дадян, Ю.А. Зеленков. — Москва: Вузовский учебник: ИНФРА-М, 2024. — 168 с. - ISBN 978-5-9558-0490-3. - Текст: электронный. - URL: <https://znanium.com/catalog/product/2122966>
2. Методы и технологии обработки больших данных [Электронный ресурс]: учебно-методическое пособие / М.М. Железнов. — Москва: Издательство МИСИ – МГСУ, 2020. ISBN 978-5-7264-2193-3

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 «Методы анализа и хранения больших данных».

* - 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.