

Аннотация рабочей программы дисциплины

К.М.01.01 «Обработка больших данных»

Направление подготовки/специальность

02.03.02 Фундаментальная информатика и информационные технологии

Курс 3 Семестр 6 Количество з.е. 4 (144 час., из них – 68,3 часов аудиторной нагрузки: лекционных 32 ч., практических 32 ч., иной контактной работы 0,3 ч 31 часов самостоятельной работы), 44,7 – контроль - экзамен.

Цель дисциплины: формирование у студентов профессиональной компетенции в области разработки и использования систем обработки и анализа больших массивов данных. Данная цель соотносится с целью образовательной программой в частности с технологий разработки специализированных программных систем, отвечающих за обработку больших данных.

Задачи дисциплины:

В результате освоения дисциплины студент должен овладеть компетенциями ОПК-2; ОПК-3; ПК-5.

Основные задачи освоения дисциплины:

Студент должен **знать** методы анализа и хранения больших объемов данных, этапы жизненного цикла обработки больших данных, языки, наиболее приспособленные для обработки и аналитики больших данных, способы организации хранения и доступа к большим данным; **уметь** выполнять элементы анализа данных и интерпретировать результаты, различать характеристики SQL и NoSql БД, формулировать алгоритмы в парадигме MapReduce, выбрать подходящий инструмент анализа больших данных, выбрать подходящую технологию хранения больших данных.; **владеть** математическими методами анализа данных, языками и компьютерными методами обработки.

Место дисциплины (модуля) в структуре образовательной программы

Курс «Обработка больших данных» относится к части блока Б1 основных дисциплин учебного плана.

Для изучения дисциплины студент должен владеть знаниями, умениями и навыками по дисциплинам:

Дискретная математика, Алгебраические структуры, Основы программирования, Алгоритмы вычислительной математики, Конструирование алгоритмов и структур данных, Теория алгоритмов и вычислительных процессов, Основы теории вероятностей и статистических методов, Алгоритмы и структуры данных, Математическая логика и теория алгоритмов, Интеллектуальный анализ данных.

Знания, получаемые при изучении дисциплины «Обработка больших данных» используются при изучении профессиональных дисциплин Распределенные задачи и алгоритмы, Программирование в компьютерных сетях, Облачные вычисления, Мультиагентные системы, а также для работ над дипломной и магистерской работой.

Результаты обучения (знания, умения, опыт, компетенции):

| № п.п | Индекс компетенции | Содержание компетенции (или ее части) | В результате изучения учебной дисциплины обучающиеся должны | | |
|-------|--------------------|--|---|---|---|
| | | | знать | уметь | владеть |
| 1. | ОПК-2 | Способен применять компьютерные/супер-компьютерные методы, современное | стандарты обработки и анализа больших данных, и | использовать современные инструментальны е и вычислитель- | способностью собирать, обрабатывать и интерпретиро- |

| | | | | | |
|--|-------|--|---|---|---|
| | | программное обеспечение, в том числе отечественного происхождения, для решения задач профессиональной деятельности | требования, связанные с созданием и использованием SQL и NoSQL систем хранения и обработки данных | ные средства (в соответствии с профилем подготовки), осуществлять постановку задач анализа данных, визуализацию интерпретацию результатов | вать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям |
| | ОПК-3 | Способен к разработке алгоритмических и программных решений в области системного и прикладного программирования, математических, информационных и имитационных моделей, созданию информационных ресурсов глобальных сетей, образовательного контента, прикладных баз данных, тестов и средств тестирования систем и средств на соответствие стандартам и исходным требованиям. | математические методы анализа данных, методы и прикладные языки для разработки программных решений в области обработки больших данных, математических, информационных и имитационных моделей, для создания информационных ресурсов глобальных сетей | выполнять сбор и анализ данных, в том числе из сети Интернет, производить интерпретацию и оценку полученных результатов | языками системного и прикладного программирования для разработки математических, информационных и имитационных моделей, для обработки информационных ресурсов глобальных сетей и прикладных баз данных. |
| | ПК-5 | Способен применять в профессиональной деятельности современные языки программирования и методы параллельной обработки данных, операционные системы, электронные библиотеки и пакеты программ, сетевые технологии | Современные языки и средства обработки данных (язык R, Python), прикладные библиотеки для анализа данных | применить современные языки (R и Python) и прикладные библиотеки для анализа данных, сформулировать научную гипотезу и проверить ее достоверность | владеет средствами сбора, обработки и анализа больших данных, средствами оценки эффективности решений |

Содержание и структура дисциплины (модуля)

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.

Разделы дисциплины, изучаемые в _6_ семестре (очная форма)

| № | Наименование разделов | Количество часов | | | |
|---|-----------------------|------------------|-------------------|-----|----------------------|
| | | Всего | Аудиторная работа | | Внеаудиторная работа |
| | | | Л | КСР | |
| | | | | | |

| | | | | | | |
|----|--|------|----|---|----|----|
| 1. | Введение в большие данные. Понятие Data Minig. Прикладные инструменты для работы с Big Data. Технология MapRaduce. Hadoop. | 12 | 4 | | 4 | 4 |
| 2. | Технологии анализа данных: Жизненный цикл анализа больших данных, стандарты. Дескриптивный анализ. | 12 | 4 | | 4 | 4 |
| 3. | Алгоритмы классификации, кластеризации. Понятие корреляции и регрессионный анализ. Тестирование гипотез. Когнитивный анализ данных. Визуализация больших данных. | 52 | 16 | 4 | 16 | 16 |
| 4. | Аналитика данных на больших графах | 12 | 4 | | 4 | 4 |
| 5. | Технологии хранения больших данных. Распределенные хранилища, NoSql хранилища, классификация и примеры. | 9 | 4 | | 4 | 1 |
| | <i>Итого по разделам дисциплины:</i> | 99 | 32 | 4 | 32 | 31 |
| | ИКР | 0,3 | | | | |
| | <i>Итого:</i> | 99,3 | | | | |
| | <i>Контроль</i> | 44,7 | | | | |
| | <i>Итого по дисциплине:</i> | 144 | | | | |

Примечание: Л – лекции, КСР – контрольные и самостоятельные работы, ЛР – лабораторные занятия, СРС – самостоятельная работа студента, Д-доклад, РГЗ – расчетно-графическое задание.

Курсовые проекты или работы: *не предусмотрены* **Вид аттестации:** экзамен.

Основная литература

- Крутиков, В.Н. Анализ данных : учебное пособие / В.Н. Крутиков, В.В. Мешечкин; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Кемеровский государственный университет». - Кемерово : Кемеровский государственный университет, 2019. - 138 с. : ил. - Библиогр. в кн. - ISBN 978-5-8353-1770-7 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=278426>
- Жуковский, О.И. Информационные технологии и анализ данных : учебное пособие / О.И. Жуковский ; Министерство образования и науки Российской Федерации, Томский Государственный Университет Систем Управления и Радиоэлектроники (ТУСУР). - Томск : Эль Контент, 2020. - 130 с. : схем., ил. - Библиогр.: с. 126. - ISBN 978-5-4332-0158-3 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=480500>
- Машинное обучение : учебник : [16+] / Е. Ю. Бутырский, В. В. Цехановский, Н. А. Жукова [и др.]. – Москва : Директ-Медиа, 2023. – 368 с. : ил., табл., схем., граф. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=701807> (дата обращения: 26.05.2024). – Библиогр. в кн. – ISBN 978-5-4499-3778-0. – DOI 10.23681/701807. – Текст : электронный.
- Баяк, О. А. Практикум по анализу данных на языках Python и R : учебное пособие для студентов, обучающихся по направлениям 38.03.01 «Экономика», 38.03.05 «Бизнес-информатика» : [16+] / О. А. Баяк, М. Р. Исаева, М. О. Самсонкин ; Финансовый университет при Правительстве Российской Федерации. – Москва : Прометей, 2023. – 100 с. : ил., табл., схем., граф. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=700938> (дата обращения: 26.05.2024). – Библиогр. в кн. – ISBN 978-5-00172-356-1. – Текст : электронный.
- Базы данных в высокопроизводительных информационных системах : учебное пособие /

авт.-сост. Е.И. Николаев ; Министерство образования и науки РФ, Федеральное государственное автономное образовательное учреждение высшего образования «Северо-Кавказский федеральный университет». - Ставрополь : СКФУ, 2016. - 163 с. : ил. - Библиогр.: с.161. ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=466799>

Автор Приходько Т.А. – кандидат технических наук, доцент кафедры вычислительных технологий