

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Воронежский государственный технический университет»

УТВЕРЖДАЮ

Декан дорожно-транспортного факультета



/ Тюнин В.Л./

21 января 2025 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Распределенные хранение и обработка пространственных данных»

Направление подготовки 05.04.03 Картография и геоинформатика

Профиль Геоинформационное моделирование

Квалификация выпускника Магистр

Нормативный период обучения 2 года

Форма обучения очная

Год начала подготовки 2026

Автор программы

Заведующий кафедрой

Базовая кафедра

кибернетики в системах

организационного

управления

Руководитель ОПОП

В.Е. Белоусов

В.Е. Белоусов

Н.И. Самбулов

Воронеж 2025

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

1.1. Цели дисциплины - формирование у обучающихся необходимых для разработки инновационных программных продуктов профессиональных компетенций, приобретение опыта в области применения методов и алгоритмов обработки данных в информационно-вычислительных системах, связанных с пространственными базами данных.

1.2. Задачи освоения дисциплины

- освоение знаний о программной реализации информационно-вычислительных систем, связанных с пространственными базами данных;
- развитие умений и навыков в области программной реализации информационно-вычислительных систем, связанных с пространственными базами данных;
- формирование компетенций и подготовка к профессиональной деятельности в области программной реализации информационно-вычислительных систем, связанных с пространственными базами данных.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина «Распределенные хранение и обработка пространственных данных» относится к дисциплинам части, формируемой участниками образовательных отношений (дисциплина по выбору) блока Б1.

3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Процесс изучения дисциплины «Распределенные хранение и обработка пространственных данных» направлен на формирование следующих компетенций:

ПК-3 - Способен выполнять технологическое обеспечение и координацию выполнения комплекса операций по развитию и модернизации существующих разноуровневых геоинформационных систем

Компетенция	Результаты обучения, характеризующие сформированность компетенции
ПК-3	знать способы интеграции компонентов и план-график решения задачи
	уметь использовать способы интеграции компонентов и план-график решения задачи
	владеть их видами профессиональной деятельности: навыками использования способов интеграции компонентов и план-график решения задачи

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины «Распределенные хранение и обработка пространственных данных» составляет 5 з.е.

Распределение трудоемкости дисциплины по видам занятий

очная форма обучения

Виды учебной работы	Всего часов	Семестры
		2
Аудиторные занятия (всего)	48	48
В том числе:		
Лекции	16	16
Лабораторные работы (ЛР)	32	32
в том числе в форме практической подготовки	16	16
Самостоятельная работа	105	105
Курсовая работа	+	+
Часы на контроль	27	27
Виды промежуточной аттестации - экзамен	+	+
Общая трудоемкость:		
академические часы	180	180
зач.ед.	5	5

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

5.1 Содержание разделов дисциплины и распределение трудоемкости по видам занятий

очная форма обучения

№ п/п	Наименование темы	Содержание раздела	Лекц	Лаб. зан.	СРС	Всего, час
1	Пространственные объектные модели	Логические конструктивы объектных моделей пространственных БД (содержание и формы). Представление измерений и системы координат. Содержание геосистем. Структура объектных моделей геосистем, объекты и их связи, простые и сложные объекты. Типы объектных форм. Строеение иерархической модели – состав объектов и группировок, структура связей.	4	6	16	26
		<i>практическая подготовка обучающихся</i>	-	4	-	4
2	Средства и функции, структуры и элементы	Структура БД и СУБД. Основные понятия и элементы: индексы, ключи, таблицы, схемы, языки. Особенности пространственных БД и СУБД. Ключевые понятия: типы структур и баз данных; множества; таблица, домен, свойство и атрибут, кортеж и отношение; отображения. Концептуальная и логическая модели данных. Модель «сущность-связь». Классификация отображений. Реляционная модель данных. Язык SQL. Спецификации данных и операторы манипулирования ими. Язык запросов, типы запросов и вычисляемые поля в запросах. Создание и редактирование БД. Использование реляционных коммерческих СУБД. Функции СУБД и их место в ГИС.	4	6	18	28
		<i>практическая подготовка обучающихся</i>	-	4	-	4
3	Создание объектных структур; цифровая карта.	Порядок проектирования пространственных БД. Модель в информационном обеспечении ГИС. Выбор модели. Адекватность модели предметной области и потребностям. Анализ потребностей. Создание схемы предметной области. Отображение схемы предметной области на схему базы данных. Организация объектов, воплощение межобъектных связей и сложных объектов. Объектная классификация и легенда карты.	2	6	18	26

		Отображение (изображение) геосистемы. Понятие цифровой и электронной карты.				
		<i>практическая подготовка обучающихся</i>	-	2	-	2
4	Хранение и индексирование пространственных данных	Пространственное индексирование. Файлы решетки. R-деревья. Декомпозиция объектов при помощи TR*-деревьев. Индекс пространственного соединения. Разработка индексов в пространственных базах данных. Инструментальные средства разработки и библиотеки повторно используемых модулей для индексов. Документирование разработанных индексов. Оценка сложности, трудоёмкости и сроков разработки индексов в пространственных базах данных. Сопровождение созданных индексов.	2	6	18	26
		<i>практическая подготовка обучающихся</i>	-	2	-	2
5	Пространственные построения.	Манипулирование пространственными данными – пространственные построения (создание буферных зон, нахождение пересечения объектов и так далее). Пространственные предикаты и работа с ними. Понятие ГИС-модели.	2	4	18	24
		<i>практическая подготовка обучающихся</i>	-	2	-	2
6	Организация и работы, средства и ресурсы.	Способы оценки качества данных и контроль ошибок. Создание каталогов метаданных. Хранение и обновление данных в ГИС. Виды источников, материалов и продуктов в цифровой картографии. Интеграция разнотипных данных. Унифицированный программный интерфейс ODBC. Сетевые и распределённые базы и обработка данных. Модель и технология клиент-сервер. Файловый сервер. Понятие базы геоданных. SQL серверы. Языки описания и манипулирования данными. Языки запросов. Стандарты. Интернет ресурсы БД.	2	4	17	23
		<i>практическая подготовка обучающихся</i>	-	2	-	2
Итого			16	32	105	153

Практическая подготовка при освоении дисциплины (модуля) проводится путем непосредственного выполнения обучающимися отдельных элементов работ, связанных с будущей профессиональной деятельностью, способствующих формированию, закреплению и развитию практических навыков и компетенций по профилю соответствующей образовательной программы на практических занятиях и (или) лабораторных работах:

№ п/п	Перечень выполняемых обучающимися отдельных элементов работ, связанных с будущей профессиональной деятельностью	Формируемые профессиональные компетенции
1	Содержание геосистем. Структура объектных моделей геосистем, объекты и их связи, простые и сложные объекты.	ПК-3
2	Классификация отображений. Реляционная модель данных.	ПК-3
3	Создание схемы предметной области	ПК-3
4	Задачи, решаемые с помощью пространственных БД, и области их использования.	ПК-3
5	Пространственные предикаты и работа с ними. Понятие ГИС-модели.	ПК-3
6	Унифицированный программный интерфейс ODBC. Сетевые и распределённые базы и обработка данных.	ПК-3

5.2 Перечень лабораторных работ

1. Восстановление навыка работы с используемыми (ранее изучавшимися) типовыми программными инструментальными средствами СУБД и МГ или первичное изучение и освоение таких (новых) средств.

2. Ознакомление с образцами пространственных БД, выбор фрагмента образца для исследования согласно заданной теме. Изучение модели, заложенной в такую БД и восстановление её структуры. Сверка с имеющимися классификаторами или восполнение их.

3. Выбор фрагмента БД для исследования. Исследование конструктива БД – атрибутивной составляющей с определением характеристик. Выявление (выверка) и/или исправление ошибок или недочётов.

4. Ознакомление с картографическим изображением исследуемой БД. Выборочное исследование соответствия между объектными информационными структурами геосистемы и её отображением.

5. Оценка исследуемой БД и проектирование дополнений к атрибутивной составляющей исследуемого фрагмента по заданной теме и в заданном объёме.

6. Сравнение исследуемой БД со схожими реализациями. Подготовка отчёта и полученных информационных продуктов в электронном виде. Защита работ.

6. ПРИМЕРНАЯ ТЕМАТИКА КУРСОВЫХ ПРОЕКТОВ (РАБОТ) И КОНТРОЛЬНЫХ РАБОТ

В соответствии с учебным планом освоение дисциплины предусматривает выполнение курсовой работы в 2 семестре для очной формы обучения.

Примерная тематика курсовой работы: «Распределённая СУБД для интегрированной обработки пространственных данных в ГИС»

Задачи, решаемые при выполнении курсовой работы:

- Форматы и структуры данных для моделирования ПРД.
- Распределённая СУБД ГИС Терра.
- Временные оценки.

Курсовая работа включает в себя графическую часть и расчетно-пояснительную записку.

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

7.1. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

7.1.1 Этап текущего контроля

Результаты текущего контроля знаний и межсессионной аттестации оцениваются по следующей системе:

- «аттестован»;
- «не аттестован».

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Аттестован	Не аттестован
ПК-3	знать способы интеграции компонентов и план-график решения задачи	укажите критерий	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	уметь использовать способы интеграции компонентов и план-график решения задачи	укажите критерий	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	владеть их видами профессиональной деятельности: навыками использования способов интеграции компонентов и план-график решения задачи	укажите критерий	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах

7.1.2 Этап промежуточного контроля знаний

Результаты промежуточного контроля знаний оцениваются в 2 семестре для очной формы обучения по четырехбалльной системе:

«отлично»;

«хорошо»;

«удовлетворительно»;

«неудовлетворительно».

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Отлично	Хорошо	Удовл.	Неудовл.
ПК-3	знать способы интеграции компонентов и план-график решения задачи	Тест	Выполнение теста на 90-100%	Выполнение теста на 80-90%	Выполнение теста на 70-80%	В тесте менее 70% правильных ответов
	уметь использовать способы интеграции компонентов и план-график решения задачи	Решение стандартных практических задач	Задачи решены в полном объеме и получены верные ответы	Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены
	владеть их видами профессиональной деятельности: навыками использования способов интеграции компонентов и план-график решения задачи	Решение прикладных задач в конкретной предметной области	Задачи решены в полном объеме и получены верные ответы	Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены

7.2 Примерный перечень оценочных средств (типовые контрольные задания или иные материалы, необходимые для оценки

знаний, умений, навыков и (или) опыта деятельности)

7.2.1 Примерный перечень заданий для подготовки к тестированию

1. Алгоритм расширения выборки

А) метод подразумевает построение наиболее вероятных «неправильных» вариантов поискового шаблона, т.е. строится множество всевозможных «ошибочных» слов, например, получающихся из исходного в результате одной операции редактирования, после чего построенные термины сравниваются на точное соответствие.

Б) алгоритмов сравнения двух строк по их звучанию, он устанавливает одинаковый индекс для строк, имеющих схожее звучание в языке согласно заданной таблице схожих по звучанию символов и их сочетаний.

В) позволяет для двух строк найти расстояние Левенштейна — минимальное количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую

Г) позволяет для двух строк найти расстояние Левенштейна — максимальное количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую, по которому определяется степень совпадения строк

Д) позволяет для двух строк найти расстояние Левенштейна — среднее количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую

2. Алгоритм Вагнера-Фишера

А) позволяет для двух строк найти расстояние Левенштейна — минимальное количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую, по которому определяется степень совпадения строк

Б) метод подразумевает построение наиболее вероятных «неправильных» вариантов поискового шаблона, т.е. строится множество всевозможных «ошибочных» слов, например, получающихся из исходного в результате одной операции редактирования, после чего построенные термины сравниваются на точное соответствие.

В) алгоритмов сравнения двух строк по их звучанию, он устанавливает одинаковый индекс для строк, имеющих схожее звучание в языке согласно заданной таблице схожих по звучанию символов и их сочетаний.

Г) позволяет для двух строк найти расстояние Левенштейна — максимальное количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую, по которому определяется степень совпадения строк

Д) позволяет для двух строк найти расстояние Левенштейна — среднее количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую

3. Система управления центра обработки данных предназначена

А) для общего управления системой, внешнего доступа, обеспечения аутентификации, авторизации и аккаунтинга и представления результатов пользователям

Б) для общего управления системой, внешнего контроля, обеспечения аутентификации, авторизации и аккаунтинга и представления результатов пользователям

В) для общего управления системой, внешнего доступа, обеспечения контроля

Г) ускорения работы системы, внешнего доступа, обеспечения аутентификации, авторизации и аккаунтинга и представления результатов пользователям

Д) для общего управления системой, внешнего доступа, обеспечения и представления результатов пользователям

4. Целью кластеризации является:

А) упрощение дальнейшей обработки данных, сокращение объема хранимых данных, выделение нетипичных объектов, построение иерархии множества объектов

Б) упрощение дальнейшей обработки данных, сокращение объема хранимых данных, выделение нетипичных объектов, построение критических диаграмм

В) упрощение дальнейшей обработки данных, сокращение объема хранимых данных, выделение основных параметров выборки, построение иерархии множества объектов

Г) проверка подчинения предполагаемому закону распределения, упрощение дальнейшей обработки данных, сокращение объема хранимых данных,

Д) упрощение дальнейшей обработки данных, сокращение объема хранимых данных, выделение нетипичных объектов, построение законов распределения

5. Метод k ближайших соседей это метрический классификатор

А) в котором берется k число ближайших соседей, и объект относится к тому классу, к которому относится большинство его соседей

Б) который основан на построении разделяющей линейной поверхности k-мерного пространства

В) который берет в расчет k параметров вероятностного распределения признака (средний и дисперсия). ближайших соседей

Г) который использует k ближайших соседей для построении разделяющей линейной поверхности

Д) который берет в расчет k параметров признака в генеральной совокупности ближайших соседей

6. Обучение без учителя:

А) в этом случае ответы не задаются и приходится искать зависимости между заданными объектами, выполняются задачи кластеризации, когда необходимо сгруппировать объекты в кластеры, пользуясь данными о сходстве парных объектов

Б) это задание пар «объект, ответ», для нахождения функциональной зависимости ответов от описаний необходимо составить алгоритм, который на входе принимает описание, а на выходе дает ответ

В) это обучение с привлечением человека, выполняющего функции учителя

Г) это обучение с привлечением специального приложения, выполняющего функции учителя

Д) это алгоритм классификации, который основан на построении разделяющей линейной поверхности

7. Установите соответствие

А) Структурированные данные - данные определяющие конкретную предметную область, упорядоченные специальным образом и организованные так, чтобы над такими данными можно было выполнить анализ

Б) Полуструктурированные данные - не соответствуют чёткой структуре таблиц и отношений в реляционных базах данных, однако такие данные содержат специальные теги и иные маркёры, позволяющие отделить семантические элементы

В) Квазиструктурированные данные - текстовые данные с неустойчивым форматом, которые для обработки инструментами требуют больших временных затрат на преобразование

8. Установите соответствие

А) Метод обучения ассоциативным правилам - метод используемый для обучения машин способам обнаружения зависимостей между данными в больших базах данных

Б) Метод классификации - методы категоризации новых данных на основе принципов, ранее применённых к уже наличествующим данным

В) Метод кластерного анализа - статистический метод классификации объектов, который приводит к разделению разнообразных групп на более мелкие группы подобных (сходных) объектов, для которых критерий подобия заранее не известен

9. Установите соответствие

А) Хранилище данных - технология работы с Big data, которая представляет собой предметно-ориентированную информационную базу данных, специально разработанную и предназначенную для подготовки отчётов и анализа данных с целью поддержки принятия решений в организации

Б) Распределенная система- это технология работы с Big data, которая

представляет собой множество компьютеров, взаимодействующих по сети и объединенных для решения общей вычислительной задачи

В) Hadoop- это технология работы с Big data, которая представляет собой свободно распространяемый набор утилит, библиотек и фреймворк для разработки и выполнения распределённых программ, работающих на кластерах из сотен и тысяч узлов

Г) MapReduce - это технология работы с Big data, которая представляет собой модель распределённых вычислений. используемая для параллельных вычислений над очень большими, вплоть до нескольких петабайт, наборами данных в компьютерных кластерах

10. Установите соответствие

А) Сетевой анализ - набор методов, используемых для описания и анализа отношений между дискретными узлами в графе или сети

Б) Распознавание образов - набор методов машинного обучения, развивающих основы и методы классификации и идентификации предметов, явлений, процессов, сигналов, ситуаций и т. п. объектов, которые характеризуются конечным набором некоторых свойств и признаков

В) Кластерграмма - метод визуализации, использующийся при кластерном анализе и показывающий, как отдельные элементы множества данных соотносятся с кластерами по мере изменения их количества

Г) Машинное обучение - класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение в процессе применения решений множества сходных задач

7.2.2 Примерный перечень заданий для решения стандартных задач

1. Алгоритм расширения выборки

А) метод подразумевает построение наиболее вероятных «неправильных» вариантов поискового шаблона, т.е. строится множество всевозможных «ошибочных» слов, например, получающихся из исходного в результате одной операции редактирования, после чего построенные термины сравниваются на точное соответствие.

Б) алгоритмов сравнения двух строк по их звучанию, он устанавливает одинаковый индекс для строк, имеющих схожее звучание в языке согласно заданной таблице схожих по звучанию символов и их сочетаний.

В) позволяет для двух строк найти расстояние Левенштейна — минимальное количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую

Г) позволяет для двух строк найти расстояние Левенштейна — максимальное количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую, по которому определяется степень совпадения строк

Д) позволяет для двух строк найти расстояние Левенштейна — среднее количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения

одной строки в другую

2. Алгоритм Вагнера-Фишера

А) позволяет для двух строк найти расстояние Левенштейна — минимальное количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую, по которому определяется степень совпадения строк

Б) метод подразумевает построение наиболее вероятных «неправильных» вариантов поискового шаблона, т.е. строится множество всевозможных «ошибочных» слов, например, получающихся из исходного в результате одной операции редактирования, после чего построенные термины сравниваются на точное соответствие.

В) алгоритмов сравнения двух строк по их звучанию, он устанавливает одинаковый индекс для строк, имеющих схожее звучание в языке согласно заданной таблице схожих по звучанию символов и их сочетаний.

Г) позволяет для двух строк найти расстояние Левенштейна — максимальное количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую, по которому определяется степень совпадения строк

Д) позволяет для двух строк найти расстояние Левенштейна — среднее количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую

3. Система управления центра обработки данных предназначена

А) для общего управления системой, внешнего доступа, обеспечения аутентификации, авторизации и аккаунтинга и представления результатов пользователям

Б) для общего управления системой, внешнего контроля, обеспечения аутентификации, авторизации и аккаунтинга и представления результатов пользователям

В) для общего управления системой, внешнего доступа, обеспечения контроля

Г) ускорения работы системы, внешнего доступа, обеспечения аутентификации, авторизации и аккаунтинга и представления результатов пользователям

Д) для общего управления системой, внешнего доступа, обеспечения и представления результатов пользователям

4. Целью кластеризации является:

А) упрощение дальнейшей обработки данных, сокращение объема хранимых данных, выделение нетипичных объектов, построение иерархии множества объектов

Б) упрощение дальнейшей обработки данных, сокращение объема хранимых данных, выделение нетипичных объектов, построение критических

диаграмм

В) упрощение дальнейшей обработки данных, сокращение объема хранимых данных, выделение основных параметров выборки, построение иерархии множества объектов

Г) проверка подчинения предполагаемому закону распределения, упрощение дальнейшей обработки данных, сокращение объема хранимых данных,

Д) упрощение дальнейшей обработки данных, сокращение объема хранимых данных, выделение нетипичных объектов, построение законов распределения

5. Метод k ближайших соседей это метрический классификатор

А) в котором берется k число ближайших соседей, и объект относится к тому классу, к которому относится большинство его соседей

Б) который основан на построении разделяющей линейной поверхности k-мерного пространства

В) который берет в расчет k параметров вероятностного распределения признака (средний и дисперсия). ближайших соседей

Г) который использует k ближайших соседей для построении разделяющей линейной поверхности

Д) который берет в расчет k параметров признака в генеральной совокупности ближайших соседей

6. Обучение без учителя:

А) в этом случае ответы не задаются и приходится искать зависимости между заданными объектами, выполняются задачи кластеризации, когда необходимо сгруппировать объекты в кластеры, пользуясь данными о сходстве парных объектов

Б) это задание пар «объект, ответ», для нахождения функциональной зависимости ответов от описаний необходимо составить алгоритм, который на входе принимает описание, а на выходе дает ответ

В) это обучение с привлечением человека, выполняющего функции учителя

Г) это обучение с привлечением специального приложения, выполняющего функции учителя

Д) это алгоритм классификации, который основан на построении разделяющей линейной поверхности

7. Установите соответствие

А) Структурированные данные - данные определяющие конкретную предметную область, упорядоченные специальным образом и организованные так, чтобы над такими данными можно было выполнить анализ

Б) Полуструктурированные данные - не соответствуют чёткой структуре таблиц и отношений в реляционных базах данных, однако такие данные содержат специальные теги и иные маркёры, позволяющие отделить

семантические элементы

В) Квазиструктурированные данные - текстовые данные с неустойчивым форматом, которые для обработки инструментами требуют больших временных затрат на преобразование

8. Установите соответствие

А) Метод обучения ассоциативным правилам - метод используемый для обучения машин способам обнаружения зависимостей между данными в больших базах данных

Б) Метод классификации - методы категоризации новых данных на основе принципов, ранее применённых к уже наличествующим данным

В) Метод кластерного анализа - статистический метод классификации объектов, который приводит к разделению разнообразных групп на более мелкие группы подобных (сходных) объектов, для которых критерий подобия заранее не известен

9. Установите соответствие

А) Хранилище данных - технология работы с Big data, которая представляет собой предметно-ориентированную информационную базу данных, специально разработанную и предназначенную для подготовки отчётов и анализа данных с целью поддержки принятия решений в организации

Б) Распределенная система- это технология работы с Big data, которая представляет собой множество компьютеров, взаимодействующих по сети и объединенных для решения общей вычислительной задачи

В) Hadoop- это технология работы с Big data, которая представляет собой свободно распространяемый набор утилит, библиотек и фреймворк для разработки и выполнения распределённых программ, работающих на кластерах из сотен и тысяч узлов

Г) MapReduce - это технология работы с Big data, которая представляет собой модель распределённых вычислений. используемая для параллельных вычислений над очень большими, вплоть до нескольких петабайт, наборами данных в компьютерных кластерах

10. Предположим, что некоторая база данных содержит поля

ФАМИЛИЯ, ГОД РОЖДЕНИЯ, ДОХОД. При поиске по условию ГОД РОЖДЕНИЯ > 1958 AND ДОХОД < 3500 будут найдены фамилии лиц:

А) имеющих доход не менее 3500, и старше тех, кто родился в 1958 году.

Б) имеющих доход менее 3500, или тех, кто родился в 1958 году и позже;

В) имеющих доход менее 3500, и родившихся в 1958 году и позже;

Г) имеющих доход менее 3500, и родившихся в 1959 году и позже;

Д) имеющих доход менее 3500, и тех, кто родился в 1958 году;

7.2.3 Примерный перечень заданий для решения прикладных задач

1. Выберите метрики качества, которые можно использовать при решении задачи классификации.

- а) коэффициент детерминации
- б) F-мера
- в) площадь под ROC-кривой
- г) доля правильных ответов
- д) площадь под PR-кривой

2. Магазин вел статистику своих продаж смартфонов в течение года. В базе имеется информация о количестве работавших сотрудников, средней, максимальной и минимальной стоимости смартфона и количестве единиц продукции, имеющейся в наличии, количестве проданных за день смартфонов и объеме дневной выручки. Требуется спрогнозировать выручку магазина на ближайший месяц. Какую задачу машинного обучения придется решать?

- а) задача реализации метода наименьших квадратов
- б) задача кластеризации
- в) задача многоклассовой классификации
- г) задача поиска аномалий
- д) задача регрессии

3. Какие функции *numpy* можно использовать для формирования числовой последовательности. Например, нужно сформировать последовательность всех четных чисел из промежутка [1; 100]. Правильных ответов может быть несколько.

- а) `linspace`
- б) `arange`
- в) `hstack`
- г) `array`

4. Дан кортеж *A*. Как можно сформировать список *B* из тех же значений? Правильных ответов может быть несколько.

- а) `B = list(A)`
- б) `B = tuple(A)`
- в) `B = [elem for elem in A]`
- г) `B = (elem for elem in A)`
- д) `B = [A[i] for i in range(len(A))]`

5. Сколько функций будет на графике, построенном командой
`import matplotlib.pyplot as plt`
`plt.plot(x, x, '-', x, 2*x, '--', x, 3*x, ':', x, 4*x, '-')`

Ответ: _____

6. Даны две строки с произвольным текстом на русском языке. Напишите код на Python, который позволит определить схожесть текстов на основе метрики косинусного расстояния.

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Косинусное расстояние – это мера похожести двух векторов. Векторами считать частоты встречаемости слов в текстах.

7. Дан список A с координатами точек (x, y) в двумерном пространстве. Каждая точка хранится кортежем.

Как отсортировать эти точки по ключу **x(убыв.) + y (убыв.)**

- а) A.sort() A.reverse()
- б) A.sort().reverse()
- в) A.sort(key=lambda elem: (-elem[0], -elem[1]))
- г) A.sort(key=lambda elem: (elem[0], elem[1]), reverse=True)
- д) A.sort(reverse=True)

8. Штатный психолог в школе заболел. Для заполнения отчета требуется срочно всех учеников школы разбить на две группы по психологическому типу: интроверты и экстраверты. Для решения задачи будут использованы данные психологического тестирования и профили активностей учеников в социальных сетях. Какая задача машинного обучения здесь будет решаться?

- а) задача кластеризации
- б) задача классификации
- в) Задача ранжирования
- г) Задача регрессии
- д) Задача визуализации

9. Какие алгоритмы из библиотеки sklearn можно использовать для решения задач восстановления регрессии. Правильных ответов может быть несколько.

- а) Ridge
- б) Lasso
- в) LinearRegression
- г) SGDRegressor
- д) MakeRegression

10. Дана строка с произвольным текстом на русском языке. Напишите код на Python, который позволит построить таблицу с частотой встречаемости гласных букв, присутствующих в данной строке. Для решения задачи обязательно использовать словарь.

Ответ: _____

7.2.4 Примерный перечень вопросов для подготовки к зачету

Не предусмотрено учебным планом

7.2.5 Примерный перечень заданий для подготовки к экзамену

1. Предмет и место дисциплины. Основные понятия и термины.
2. Задачи и области применения пространственных БД (предметы, темы и потребители).
3. Развёрнутое понятие геосистемы и объекта (геообъекта).
4. Конструктивы пространственного моделирования (содержание и формы).
5. Представление измерений и системы координат.
6. Содержание геосистем.
7. Структура объектных геосистем.
8. Типы объектных моделей (геосистем) по строению.
9. Строение иерархической модели – состав группировок и структура связей.
10. Понятие о БД и СУБД. Индексирование данных и ключи. Таблицы.
11. Схема данных и схема БД. Языки описания и манипулирования данными.
12. Особенности пространственных БД и СУБД.
13. Концептуальная и логическая модели данных.
14. Модель «сущность-связь». Классификация отображений.
15. Реляционная модель данных.
16. Язык SQL. Спецификации данных и операторы манипулирования ими.
17. Язык запросов, типы запросов и вычисляемые поля в запросах.
18. Создание и редактирование БД. Использование реляционных СУБД.
19. Функциональные возможности пространственных БД и СУБД и их место в ГИС.
20. Привлечение средств МГ для картографического отображения.
21. Атрибутивная и позиционная части; взаимоувязка и изображение.
22. Растровая и векторная модели МГ, их создание и редактирование.
23. Этапы проектирования пространственной БД.
24. Модель в информационном обеспечении ГИС.
25. Выбор модели. Адекватность модели. Анализ потребностей.
26. Создание схемы предметной области и отображение её на схему БД.
27. Организация простых и сложных объектов, воплощение связей.
28. Объектная классификация и легенда карты.
29. Изображение геосистемы. Понятие цифровой и электронной карты.
30. Аналитические запросы и операции измерений в пространственной СУБД.
31. Пространственные построения и манипулирование данными.
32. Пространственные предикаты.
33. Способы оценки качества данных и контроль ошибок. Создание каталогов метаданных.
34. Хранение и обновление данных в ГИС.

35. Виды источников, материалов и продуктов в цифровой картографии.
36. Интеграция разнотипных данных.
37. Унифицированный программный интерфейс ODBC. Технические характеристики.
38. Сетевые и распределённые базы и обработка данных. Модель и технология клиентсервер. Файловый сервер.
39. Понятие базы геоданных.
40. SQL серверы. Языки описания и манипулирования данными. Языки запросов QBE и SQL.
41. Стандарты OGC и их поддержка в пространственной СУБД.
42. Интернет-ресурсы с пространственными БД по протоколам OGC.

7.2.6. Методика выставления оценки при проведении промежуточной аттестации

Экзамен проводится по тест-билетам, каждый из которых содержит 10 вопросов и задачу. Каждый правильный ответ на вопрос в тесте оценивается 1 баллом, задача оценивается в 10 баллов (5 баллов верное решение и 5 баллов за верный ответ). Максимальное количество набранных баллов – 20.

1. Оценка «Неудовлетворительно» ставится в случае, если студент набрал менее 6 баллов.

2. Оценка «Удовлетворительно» ставится в случае, если студент набрал от 6 до 10 баллов

3. Оценка «Хорошо» ставится в случае, если студент набрал от 11 до 15 баллов.

4. Оценка «Отлично» ставится, если студент набрал от 16 до 20 баллов.

7.2.7 Паспорт оценочных материалов

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции	Наименование оценочного средства
1	Пространственные объектные модели	ПК-3	Тест, защита лабораторных работ, требования к курсовой работе, зачет
2	Средства и функции, структуры и элементы	ПК-3	Тест, защита лабораторных работ, требования к курсовой работе, зачет
3	Создание объектных структур; цифровая карта.	ПК-3	Тест, защита лабораторных работ, требования к курсовой работе, зачет
4	Хранение и индексирование пространственных данных	ПК-3	Тест, защита лабораторных работ, требования к курсовой работе, зачет
5	Пространственные построения.	ПК-3	Тест, защита лабораторных работ, требования к курсовой работе, зачет
6	Организация и работы, средства и ресурсы.	ПК-3	Тест, защита лабораторных работ, требования к курсовой работе, зачет

7.3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Тестирование осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных тест-заданий на бумажном носителе. Время тестирования 30 мин. Затем осуществляется проверка теста экзаменатором и выставляется оценка согласно методики выставления оценки при проведении промежуточной аттестации.

Решение стандартных задач осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методики выставления оценки при проведении промежуточной аттестации.

Решение прикладных задач осуществляется, либо при помощи компьютерной системы тестирования, либо с использованием выданных задач на бумажном носителе. Время решения задач 30 мин. Затем осуществляется проверка решения задач экзаменатором и выставляется оценка, согласно методики выставления оценки при проведении промежуточной аттестации.

Защита курсовой работы, курсового проекта или отчета по всем видам практик осуществляется согласно требованиям, предъявляемым к работе, описанным в методических материалах. Примерное время защиты на одного студента составляет 20 мин.

8 УЧЕБНО МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

8.1 Перечень учебной литературы, необходимой для освоения дисциплины

1. Геоинформационные системы : учебное пособие / составители О. Л. Гиниятуллина, Т. А. Хорошева. — Кемерово : КемГУ, 2018. — 122 с. — ISBN 978-5-8353-2232-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/120040>
2. Жуковский, О. И. Геоинформационные системы : учебное пособие / О. И. Жуковский. — Томск : Томский государственный университет систем управления и радиоэлектроники, Эль Контент, 2014. — 130 с. — ISBN 978-5-4332-0194-1. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/72081.html>
3. Инженерная геодезия и геоинформатика. Краткий курс : учебник для вузов / М. Я. Брынь, Е. С. Богомоллова, В. А. Коугия [и др.] ; под редакцией В. А. Коугия. — 4-е изд., стер. — Санкт-Петербург : Лань, 2025. — 288 с. — ISBN 978-5-507-50468-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/438974>
4. Подрядчикова, Е. Д. Инструментальные средства ГИС : учебное пособие / Е. Д. Подрядчикова. — Тюмень : ТюмГНГУ, 2018. — 86 с. — ISBN 978-5-9961-1887-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/138256>

8.2 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень лицензионного программного обеспечения, ресурсов информационно-телекоммуникационной сети «Интернет», современных профессиональных баз данных и информационных справочных систем:

Лицензионное программное обеспечение

1. WIN HOME 10 32-bit/64-bit All Lng PK Lic Online DwnLd NR
2. nanoCAD

Свободное ПО

1. 7zip
2. Adobe Acrobat Reader
3. Google Chrome
4. HeidiSQL
5. HK-Software IBExpert Personal Edition
6. LibreOffice
7. Moodle
8. QGIS
9. SQLite
10. STDU Viewer
11. WinDjView

Информационные справочные системы

1. Образовательный портал ВГТУ
<https://old.education.cchgeu.ru/>
2. КонсультантПлюс правовая поддержка
<http://www.consultant.ru/>
3. Электронно-библиотечная система Лань
<https://e.lanbook.com/>
4. База данных «Цифровая библиотека IPRsmart (IPRsmart ONE)»
<http://www.iprbookshop.ru/>
5. Natural Earth Data:

Предлагает векторные и растровые картографические данные в различных масштабах, идеально подходящие для исторических и политических карт.

<https://www.naturalearthdata.com/downloads/>

6. USGS Earth Explorer:

Предоставляет доступ к спутниковым снимкам, аэрофотосъемке и наборам данных о земле.

<https://earthexplorer.usgs.gov/>

7. Esri Open Data Hub:

Платформа для доступа к широкому спектру географических данных.

<https://hub.arcgis.com/search>

8. OpenStreetMap:

Совместный проект по созданию бесплатной редактируемой карты мира.

<https://gisgeography.com/openstreetmap-download-osm-data/>

9. Центр социально-экономических данных и приложений НАСА (SEDAC):

Сосредоточен на взаимодействии человека с окружающей средой.

<https://earthdata.nasa.gov/centers/sedac-daac>

10. Открытая топография:

Специализируется на наборах данных высокого разрешения о земной поверхности, в основном на топографических данных.

<https://opentopography.org/>

11. UNEP Environmental Data Explorer:

Содержит наборы данных, относящихся к экологическим исследованиям, от Программы ООН по окружающей среде.

<https://www.unep.org/publications-data>

12. ArcGIS Living Atlas of the World:

Это крупнейшая коллекция географической информации со всего мира. Он включает карты, приложения, слои данных и многое другое.

<https://livingatlas.arcgis.com/en/home/>

Terra Populus:

Интегрирует данные о населении и окружающей среде.

<https://terra.ipums.org/>

9 МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

Реализация дисциплины «Распределенные хранение и обработка пространственных данных» требует наличия учебной аудитории для проведения учебных занятий

Оборудование учебной аудитории: комплект учебной мебели:

- рабочее место преподавателя (стол, стул);
- рабочие места обучающихся (столы, стулья);

Учебная аудитория для проведения практических работ Лаборатория "Компьютерный класс"/ Лаборатория "Математической обработки результатов геодезических измерений, информационного обеспечения кадастра"

Оборудование учебного кабинета: комплект учебной мебели:

- рабочее место преподавателя (стол, стул);
- рабочие места обучающихся (столы, стулья);

Технические средства обучения:

- Персональные компьютеры с установленным ПО, подключенные к сети Интернет – 14 шт.

Помещение для самостоятельной работы «Аудитория для групповых и

индивидуальных консультаций/ Аудитория для самостоятельной работы».

Оборудование кабинета: комплект учебной мебели:

- рабочее место преподавателя (стол, стул);
- рабочие места обучающихся (столы, стулья)

Технические средства обучения:

- интерактивная доска Trace Board TS6080B;

персональный компьютер с возможностью подключения к сети "Интернет" и обеспечением доступа к электронной информационно-образовательной среде вуза

10. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

По дисциплине «Распределенные хранение и обработка пространственных данных» читаются лекции, проводятся лабораторные работы, выполняется курсовая работа.

Основой изучения дисциплины являются лекции, на которых излагаются наиболее существенные и трудные вопросы, а также вопросы, не нашедшие отражения в учебной литературе.

Лабораторные работы выполняются на лабораторном оборудовании в соответствии с методиками, приведенными в указаниях к выполнению работ.

Методика выполнения курсовой работы изложена в учебно-методическом пособии. Выполнять этапы курсовой работы должны своевременно и в установленные сроки.

Контроль усвоения материала дисциплины производится проверкой курсовой работы, защитой курсовой работы.

Вид учебных занятий	Деятельность студента
Лекция	Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; пометать важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначение вопросов, терминов, материала, которые вызывают трудности, поиск ответов в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на лекции или на практическом занятии.
Лабораторная работа	Лабораторные работы позволяют научиться применять теоретические знания, полученные на лекции при решении конкретных задач. Чтобы наиболее рационально и полно использовать все возможности лабораторных для подготовки к ним необходимо: следует разобрать лекцию по соответствующей теме, ознакомиться с соответствующим разделом учебника, проработать дополнительную литературу и источники, решить задачи и выполнить другие письменные задания.
Самостоятельная работа	Самостоятельная работа студентов способствует глубокому усвоения учебного материала и развитию навыков самообразования. Самостоятельная работа предполагает следующие составляющие: - работа с текстами: учебниками, справочниками, дополнительной

	литературой, а также проработка конспектов лекций; - выполнение домашних заданий и расчетов; - работа над темами для самостоятельного изучения; - участие в работе студенческих научных конференций, олимпиад; - подготовка к промежуточной аттестации.
Подготовка к промежуточной аттестации	Готовиться к промежуточной аттестации следует систематически, в течение всего семестра. Интенсивная подготовка должна начаться не позднее, чем за месяц-полтора до промежуточной аттестации. Данные перед экзаменом три дня эффективнее всего использовать для повторения и систематизации материала.

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

№ п/п	Перечень вносимых изменений	Дата внесения изменений	Подпись заведующего кафедрой, ответственной за реализацию ОПОП
----------	-----------------------------	----------------------------	--