



Федеральное агентство по рыболовству  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Калининградский государственный технический университет»  
(ФГБОУ ВО «КГТУ»)

УТВЕРЖДАЮ  
И.о. директора института

Фонд оценочных средств  
(приложение к рабочей программе модуля)  
**«СБОР, ХРАНЕНИЕ И АНАЛИЗ БОЛЬШИХ ДАННЫХ»**

основной профессиональной образовательной программы бакалавриата  
по направлению подготовки

**09.03.02 ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ**

**Профиль «Проектирование корпоративных информационных систем»**

ИНСТИТУТ  
РАЗРАБОТЧИК

цифровых технологий  
кафедра прикладной информатики

## 1 РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ, ПЕРЕЧЕНЬ ОЦЕНОЧНЫХ СРЕДСТВ И КРИТЕРИИ ОЦЕНИВАНИЯ

### 1.1 Результаты освоения дисциплины

Таблица 1 – Планируемые результаты обучения по дисциплине, соотнесенные с установленными компетенциями

Код и наименование компетенции	Дисциплина	Результаты обучения (владения, умения и знания), соотнесенные с компетенциями
ПК-2 «Способен осуществлять организационно- экономическое и технологическое сопровождение цифровой трансформации документированных сфер деятельности организации»	Сбор, хранение и анализ больших данных	<p><u>знать:</u> 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;</p> <p><u>уметь:</u> 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;</p>

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

1.2 К оценочным средствам текущего контроля успеваемости относятся:

- тестовые задания открытого и закрытого типов;
- типовые задания по контрольной работе для заочной формы обучения.

Промежуточная аттестация в форме зачета проходит по результатам прохождения всех видов текущего контроля успеваемости. В отдельных случаях (при не прохождении всех видов текущего контроля) зачет может быть проведен в виде тестирования.

1.3 Критерии оценки результатов освоения дисциплины

Универсальная система оценивания результатов обучения включает в себя системы оценок: 1) «отлично», «хорошо», «удовлетворительно», «неудовлетворительно»; 2) «зачтено», «не зачтено»; 3) 100 – балльную/процентную систему и правило перевода оценок в пятибалльную систему (табл. 2).

Таблица 2 – Система оценок и критерии выставления оценки

Система оценок	2	3	4	5
	0-40%	41-60%	61-80 %	81-100 %
	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»
Критерий	«не зачтено»	«зачтено»		
<b>1 Системность и полнота знаний в отношении изучаемых объектов</b>	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
<b>2 Работа с информацией</b>	Не в состоянии находить необходимую информацию, либо в состоянии находить отдельные фрагменты информации в рамках поставленной задачи	Может найти необходимую информацию в рамках поставленной задачи	Может найти, интерпретировать и систематизировать необходимую информацию в рамках поставленной задачи	Может найти, систематизировать необходимую информацию, а также выявить новые, дополнительные источники информации в рамках поставленной задачи
<b>3 Научное осмысление изучаемого явления, процесса, объекта</b>	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 задаче данные	В состоянии осуществлять систематический и научно-корректный анализ предоставленной информации, вовлекает в исследование новые релевантные поставленной задаче данные, предлагает новые ракурсы поставленной

Система оценок	2	3	4	5
	0-40%	41-60%	61-80 %	81-100 %
Критерий	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»
	«не зачтено»	«зачтено»		
				задачи
<b>4 Освоение стандартных алгоритмов решения профессиональных задач</b>	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

## 2 ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Компетенция ПК-2 «Способен осуществлять организационно-экономическое и технологическое сопровождение цифровой трансформации документированных сфер деятельности организации»

### Тестовые задания закрытого типа:

1. Объём накопленных человечеством цифровых данных на 2012 год измеряется:

1. петабайтами
- 2. зеттабайтами**
3. эксабайтами
4. йоттабайтами

2. Укажите фактор, способствовавший появлению тренда больших данных

- 1. маркетинговые кампании крупных корпораций**
- 2. снижение издержек на хранение данных**
3. появление новых технологий обработки потоковых данных
4. выпуск баз данных с обработкой данных в памяти

3. Какие вероятные разочарования тренда больших данных?

- 1. из-за угрозы безопасности личной жизни (privacy) граждан будут усложнены процедуры сбора данных, что приведёт к падению ценности больших данных**
2. из-за угрозы безопасности личной жизни (privacy) граждан будут

упрощены процедуры сбора данных, что приведёт к падению ценности больших данных

3. выпуск баз данных с обработкой данных в памяти

4. нет

4. Для машинного обучения подходят данные

1. Любых форматов в цифровом виде

2. Числовые типа int

3. Бинарные

**4. Предварительно подготовленные, очищенные от ошибок, пропусков и выбросов, а также нормализованные и представленные в виде числовых векторов**

5. Перечислите четыре основных характеристики Big Data:

1. Virtualization, Volume, Variability, Vehicle

**2. Variety, Velocity, Volume, Value**

3. Verification, Volume, Velocity, Visualization

4. Video, Value, Variety, Volume

6. Принцип MapReduce состоит в том, чтобы

**1. производить вычисления на узлах, где информация изначально была сохранена**

**2. использовать вычислительные мощности систем хранения**

3. использовать функциональное программирование для решения задач массивно-параллельной обработки

4. нет правильного ответа

7. В чём преимущество колоночно-ориентированных СУБД?

1. они позволяют выполнять более сложные SQL-запросы по сравнению с реляционными СУБД

**2. они позволяют динамически дополнять содержание записей новыми полями**

3. они имеют более гибкие возможности аналитики

4. они позволяют эффективно делать межколоночные сравнения

### **Тестовые задания открытого типа:**

8. Hadoop – это: \_\_\_\_\_

**Ответ: набор утилит, и программный каркас для выполнения распределённых программ, работающих на кластерах**

9. Методология \_\_\_\_\_ считается стандартным процессом в Data Science и состоит из 6 основных этапов: понимание бизнеса, понимание данных, подготовка данных, моделирование, оценка, развертывание

**Ответ: CRISP-DM (CRoss Industry Standard Process for Data Mining\*)**

10. Фаза методологии CRISP-DM, заключающаяся в отборе данных с обоснованием их включения/исключения, а также в очистке и преобразовании данных в формат, пригодный для анализа, называется: \_\_\_\_\_.

**Ответ: подготовка данных**

11. В отличие от другого широко используемого стандарта Data Mining, CRISP-DM, \_\_\_\_\_ фокусируется на задачах моделирования, не затрагивая бизнес-аспекты.

**Ответ: SEMMA**

12. Основной вклад методологии \_\_\_\_\_ – это включение в рассмотрение задачи ресурсообеспечения (resourcing) для повышения эффективности процесса обнаружения знаний.

**Ответ: KDD Process (KDD Roadmap\*)**

13. Информация, автоматически генерируемая компьютером, процессом, приложением или устройством без вмешательства человека – \_\_\_\_\_ данные.

**Ответ: машинные**

14. Файл \_\_\_\_\_ — это формат файла, созданный Microsoft для их Office Excel 2007 и более поздних версий. Он основан на формате Open XML и используется для хранения электронных таблиц с поддержкой формул, таблиц, диаграмм и макросов. Файлы этого формата обычно используются для анализа данных, финансового моделирования и визуализации данных. Их также можно использовать для хранения других типов информации, например изображений, презентаций и документов.

**Ответ: XLSX**

15. Файл формата \_\_\_\_\_ представляет собой файл данных, который представляет собой приложение, широко используемое исследователями рынка, исследователями в области здравоохранения, исследовательскими компаниями, правительством, исследователями в области образования, маркетинговыми организациями, сборщиками данных для статистического анализа. Он сохраняется в проприетарном двоичном формате и состоит из набора данных, а также словаря, который представляет набор данных, сохраняет данные в строках и столбцах.

**Ответ: SAV**

16. \_\_\_\_\_ — текстовый формат, предназначенный для представления табличных данных. Строка таблицы соответствует строке текста, которая содержит одно или несколько полей, разделенных запятыми.

**Ответ: CSV**

17. \_\_\_\_\_ — это текстовый формат представления структурированных данных. Он помогает передавать данные между разными веб-приложениями, клиентской и серверной частью одного приложения или его компонентами, даже если они написаны на разных языках программирования.

**Ответ: JSON**

18. \_\_\_\_\_ — это язык запросов, используемый для навигации и поиска информации в XML-документах. Он позволяет точно указать путь к элементам, атрибутам и текстовым данным, которые нужно найти в структуре XML файлов и получить данные при необходимости.

**Ответ: XPath (XML Path Language\*)**

19. \_\_\_\_\_ – это библиотека для парсинга HTML и XML документов в Python.

**Ответ: BeautifulSoup**

20. \_\_\_\_\_ постоянно генерируются различными источниками данных, например датчиками. Для извлечения полезных сведений в режиме реального времени данные должны быть очищены и преобразованы. Хорошим примером может служить финансовая организация, отслеживающая изменения на фондовом рынке в режиме реального времени.

**Ответ: потоковые данные**

21. \_\_\_\_\_ позволяет производить обработку большого объёма данных в пределах одного вызова, но разбивая выполнение на подзапросы.

**Ответ: Пакетная обработка данных (Batch, Batch API\*)**

22. К инструментальным средствам для \_\_\_\_\_ относятся Apache Kafka, Flink, NiFi, Storm, Spark Structured Streaming и основанные на них платформы.

**Ответ: потоковой обработки**

23. К инструментальным средствам для \_\_\_\_\_ относятся Apache Hadoop, Apache Airflow, Apache Spark и основанные на них платформы.

**Ответ: пакетной обработки**

24. Инструменты \_\_\_\_\_ используются для интеграции данных, чтобы удовлетворить требованиям систем управления реляционными базами данных и/или традиционных хранилищ данных с поддержкой OLAP (online analytical processing, аналитической онлайн-обработки)

**Ответ: ETL**

25. \_\_\_\_\_ — это процесс интеграции данных, при котором из различных источников извлекается сырая информация (в её исходных форматах), загружается непосредственно в центральный репозиторий, будь то облачное хранилище данных, озеро

данных или data lakehouse, где происходит преобразование данных в подходящие форматы для дальнейшего анализа и отчётности.

**Ответ: ELT**

26. API — это \_\_\_\_\_, то есть способ взаимодействия какого-то программного кода с набором каких-то программных компонентов, с помощью которых одна компьютерная программа (например, бот или сайт) может использовать другую программу.

**Ответ: программный интерфейс (программный интерфейс приложения\*)**

27. \_\_\_\_\_ — это фреймворк для вычисления некоторых наборов распределенных задач, реализующий двухуровневую концепцию с хранением промежуточных данных на накопителях.

**Ответ: MapReduce**

28. На \_\_\_\_\_-шаге происходит предварительная обработка входных данных. Для этого один из компьютеров (называемый главным узлом — master node) получает входные данные задачи, разделяет их на части и передает другим компьютерам (рабочим узлам — worker node) для предварительной обработки.

**Ответ: Map**

29. На \_\_\_\_\_-шаге происходит свёртка предварительно обработанных данных. Главный узел получает ответы от рабочих узлов и на их основе формирует результат — решение задачи.

**Ответ: Reduce**

30. В фреймворк \_\_\_\_\_ входят:

ядро, которое отвечает за хранение данных, управление памятью, распределение и отслеживание задач в кластере;

MLlib — набор библиотек для машинного обучения;

SQL — компонент, который отвечает за запрос данных;

GraphX — модуль для работы с графами (абстрактными представлениями связей между множеством объектов);

Streaming — средство для обработки потоков Big Data в реальном времени.

**Ответ: Spark (Apache Spark\*)**

### **3 ТИПОВЫЕ ЗАДАНИЯ НА КОНТРОЛЬНУЮ РАБОТУ, КУРСОВУЮ РАБОТУ/КУРСОВОЙ ПРОЕКТ, РАСЧЕТНО-ГРАФИЧЕСКУЮ РАБОТУ**

Данный вид контроля по дисциплине и предусмотрен учебным планом только для заочной формы обучения.

Темы контрольных работ приведены ниже.

1. Методы сбора и обработки данных в больших информационных системах.
2. Технологии машинного обучения для анализа больших данных.
3. Применение графовых моделей для анализа связей между объектами.
4. Методы визуализации и представления результатов анализа больших данных.
5. Безопасность и защита данных в контексте анализа больших данных.
6. Технологии анализа данных: интеллектуальный анализ данных, алгоритмы кластеризации и классификации.
7. Методы сжатия больших данных: метод главных компонент.
8. Иерархические методы кластерного анализа больших данных.
9. Алгоритмы машинного обучения: линейная регрессия, нейросети, деревья решений.
10. Программное обеспечение для анализа и хранения больших данных: Apache Hadoop, MongoDB, MapReduce, R.

**4 СВЕДЕНИЯ О ФОНДЕ ОЦЕНОЧНЫХ СРЕДСТВ И ЕГО СОГЛАСОВАНИИ**

Фонд оценочных средств для аттестации по дисциплине «Сбор, хранение и анализ больших данных» представляет собой компонент основной профессиональной образовательной программы бакалавриата по направлению подготовки 09.03.02 Информационные системы и технологии, профиль «Проектирование корпоративных информационных систем»

Преподаватель-разработчик – к.т.н. Т. В. Снытникова

Фонд оценочных средств рассмотрен и одобрен заведующим кафедрой прикладной информатики.

Заведующий кафедрой



М.В. Соловей

Фонд оценочных средств рассмотрен и одобрен методической комиссией института цифровых технологий (протокол №5 от 29.08.2024 г).

Председатель методической комиссии



О.С. Витренко