

Федеральное государственное бюджетное образовательное учреждение высшего
образования
«Сахалинский государственный университет»

Кафедра информатики

УТВЕРЖДЕН
на заседании кафедры
«22» мая 2025 г., протокол № 9
И.О. заведующего кафедрой



Осипов Г.С.

**ФОНД
ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

***Б1.В.ДВ.07.01 Технологии организации и управления инфраструктурой
больших данных***

Направление подготовки
09.03.03 Прикладная информатика

профиль
Автоматизированные системы обработки информации и управления

Уровень высшего образования
БАКАЛАВРИАТ

Южно-Сахалинск
2025 г.

1. Формируемые компетенции и индикаторы их достижения по дисциплине

Код компетенции	Содержание компетенции	Код и наименование индикатора достижения компетенции
ПКС-2.	Способен проводить формализацию предметной области с целью создания информационной системы	ПКС-2.1 - Знает требования к компьютерному программному обеспечению; виды технической спецификации на программные компоненты и их взаимодействие; методы проектирование компьютерного программного обеспечения ПКС-2.2 – Умеет применять требования к компьютерному программному обеспечению; разрабатывать технические спецификации на программные компоненты и их взаимодействие; применять методы проектирования компьютерного программного обеспечения; ПКС-2.3 – Владеет методами разработки требований к компьютерному программному обеспечению, технических спецификаций на программные компоненты, методами проектирования компьютерного программного обеспечения.

2. Паспорт фонда оценочных средств по дисциплине (модулю)

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
1.	Тема 1. Анализ систем обработки больших данных	ПКС-2	задания к лабораторным работам
2.	Тема 2. Современные методы визуализации информации Big Data	ПКС-2	задания к лабораторным работам
3.	Тема 3 Детектирование выбросов и аномалий. Очистка данных и технологии регуляризации	ПКС-2	задания к лабораторным работам
4.	Тема 4 Технологии кластеризации и классификации. Нейронные сети. Генетические алгоритмы	ПКС-2	задания к лабораторным работам
5.	Тема 5 Выделение особенностей (Feature detection); нормализация данных. Нечеткие множества. Байесовы сети		
	зачет	ПКС-2	вопросы к зачету.

3. Оценочные средства

Тестирование является одним из средств формального контроля качества обучения – методе, основанном на стандартизированных заданиях, которые позволяют оценить знания, умения и навыки испытуемого

Требования к составлению тестового задания

Тесты должны быть построены по методике, обеспечивающей выполнение требований соответствующего федерального государственного образовательного стандарта.

В тестовых заданиях используются четыре типа вопросов:

– **закрытая** форма – является наиболее распространенной и предлагает несколько альтернативных ответов на поставленный вопрос. Тестовое задание, содержащее вопрос в закрытой форме, включает в себя один или несколько правильных ответов.

Закрытая форма вопросов используется также в тестах-задачах с выборочными ответами. В тестовом задании в этом случае сформулированы условие задачи и все необходимые исходные данные, а в ответах представлены несколько вариантов результата решения в числовом или буквенном виде.

Обучающийся должен решить задачу и показать, какой из представленных ответов он получил.

– **открытая** форма – вопрос представляет собой утверждение, которое необходимо дополнить. Данная форма может быть представлена в тестовом задании, например, в виде словесного текста, формулы (уравнения), графика, в которых пропущены существенные составляющие – части слова или буквы, условные обозначения, линии или изображения элементов схемы или графика.

Обучающийся должен по памяти вставить соответствующие элементы в указанные места («пропусти»).

– **установление соответствия** – в данном случае обучающемуся предлагают два списка, между элементами которых следует установить соответствие;

– **установление последовательности** – предполагает необходимость установить правильную последовательность предлагаемого списка слов или фраз.

Примерный тест

Задания для проверки сформированности компетенций:

ПКС-2 Способен проводить формализацию предметной области с целью создания информационной системы

Множественный выбор (один верный)

1 Выберите правильный ответ:

Какая из перечисленных характеристик НЕ относится к Big Data?

1. Volume (объем)
2. Velocity (скорость)
3. Veracity (достоверность)
4. Variety (разнообразие)

Ответ: 3

2 Выберите правильный ответ:

Какой инструмент чаще всего используется для обработки потоковых данных?

5. Hadoop HDFS
6. Apache Kafka
7. Microsoft Excel
8. Oracle Database

Ответ 2

3 Выберите правильный ответ:

Какой язык программирования наиболее популярен для анализа больших данных?

1. C++
2. Python
3. HTML
4. CSS

Ответ 2

4 Выберите правильный ответ:

Как называется процесс обнаружения полезных шаблонов в больших данных?

1. Data entry
2. Data mining
3. Data cleaning

4. Data visualization

Ответ 2

5 Выберите правильный ответ:

Какой формат данных НЕ является типичным для Big Data?

1. JSON
2. CSV
3. DOCX
4. Parquet

Ответ 3

6 Выберите правильный ответ:

Что означает термин "ETL" в обработке данных?

1. Extract, Transform, Load
2. Extract, Transfer, Load
3. Enter, Test, Leave
4. Export, Translate, Link

Ответ 2

7 Выберите правильный ответ:

Какой облачный сервис предоставляет инструменты для работы с Big Data?

1. Adobe Photoshop
2. AWS (Amazon Web Services)
3. Microsoft Word
4. Zoom

Ответ 2

8 Выберите правильный ответ:

Какой тип базы данных наиболее подходит для хранения неструктурированных данных?

1. Реляционная БД
2. NoSQL БД
3. Таблицы Excel
4. Текстовые файлы

Ответ 2

9 Выберите правильный ответ:

Что такое «машинное обучение» в контексте Big Data?

1. Ручной ввод данных
2. Алгоритмы, которые улучшают свою работу через опыт
3. Печать документов
4. Сканирование изображений

Ответ 2

10 Выберите правильный ответ:

Какой инструмент используется для визуализации больших данных?

1. Блокнот
2. Tableau
3. Калькулятор
4. Браузер

Ответ 2

11 Выберите правильный ответ:

Что такое «реальное время» в обработке Big Data?

1. Данные обрабатываются через месяц
2. Данные обрабатываются мгновенно

3. Данные никогда не обрабатываются
4. Данные печатаются на бумаге

Ответ 2

12 Выберите правильный ответ:

Как называется процесс очистки данных перед анализом?

1. Data visualization
2. Data wrangling
3. Data deletion
4. Data ignoring

Ответ 2

13 Выберите правильный ответ:

Какой метод используется для уменьшения объема данных без потери важной информации?

1. Data duplication
2. Data compression
3. Data expansion
4. Data corruption

Ответ 2

14 Выберите правильный ответ:

Какой алгоритм машинного обучения используется для классификации данных?

1. Линейная регрессия
2. Дерево решений
3. Генерация случайных чисел
4. Сортировка пузырьком

Ответ 2

15 Выберите правильный ответ:

Какой показатель характеризует скорость обработки данных?

1. Объем хранилища
2. Пропускная способность
3. Количество сотрудников
4. Размер экрана монитора

Ответ 3

16 Выберите правильный ответ:

Как называется процесс объединения данных из разных источников?

1. Data separation
2. Data integration
3. Data deletion
4. Data hiding

Ответ 2

17 Выберите правильный ответ:

Какой инструмент используется для SQL-запросов к большим данным?

1. Apache Hadoop
2. Apache Spark
3. Microsoft Paint
4. Adobe Reader

Ответ 2

18 Выберите правильный ответ:

Какие из перечисленных характеристик наиболее существенны при работе с

экономическими данными в больших объемах?

1. Малый объем данных и простота анализа
2. Медленная скорость обработки данных
3. Высокая скорость обработки данных и масштабируемость
4. Низкая точность данных

Ответ 3

19 Выберите правильный ответ:

Какие из перечисленных методов являются типичными для анализа временных рядов в экономике с использованием больших данных?

1. Анализ случайных величин
2. Геостатистика
3. Временные ряды и статистика
4. Анализ текстовых данных

Ответ 3

20 Выберите правильный ответ:

Какая роль машинного обучения в обработке больших данных в экономике?

1. Машинное обучение не применимо к экономическим данным
2. Только для создания анимации и графики
3. Используется исключительно для разработки веб-приложений
4. Прогнозирование и выявление закономерностей в данных

Ответ 4

21 Выберите правильный ответ:

Какой из следующих форматов обычно используется для хранения больших объемов экономических данных?

1. Текстовые файлы (.txt)
2. Аудиофайлы (.mh23)
3. Базы данных формата NoSQL
4. Графические файлы (.jh2g)

Ответ 3

22 Выберите правильный ответ:

Какой процесс отвечает за объединение данных из разных источников для создания общего набора данных для анализа?

1. Декомпозиция данных
2. Классификация данных
3. Data Integration
4. Хранение данных

Ответ 3

23 Выберите правильный ответ:

Что представляет собой понятие «Map2Reduce» в контексте обработки больших данных?

1. Алгоритм сортировки данных
2. Технология сжатия файлов
3. Функция для удаления дубликатов
4. Модель программирования для обработки параллельных данных

Ответ 4

24 Выберите правильный ответ:

Какие из следующих методов используются для обеспечения безопасности больших данных?

1. Использование открытых сетей для передачи данных
2. Хранение данных в неструктурированном формате

3. Шифрование данных
4. Публичный доступ к данным без ограничений

Ответ 3

25 Выберите правильный ответ:

Что представляет собой «Data Warehousing» в контексте больших данных?

1. Процесс сжатия данных
2. Хранение и управление большим объемом структурированных данных
3. Моделирование данных
4. Централизованное хранилище данных для поддержки анализа и отчетности

Ответ 4

26 Выберите правильный ответ:

Что означает термин «Data Mining» в контексте больших данных?

1. Процесс уничтожения данных
2. Этап развертывания данных
3. Извлечение интересных и ранее неизвестных шаблонов из данных
4. Методология сбора данных

Ответ 3

27 Выберите правильный ответ:

Что означает «Big Data Velocity»?

1. Способность обрабатывать данные большого объема
2. Скорость передачи данных
3. Скорость поступления и обработки потоков данных в реальном времени
4. Частота анализа данных

Ответ 3

28 Выберите правильный ответ:

Какие типы данных включаются в «Big Data Variety»?

1. Только структурированные данные
2. Только текстовые данные
3. Только числовые данные
4. Разнообразие структурированных и неструктурированных данных

Ответ 4

29 Выберите правильный ответ:

Какая особенность «Big Data Volume»?

1. Большое количество аналитических инструментов
2. Большой объем хранимых данных
3. Большая скорость обработки данных
4. Огромные объемы данных, превышающие возможности традиционных баз данных

Ответ 4

30 Выберите правильный ответ:

Какие типы данных обычно анализируются в рамках исследований с использованием больших данных?

1. Только числовые данные
2. Текстовые данные
3. Изображения и звуки
4. Разнообразные данные, включая числовые, текстовые и временные ряды

Ответ 4

31 Выберите правильный ответ:

Какое из следующих утверждений о характеристиках больших данных является верным?

1. Большие данные всегда имеют структурированный формат

2. Обработка больших данных обязательно требует использования одной базы данных
3. Большие данные характеризуются объемом, разнообразием и скоростью обработки
4. Большие данные и малые данные идентичны по своим характеристикам

Ответ 3

32 Выберите правильный ответ:

Какова роль алгоритмов машинного обучения в области анализа больших данных?

1. Алгоритмы машинного обучения не применимы к большим данным
2. Они используются только для создания баз данных
3. Алгоритмы машинного обучения помогают выявлять закономерности и делать прогнозы на основе данных
4. Они применяются только в кибербезопасности

Ответ 3

33 Выберите правильный ответ:

Какова роль распределенных баз данных в обработке больших данных?

1. Они не используются в обработке больших данных
2. Распределенные базы данных обеспечивают масштабируемость и параллельную обработку данных
3. Они применяются только для хранения изображений
4. Распределенные базы данных используются только в медицинских исследованиях

Ответ 2

Критерии оценки знаний при проведении тестирования

Отметка **«отлично»** выставляется при условии правильного ответа не менее чем 85% тестовых заданий;

Отметка **«хорошо»** выставляется при условии правильного ответа не менее чем 70 % тестовых заданий;

Отметка **«удовлетворительно»** выставляется при условии правильного ответа не менее 50 %;

Отметка **«неудовлетворительно»** выставляется при условии правильного ответа менее чем на 50 % тестовых заданий.

Результаты текущего контроля используются при проведении промежуточной аттестации