

МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»

УТВЕРЖДАЮ:

Директор ИУЦТ



С.П. Вакуленко

30 апреля 2020 г.

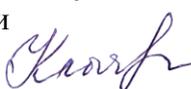
Кафедра «Вычислительные системы, сети и информационная
безопасность»

Автор Давыдовский Михаил Альбинович, к.т.н., доцент

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Информационная аналитика и обработка больших данных

Направление подготовки:	09.04.01 – Информатика и вычислительная техника
Магистерская программа:	Компьютерные сети и технологии
Квалификация выпускника:	Магистр
Форма обучения:	очная
Год начала подготовки	2020

Одобрено на заседании Учебно-методической комиссии института Протокол № 4 30 апреля 2020 г. Председатель учебно-методической комиссии  Н.А. Клычева	Одобрено на заседании кафедры Протокол № 15 27 апреля 2020 г. Заведующий кафедрой  Б.В. Желенков
---	--

Москва 2020 г.

1. ЦЕЛИ ОСВОЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ

Целью преподавания дисциплины «Информационная аналитика и обработка больших данных» является изучение современных методов и программных средств, используемых при анализе и обработке больших данных. В результате изучения дисциплины студенты должны научиться применять алгоритмы поиска в данных новых знаний, необходимых для принятия решений в различных сферах человеческой деятельности. Студенты должны изучить методы и средства, используемые для хранения и обработки больших баз данных. Основной целью изучения учебной дисциплины «Информационная аналитика и обработка больших данных» является формирование компетенций в области разработки моделей и анализа данных для задач профессиональной деятельности следующих типов:

- научно-исследовательский;
- проектный.

Дисциплина предназначена для получения знаний для решения следующих профессиональных задач (в соответствии с типами задач профессиональной деятельности):

научно-исследовательский:

участие в фундаментальных и прикладных исследованиях в области связи, информационных и коммуникационных технологий;

разработка планов, программ и методик проведения исследований объектов профессиональной деятельности;

проектный:

проектирование, разработка, модернизация средств вычислительной техники и информационных систем.

2. МЕСТО УЧЕБНОЙ ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Учебная дисциплина "Информационная аналитика и обработка больших данных" относится к блоку 1 "Дисциплины (модули)" и входит в его вариативную часть.

2.1. Наименования предшествующих дисциплин

2.2. Наименование последующих дисциплин

Результаты освоения дисциплины используются при изучении последующих учебных дисциплин:

2.2.1. Преддипломная практика

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ), СООТНЕСЕННЫЕ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

В результате освоения дисциплины студент должен:

№ п/п	Код и название компетенции	Ожидаемые результаты
1	ПКР-2 Применение перспективных методов исследования и решения профессиональных задач на основе знания мировых тенденций развития вычислительной техники и информационных технологий	ПКР-2.1 Знать методы исследования и решения профессиональных задач;мировые тенденции развития вычислительной техники;знать перспективные тенденции развития информационных технологий. ПКР-2.2 Уметь применять перспективные методы исследования для решения. ПКР-2.3 Владеть навыками применения перспективных методов исследования и решения профессиональных задач на основе знания мировых тенденций развития вычислительной техники и информационных технологий.

4. ОБЪЕМ ДИСЦИПЛИНЫ (МОДУЛЯ) В ЗАЧЕТНЫХ ЕДИНИЦАХ И АКАДЕМИЧЕСКИХ ЧАСАХ

4.1. Общая трудоемкость дисциплины составляет:

4 зачетные единицы (144 ак. ч.).

4.2. Распределение объема учебной дисциплины на контактную работу с преподавателем и самостоятельную работу обучающихся

Вид учебной работы	Количество часов	
	Всего по учебному плану	Семестр 2
Контактная работа	36	36,15
Аудиторные занятия (всего):	36	36
В том числе:		
лекции (Л)	18	18
лабораторные работы (ЛР)(лабораторный практикум) (ЛП)	18	18
Самостоятельная работа (всего)	72	72
Экзамен (при наличии)	36	36
ОБЩАЯ трудоемкость дисциплины, часы:	144	144
ОБЩАЯ трудоемкость дисциплины, зач.ед.:	4.0	4.0
Текущий контроль успеваемости (количество и вид текущего контроля)	ПК1, ПК2	ПК1, ПК2
Виды промежуточной аттестации (экзамен, зачет)	ЭК	ЭК

4.3. Содержание дисциплины (модуля), структурированное по темам (разделам)

№ п/п	Семестр	Тема (раздел) учебной дисциплины	Виды учебной деятельности в часах/ в том числе интерактивной форме						Формы текущего контроля успеваемости и промежуточной аттестации
			Л	ЛР	ПЗ/ТП	КСР	СР	Всего	
1	2	3	4	5	6	7	8	9	10
1	2	Раздел 1 Наука о данных (data science)	4				18	22	
2	2	Тема 1.1 Этапы процесса data science	2					2	
3	2	Тема 1.2 Машинное обучение	2					2	
4	2	Раздел 2 Большие данные	8	6			22	36	
5	2	Тема 2.1 Методы обработки больших объемов данных	4					4	
6	2	Тема 2.2 Распределенное хранение данных в системах NoSQL	4	6				10	ПК1, Защита л.р.
7	2	Раздел 3 Анализ больших данных	6	12			32	50	
8	2	Тема 3.1 Система анализа Spark	2	4				6	
9	2	Тема 3.2 Графовые базы данных	2	4				6	ПК2, Защита л.р.
10	2	Тема 3.3 Глубокий анализ текста	2	4				6	
11	2	Раздел 4 Итоговая аттестация						36	ЭК
12		Всего:	18	18			72	144	

4.4. Лабораторные работы / практические занятия

Практические занятия учебным планом не предусмотрены.

Лабораторные работы предусмотрены в объеме 18 ак. ч.

№ п/п	№ семестра	Тема (раздел) учебной дисциплины	Наименование занятий	Всего часов/ из них часов в интерактивной форме
1	2	3	4	5
1	2	РАЗДЕЛ 2 Большие данные Тема: Распределенное хранение данных в системах NoSQL	Лабораторная работа 1. Создание базы данных в системе Cassandra.	6
2	2	РАЗДЕЛ 3 Анализ больших данных Тема: Система анализа Spark	Лабораторная работа 2. Анализ данных в системе Spark.	4
3	2	РАЗДЕЛ 3 Анализ больших данных Тема: Графовые базы данных	Лабораторная работа 3. Анализ данных с использованием графа	4
4	2	РАЗДЕЛ 3 Анализ больших данных Тема: Глубокий анализ текста	Лабораторная работа 4. Классификация текстовых сообщений	4
ВСЕГО:				18/0

4.5. Примерная тематика курсовых проектов (работ)

Курсовой проект (работа) не предусмотрен учебным планом.

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Преподавание дисциплины «Информационная аналитика и обработка больших данных» осуществляется в форме лекций, лабораторных работ, самостоятельной работы студентов. Лекции проводятся в форме мультимедиа-лекций, на которых демонстрируются презентации. Студенты имеют возможность ознакомиться с материалами презентации до начала лекции.

Лабораторные занятия проводятся в компьютерном классе с установленным программным обеспечением, необходимым для разработки индивидуальных заданий. На лабораторных занятиях выполняются индивидуальные задания, демонстрируются готовые части выполненных заданий и отчета по заданию. Лабораторные работы проводятся с использованием интерактивных технологий. Разработка проектов по индивидуальным заданиям ведется с применением систем управления базами данных и добычи данных. Самостоятельная работа студента организована с использованием традиционных видов работы и интерактивных технологий. К традиционным видам работы относятся отработка лекционного материала и отработка отдельных тем по учебной литературе. К интерактивным (диалоговым) технологиям относится отработка отдельных тем с использованием электронных информационных ресурсов и разработка индивидуальных заданий в программных системах проведения анализа и хранения больших данных. Оценка полученных знаний, умений и навыков основана на модульно-рейтинговой технологии. Весь курс разбит на 3 раздела, представляющих собой логически завершённый объём учебной информации. Фонды оценочных средств освоенных компетенций включают как вопросы теоретического характера для оценки знаний, так и задания практического содержания (индивидуальные задания по лабораторным работам) для оценки умений и навыков. Теоретические знания проверяются на экзамене и в ходе проверки отчетов по выполненным лабораторным работам.

Проведении занятий по дисциплине (модулю) возможно с применением электронного обучения и дистанционных образовательных технологий, реализуемые с применением информационно-телекоммуникационных сетей при опосредованном (на расстоянии) взаимодействии обучающихся и педагогических работников.

В процессе проведения занятий с применением электронного обучения и дистанционных образовательных технологий применяются современные образовательные технологии, такие как (при необходимости):

- использование современных средств коммуникации;
- электронная форма обмена материалами;
- дистанционная форма групповых и индивидуальных консультаций;
- использование компьютерных технологий и программных продуктов, необходимых для сбора и систематизации информации, проведения требуемых программой расчетов и т.д.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

№ п/п	№ семестра	Тема (раздел) учебной дисциплины	Вид самостоятельной работы студента. Перечень учебно-методического обеспечения для самостоятельной работы	Всего часов
1	2	3	4	5
1	2	РАЗДЕЛ 1 Наука о данных (data science)	1. Изучение учебной литературы: [1, стр. 75-79], [2, стр. 134-141], [6, стр. 182-189]	18
2	2	РАЗДЕЛ 2 Большие данные	1. Изучение системы Cassandra. 2. Изучение учебной литературы: [3, стр. 131-166]	22
3	2	РАЗДЕЛ 3 Анализ больших данных	1. Изучение системы Spark. 2. Изучение учебной литературы: [4, стр. 159-163], [5, стр. 120-127].	32
ВСЕГО:				72

7. ПЕРЕЧЕНЬ ОСНОВНОЙ И ДОПОЛНИТЕЛЬНОЙ ЛИТЕРАТУРЫ, НЕОБХОДИМОЙ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

7.1. Основная литература

№ п/п	Наименование	Автор (ы)	Год и место издания Место доступа	Используется при изучении разделов, номера страниц
1	Большие данные. Методы интеллектуального анализа	Емельченков Е.П., Курицына Е.Н., Лагуткин Д.А.	Системы компьютерной математики и их приложения. 2013. № 14., https://elibrary.ru/item.asp?id=20588343 , 2013	1[75-79]
2	Обзор архитектур и методов машинного обучения для анализа больших данных	Аксютин Е.М., Белов Ю.С.	Электронный журнал: наука, техника и образование. 2016. № 1 (5). https://elibrary.ru/item.asp?id=26747505 , 2016	1 [134-141]
3	Исследование и развитие метода декомпозиции для анализа пространственных данных	Золотов В.А., Семенов В.А.	Труды Института системного программирования РАН. 2013. Т. 25., https://elibrary.ru/item.asp?id=21134920 , 2013	2[131-166]
4	Большие данные. Проблемы анализа	Цветкова Ю.С., Асмандиярова П.С., Лебедева М.Ю.	Вопросы науки. 2015. Т. 3. https://elibrary.ru/item.asp?id=23331377 , 2015	3[159-163]
5	Технологии для обработки и анализа больших данных	Черепков Е.А., Рыбкин С.В.	Электронный журнал: наука, техника и образование. 2016. № 4 (9). https://elibrary.ru/item.asp?id=27664314 , 2016	3[120-127]

7.2. Дополнительная литература

№ п/п	Наименование	Автор (ы)	Год и место издания Место доступа	Используется при изучении разделов, номера страниц
6	Применение data science в геймификации бизнес и HR процессов организации	Шиховцова А.И., Виниченко М.В.	2016, Новое поколение, Издательство: Макушкин Сергей Анатольевич, ISSN: 2500-3569, https://elibrary.ru/item.asp?id=26025271 , 2016	1[182-189]

8. ПЕРЕЧЕНЬ РЕСУРСОВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ "ИНТЕРНЕТ", НЕОБХОДИМЫЕ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

- <http://library.miit.ru/> - электронно-библиотечная система Научно-технической библиотеки МИИТ
- <https://ru.wikipedia.org> - Википедия
- www.citforum.ru – материалы по информационным технологиям
- <http://cassandra.apache.org/> - сайт системы Cassandra

- <http://spark.apache.org/> - сайт системы Spark
- <https://rapidminer.com/> - сайт системы Rapid Miner

9. ПЕРЕЧЕНЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ, ИСПОЛЬЗУЕМЫХ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

Используемые информационные технологии:

- объектно-ориентированное проектирование;
- структурный подход к проектированию ИС;
- сервис-ориентированная архитектура.

Поисковые системы:

- Google;
- Яндекс.

Для выполнения курсовой работы требуется следующее программное обеспечение:

- Язык программирования Java,
- Среда Eclipse (лицензия – Eclipse Public License (EPL) Open-Source License)

При организации обучения по дисциплине (модулю) с применением электронного обучения и дистанционных образовательных технологий необходим доступ каждого студента к информационным ресурсам – библиотечному фонду Университета, сетевым ресурсам и информационно-телекоммуникационной сети «Интернет».

В случае проведения занятий с применением электронного обучения и дистанционных образовательных технологий может понадобиться наличие следующего программного обеспечения (или их аналогов): ОС Windows, Microsoft Office, Интернет-браузер, Microsoft Teams и т.д.

В образовательном процессе, при проведении занятий с применением электронного обучения и дистанционных образовательных технологий, могут применяться следующие средства коммуникаций: ЭИОС РУТ(МИИТ), Microsoft Teams, электронная почта, скайп, Zoom, WhatsApp и т.п.

10. ОПИСАНИЕ МАТЕРИАЛЬНО ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

Учебная аудитория для проведения занятий лекционного типа, групповых и индивидуальных консультаций №1329.

Проектор для вывода изображения на экран для студентов, акустическая система, место для преподавателя оснащенное компьютером (CPU Corei3, 8GBRAM, 1Tb HDD, GeForceGTSeries). Аудитория подключена к интернету МИИТ.

Учебная аудитория для проведения занятий лекционного типа, практических занятий, лабораторных работ №1330.

Аудиовизуальное оборудование для аудитории, АРМ управляющий, проектор, экран, 25 персональных компьютеров, 25 мониторов, 1 принтер, доска учебная. Аудитория подключена к интернету МИИТ.

В случае проведения занятий с применением электронного обучения и дистанционных образовательных технологий необходимо наличие компьютерной техники, для организации коллективных и индивидуальных форм общения педагогических работников со студентами, посредством используемых средств коммуникации.

Допускается замена оборудования его виртуальными аналогами.

11. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

1. В лекционном курсе рассматриваются основные вопросы по данной дисциплине. Дополнительные вопросы, необходимые студентам при выполнении своих индивидуальных заданий, изучаются студентами самостоятельно и контролируются преподавателем.
2. Задания по всем лабораторным работам выдаются студентам в начале семестра, чтобы студенты имели возможность самостоятельно изучить дополнительные теоретические сведения, необходимые им при выполнении индивидуальных заданий, и спланировать график выполнения заданий с учетом их специфики.
3. Прежде чем приступить к выполнению конкретного задания студент должен изучить:
 - материалы лекций по теме задания;
 - дополнительные материалы, относящиеся к специфике индивидуального задания;
 - программные средства, используемые при выполнении задания.
4. Выполнение индивидуальных заданий и их сдача осуществляется по определенному графику и учитывается при периодической аттестации студентов.
5. Лекции по дисциплине, подготовленные в электронном виде, рекомендуется выдавать студентам в начале семестра с целью лучшего освоения материала и возможности досрочного изучения вопросов, необходимых для выполнения индивидуальных заданий.
6. Индивидуальные задания, требующие разработки сложных программных систем, могут выдаваться на группу студентов, но при этом необходимо контролировать знание каждым студентом всего задания в целом.
7. знание каждым студентом всего задания в целом.
8. Для полноценного освоения дисциплины необходимо:
 - посещение лекций и практических занятий;
 - изучение лекционного материала;
 - освоение теоретического материала, вынесенного на самостоятельное изучение, по предложенным источникам (литература, интернет-ресурсы);
 - изучение программного обеспечения, необходимого, для выполнения индивидуальных заданий;
 - консультации с преподавателем в ходе выполнения индивидуальных заданий и обсуждение промежуточных результатов выполнения индивидуальных заданий;
 - своевременное выполнение индивидуальных заданий;
 - своевременное предоставление отчетов по индивидуальным заданиям и защита выполненных работ.