

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Уральский федеральный университет имени первого Президента России Б.Н. Ельцина»

УТВЕРЖДАЮ
Директор по образовательной
деятельности

_____ С.Т. Князев
«__» _____

РАБОЧАЯ ПРОГРАММА МОДУЛЯ

Код модуля	Модуль
1161481	Обработка больших объемов данных

Екатеринбург

Перечень сведений о рабочей программе модуля	Учетные данные
Образовательная программа 1. Управление техносферными рисками	Код ОП 1. 20.04.01/33.01
Направление подготовки 1. Техносферная безопасность	Код направления и уровня подготовки 1. 20.04.01

Программа модуля составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Комоцкий Евгений Игоревич	без ученой степени, без ученого звания	Ассистент	Базовая кафедра "Аналитика больших данных и методы видеопроцессинга"
2	Федотов Олег Васильевич	без ученой степени, без ученого звания	Старший преподаватель	интеллектуальных информационных технологий

Согласовано:

Управление образовательных программ

Р.Х. Токарева

1. ОБЩАЯ ХАРАКТЕРИСТИКА МОДУЛЯ Обработка больших объемов данных

1.1. Аннотация содержания модуля

Модуль “Обработка больших объемов данных” позволяет студентам изучить комплексы инструментов бизнес-аналитики для обработки, преобразования, анализа и визуализации данных. Изучение дисциплины модуля позволит студентам овладеть необходимыми знаниями и умениями для практического освоения методов создания и разработки систем управления базами данных.

1.2. Структура и объем модуля

Таблица 1

№ п/п	Перечень дисциплин модуля в последовательности их освоения	Объем дисциплин модуля и всего модуля в зачетных единицах
1	Обработка больших объемов данных	3
ИТОГО по модулю:		3

1.3. Последовательность освоения модуля в образовательной программе

Пререквизиты модуля	Не предусмотрены
Постреквизиты и кореквизиты модуля	Не предусмотрены

1.4. Распределение компетенций по дисциплинам модуля, планируемые результаты обучения (индикаторы) по модулю

Таблица 2

Перечень дисциплин модуля	Код и наименование компетенции	Планируемые результаты обучения (индикаторы)
1	2	3
Обработка больших объемов данных	ПК-6 - Способен использовать цифровые решения для работы с потоками ESG-данных и составления отчетности организации	З-2 - Описывать современные техники и методики сбора данных для решения управленческих и исследовательских задач З-5 - Описать технологии хранения и обработки больших данных в организации: базы данных, хранилища данных, распределенную и параллельную обработку данных, вычисления в оперативной памяти У-2 - Выбирать оптимальные методы обработки и анализа данных, в том числе используя интеллектуальные

		<p>информационно-аналитические системы при решении управленческих и исследовательских задач</p> <p>У-5 - Анализировать данные с использованием средств языка программирования Python и объяснить полученные результаты</p> <p>У-6 - Выбирать оптимальные современные методы и инструментальные средства анализа больших данных, методы интерпретации и визуализации больших данных</p> <p>П-2 - Создавать отчеты, рефераты, статьи, оформленные в соответствии с предъявляемыми требованиями с использованием интеллектуальных информационно-аналитических систем</p>
--	--	---

1.5. Форма обучения

Обучение по дисциплинам модуля может осуществляться в очной формах.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Обработка больших объемов данных

Рабочая программа дисциплины составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Комоцкий Евгений Игоревич	без ученой степени, без ученого звания	Ассистент	Базовая кафедра "Аналитика больших данных и методы видеоанализа"
2	Федотов Олег Васильевич	без ученой степени, без ученого звания	Старший преподаватель	интеллектуальных информационных технологий

Рекомендовано учебно-методическим советом института Фундаментального образования

Протокол № 5 от 25.05.2023 г.

1. СОДЕРЖАНИЕ И ОСОБЕННОСТИ РЕАЛИЗАЦИИ ДИСЦИПЛИНЫ

Авторы:

- **Комоцкий Евгений Игоревич, Ассистент, Базовая кафедра ”Аналитика больших данных и методы видеоанализа”**
- **Федотов Олег Васильевич, Старший преподаватель, интеллектуальных информационных технологий**

1.1. Технологии реализации, используемые при изучении дисциплины модуля

- Традиционная (репродуктивная) технология
- Разноуровневое (дифференцированное) обучение
 - Базовый уровень

**Базовый I уровень – сохраняет логику самой науки и позволяет получить упрощенное, но верное и полное представление о предмете дисциплины, требует знание системы понятий, умение решать проблемные ситуации. Освоение данного уровня результатов обучения должно обеспечить формирование запланированных компетенций и позволит обучающемуся на минимальном уровне самостоятельности и ответственности выполнять задания;*

Продвинутый II уровень – углубляет и обогащает базовый уровень как по содержанию, так и по глубине проработки материала дисциплины. Это происходит за счет включения дополнительной информации. Данный уровень требует умения решать проблемы в рамках курса и смежных курсов посредством самостоятельной постановки цели и выбора программы действий. Освоение данного уровня результатов обучения позволит обучающемуся повысить уровень самостоятельности и ответственности до творческого применения знаний и умений.

1.2. Содержание дисциплины

Таблица 1.1

Код раздела, темы	Раздел, тема дисциплины*	Содержание
P1	Большие данные и методы их обработки	Определение больших данных. Источники больших данных. Задачи, для решения которых требуются большие данные. Особенности технологий обработки больших данных.
P2	Жизненный цикл приложений обработки больших объемов данных	Постановка задачи анализа данных. Подготовка набора данных. Очистка данных. Выбор алгоритма анализа данных. Оценка качества работы алгоритма анализа данных. Принятие решений на основе данных. Разработка приложения обработки больших данных.
P3	Распределенные системы хранения больших данных	Распределенная файловая система HDFS. Архитектура файловой системы HDFS. Обеспечение надежности хранения данных и производительности. Импорт/экспорт данных в распределенную файловую систему.
P4	Инструменты распределенной обработки больших объемов данных	Экосистема Apache Hadoop. Технология MapReduce. Кластер Apache Hadoop. Запуск задач MapReduce в Hadoop. Менеджер ресурсов YARN. Интеграция MapReduce и HDFS.
P5	Анализ больших данных с помощью Apache Spark	Алгоритмы обработки больших данных с помощью Apache Spark. Система обработки больших данных Apache Spark. Интеграция Apache Spark с экосистемой Hadoop. Отличия Spark от Map/Reduce. Трансформации Spark. Действия Spark. Resilient Distributed Datasets (RDD).

		<p>Разработка программ Spark на Python с помощью PySpark. Командная строка Spark. Разработка программ Spark на Python с помощью PySpark. Подключение к кластеру Hadoop из PySpark. Запуск задач PySpark в кластере Hadoop. Анализ выходные данных результатов работы программ PySpark.</p> <p>Использование DataFrame API в Apache Spark. Spark DataFrame API в Python. Отличия от операций с RDD. Загрузка данных в DataFrame, фильтрация и обработка данных. Запись данных в файлы формата parquet. Оптимизация производительности работы программ PySpark с DataFrame.</p>
P6	Анализ больших данных с помощью SQL	<p>P6 Анализ больших данных с помощью SQL Основы SQL. Язык запросов SQL. Оператор выбора данных SELECT. Объединение данных из нескольких таблиц. Функции SQL.</p> <p>Представление больших данных в виде таблиц. Представление данных в распределенной файловой системе HDFS с помощью таблиц. Таблицы в Apache Hive. Таблицы для SQL в Apache Spark.</p> <p>Анализ больших данных с помощью SQL. SQL в Apache Hive. Командная строка Hive. Запуск запросов SQL в Hive. SQL в Apache Spark. Запуск запросов SQL в Apache Spark. Сравнение возможностей SQL и DataFrame API в Apache Spark.</p> <p>Импорт/экспорт табличных данных в базы данных. Импорт/экспорт данных из реляционных СУБД с помощью Apache Sqoop. Работа с реляционными СУБД средствами Apache Spark SQL.</p>
P7	Построение отчетов по результатам анализа больших данных	<p>Построение интерактивных отчетов (dashboards). Способы визуализации данных.</p> <p>Сравнительный анализ современных систем Business Intelligence, Tableau; QlikView; Prognoz Platform; Power BI.</p>

1.3. Направление, виды воспитательной деятельности и используемые технологии

Направления воспитательной деятельности сопрягаются со всеми результатами обучения компетенций по образовательной программе, их освоение обеспечивается содержанием всех дисциплин модулей.

1.4. Программа дисциплины реализуется на государственном языке Российской Федерации .

2. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Обработка больших объемов данных

Электронные ресурсы (издания)

1. Комлева, Н. В.; Структуры и алгоритмы компьютерной обработки данных : учебное пособие.; Московский государственный университет экономики, статистики и информатики, Москва; 2004; <https://biblioclub.ru/index.php?page=book&id=93226> (Электронное издание)

2. Мейер, Б., Б.; Инструменты, алгоритмы и структуры данных; Национальный Открытый Университет «ИНТУИТ», Москва; 2016; <https://biblioclub.ru/index.php?page=book&id=429033> (Электронное издание)

3. ; Интеллектуальные информационные системы и технологии : учебное пособие.; Тамбовский государственный технический университет (ТГТУ), Тамбов; 2013; <https://biblioclub.ru/index.php?page=book&id=277713> (Электронное издание)

Печатные издания

1. Мазуров, В. Д.; Метод комитетов в задачах оптимизации и классификации; Наука, Москва; 1990 (6 экз.)
2. Редько, В. Г., Малинецкий, Г. Г.; Эволюция, нейронные сети, интеллект. Модели и концепции эволюционной кибернетики; URSS, Москва; 2015 (2 экз.)
3. Редько, В. Г.; Эволюция, нейронные сети, интеллект: Модели и концепции эволюционной кибернетики; [КомКнига], Москва; 2005 (1 экз.)
4. Котов, В. М.; Алгоритмы и структуры данных : [учебное пособие для вузов].; БГУ, Минск; 2011 (1 экз.)

Профессиональные базы данных, информационно-справочные системы

1. Электронный научный архив УрФУ <https://elar.urfu.ru/>
2. Зональная научная библиотека (УрФУ) - <http://lib.urfu.ru/>
3. Портал информационно-образовательных ресурсов УрФУ <https://study.urfu.ru/>
4. Электронно-библиотечная система «Лань» – <https://e.lanbook.com/>
5. Университетская библиотека ONLINE – <https://biblioclub.ru/>
6. Электронно-библиотечная система "Библиокомплектатор" (IPRbooks)
7. <http://www.bibliocomplectator.ru/available>
8. Электронные информационные ресурсы Российской государственной библиотеки <https://www.rsl.ru/>
9. Научная электронная библиотека «КиберЛенинка» <https://cyberleninka.ru/>

Материалы для лиц с ОВЗ

Весь контент ЭБС представлен в виде файлов специального формата для воспроизведения синтезатором речи, а также в тестовом виде, пригодном для прочтения с использованием экранной лупы и настройкой контрастности.

Базы данных, информационно-справочные и поисковые системы

1. Applied Science & Technology Source. EBSCO publishing <http://search.ebscohost.com>
2. Wiley Online Library <http://onlinelibrary.wiley.com/>
3. Гугл Академия <https://scholar.google.ru/>
4. Academic Search Ultimate EBSCO publishing – <http://search.ebscohost.com>
5. eBook Collections Springer Nature – <https://link.springer.com/>

3. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Обработка больших объемов данных

Сведения об оснащённости дисциплины специализированным и лабораторным оборудованием и программным обеспечением

Таблица 3.1

№ п/п	Виды занятий	Оснащённость специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения
1	Лекции	Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов Рабочее место преподавателя Доска аудиторная Периферийное устройство Подключение к сети Интернет	Office Professional 2003 Win32 Russian CD-ROM Бесплатно распространяемое программное обеспечение: Python – https://www.python.org/ Веб - среда разработки для языка программирования Python: google colab - https://colab.research.google.com/ Anaconda solution - https://www.anaconda.com/ Система распределенной обработки данных Apache Spark – https://spark.apache.org/ Кластер HDInsight от Microsoft Azure
2	Практические занятия	Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов Рабочее место преподавателя Доска аудиторная Периферийное устройство Персональные компьютеры по количеству обучающихся Подключение к сети Интернет	Office Professional 2003 Win32 Russian CD-ROM Бесплатно распространяемое программное обеспечение: Python – https://www.python.org/ Веб - среда разработки для языка программирования Python: google colab - https://colab.research.google.com/

			<p>Anaconda solution - https://www.anaconda.com/ Система распределенной обработки данных Apache Spark – https://spark.apache.org/ Кластер HDInsight от Microsoft Azure</p>
3	Консультации	<p>Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов</p> <p>Рабочее место преподавателя</p> <p>Подключение к сети Интернет</p>	Не требуется
4	Текущий контроль и промежуточная аттестация	<p>Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов</p> <p>Рабочее место преподавателя</p> <p>Персональные компьютеры по количеству обучающихся</p>	<p>Office Professional 2003 Win32 Russian CD-ROM</p> <p>Бесплатно</p> <p>распространяемое программное обеспечение:</p> <p>Python – https://www.python.org/</p> <p>Веб - среда разработки для языка программирования Python: google colab - https://colab.research.google.com/</p> <p>Anaconda solution - https://www.anaconda.com/ Система распределенной обработки данных Apache Spark – https://spark.apache.org/ Кластер HDInsight от Microsoft Azure</p>
5	Самостоятельная работа студентов	<p>Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов</p> <p>Подключение к сети Интернет</p>	<p>Office Professional 2003 Win32 Russian CD-ROM</p> <p>Бесплатно</p> <p>распространяемое программное обеспечение:</p>

			<p>Python – https://www.python.org/ Веб - среда разработки для языка программирования Python: google colab - https://colab.research.google.com/ Anaconda solution - https://www.anaconda.com/ Система распределенной обработки данных Apache Spark – https://spark.apache.org/ Кластер HDInsight от Microsoft Azure</p>
--	--	--	--