

МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»
(РУТ (МИИТ))



Рабочая программа дисциплины (модуля),
как компонент образовательной программы
высшего образования - программы магистратуры
по направлению подготовки
09.04.01 Информатика и вычислительная техника,
утвержденной первым проректором РУТ (МИИТ)
Тимониным В.С.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Информационная аналитика и обработка больших данных

Направление подготовки: 09.04.01 Информатика и вычислительная техника

Направленность (профиль): Компьютерные сети и технологии

Форма обучения: Очная

Рабочая программа дисциплины (модуля) в виде
электронного документа выгружена из единой
корпоративной информационной системы управления
университетом и соответствует оригиналу

Простая электронная подпись, выданная РУТ (МИИТ)
ID подписи: 4196
Подписал: заведующий кафедрой Желенков Борис
Владимирович
Дата: 20.10.2022

1. Общие сведения о дисциплине (модуле).

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. Планируемые результаты обучения по дисциплине (модулю).

Перечень формируемых результатов освоения образовательной программы (компетенций) в результате обучения по дисциплине (модулю):

ОПК-8 - Способен осуществлять эффективное управление разработкой программных средств и проектов.;

ПК-5 - Владение методами и алгоритмами решения задач обработки данных;

ПК-7 - Применение перспективных методов исследования и решения профессиональных задач на основе знания мировых тенденций развития вычислительной техники и информационных технологий.

Обучение по дисциплине (модулю) предполагает, что по его результатам обучающийся будет:

Знать:

-основные понятия больших баз данных и информационной аналитики.

Уметь:

-разрабатывать схемы и запросы для NoSQL баз данных;
- строить модели анализа данных.

Владеть:

Владеть:

-языками запросов к большим базам данных;
- языками программирования для решения аналитических задач, системами аналитики данных.

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 4 з.е. (144 академических часа(ов)).

3.2. Объем дисциплины (модуля) в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Тип учебных занятий	Количество часов	
	Всего	Сем. №2
Контактная работа при проведении учебных занятий (всего):	48	48
В том числе:		
Занятия лекционного типа	32	32
Занятия семинарского типа	16	16

3.3. Объем дисциплины (модуля) в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации составляет 96 академических часа (ов).

3.4. При обучении по индивидуальному учебному плану, в том числе при ускоренном обучении, объем дисциплины (модуля) может быть реализован полностью в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации.

4. Содержание дисциплины (модуля).

4.1. Занятия лекционного типа.

№ п/п	Тематика лекционных занятий / краткое содержание
1	<p>1. Большие данные, анализ данных и наука о данных Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - основные понятия - источники больших данных, - анализ данных, - типы анализируемых данных, - процесс обработки данных. <p>2. Система управления базами данных Cassandra Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - установка системы, - работа с системой из командной строки, - модель данных, - пространство ключей, - таблицы и материализованные представления, - операции обновления данных, - выполнение запросов. <p>3. Распределение данных в системе Cassandra Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - параллельные архитектуры, - масштабирование баз данных, - распределение данных в Cassandra, - теорема CAP, - обнаружение отказов, - репликация данных. <p>4. Установка систем Spyder и Spark Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - установка Spyder, - установка pip, - установка Spark. <p>5. Система анализа Spark Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - использование Spark в Spyder,

№ п/п	Тематика лекционных занятий / краткое содержание
	<ul style="list-style-type: none"> - наборы RDD, - выполнение программы, - Spark и Cassandra, - работа с набором данных в Spyder- <p>6. Spark SQL</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - выполнение запросов, - соединение таблиц, - функции агрегирования. <p>7. Технология MapReduce</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - процесс вычислений, - функция map, - функция reduce, - умножение матрицы на вектор и на матрицу, - реляционная алгебра.- <p>8. Поиск похожих документов</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - сходство множеств, - представление документа в виде множеств, - шинглы, - матричное представление множеств, - минхэш-сигнатуры, - алгоритм определения похожих документов, - метрики. <p>9. Линейная регрессия</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - машинное обучение, - оценка модели, - библиотека функций машинного обучения, - линейная регрессия. - обучающий и тестовый наборы, - предсказание значений.

4.2. Занятия семинарского типа.

Лабораторные работы

№ п/п	Наименование лабораторных работ / краткое содержание
1	<p>1. Разработка базы данных в системе Cassandra. В результате выполнения лабораторной работы студент получает навык проектирования базы данных и выполнения запросов в NoSQL си стеме управления базами данных Cassandra</p> <p>2. Анализ данных в системе Spark. Выполнение запросов, использующих функции агрегирования В результате выполнения лабораторной работы студент получает навык выполнения запросов в</p>

№ п/п	Наименование лабораторных работ / краткое содержание
	<p>системе Spark к данным, хранящимся в базе данных системы Cassandra</p> <p>3. Анализ данных в системе Spark. Выявление похожих объектов В результате выполнения лабораторной работы студент получает навык сравнения документов в системе Spark и выявления похожих документов</p> <p>4 Анализ данных в системе Spark. Линейная регрессия В результате выполнения лабораторной работы студент получает навык использования алгоритма линейной регрессии для прогнозирования значений данных</p>

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	<p>1. Изучение документации по системам Cassandra, Spyder, Spark</p> <p>2. Анализ и дополнительная проработка лекционного материала</p> <p>3. Подготовка к лабораторным работам</p> <p>4. Изучение учебной литературы из приведенных источников</p>
2	Подготовка к промежуточной аттестации.
3	Подготовка к текущему контролю.

5. Перечень изданий, которые рекомендуется использовать при освоении дисциплины (модуля).

№ п/п	Библиографическое описание	Место доступа
1	Юре Лесковец, Ананд Раджараман, Джеффри Д. Ульман. Анализ больших наборов данных. - М.: ДМК Пресс, 2016. - 498 с.	https://e.lanbook.com/book/93571 (дата обращения: 03.10.2022)
2	Емельченков Е.П., Курицына Е.Н., Лагуткин Д.А. Большие данные. Методы интеллектуального анализа. Системы компьютерной математики и их приложения. 2013. № 14	https://elibrary.ru/item.asp?id=20588343 (дата обращения: 03.10.2022)
3	Аксютин Е.М., Белов Ю.С. Обзор архитектур и методов машинного обучения для анализа больших данных. Электронный журнал: наука, техника и образование. 2016. № 1	https://elibrary.ru/item.asp?id=26747505 (дата обращения: 03.10.2022)
4	Цветкова Ю.С., Асмандиярова П.С., Лебедева М.Ю. Большие данные. Проблемы анализа. Вопросы науки. 2015. Т. 3	https://elibrary.ru/item.asp?id=23331377 (дата обращения: 03.10.2022)
5	Черепков Е.А., Рыбкин С.В. Технологии для обработки и анализа больших данных. Электронный журнал: наука, техника и образование. 2016. № 4	https://elibrary.ru/item.asp?id=27664314 (дата обращения: 03.10.2022)

6. Перечень современных профессиональных баз данных и информационных справочных систем, которые могут использоваться при освоении дисциплины (модуля).

- Научная электронная библиотека (<http://elibrary.ru>)
- Электронно-библиотечная система Научно-технической библиотеки МИИТ (<http://library.miit.ru>)
- Википедия (<https://ru.wikipedia.org>)
- Материалы по информационным технологиям (www.citforum.ru)
- Сайт системы Cassandra (<http://cassandra.apache.org>)
- Сайт системы Spark (<http://spark.apache.org>)

7. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства, необходимого для освоения дисциплины (модуля).

- Язык программирования Python,
- Система баз данных Apache Cassandra (лицензия – свободно-распространяемое ПО с открытым кодом)
- Система анализа данных Spark (лицензия – свободно-распространяемое ПО с открытым кодом)

8. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю).

- Учебная аудитория для проведения занятий лекционного типа, групповых и индивидуальных консультаций .

Проектор для вывода изображения на экран для студентов, акустическая система, место для преподавателя оснащенное компьютером). Аудитория подключена к интернету МИИТ.

- Учебная аудитория для проведения занятий лекционного типа, практических занятий, лабораторных работ.

Аудиовизуальное оборудование для аудитории, АРМ управляющий, проектор, экран, персональные компьютеры ,мониторы, 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.

9. Форма промежуточной аттестации:

Экзамен во 2 семестре.

10. Оценочные материалы.

Оценочные материалы, применяемые при проведении промежуточной аттестации, разрабатываются в соответствии с локальным нормативным актом РУТ (МИИТ).

Авторы:

доцент, доцент, к.н. кафедры
«Вычислительные системы, сети и
информационная безопасность»

М.А. Давыдовский

Согласовано:

Заведующий кафедрой ВССиИБ

Б.В. Желенков

Председатель учебно-методической
комиссии

Н.А.Клычева