

Министерство науки и высшего образования РФ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Национальный исследовательский университет «МЭИ»

Направление подготовки/специальность: 27.04.04 Управление в технических системах

Наименование образовательной программы: Интеллектуальные технологии управления в технических системах, обработка и анализ данных

Уровень образования: высшее образование - магистратура

Форма обучения: Очная

Рабочая программа дисциплины
ИНТЕЛЛЕКТУАЛЬНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ

Блок:	Блок 1 «Дисциплины (модули)»
Часть образовательной программы:	Часть, формируемая участниками образовательных отношений
№ дисциплины по учебному плану:	Б1.Ч.05
Трудоемкость в зачетных единицах:	3 семестр - 5;
Часов (всего) по учебному плану:	180 часов
Лекции	3 семестр - 32 часа;
Практические занятия	не предусмотрено учебным планом
Лабораторные работы	3 семестр - 16 часов;
Консультации	проводится в рамках часов аудиторных занятий
Самостоятельная работа	3 семестр - 131,7 часа;
в том числе на КП/КР	не предусмотрено учебным планом
Иная контактная работа	проводится в рамках часов аудиторных занятий
включая: Тестирование Лабораторная работа	
Промежуточная аттестация:	
Зачет с оценкой	3 семестр - 0,3 часа;

Москва 2024

ПРОГРАММУ СОСТАВИЛ:

Преподаватель

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Мохов А.С.
	Идентификатор	R55ae9104-MokhovAS-2434a28b

А.С. Мохов

СОГЛАСОВАНО:

Руководитель
образовательной программы

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Бобряков А.В.
	Идентификатор	R2c90f415-BobriakovAV-70dec1fa

А.В. Бобряков

Заведующий выпускающей
кафедрой

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Бобряков А.В.
	Идентификатор	R2c90f415-BobriakovAV-70dec1fa

А.В. Бобряков

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель освоения дисциплины: изучение стека современных технологий, применяемых для хранения, доступа и, обработки больших объемов данных в информационных системах, а также для проведения анализа этих данных с применением методов и алгоритмов машинного обучения

Задачи дисциплины

- освоение методов обработки больших объемов данных;
- приобретение навыков работы с современными технологиями, применяемыми в системах хранения, доступа и обработки больших объемов данных;
- приобретение навыков самостоятельно формировать, управлять и обслуживать системы обработки и анализа больших объемов данных.

Формируемые у обучающегося **компетенции** и запланированные **результаты обучения** по дисциплине, соотнесенные с **индикаторами достижения компетенций**:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Запланированные результаты обучения
РПК-2 Способен разрабатывать и применять информационные системы обработки и анализа данных для автоматизации процессов управления в сложных технических и организационно-технических системах	ИД-1рпк-2 Демонстрирует умение организовывать экспериментальные исследования и сбор экспертной информации, проводить анализ и предварительную обработку данных с применением автоматизированных информационных систем, выбирать обоснованные способы обеспечения защиты данных	знать: - методы снижения размерности выборки большого объема для проведения их анализа. уметь: - применять современные технологии для решения задачи обработки и анализа больших данных с целью увеличения качества.
РПК-2 Способен разрабатывать и применять информационные системы обработки и анализа данных для автоматизации процессов управления в сложных технических и организационно-технических системах	ИД-2рпк-2 Может разрабатывать информационные и информационно-аналитические системы автоматизации процессов управления в сложных технических и организационно-технических системах	знать: - современные подходы и инструменты работы с большим объемом данных. уметь: - применять модель MapReduce для проведения вычислений над большими объемами данных.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ ВО

Дисциплина относится к основной профессиональной образовательной программе Интеллектуальные технологии управления в технических системах, обработка и анализ данных (далее – ОПОП), направления подготовки 27.04.04 Управление в технических системах, уровень образования: высшее образование - магистратура.

Базируется на уровне высшего образования (бакалавриат, специалитет).

Результаты обучения, полученные при освоении дисциплины, необходимы при выполнении выпускной квалификационной работы.

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

3.1 Структура дисциплины

Общая трудоемкость дисциплины составляет 5 зачетных единиц, 180 часов.

№ п/п	Разделы/темы дисциплины/формы промежуточной аттестации	Всего часов на раздел	Семестр	Распределение трудоемкости раздела (в часах) по видам учебной работы										Содержание самостоятельной работы/ методические указания	
				Контактная работа							СР				
				Лек	Лаб	Пр	Консультация		ИКР		ПА	Работа в семестре	Подготовка к аттестации /контроль		
КПР	ГК	ИККП	ТК												
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	Подготовка данных в процессе Data Mining в условии больших данных	59	3	8	4	-	-	-	-	-	-	47	-	<p><u>Подготовка к текущему контролю:</u> Подготовка к тестированию. Для прохождения тестирования необходимо предварительно изучить тему тестирования, а также изучить вопросы, связанные с проблемами обработки больших данных и методами сэмпинга и снижения размерности по изученному в разделе "Подготовка данных в процессе Data Mining в условии больших данных" материалу.</p> <p><u>Подготовка к лабораторной работе:</u> Подготовка к лабораторной работе №1. Для выполнения заданий по лабораторной работе необходимо предварительно изучить тему и задачи выполнения лабораторной работы, а так же изучить вопросы, связанные с методами сэмпинга и снижения размерности по изученному в разделе "Подготовка данных в процессе Data Mining в условии больших данных" материалу.</p> <p><u>Изучение материалов литературных источников:</u> [1], 195-212 [2], 279-287, 339-344 [3], 310-341</p>	
1.1	Введение в проблему больших данных	21		2	2	-	-	-	-	-	-	17	-		
1.2	Обработка данных в условиях Big Data	38		6	2	-	-	-	-	-	-	-	30		-
2	Модель MapReduce обработки больших объемов данных	62.7		10	8	-	-	-	-	-	-	-	44.7		-

2.1	Технологии работы с большими данными	20.7	4	2	-	-	-	-	-	-	14.7	-	необходимо предварительно изучить тему и задачи выполнения лабораторной работы, а также изучить вопросы, связанные с технологией MapReduce и файловой системой HDFS по изученному в разделе "Модель MapReduce обработки больших объемов данных" материалу. <u>Изучение материалов литературных источников:</u> [1], 133-160
2.2	Файловая система HDFS	19	2	2	-	-	-	-	-	-	15	-	
2.3	Технология MapReduce	23	4	4	-	-	-	-	-	-	15	-	
3	Современные технологии обработки больших объемов данных	58	14	4	-	-	-	-	-	-	40	-	<u>Подготовка к текущему контролю:</u> Подготовка к тестированию. Для прохождения тестирования необходимо предварительно изучить тему тестирования, а также изучить вопросы, связанные с инструментами хранения и технологиями обработки больших данных по изученному в разделе "Современные технологии обработки больших объемов данных" материалу. <u>Подготовка к лабораторной работе:</u> Подготовка к лабораторной работе №3. Для выполнения заданий по лабораторной работе необходимо предварительно изучить тему и задачи выполнения лабораторной работы, а также изучить вопросы, связанные с технологиями обработки Big Data по изученному в разделе "Современные технологии обработки больших объемов данных" материалу. <u>Изучение материалов литературных источников:</u> [1], 145-170
3.1	Инструменты хранения и обработки больших коллекций	29	7	2	-	-	-	-	-	-	20	-	
3.2	Технологии обработки Big Data	29	7	2	-	-	-	-	-	-	20	-	
	Зачет с оценкой	0.3	-	-	-	-	-	-	-	0.3	-	-	
	Всего за семестр	180.0	32	16	-	-	-	-	-	0.3	131.7	-	
	Итого за семестр	180.0	32	16	-	-	-	-	-	0.3	131.7	-	

Примечание: Лек – лекции; Лаб – лабораторные работы; Пр – практические занятия; КПП – аудиторные консультации по курсовым проектам/работам; ИККП – индивидуальные консультации по курсовым проектам/работам; ГК- групповые консультации по разделам

дисциплины; СР – самостоятельная работа студента; ИКР – иная контактная работа; ТК – текущий контроль; ПА – промежуточная аттестация

3.2 Краткое содержание разделов

1. Подготовка данных в процессе Data Mining в условиях больших данных

1.1. Введение в проблему больших данных

Задачи Data Mining. Кто такие Data Scientist и Data Engineer. Проблематика Big Data.

1.2. Обработка данных в условиях Big Data

Методы обработки данных. Методы снижения размерности. Проклятие размерности. Методы выделения признаков: метод главных компонент, Анализ независимых компонент. Методы отбора признаков: жадные алгоритмы отбора признаков, теоретико-информационные методы, статистические методы. Сэмплинг в условиях больших данных – построение обучающих множеств (Active learning, semi-supervised sampling).

2. Модель MapReduce обработки больших объемов данных

2.1. Технологии работы с большими данными

Облачные вычисления. Идеи Big Data (BD). Масштабирование горизонтальное и вертикальное. Технология MapReduce (MR). Экосистема Hadoop.

2.2. Файловая система HDFS

Файловая система GFS, HDFS. Для каких задач подходит файловая система HDFS. Основные компоненты HDFS: NameNode, DataNode. Файлы и блоки в HDFS. HDFS Shell и основные команды. Особенности при работе с большими данными – репликация, запись данных.

2.3. Технология MapReduce

Технология MR – классическая схема. Управление командами – JobTracker, MapTracker. На каких машинах лучше запускать задания. Этапы MR: Map, Reduce, Combine, Partition. Программный интерфейс приложения Hadoop. Примеры алгоритмов работы с данными по технологии MR. Графы в MR.

3. Современные технологии обработки больших объемов данных

3.1. Инструменты хранения и обработки больших коллекций

Высокоуровневая платформа Pig для анализа и обработки больших коллекций. Hive - SQL-подобная СУБД платформы Hadoop. NoSQL технологии. Отличия от реляционных баз данных. Представители NoSQL-СУБД: HBase, Cassandra.

3.2. Технологии обработки Big Data

ZooKeeper – сервис для управления и координации распределенных систем. Технологии Big Data на языке Python.

3.3. Темы практических занятий

не предусмотрено

3.4. Темы лабораторных работ

1. Применение современных библиотек для анализа больших объемов данных;
2. Применение модели MapReduce для решения задач анализа текстовых данных;
3. Влияние методов сэмплинга и снижения размерности на качество решения задач

машинного обучения.

3.5 Консультации

Текущий контроль (ТК)

1. Консультации направлены на получение индивидуального задания для выполнения контрольных мероприятий по разделу "Подготовка данных в процессе Data Mining в условиях больших данных".
2. Консультации направлены на получение индивидуального задания для выполнения контрольных мероприятий по разделу "Модель MapReduce обработки больших объемов данных".
3. Консультации направлены на получение индивидуального задания для выполнения контрольных мероприятий по разделу "Современные технологии обработки больших объемов данных".

3.6 Тематика курсовых проектов/курсовых работ

Курсовой проект/ работа не предусмотрены

3.7. Соответствие разделов дисциплины и формируемых в них компетенций

Запланированные результаты обучения по дисциплине (в соответствии с разделом 1)	Коды индикаторов	Номер раздела дисциплины (в соответствии с п.3.1)			Оценочное средство (тип и наименование)
		1	2	3	
Знать:					
методы снижения размерности выборок большого объема для проведения их анализа	ИД-1 _{РПК-2}	+			Лабораторная работа/Защита лабораторной работы №1 Тестирование/Контрольная работа №1
современные подходы и инструменты работы с большим объемом данных	ИД-2 _{РПК-2}		+	+	Тестирование/Контрольная работа №2
Уметь:					
применять современные технологии для решения задачи обработки и анализа больших данных с целью увеличения качества	ИД-1 _{РПК-2}			+	Лабораторная работа/Защита лабораторной работы №3
применять модель MapReduce для проведения вычислений над большими объемами данных	ИД-2 _{РПК-2}		+		Лабораторная работа/Защита лабораторной работы №2

4. КОМПЕТЕНТНОСТНО-ОРИЕНТИРОВАННЫЕ ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ КОНТРОЛЯ ОСВОЕНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (ТЕКУЩИЙ КОНТРОЛЬ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ДИСЦИПЛИНЕ)

4.1. Текущий контроль успеваемости

3 семестр

Форма реализации: Компьютерное задание

1. Защита лабораторной работы №1 (Лабораторная работа)
2. Защита лабораторной работы №2 (Лабораторная работа)
3. Защита лабораторной работы №3 (Лабораторная работа)
4. Контрольная работа №1 (Тестирование)
5. Контрольная работа №2 (Тестирование)

Балльно-рейтинговая структура дисциплины является приложением А.

4.2 Промежуточная аттестация по дисциплине

Зачет с оценкой (Семестр №3)

Оценка определяется по совокупности результатов текущего контроля успеваемости в соответствии с Положением о балльно-рейтинговой системе для студентов НИУ «МЭИ».

В диплом выставляется оценка за 3 семестр.

Примечание: Оценочные материалы по дисциплине приведены в фонде оценочных материалов ОПОП.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1 Печатные и электронные издания:

1. Просто о больших данных : пер. с англ. / Д. Гурвиц, и др. – М. : Эксмо, 2015 . – 400 с. – (Библиотека Сбербанка ; Т.58) . - ISBN 978-5-699-85807-1 .;
2. Маннинг, К. Д. Введение в информационный поиск : пер. с англ. / К. Д. Маннинг, П. Рагхаван, Х. Шютце . – М. : Вильямс, 2011 . – 528 с. - ISBN 978-5-8459-1623-5 .;
3. Флах П.- "Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных", Издательство: "ДМК Пресс", Москва, 2015 - (400 с.) http://e.lanbook.com/books/element.php?pl1_id=69955.

5.2 Лицензионное и свободно распространяемое программное обеспечение:

1. Office / Российский пакет офисных программ;
2. Windows / Операционная система семейства Linux;
3. VirtualBox;
4. Apache Cassandra;
5. Apache Hadoop;
6. Apache Hbase;
7. Apache Pig;
8. Apache Spark;
9. Apache ZooKeeper;
10. Jupiter Notebook;
11. ОС CentOS.

5.3 Интернет-ресурсы, включая профессиональные базы данных и информационно-справочные системы:

1. ЭБС Лань - <https://e.lanbook.com/>
2. ЭБС "Университетская библиотека онлайн" - http://biblioclub.ru/index.php?page=main_ub_red
3. Научная электронная библиотека - <https://elibrary.ru/>
4. База данных Web of Science - <http://webofscience.com/>
5. База данных Scopus - <http://www.scopus.com>
6. Национальная электронная библиотека - <https://rusneb.ru/>
7. ЭБС "Консультант студента" - <http://www.studentlibrary.ru/>
8. Электронная библиотека МЭИ (ЭБ МЭИ) - <http://elib.mpei.ru/login.php>
9. Портал открытых данных Российской Федерации - <https://data.gov.ru>
10. База открытых данных Министерства труда и социальной защиты РФ - <https://rosmintrud.ru/opendata>
11. База открытых данных профессиональных стандартов Министерства труда и социальной защиты РФ - <http://profstandart.rosmintrud.ru/obshchiy-informatsionnyy-blok/natsionalnyy-reestr-professionalnykh-standartov/>
12. База открытых данных Министерства экономического развития РФ - <http://www.economy.gov.ru>
13. База открытых данных Росфинмониторинга - <http://www.fedsfm.ru/opendata>
14. Электронная открытая база данных "Polpred.com Обзор СМИ" - <https://www.polpred.com>
15. Информационно-справочная система «Кодекс/Техэксперт» - <Http://proinfosoft.ru;>
<http://docs.cntd.ru/>
16. Национальный портал онлайн обучения «Открытое образование» - <https://openedu.ru>
17. Официальный сайт Федерального агентства по техническому регулированию и метрологии - <http://protect.gost.ru/>
18. Открытая университетская информационная система «РОССИЯ» - <https://uisrussia.msu.ru>

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Тип помещения	Номер аудитории, наименование	Оснащение
Учебные аудитории для проведения лекционных занятий и текущего контроля	М-307, Учебная аудитория	стол преподавателя, стол учебный, стул, доска меловая, компьютерная сеть с выходом в Интернет, мультимедийный проектор, экран
Учебные аудитории для проведения лабораторных занятий	М-304а/1, Учебная лаборатория моделирования систем и анализа данных	стол преподавателя, стол компьютерный, стул, компьютерная сеть с выходом в Интернет, доска маркерная, компьютер персональный
Учебные аудитории для проведения промежуточной аттестации	М-307, Учебная аудитория	стол преподавателя, стол учебный, стул, доска меловая, компьютерная сеть с выходом в Интернет, мультимедийный проектор, экран
Помещения для самостоятельной работы	НТБ-201, Компьютерный читальный зал	стол компьютерный, стул, стол письменный, вешалка для одежды, компьютерная сеть с выходом в Интернет, компьютер персональный, принтер, кондиционер
Помещения для консультирования	М-305, Преподавательская каф. "УиИ"	кресло рабочее, стол, стул, шкаф для документов, шкаф для одежды,

		компьютерная сеть с выходом в Интернет, компьютер персональный
Помещения для хранения оборудования и учебного инвентаря	М-309, Кладовая	стол, стул, шкаф для хранения инвентаря
	М-301/1, Кладовая	стул

БАЛЛЬНО-РЕЙТИНГОВАЯ СТРУКТУРА ДИСЦИПЛИНЫ

Интеллектуальные информационные системы

(название дисциплины)

3 семестр

Перечень контрольных мероприятий текущего контроля успеваемости по дисциплине:

- КМ-1 Контрольная работа №1 (Тестирование)
 КМ-2 Защита лабораторной работы №1 (Лабораторная работа)
 КМ-3 Защита лабораторной работы №2 (Лабораторная работа)
 КМ-4 Контрольная работа №2 (Тестирование)
 КМ-5 Защита лабораторной работы №3 (Лабораторная работа)

Вид промежуточной аттестации – Зачет с оценкой.

Номер раздела	Раздел дисциплины	Индекс КМ:	КМ-1	КМ-2	КМ-3	КМ-4	КМ-5
		Неделя КМ:	3	6	12	13	15
1	Подготовка данных в процессе Data Mining в условии больших данных						
1.1	Введение в проблему больших данных		+	+			
1.2	Обработка данных в условиях Big Data		+	+			
2	Модель MapReduce обработки больших объемов данных						
2.1	Технологии работы с большими данными				+	+	
2.2	Файловая система HDFS				+	+	
2.3	Технология MapReduce				+	+	
3	Современные технологии обработки больших объемов данных						
3.1	Инструменты хранения и обработки больших коллекций					+	+
3.2	Технологии обработки Big Data					+	+
Вес КМ, %:			10	20	30	10	30