Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Баламирзоев Назим Лиодинович
Должность: Ректор
Дата подписания: 10.03.2026 14:01:29
Уникальный программный ключ:
5cf0d6f89e80f49a334f6a4ba58e91f3326b9926

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
Институт кибербезопасности и цифровых технологий

Региональный партнёр
ФГБОУ ВО
«Дагестанский государственный технический университет»

УТВЕРЖДАЮ
И.о. ректора ФГБОУ ВО «ДГТУ»
_____ Н.Л. Баламирзоев
«____» _____ 2023 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Б1.В.08 ОБРАБОТКА БОЛЬШИХ ДАННЫХ

Направление подготовки: 09.03.01 «Информатика и вычислительная техника

Направленность (профиль подготовки): «Прикладной искусственный интеллект»

Квалификация выпускника: бакалавр

Форма обучения очная

Махачкала 2023

Программа составлена в соответствии с требованиями ФГОС ВО с учетом рекомендаций ОПОП ВО по направлению подготовки 09.03.01. – Информатика и вычислительная техника, профилю «Прикладной искусственный интеллект»

Разработчик _____ Магомедов И.А., ктн, доцент
подпись ФИО уч. степень, уч. звание)

05.09.2023г.

Зав. кафедрой, за которой закреплена дисциплина (модуль)

_____ Гасанова Н.М., к.э.н., доцент
подпись (ФИО уч. степень, уч. звание)

05.09. 2023г.

Программа одобрена на заседании выпускающей кафедры УиИТСиВТ

от 12.09.2023 г., протокол № 1

Зав. выпускающей кафедрой по данному направлению (специальности, профилю)

_____ Гасанова Н.М., к.э.н., доцент
подпись (ФИО уч. степень, уч. звание)

12.09.2023 г.

Программа одобрена на заседании Методического совета факультета компьютерных технологий, вычислительной техники и энергетики от 22.09.2023 года, протокол № 1.

Председатель Методического совета факультета КТиЭ

_____ Исабекова Т.И. , к.ф.-м. н., доцент
подпись (ФИО уч. степень, уч. звание)

«22» 09. 2023 г.

Декан факультета _____ Ш.А. Юсуфов
подпись ФИО

Начальник УО _____ Э.В. Магомаева
подпись ФИО

1. Цели освоения дисциплины

Целями освоения дисциплины «Обработка больших данных» являются приобретение обучающимися углубленных знаний современного состояния науки о данных (Data Science) и практических навыков в области решения задач обработки больших данных (Big Data), а также технологий их сбора, хранения, представления.

2. Место дисциплины в структуре ОПОП бакалавриата

2.1 Дисциплина относится к дисциплинам Блока 1 (Б1) образовательной программы, к части, формируемой участниками образовательных отношений (Б1.В).

2.2. Минимальные требования к «входным» знаниям, необходимым для успешного усвоении данной дисциплины - удовлетворительное усвоение программ по следующим дисциплинам бакалавриата:

- «Программирование»;
- «Современные информационные технологии»;
- «Информационные технологии в профессиональной деятельности»;
- «Базы данных и знаний программно-аппаратных комплексов вычислительных систем».

2.3 Компетенции, приобретенные в ходе изучения дисциплины, могут быть использованы при прохождении различных видов практик и готовят бакалавра к написанию выпускной квалификационной работы.

3. Результаты освоения дисциплины «Технологии больших данных»

Процесс изучения дисциплины направлен на формирование элементов следующих компетенций в соответствии с ФГОС ВО по данному направлению:

Коды компетенции	Наименование компетенции	Индикатор достижения компетенции	В результате освоения дисциплины обучающийся **должен**
ПК-6	Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта	ПК 6.1 Осуществляет поиск данных в открытых источниках, специализированных библиотеках и репозиториях	**Знать:** - концептуальные положения современной науки о данных (Data Science); - модели представления больших данных и способы организации структур их хранения; - области применения больших данных и их преимущества; - основные источники поиска и добычи больших данных; - основы логической и физической организации распределенных файловых систем для хранения больших данных; - основы обработки и анализа потоковых данных; - методы машинного обучения и

			интеллектуального анализа больших данных. - этапы обработки и аналитики больших данных; - принципы формирования поисковых запросов для открытых источников, специализированных библиотек и репозиториев. **Уметь:** - применять методы обработки и интерпретации больших данных; - применять методы обработки потоковых данных; - использовать распределенную инфраструктуру для поиска в открытых источниках, специализированных библиотеках и репозиториях; - решать задачи анализа ссылок и потоков данных; - решать задачи частых предметных наборов; - использовать основные алгоритмы кластеризации и поиска похожих объектов в структурах больших данных; - применять языки программирования, скриптовые языки и специализированное ПО для обработки больших данных. **Владеть:** - основными современными технологиями поиска данных в открытых источниках, специализированных библиотеках и репозиториях; - языком формирования поисковых запросов; - навыками представления

			и обработки больших социальных графов; - технологией настройки поисковых критериев и применения автоматических поисковых систем; - современными технологиями хранения и обработки больших данных.

4. Структура и содержание дисциплины **Б1.В.08 ОБРАБОТКА БОЛЬШИХ ДАННЫХ**

4.1. Структура дисциплины

Общая трудоемкость дисциплины составляет 5 зачетных единицы, 180 часа.

Форма обучения	Семестр	Общая трудоемкость по дисциплине (ЗЕТ/ в часах)	ЛК, час	ПЗ, час	ЛР, час	СРС. час	Контр., час	Контроль
Очно	6	5/180	34		34	76	36	Экзамен

4.2 Содержание дисциплины **Б1.В.08 ОБРАБОТКА БОЛЬШИХ ДАННЫХ**

Общая трудоемкость дисциплины составляет 5 зачетных единицы, 180 часа.

№ п/п	Раздел дисциплины, тема лекции и вопросы	Энная форма				Заочная форма			
		ЛК	ПЗ	ЛБ	СР	ЛК	ПЗ	ЛБ	СР
1	2	3	4	5	6	7	8	9	10
1.	Раздел 1. Введение в проблематику сбора и обработки больших данных (Big Data). Тема 1.1 Цели, задачи, методы и области применения методов обработки больших данных. Обзор современного положения дел в сфере сбора и обработки больших данных. Основные проблемы, связанные с обработкой больших данных.	2		2	4	2		2	1
2.	Тема 1.2 Источники больших данных. Наука о данных (Data Science) – история возникновения, этапы развития, цели, методы и области применения.	2		2	4				1
3.	Тема 1.3 Сбор больших данных Проблемы поиска и представления больших данных. Задача предварительной оценки и подготовки данных. Структурированные и неструктурированные данные. Виды больших данных	2		2	4	2		2	1
4.	Понятие статистического моделирования. Задача машинного обучения. Добыча данных как алгоритмическая задача. Проблема обобщения данных и выделения признаков. Статистические пределы сбора больших данных.	2		2	4				1
5.	Раздел 2. Программный стек для обработки больших данных. Тема 2.1. Распределенные файловые системы. Логическая и физическая организация вычислительных узлов. Обзор существующих распределенных файловых систем. Принципы организации больших файловых систем.	2		2	4	2		2	1

6.	Тема 2.2 Технология вычислений MapReduce. Схема вычислений по технологии MapReduce. Задачи-распределители Map. Группировка по ключу. Задачи-редукторы Reduce. Комбинаторы функций Reduce.	2		2	4				1
7.	Тема 2.3 Алгоритмы обработки больших данных на основе MapReduce. Основные алгоритмы и операции с использованием MapReduce: матричные операции, операции реляционной алгебры, вычисление выборок и проекций, вычисление объединения/пересечения/разности, вычисление естественного соединения, вычисление группировки/агрегирования.	2		2	4	2		2	1
8.	Тема 2.4 Обобщения и модификации технологии MapReduce. Основные алгоритмы и операции с использованием MapReduce: матричные операции, операции реляционной алгебры, вычисление выборок и проекций, вычисление объединения/пересечения/разности, вычисление естественного соединения, вычисление группировки/агрегирования.	2		2	4				1
9.	Раздел 3. Задачи обработки больших данных Тема 3.1 Задачи поиска похожих объектов Постановка задачи о сходстве объектов. Понятие меры расстояния (метрики). Сходства множеств по Жаккару. Сходство текстовых документов.	2		2	4				1

10.	Тема 3.2 Задачи анализа потоков данных Системы управления потоками данных. Проблема выборки данных из потока. Фильтрация потоков. Задача подсчета элементов в потоке. Задача вычисления распределения частоты элементов в потоке (задача оценивания моментов).	2		2	4				1
11.	Тема 3.3 Задачи анализа ссылок. Алгоритмы PageRank и TrustRank Постановка задачи анализа ссылок в поисковых системах. Функция PageRank и особенности ее алгоритмической реализации.	2		2	4				1
12.	Графовая структура веб-пространства. Проблема тупиков и паучьих ловушек. Варианты использование PageRank в поисковой системе. Эффективное вычисление PageRank. Тематический PageRank и алгоритм TrustRank. Алгоритм HITS.	2		2	4				1
13.	Тема 3.4 Задачи частых предметных наборов Понятие частого предметного набора. Модель корзины покупок. Алгоритм Apriori. Использование оперативной памяти для подсчета предметных наборов. Монотонность предметных наборов.	2		2	4				1
14.	Тема 3.5 Задачи кластеризации больших данных Обзор методов и стратегий кластеризации. Классификация алгоритмов кластеризации. Иерархическая кластеризация в евклидовом пространстве и оценка ее эффективности.	2		2	6				1

15.	Иерархическая кластеризация в неевклидовых пространствах. Применение алгоритма k-средних для кластеризации больших данных. Алгоритм кластеризации больших наборов данных CURE (Clustering Using REpresentatives). Кластеризация в неевклидовых пространствах (общий обзор методов). Кластеризация для потоков и параллелизм.	2		2	6				1
16.	Тема 3.6 Задачи представления и обработки больших графов Графовые модели социальных сетей. Задача кластеризации графа социальной сети. Метрики для графов социальных сетей.	2		2	6				2
17.	Алгоритм Гирвана-Ньюмана. Промежуточность и ее использование для нахождения сообществ. Прямое нахождение сообществ. Задача разрезания графа социальной сети. Нахождение пересекающихся сообществ. Метод анализа графов социальных сетей Simrank.	2		2	6	1		1	2

Форма текущего контроля успеваемости (по срокам текущих аттестаций в семестре)	Входная контр, работа; 1 аттестация 1-5 тема 2 аттестация 3-9 тема 3 аттестация 10-17 тема				Входная контр, работа; контрольная работа			
Форма промежуточной аттестации (по семестрам)	Экзамен 1 ЗЕТ (36 часов)				*Экзамен*			
ИТОГО	34	34		76	9	9		19

4.3 Перечень и содержание лабораторных занятий.

№ п/п	№ темы	Наименование лабораторных работ	Кол. ч
1	1.1, 1.2, 1.3.	Подготовка и сбор больших данных средствами языка Python и библиотеки PANDAS.	8
2	2.1-2.4	Изучение алгоритмов обработки больших данных на основе MapReduce.	8
3	3.1	Решение задач поиска похожих объектов.	2

4	3.2	Решение задач анализа потоков данных.	2
5	3.3	Исследование алгоритмов PageRank и TrustRank.	4
6	3.4	Решение задач частых предметных наборов	2
7	3.5	Решение задач кластеризации больших данных	4
8	3.6	Изучение алгоритмов обработки больших графов	4
		Итого	34

5. Образовательные технологии

5.1 При чтении лекций используется мультимедийный компьютерный проектор. Используются активные и интерактивные методы обучения: проблемный подход, интеллектуальные разминки в начале лекции, коллективные обсуждения возникающих вопросов в ходе лекции, работа в парах.

При проведении лабораторных работ используются интерактивные методы обучения: работа в малых группах, кейс-метод, коллективное решение творческих задач. Используются индивидуальные задания для малых групп.

Для дополнительных вопросов, ответов и обсуждений используется форум в разделе дисциплины ЭИОС.

5.2 При выполнении лабораторных работ используется компьютерный класс, оснащенный персональными компьютерами, объединенными в локальную сеть с выходом в Internet. При изучении материалов лабораторного практикума используются информационно-образовательные ресурсы и программное обеспечение с сайта ЭИОС (moodle.pnzgu.ru) и файл-сервера кафедры САПР (диски M и T, сетевой узел cad-filer, IP: 172.16.72.254).

5.3 При организации самостоятельной работы студентов и, при необходимости, при проведении аудиторных занятий используются /могут быть использованы дистанционные образовательные технологии.

5.4. При организации самостоятельной работы также используются информационно-образовательные ресурсы и программное обеспечение с сайта ЭИОС (moodle.pnzgu.ru) и файл-сервера кафедры САПР (диски M и T, сетевой узел cad-filer, IP: 172.16.72.254).

5.5. Для промежуточного и итогового контроля используются средства электронного тестирования (система Electa и/или Moodle).

5.6. В целях реализации индивидуального подхода к обучению студентов, осуществляющих учебный процесс по собственной траектории в рамках индивидуального рабочего плана, изучение данной дисциплины базируется на следующих возможностях: обеспечение внеаудиторной работы со студентами (в том числе, в электронной образовательной среде с использованием соответствующего программного обеспечения), дистанционные формы обучения, возможности сторонних Интернет-ресурсов, индивидуальные консультации (очные или удаленные на базе форума в разделе дисциплины ЭИОС).

5.7. В качестве других видов контактной работы запланированы консультации при подготовке и проведении текущей и промежуточной аттестации.

6. Учебно-методическое обеспечение самостоятельной работы студентов. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины.

6.1. План самостоятельной работы студентов

№ нед.	Тема	Вид самостоятельной работы	Задание	Рекомендуемые источники	Кол-во часов
1	Тема 1.1 Цели,	Подготовка к	Изучение истории	- поисковые	5

	задачи, методы и области применения методов обработки больших данных	аудиторным занятиям.	происхождения науки о данных (Data Science) и самостоятельный поиск с последующим анализом сфер применения технологии Big Data.	системы Internet; - рекомендуемая литература	
2	Тема 1.2 Источники больших данных	Подготовка к аудиторным занятиям.	Выбор преимущественных источников данных по тематическому заданию преподавателя.	- поисковые системы Internet.	5
3-4	Тема 1.3 Сбор больших данных	Подготовка к аудиторным занятиям.	Самостоятельная работа над оценкой качества и объема данных. В качестве источников берутся результаты самостоятельной работы по теме 1.2	- поисковые системы Internet; - рекомендуемая литература; - материалы лекционного курса	5
5	Тема 2.1. Распределенные файловые системы.	Подготовка к аудиторным занятиям.	Самостоятельная работа по изучению и практическому закреплению навыков взаимодействия с файлами в файловой системе HDFS.	- рекомендуемая литература; - материалы лекционного курса	5
6	Тема 2.2 Технология вычислений MapReduce.	Подготовка к аудиторным занятиям.	Самостоятельная разработка MapReduce-программы подсчета слов для выбранного web-ресурса (исследуемый ресурс, количество распределителей и редукторов – по заданию преподавателя).	- заданный web-ресурс; - рекомендуемая литература; - материалы лекционного курса.	5
7	Тема 2.3 Алгоритмы обработки больших данных на	Подготовка к аудиторным занятиям.	Самостоятельное проектирование алгоритмов MapReduce для обработки файлов с	- рекомендуемая литература; - материалы лекционного	6,3

	основе MapReduce.		наборами целых чисел (по заданию преподавателя).	курса.	
8	Тема 2.4 Обобщения и модификации технологии MapReduce	Подготовка к аудиторным занятиям.	Самостоятельное практическое решение задачи по нахождению математического ожидания полного времени выполнения задания в вычислительном кластере (параметры вычислительного кластера задаются преподавателем).	- рекомендуемая литература; - материалы лекционного курса.	5
9	Тема 3.1 Задачи поиска похожих объектов	Подготовка к аудиторным занятиям.	Самостоятельная практическая тренировка по вычислению коэффициента Жаккара заданных наборов мультимножеств (количество и параметры мультимножеств задаются преподавателем).	- рекомендуемая литература; - материалы лекционного курса.	5
10	Тема 3.2 Задачи анализа потоков данных	Подготовка к аудиторным занятиям.	Самостоятельное построение потоковой модели данных для заданного гипотетического web-сайта (параметры web-сайта и их посещаемость задаются преподавателем).	- рекомендуемая литература; - материалы лекционного курса.	6
11-12	Тема 3.3 Задачи анализа ссылок. Алгоритмы PageRank и TrustRank	Подготовка к аудиторным занятиям.	Самостоятельное вычисление PageRank всех страниц по заданному графу переходов (структура графа переходов задается преподавателем).	- рекомендуемая литература; - материалы лекционного курса.	6

13	Тема 3.4 Задачи частых предметных наборов	Подготовка к аудиторным занятиям.	Самостоятельная оценка достоверности заданных ассоциативных правил для набора входных данных (ассоциативные правила и параметры входных данных задаются преподавателем).	- рекомендуемая литература; - материалы лекционного курса.	6
14-15	Тема 3.5 Задачи кластеризации больших данных	Подготовка к аудиторным занятиям.	Самостоятельное решение задачи по выполнению иерархической кластеризации заданного одномерного множества точек (множество точек и параметры кластеризации задаются преподавателем)	- рекомендуемая литература; - материалы лекционного курса.	6
16-17	Тема 3.6 Задачи представления и обработки больших графов	Подготовка к аудиторным занятиям.	Самостоятельное нахождение количества кратчайших путей заданного социального графа С помощью алгоритма Гирвана-Ньюмана (структура социального графа задается преподавателем)	- рекомендуемая литература; - материалы лекционного курса.	6

6.2. Методические указания по организации самостоятельной работы студентов

Планируются следующие виды самостоятельной работы (внеаудиторной) относятся:

- подготовка к выполнению лабораторных работ,
- оформление отчётов по лабораторным работам,
- изучение рекомендованной литературы для сдачи лабораторных работ,
- изучение рекомендованной литературы при подготовке к зачету.

6.3. Материалы для проведения текущего и промежуточного контроля знаний студентов

Контроль освоения компетенций

№ п\п	Вид контроля	Контролируемые темы (разделы)	Компетенции, компоненты которых контролируются
1	Текущий: собеседование при защите лаб. работ, проверка заданий для самостоятельной работы. Промежуточный: сдача экзамена.	Раздел 1. Введение в проблематику сбора и обработки больших данных (Big Data).	ПК-6
2	Текущий: собеседование при защите лаб. работ, проверка заданий для самостоятельной работы. Промежуточный: сдача экзамена.	Раздел 2. Программный стек для обработки больших данных.	ПК-6
3	Текущий: собеседование при защите лаб. работ, проверка заданий для самостоятельной работы. Промежуточный: сдача экзамена.	Раздел 3. Задачи обработки больших данных	ПК-6

Материалы для проведения текущего контроля знаний и промежуточной аттестации составляют отдельный документ – Фонд оценочных средств по дисциплине «Обработка больших данных».
Демонстрационные варианты оценочных средств для каждого вида контроля смотри moodle.pnzgu.ru.

7. Учебно-методическое и материально-техническое обеспечение дисциплины «Обработка больших данных»

а) учебная литература:

1. Лебедев А.С. Методы Big Data [Электронный ресурс]: Учебно-методическое пособие / Лебедев А.С., Магомедов Ш.Г. – М.: МИРЭА – Российский технологический университет, 2021.
2. Адлер Ю.П., Черных Е.А. Статистическое управление процессами. «Большие данные»: учебное пособие, М.: Издательство "МИСИС", 2016 – 52 с. ЭБС «Лань» https://e.lanbook.com/book/93652
3. Майер-Шенбергер В., Кукьер К. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим — М.: Издательство "Манн, Иванов и Фербер", 2014.— 240 с. — ЭБС «Лань» https://e.lanbook.com/book/62171

4. Форман Дж. Много цифр: Анализ больших данных при помощи Excel - М.: Издательство "Альпина Паблишер", 2016.— 461 с. — ЭБС «Лань» https://e.lanbook.com/book/87871
5. Эрик Редмонд, Джим. Р. Уилсон Семь баз данных за семь недель. Введение в современные базы данных и идеологию NoSQL – М.: ДМК Пресс, 2013 – 384 с. — ЭБС «Лань» https://e.lanbook.com/book/58690

б) Интернет-ресурсы
Материалы сайта https://machinelearning.ru

в) Программное обеспечение
Свободно-распространяемый набор утилит, библиотек и Фреймворк Hadoop (Apache Software Foundation)
Среда программирования Anacondas (свободно-распространяемое ПО)

г) Другое материально-техническое обеспечение
ПЭВМ, с 64-разрядной операционной системой Windows.

Сведения о переутверждении программы на очередной учебный год и регистрации изменений

Учебный год	Решение кафедры (№ протокола, дата)	Внесенные изменения	Подпись зав. кафедрой