

Федеральное государственное образовательное учреждение
высшего образования
«КАЛИНИНГРАДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

Т. В. Снытникова

СБОР, ХРАНЕНИЕ И АНАЛИЗ БОЛЬШИХ ДАННЫХ

Учебно-методическое пособие по изучению дисциплины
для студентов бакалавриата по направлению подготовки
09.03.02 Информационные системы и технологии,
профиль программы
«Проектирование корпоративных информационных систем»

Калининград
Издательство ФГБОУ ВО «КГТУ»
2025

УДК 004(075)

Рецензент

кандидат педагогических наук,
доцент кафедры прикладной информатики института цифровых
технологий ФГБОУ ВО «Калининградский государственный технический
университет» Е. Ю. Заболотнова

Снытникова, Т. В.

Сбор, хранение и анализ больших данных: учеб.-метод. пособие по изучению дисциплины для студентов бакалавриата по направлению подготовки 09.03.02 Информационные системы и технологии, профиль программы «Проектирование корпоративных информационных систем» / Т. В. Снытникова. – Калининград: Изд-во ФГБОУ ВО «КГТУ», 2025. – 19 с.

В учебно-методическом пособии приведен тематический план по дисциплине и даны методические указания по её самостоятельному изучению, подготовке к лабораторным занятиям, подготовке и сдаче зачета, выполнению самостоятельной работы.

Пособие подготовлено в соответствии с требованиями утвержденной рабочей программы дисциплины направления подготовки 09.03.02 Информационные системы и технологии.

Табл. 4, список лит. – 12 наименований

Учебно-методическое пособие рассмотрено и одобрено в качестве локального электронного методического материала кафедрой прикладной информатики 11 ноября 2024 г., протокол № 11

Учебно-методическое пособие по изучению дисциплины рекомендовано к использованию в качестве локального электронного методического материала в учебном процессе методической комиссией института цифровых технологий ФГБОУ ВО «Калининградский государственный технический университет» 21 января 2025 г., протокол № 1

УДК 004(075)

© Федеральное государственное
бюджетное образовательное учреждение
высшего образования
«Калининградский государственный
технический университет», 2025 г.
© Снытникова Т. В., 2025 г.

ОГЛАВЛЕНИЕ

1. Введение.....	4
2. Тематический план	6
3. Содержание дисциплины.....	6
4. Методические указания по проведению лабораторных занятий.....	10
5. Методические указания по выполнению самостоятельной работы	11
6. Методические указания по проведению занятий и освоению дисциплины... ..	11
7. Методические указания по выполнению контрольной работы.....	12
8. Требования к аттестации по дисциплине.....	13
8.1 Текущая аттестация	13
8.2. Промежуточная аттестация по дисциплине	16
9. Заключение.....	16
10. Библиографический список.....	17

1. ВВЕДЕНИЕ

Данное учебно-методическое пособие предназначено для студентов направления подготовки 09.03.02 «Информационные системы и технологии», изучающих дисциплину «Сбор, хранение и анализ больших данных».

Целью освоения дисциплины «Сбор, хранение и анализ больших данных» является формирование у обучающихся практических навыков, связанных с применением в своей профессиональной деятельности современных технологий в области обработки больших данных, решать задачи сбора, организации, хранения, анализа больших данных. На практике получить навыки разработки алгоритмов, программных модулей и моделей для обработки больших данных. Полученные знания могут быть использованы в профессиональной деятельности для обработки больших данных.

Задачи изучения дисциплины:

- формирование базовых приемов функционального и логического программирования для обработки больших данных;
- приобретение теоретических знаний и практических навыков по современным методам проектирования и управления хранилищами данных;
- приобретение практических навыков анализа и интерпретации больших данных.

В результате изучения дисциплины студент должен:

знать:

- основные платформы, технологии и инструментальные программно-аппаратные средства для реализации информационных систем;
- современные методы проектирования программного обеспечения, позволяющие вести разработку прототипов ИС на базе типовой ИС в рамках выполнения работ по созданию (модификации) и сопровождению ИС;
- основные приемы функционального и логического программирования;

уметь:

- применять современные технологии для реализации информационных систем;
- применять современные методы проектирования программного обеспечения, позволяющие вести прототипов ИС на базе типовой ИС в рамках выполнения работ по созданию (модификации) и сопровождению ИС;
- применять основные приемы функционального и логического программирования для создания программного кода ИС в рамках выполнения работ по созданию (модификации) и сопровождению ИС;

владеть:

- навыками использования технологий, применения инструментальных программно-аппаратных средств реализации информационных систем;
- современными методами проектирования программного обеспечения, позволяющие вести разработку прототипов ИС на базе типовой ИС в рамках выполнения работ по созданию (модификации) и сопровождению ИС;
- основными приемами функционального и логического программирования для создания программного кода ИС в рамках выполнения работ по созданию (модификации) и сопровождению ИС.

Дисциплина «Сбор, хранение и анализ больших данных» входит в состав профессионального модуля части, формируемой участниками образовательных отношений образовательной программы бакалавриата по направлению подготовки 09.03.02 Информационные системы и технологии.

Для успешного освоения дисциплины, в соответствии с учебным планом, ей предшествуют такие дисциплины, как: «Программирование на языках высокого уровня», «Передача данных и сетевые технологии», «Базы данных».

Результаты освоения дисциплины могут быть использованы при выполнении выпускной квалификационной работы, а также в дальнейшей профессиональной деятельности.

Далее в пособии представлен тематический план, содержащий перечень изучаемых тем и лабораторных занятий, мероприятий текущей аттестации и отводимое на них аудиторное время (занятия в соответствии с расписанием) и самостоятельную работу. При формировании личного образовательного плана на семестр следует оценивать рекомендуемое время на изучение дисциплины, возможно, вам потребует больше времени на выполнение отдельных заданий или проработку отдельных тем.

В разделе «Содержание дисциплины» приведены подробные сведения об изучаемых вопросах, по которым вы можете ориентироваться в случае пропуска каких-то занятий, а также методические рекомендации преподавателя для самостоятельной подготовки, каждая тема имеет ссылки на литературу (или иные информационные ресурсы), а также контрольные вопросы для самопроверки.

Раздел «Требования к аттестации по дисциплине» содержит описание обязательных мероприятий контроля самостоятельной работы и усвоения разделов или отдельных тем дисциплины. Далее изложены требования к завершающей аттестации – экзамену.

Помимо данного пособия, студентам следует использовать материалы, размещенные в соответствующем данной дисциплине разделе ЭИОС, в которые более оперативно вносятся изменения для адаптации дисциплины под конкретную группу.

2. ТЕМАТИЧЕСКИЙ ПЛАН

Общая трудоемкость дисциплины составляет 3 зачетные единицы (ЗЕТ), т. е. 108 академических часов контактной (лекционных и лабораторных занятий) и самостоятельной работы студента, в том числе связанной с текущей и промежуточной (заключительной) аттестацией по дисциплине.

Формы аттестации по дисциплине:

- очная форма, пятый семестр – зачет;
- заочная форма, пятый семестр – контрольная работа, зачет.

Тематический план лекционных занятий для очной и заочной формы обучения приведен в таблице 1

Таблица 1 – Тематический план лекционных занятий

	Тема лекционного занятия	Объем учебной работы	
		очная форма, ч	заочная форма, ч
Тема 1	Введение в большие данные.	8	1
Тема 2	Модели жизненного цикла аналитики данных.	8	1
Тема 3	Сбор и хранение Big Data.	8	2
Тема 4	Анализ больших данных.	8	2
	ИТОГО	32	6

3. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Содержательно структура дисциплины представлена четырьмя тематическими блоками (темами).

Тема 1. Введение в большие данные

Перечень изучаемых вопросов

Цели и задачи дисциплины. Место дисциплины в структуре образовательной программы. Планируемые результаты освоения дисциплины.

Понятие Data Mining. Прикладные инструменты для работы с Big Data. Технология MapReduce. Hadoop.

Вопросы конфиденциальности, защиты персональных данных, правовые аспекты сбора и обработки информации, этические дилеммы при использовании больших данных.

Методические указания к изучению

Знакомство с ключевыми понятиями и терминами, такими как Data Mining, Big Data, MapReduce и Hadoop, является первым и необходимым шагом в изучении дисциплины. Игнорирование этих понятий может привести к непониманию базовых принципов работы с большими данными и затруднить дальнейшее обучение. Кроме того, важно уделить внимание вопросам конфиденциальности, защиты персональных данных и этическим аспектам использования больших данных, так как пренебрежение этими вопросами может повлечь за собой серьезные последствия, как юридические, так и моральные.

Рекомендуемая литература: [1, гл. 1, 7].

Контрольные вопросы

1. Что такое большие данные? Назовите четыре ключевых характеристики больших данных.
2. Назовите основные подходы к анализу больших данных.
3. Какую роль играют технологии Hadoop и Spark в обработке больших данных?
4. Почему важно учитывать этические аспекты при работе с большими данными?

Тема 2. Модели жизненного цикла аналитики данных

Перечень изучаемых вопросов

Эволюция стандартизованных моделей (методологии), задающие этапы процесса интеллектуального анализа данных (KDD&DM).

Методологии SEMMA и KDD Process.

Методология CRISP DM.

Методические указания к изучению

Понимание эволюционных процессов стандартизованных моделей (методологий) является важной основой для дальнейшего освоения этапов интеллектуального анализа данных (KDD&DM). Детальное изучение различных методологий, таких как SEMMA, KDD Process и CRISP DM, помогает выбирать наиболее подходящую для конкретной ситуации выбору подхода в конкретных аналитических задачах. Это позволит избежать ошибок и повысить эффективность работы с большими данными.

Рекомендуемая литература: [2, гл. 2]; [6].

Контрольные вопросы

1. В чем заключается методология CRISP-DM?
2. В чем заключается методология SEMMA?
3. В чем заключается методология KDD Process?
4. В чем отличие вышеперечисленных методологий? Дайте свою оценку каждому методу. Сделайте вывод о том, какая методология для каких предметных областей будет лучшей.

Тема 3. Сбор и хранение Big Data

Перечень изучаемых вопросов

Источники данных. Проблемы при сборе данных.

Методы сбора данных: прямое подключение к источникам, использование API, скрейпинг веб-сайтов.

Специфика работы с потоковыми данными и batch в Big Data. Подготовка данных. Процессы ETL и ELT: ключевые этапы процессов, подробное сравнение, области применения.

Методические указания к изучению

Для реализации методов сбора и хранения данных можно использовать различные инструменты и подходы. Необходимо рассмотреть разнообразные источники данных, такие как социальные сети, интернет-трафик, сенсоры и другие, а также определите основные проблемы, возникающие при их сборе: неполные данные, ошибки измерений и отсутствие стандартов.

Подготовка включает очистку данных от шумов, заполнение пропущенных значений, нормализацию и преобразование данных в нужный формат. Без качественной подготовки данных невозможно получить точные и надежные результаты анализа. Здесь также важно учитывать, какие методы очистки и нормализации подходят для конкретного типа данных и какую роль играет подготовка данных в общем процессе анализа.

Для автоматизации сбора данных можно использовать различные методы, такие как прямое подключение к источникам, использование API и скрейпинг веб-сайтов. При этом нужно знать их преимущества и недостатки, а также ситуации, в которых они наиболее эффективны.

Необходимо рассмотреть специфику работы с потоковыми данными и пакетными данными. Каждый из этих подходов имеет свои особенности и применяется в зависимости от

требований бизнеса и характера данных. Важно понимать, как эти два типа данных влияют на архитектуру хранилищ данных и выбор инструментов для их обработки.

Также стоит обратить внимание на изучение процессов ETL и ELT, так как правильный выбор между ними влияет на производительность и точность обработки данных. Для глубокого понимания различий между ETL и ELT необходимо рассмотреть каждый шаг отдельно, оценить их сильные и слабые стороны, а также определить, какой из подходов лучше подходит для конкретных задач.

Рекомендуемая литература: [3, гл. 2].

Контрольные вопросы

1. Приведите классификацию методов и алгоритмов адаптации.
2. Назовите особенности беспоисковых адаптивных систем.
3. В чем отличие БАС с эталонной моделью и идентификатором?

Тема 4. Анализ больших данных

Перечень изучаемых вопросов

Технология MapReduce: основные компоненты, алгоритм работы, преимущества и недостатки. Примеры применения MapReduce. Инструменты и платформы, поддерживающие технологию.

Технология Spark: архитектура, основные компоненты, особенности работы Spark. Преимущества и недостатки Spark, примеры использования.

Методические указания к изучению

Для решения задач обработки больших данных разработаны специальные технологии, такие как MapReduce и Spark. Эти технологии позволяют эффективно обрабатывать огромные массивы информации, обеспечивая высокую производительность и масштабируемость.

Необходимо рассмотреть основные компоненты и алгоритм работы технологии MapReduce, её преимущества и недостатки и проанализировать инструменты и платформы, ее поддерживающие.

В рамках изучения технологии Spark следует рассмотреть архитектуру, основные компоненты и особенности работы Spark, проанализировать её преимущества и недостатки, а также примеры использования.

Рекомендуемая литература: [1]; [2, гл. 5]; [5].

Контрольные вопросы

1. Каковы ключевые различия между технологией Spark и MapReduce?

2. Опишите процесс работы функции Mapper и Reducer в рамках технологии MapReduce.

3. Объясните архитектуру Apache Spark и роль Resilient Distributed Dataset (RDD).

4. Назовите основные недостатки технологий Spark и MapReduce и предложите возможные пути их устранения.

4. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРОВЕДЕНИЮ ЛАБОРАТОРНЫХ ЗАНЯТИЙ

Особое место в структуре дисциплины занимает практикум, включающий в себя следующие лабораторные работы (таблица 2).

Таблица 2 – Объем (трудоемкость освоения) и структура ЛЗ

Номер темы	Содержание лабораторного занятия	Очная форма, ч	Заочная форма, ч
1	Введение в библиотеку NumPy	2	-
1	Манипуляция над данными с помощью пакета Pandas	6	
1	Визуализация с помощью библиотеки Matplotlib	2	
2	Методология CRISP DM	2	1
3	Работа с xlsx, sav, csv-файлами	4	1
3	Парсинг web-страниц с помощью запросов xpath и BeautifulSoup	4	2
3	Автоматизация работы с публичными данными, доступными через общедоступные API	4	2
4	Подключение PySpark и Kaggle в Google Colab	2	-
4	Знакомство с Spark API	10	-
ИТОГО:		32	6

Лабораторные занятия по дисциплине проводятся в компьютерных классах, оснащенных персональными компьютерами с программным обеспечением.

Студент в ходе лабораторного практикума согласно методическим указаниям и заданию преподавателя изучает методологии, при проведении проектов анализа данных, осваивает методы работы с данными в разных форматах, способы сбора и обработки больших данных, а также изучает инструментария для визуализации результатов обработки. Защита лабораторной работы проводится на основании выполненного задания к лабораторной работе, оформленного отчета, а также ответа на контрольные вопросы.

5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Самостоятельная работа студентов по дисциплине, а также работа в ЭИОС университета может проводиться в компьютерных классах кафедры прикладной информатики, оснащенных персональными компьютерами с выходом в сеть Интернет.

Самостоятельная работа студента включает в себя освоение теоретического учебного материала (в том числе подготовка к лабораторным занятиям, оформление работ, подготовка к защите лабораторных работ).

Наряду с проработкой лекционного материала и подготовкой к лабораторным занятиям, студент заочной формы обучения обязан выполнить индивидуальную контрольную работу

В качестве задания для контрольной работы студентов заочной формы обучения выбираются (по указанию преподавателя) два вопроса из примерного перечня контрольных вопросов по дисциплине (п. 7).

6. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРОВЕДЕНИЮ ЗАНЯТИЙ И ОСВОЕНИЮ ДИСЦИПЛИНЫ

Для успешного освоения дисциплины необходимо ознакомиться с основными принципами сбора, хранения и анализа больших данных.

При разработке образовательной технологии организации учебного процесса основной упор сделан на соединение активной и интерактивной форм обучения. Интерактивная форма позволяет студентам проявить самостоятельность в освоении теоретического материала и овладении практическими навыками, формирует интерес и позитивную мотивацию к учебе.

При изучении дисциплины внимание студентов постоянно акцентируется не только на теоретическом аспекте работы с большими данными, но и их практическом применении в современных высокотехнологичных производствах.

Для планирования работы студента в начале семестра производится выдача тем для самостоятельного изучения, определяются источники информации и график проведения текущего контроля. В качестве источников информации рекомендуется наряду с учебными пособиями использовать периодические издания (журналы) из области профессиональной деятельности.

В ходе лекционных занятий студенту следует вести конспектирование учебного материала. При самостоятельном изучении заданных преподавателем тем рекомендуется вносить основные материалы по ним в тот же конспект лекций в соответствии с рекомендованным порядком следования учебного материала.

При проведении занятий в интерактивной форме важно участвовать в процессе обсуждения и решения поставленных задач сбора и анализа больших данных, задавать преподавателю вопросы с целью уяснения теоретических положений, области их применения, разрешения спорных ситуаций.

На лекциях и лабораторных занятиях изложению нового материала предшествуют обсуждение предыдущей темы с целью восстановления и закрепления студентами изученного теоретического и практического материала и ответы на вопросы студентов. В конце лекции выделяется время для ответов на вопросы по текущему материалу и его обсуждению. На лабораторных занятиях используется разбор конкретных технологий работы с большими данными. Активность студентов и проявленные знания при обсуждении материала учитываются при текущей и промежуточной (заключительной) аттестации по дисциплине.

Самостоятельная работа студентов проводится с целью:

- систематизации и закрепления полученных теоретических знаний и практических умений студентов;
- углубления и расширения теоретических знаний;
- формирования умений использовать специальную литературу.

Для закрепления изученного материала, определения «пробелов» в знаниях студентов, а также для текущего контроля используются индивидуальные контрольные задания (два на весь курс). Индивидуальные контрольные задания выполняются студентом на 8 и 11 учебных неделях семестра в ходе самостоятельной работы по дисциплине, проверяются преподавателем и при необходимости на лекциях разбираются типичные ошибки при их выполнении.

Тестовые задания для проведения зачетного тестирования приведены в фонде оценочных средств по дисциплине.

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ КОНТРОЛЬНОЙ РАБОТЫ

Контрольная работа выполняется студентами заочной формы обучения по вариантам.

При подготовке к контрольной работе рекомендуется ознакомиться с предоставленным списком вопросов. Для каждого вопроса важно составить краткий план ответа, опираясь на материалы лекций, конспекты, источники из списка рекомендованной литературы. Для более углубленной проработки вопроса можно использовать дополнительные источники.

Вариант контрольной работы содержит два вопроса из перечня.

Перечень вопросов для контрольной работы:

1. Что такое большие данные? Назовите четыре ключевых характеристики больших данных.
2. Назовите основные подходы к анализу больших данных.
3. Какую роль играют технологии Hadoop и Spark в обработке больших данных?
4. Почему важно учитывать этические аспекты при работе с большими данными?
5. В чем заключается методология CRISP-DM?
6. В чем заключается методология SEMMA?
7. В чем заключается методология KDD Process?
8. Назовите методы сбора данных, дайте описание каждому.
9. Объясните специфику работы с потоковыми данными и batch в Big Data.
10. Процессы ETL и ELT: ключевые этапы процессов, подробное сравнение, области применения.
11. Технология MapReduce: основные компоненты, алгоритм работы. Преимущества и недостатки MapReduce, примеры использования.
12. Технология Spark: архитектура, основные компоненты, особенности работы Spark. Преимущества и недостатки Spark, примеры использования.

8 ТРЕБОВАНИЯ К АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ

8.1 Текущая аттестация

Текущая аттестация (текущий контроль) проводится с целью оценки освоения теоретического учебного материала, в том числе в рамках самостоятельной работы студента (п. 6).

Контроль на лекциях и лабораторных занятиях производится в виде тестирования или устного опроса.

Типовые контрольные вопросы для устного опроса по темам приведены в п. 3 настоящего пособия. Тестовые задания представлены в фонде оценочных средств (приложение к рабочей программе дисциплины).

Положительная оценка («зачтено») по результатам каждого контроля (опроса) выставляется в соответствии с универсальной системой оценивания, приведенной в таблице 3. В случае получения оценки «не зачтено» студент должен пройти повторный контроль по данной теме в ходе последующих консультаций.

Таблица 3 – Система оценок и критерии выставления оценки при прохождении контроля (опроса)

Критерий	Система оценок			
	«не зачтено»	«зачтено»		
1 Системность и полнота знаний в отношении изучаемых объектов	Обладает частичными и разрозненными знаниями, которые не может научно-корректно связывать между собой (только некоторые из которых может связывать между собой)	Обладает минимальным набором знаний, необходимым для системного взгляда на изучаемый объект	Обладает набором знаний, достаточным для системного взгляда на изучаемый объект	Обладает полнотой знаний и системным взглядом на изучаемый объект

Текущий контроль в виде защиты лабораторных работ проводится на лабораторном практикуме, целью которого является формирование умений и навыков по сбору, хранению и анализу больших данных. Оценка результатов выполнения задания по каждой лабораторной работе производится при представлении студентом отчета по лабораторной работе и на основании ответов студента на вопросы по тематике лабораторной работы. Студент, самостоятельно выполнивший задание, продемонстрировавший знание использованных им технических средств, алгоритмов и языков программирования задачи, получает по лабораторной работе оценку «зачтено».

С целью контроля качества самостоятельной работы студентов заочной формы запланировано выполнение и защита контрольной работы. Система оценивания и критерии оценки контрольной работы приведены в таблице 4.

Таблица 4 – Система оценивания и критерии оценки контрольной работы

Критерий	Система оценок			
	«не зачтено»	«зачтено»		
1 Работа с информацией	Не в состоянии находить	Может найти необходимую	Может найти, интерпретировать	Может найти,

Критерий	Система оценок			
	«не зачтено»	«зачтено»		
	необходимую информацию, либо в состоянии находить отдельные фрагменты информации в рамках поставленной задачи	информацию в рамках поставленной задачи	и систематизировать необходимую информацию в рамках поставленной задачи	систематизировать необходимую информацию, а также выявить новые, дополнительные источники информации в рамках поставленной задачи
2 Научное осмысление изучаемого явления, процесса, объекта	Не может делать научно-корректных выводов из имеющихся у него сведений, в состоянии проанализировать только некоторые из имеющихся у него сведений	В состоянии осуществлять научно-корректный анализ предоставленной информации	В состоянии осуществлять систематический и научно-корректный анализ предоставленной информации, вовлекает в исследование новые релевантные задачи данные	В состоянии осуществлять систематический и научно-корректный анализ предоставленной информации, вовлекает в исследование новые релевантные поставленной задаче данные, предлагает новые ракурсы поставленной задачи
3 Освоение стандартных алгоритмов решения профессиональных задач	В состоянии решать только фрагменты поставленной задачи в соответствии с заданным алгоритмом, не освоил предложенный алгоритм, допускает ошибки	В состоянии решать поставленные задачи в соответствии с заданным алгоритмом	В состоянии решать поставленные задачи в соответствии с заданным алгоритмом, понимает основы предложенного алгоритма	Не только владеет алгоритмом и понимает его основы, но и предлагает новые решения в рамках поставленной задачи

8.2 Промежуточная аттестация по дисциплине

Промежуточная (заключительная) аттестация по дисциплине предусматривает проведение зачета (зачетного тестирования).

К зачету допускаются студенты:

- выполнившие и защитившие все предусмотренные лабораторные работы (получившие положительную оценку по результатам лабораторного практикума);
- выполнившие контрольную работу (получившие оценку «зачтено» по контрольной работе) – для студентов заочной формы.

Зачет может проводиться как в традиционной форме, так и в виде тестирования. Тестовые задания для проведения тестирования приведены в фонде оценочных средств по дисциплине.

Тестовые задания для проведения зачетного тестирования приведены в фонде оценочных средств по дисциплине.

Система оценивания и критерии выставления оценок по зачету (зачетному тестированию) приведена в таблице 4.

9. ЗАКЛЮЧЕНИЕ

Современные информационные системы и технологии производят и потребляют огромные объемы данных, которые становятся все более значимыми для бизнеса, науки и повседневной жизни.

Сбор, хранение и анализ больших данных стали неотъемлемыми элементами современных информационных систем и технологий. Они способствуют улучшению принятия решений, оптимизации бизнес-процессов, персонализации клиентского опыта, прогнозированию и обеспечению безопасности. В условиях стремительно растущего объема данных, умение эффективно управлять ими и извлекать из них полезные знания становится ключевым фактором успеха для организаций любого масштаба и сферы деятельности.

Освоение дисциплины «Сбор, хранение и анализ больших данных» является одним из основополагающих шагов к формированию будущего специалиста в области информационных систем и технологий. Приобретенные в ходе изучения дисциплины знания, умения и навыки будут углубляться и совершенствоваться в процессе дальнейшего обучения и могут быть применены в профессиональной деятельности.

10. БИБЛИОГРАФИЧЕСКИЙ СПИСОК

Основная литература

1. Макшанов, А. В. Большие данные. Big Data / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. – 4-е изд., стер. – Санкт-Петербург: Лань, 2024. – 188 с. – Режим доступа: для авториз. пользователей. – Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/362318> (дата обращения: 27.07.2024). — ISBN 978-5-507-47346-5. – Текст : электронный.

2. Нурматова, Е. В. Управление большими базами данных и высоконагруженными системами: учеб. пособие / Е. В. Нурматова, Р. Ф. Халабия, Л. В. Бунина. – Москва: РТУ МИРЭА, 2019. – 120 с. – Режим доступа: для авториз. пользователей. – Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/171496> (дата обращения: 05.07.2024). – Текст : электронный.

3. Хачумов, М. В. Введение в интеллектуальный анализ данных: учеб. пособие / М. В. Хачумов. – Москва: РТУ МИРЭА, 2023. – 123 с. – Режим доступа: для авториз. пользователей. – Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/398240> (дата обращения: 05.07.2024). — ISBN 978-5-7339-2073-3. – ISBN 978-5-7339-2073-3. – Текст : электронный.

4. Пальмов, С. В. Основы сбора и обработки больших данных: учеб. пособие / С. В. Пальмов. – Самара: ПГУТИ, 2023. – 285 с. – Режим доступа: для авториз. пользователей. – Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/411830> (дата обращения: 16.07.2024). – Текст : электронный.

Дополнительная литература

5. Ланских, Ю. В. Введение в большие данные: учеб. пособие / Ю. В. Ланских, В. Г. Ланских, К. В. Родионов. – Киров: ВятГУ, 2023. – 172 с. – Режим доступа: для авториз. пользователей. – Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/408566> (дата обращения: 27.07.2024). – Текст : электронный.

6. Алетдинова, А. А. Интеллектуальный анализ больших данных: учеб. пособие / А. А. Алетдинова, М. Ш. Муртазина. – Новосибирск: НГТУ, 2023. – 66 с. – Режим доступа: для авториз. пользователей. – Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/404567> (дата обращения: 27.07.2024). – ISBN 978-5-7782-4899-1. – Текст : электронный.

7. Агафонов, А. А. Основы технологий баз данных: учеб. пособие / А. А. Агафонов, А. М. Белов. – Самара: Самарский университет, 2023. – 304 с. – ISBN 978-5-7883-1915-5. –

Режим доступа: для авториз. пользователей. – Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/406457> (дата обращения: 16.07.2024). – Текст : электронный.

8. Горожанина, Е. И. Высокопроизводительные вычисления и анализ больших данных: учеб. пособие / Е. И. Горожанина. – Самара: ПГУТИ, 2022. – 132 с. – Режим доступа: для авториз. пользователей. – Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/411386> (дата обращения: 27.07.2024). – Текст : электронный.

Учебно-методические пособия

1. Наука о данных и аналитика больших объемов данных: методические рекомендации / составители А. С. Копырин, Д. И. Попов. – Сочи: СГУ, 2023. – 32 с. – Режим доступа: для авториз. пользователей. – Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/417155> (дата обращения: 04.08.2024). – Текст: электронный.

2. Железнов, М. М. Методы и технологии обработки больших данных: учеб.-метод. пособие / М. М. Железнов. – Москва: МИСИ – МГСУ, 2020. – 46 с. – Режим доступа: для авториз. пользователей. – Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/145102> (дата обращения: 04.08.2024). – ISBN 978-5-7264-2193-3. – Текст : электронный.

3. Лебедев, А. С. Методы Big Data: учеб.-метод. пособие / А. С. Лебедев, Ш. Г. Магомедов. – Москва: РТУ МИРЭА, 2021. – 91 с. – Режим доступа: для авториз. пользователей. – Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/182452> (дата обращения: 04.08.2024). – Текст : электронный.

4. Бутаков, Н. А. Обработка больших данных с Apache Spark: учеб.-метод. пособие / Н. А. Бутаков, М. В. Петров, Д. Насонов. – Санкт-Петербург: Университет ИТМО, 2019. – 52 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=566771> (дата обращения: 04.07.2024). – Текст : электронный

Интернет-ресурсы

Российская образовательная платформа и конструктор бесплатных открытых онлайн-курсов и уроков – <https://stepik.org>

Образовательная платформа – <https://openedu.ru/>

Локальный электронный методический материал

Татьяна Валентиновна Снытникова

СБОР, ХРАНЕНИЕ И АНАЛИЗ БОЛЬШИХ ДАННЫХ

Редактор С. Кондрашова
Корректор Т. Звада

Уч.-изд. л. 1,3. Печ. л. 1,2

Издательство федерального государственного бюджетного образовательного
учреждения высшего образования
«Калининградский государственный технический университет».
236022, Калининград, Советский проспект, 1