

Министерство науки и высшего образования Российской Федерации

федеральное государственное бюджетное образовательное учреждение
высшего образования
РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ
УНИВЕРСИТЕТ

Кафедра Прикладной информатики

Рабочая программа дисциплины

Обработка и анализ данных

Основная профессиональная образовательная программа
высшего образования по направлению подготовки

09.03.03 Прикладная информатика

Направленность (профиль):

Прикладные информационные системы и технологии

Уровень:

Бакалавриат

Форма обучения

Очная

Согласовано
Руководитель ОПОП

Яготинцева Н.В. Яготинцева Н.В.

Утверждаю

Председатель УМС И.И. Палкин И.И. Палкин

Рекомендована решением

Учебно-методического совета

22 10 2019 г., протокол № 2

Рассмотрена и утверждена на заседании кафедры

22 сентября 2019 г., протокол № 2

Зав. кафедрой Истомин Е.П. /Истомин Е.П.

Авторы-разработчики:

Колбина О.Н. /Колбина О.Н.

Яготинцева Н.В. /Яготинцева Н.В.

Полковник И.И. /Полковник И.И.

Санкт-Петербург 2019

1. Цель и задачи освоения дисциплины

Цель дисциплины – формирование у студентов знаний по теоретическим аспектам организации хранения и использования больших объемов данных в информационных системах, основным вопросам проектирования, создания и использования баз данных.

Основные задачи дисциплины:

Приобретения умений и навыков проводить обработку и классификацию данных, использовать регрессионный и кластерный анализ для решения нестандартных задач

2. Место дисциплины в структуре основной профессиональной образовательной программы

Дисциплина относится к обязательной части образовательной программы. Изучение дисциплины требует входных компетенций, знаний, умений и навыков, предусмотренных следующими курсами:

- Информатика и программирование
- Информационные системы и технологии
- Управление IT-инфраструктурой предприятия

3. Перечень планируемых результатов обучения

Процесс изучения дисциплины направлен на формирование компетенции ОПК-1, ОПК-7

Таблица 1.

Общепрофессиональные компетенции

Код и наименование общепрофессиональной компетенции	Код и наименование индикатора достижения общепрофессиональной компетенции
ОПК-1. Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности	ИД-1 _{ОПК-1} Использует положения, законы и методы естественнонаучных дисциплин, общеинженерных знаний и математики для решения стандартных задач в области прикладной информатики. ИД-2 _{ОПК-1} Обосновывает и применяет методы математического анализа и моделирования для решения профессиональных задач ИД-3 _{ОПК-1} Проводит теоретические и экспериментальные исследования для решения задач в области прикладной информатики.
ОПК-7. Способен разрабатывать алгоритмы и программы, пригодные для практического применения	ИД-1 _{ОПК-7} Знает и использует основные инструментальные средства для программирования систем ИД-2 _{ОПК-7} Разрабатывает алгоритмы работы системы ИД-3 _{ОПК-7} Способен анализировать информацию для ее дальнейшего использования в информационных системах

4. Структура и содержание дисциплины

4.1. Объем дисциплины

Объем дисциплины составляет 6 зачетные единицы, 216 академических часа.

Таблица 2.

Объем дисциплины по видам учебных занятий в академических часах

Объем дисциплины	Всего часов
	Очная форма обучения
Объем дисциплины	216
Контактная работа обучающихся с преподавателем (по видам аудиторных учебных занятий) – всего:	84
в том числе:	-
лекции	28
занятия семинарского типа:	
лабораторные занятия	56
Самостоятельная работа (далее – СРС) – всего:	132
в том числе:	-
курсовая работа	-
контрольная работа	
Вид промежуточной аттестации	зачет/экзамен

4.2. Структура дисциплины

Таблица 3.

Структура дисциплины для очной формы обучения

№	Раздел дисциплины	С е м е с т р	Виды учебной работы, в т.ч. самостоятельная работа студентов, час.			Формы текущего контроля успеваемости	Формируемые компетенции	Индикаторы достижения компетенций
			Ле кц ии	Ла бо ра то рн ые ра бо ты	СРС			
1	Типы и структуры данных	1	2	2	10	Сдача лабораторных работ	ОПК-1,	ИД-3ОПК-1
2	Проектирование и моделирование структур хранения данных	1	4	8	18	Сдача лабораторных работ	ОПК-7, ОПК-1	ИД-1ОПК-1 ИД-2ОПК-1 ИД-3ОПК-1 ИД-1ОПК-7 ИД-2ОПК-7 ИД-3ОПК-7

3	Модели данных.	1	4	6	18	Сдача лабораторных работ	ОПК-1	ИД-1 _{ОПК-1} ИД-2 _{ОПК-1} ИД-3 _{ОПК-1}
4	Реляционные базы данных	1	4	12	20	Сдача лабораторных работ	ОПК-1, ОПК-7,	ИД-1 _{ОПК-2} ИД-2 _{ОПК-2} ИД-3 _{ОПК-2} ИД-1 _{ОПК-7} ИД-2 _{ОПК-7} ИД-3 _{ОПК-7}
5	Основы организации хранилищ данных	2	4	8	22	Сдача лабораторных работ	ОПК-7	ИД-1 _{ОПК-2} ИД-2 _{ОПК-2} ИД-3 _{ОПК-2}
	Технологии работы с данным в web-среде	2	6	10	22	Сдача лабораторных работ	ОПК-7,	ИД-1 _{ОПК-2} ИД-2 _{ОПК-2} ИД-3 _{ОПК-2}
7	Основы анализа данных	2	4	10	22	Сдача лабораторных работ	ОПК-1, ОПК-7	ИД-1 _{ОПК-2} ИД-2 _{ОПК-2} ИД-3 _{ОПК-2} ИД-1 _{ОПК-7} ИД-2 _{ОПК-7} ИД-3 _{ОПК-7}
	ИТОГО	-	28	56	132			

4.3. Содержание разделов дисциплины

Тема 1. Типы и структуры данных.

Понятие типа данных. Встроенные типы данных. Уточняемые типы данных. Перечисляемые типы данных. Конструируемые типы данных. . Массивы. Записи: . Записи с вариантами; . Множества; Указатели; . Динамическое распределение памяти и списки. Абстрактные (определяемые пользователями) типы данных: . Представление типа; Реализация типа; Инкапсуляция; Наследование типов; Разновидности полиморфизма. Типы и структуры данных, применяемые в реляционных базах данных. Типы и структуры данных, применяемые в объектно-реляционных базах данных: Строчные типы данных; Наследование таблиц и семантика включения; Типы коллекций; Объектные типы данных.

Тема 2. Проектирование и моделирование структур хранения данных.

Описание информационного представления предметной области. ER-диаграмма, Построение концептуальной модели в виде ER-диаграммы. Основные этапы построения. Моделирование локальных представлений. Объединение локальных моделей. Ограничения целостности.

Тема 3. Модели данных.

Представление концептуальной модели средствами модели данных СУБД. Типовые модели данных СУБД и представление концептуальной модели, Сетевая модель данных. Иерархическая модель данных. Реляционная модель данных. Многомерная модель данных. Средства автоматизированного проектирования концептуальной модели.

Тема 4. Реляционные базы данных.

Формализация реляционной модели. Формализованное описание отношений и схемы отношений. Манипулирование данными в реляционной модели. Операции реляционной алгебры.

Тема 5. Основы организации хранилищ данных.

Описание хранилищ данных. Архитектура хранилища данных. Информационные потоки в хранилище данных. Инструменты и технологии хранилищ данных. Магазины

данных.

Тема 6. Технологии работы с данным в web-среде.

Разработка web-интерфейсов к базам данных. Установка соединения. Запись данных в базу данных. Представление информации из базы данных в виде форм. Проектирование и разработка баз данных для построения динамического web-сайта.

Тема 7. Основы анализа данных

Восстановление функциональных закономерностей из эмпирических данных, Основы стохастики. Стохастические модели. Алгоритмы и структуры данных.

4.4. Содержание лабораторных работ

Таблица 4.

Содержание лабораторных занятий для очной формы обучения

№ темы дисциплины	Тематика лабораторных занятий	Всего часов
1	Сбор и подготовка данных	2
2,3,4	Проектирование базы данных	26
5	Создание, наполнение, подключение хранилища данных	8
6	Создание базы данных MySQL. Простой вывод данных.	10
7	Регрессионный анализ	10

5. Перечень учебно-методического обеспечения самостоятельной работы обучающихся по дисциплине

Методические рекомендации по организации самостоятельной работы обучающихся.

6. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации по итогам освоения дисциплины

6.1. Текущий контроль

Текущий контроль проводится в форме демонстрации преподавателю результатов лабораторной работы.

Примерное задание на лабораторную работу:

Лабораторная работа «Сбор и подготовка данных»

Целями работы являются: - ознакомление со структурой источников открытых данных, изучение способов хранения и представления данных; - приобретение навыка построения системы сбора данных.

Задачи:

- 1 Исследование наборов данных, представленных на портале открытых данных data.gov.ru.
- 2 Исследование наборов данных, представленных на портале data.worldbank.org.
- 3 Построение автоматизированной системы сбора данных.

В качестве инструментального средства используется программное обеспечение Microsoft Excel.

Критерии оценивания:

Лабораторная работа принимается в формате зачтено/ не зачтено.

Зачтено, если задание выполнено полностью, в представленном отчете обоснованно получено правильное выполненное задание.

Не зачтено, если задания выполнены частично или не выполнено.

6.2. Промежуточная аттестация

Форма промежуточной аттестации по дисциплине – экзамен.

Форма проведения экзамена: *устно по билетам*

Перечень вопросов для подготовки к зачету:

ОПК-1

1. Понятие типа данных.
2. Встроенные типы данных
3. Уточняемые типы данных
4. Перечисляемые типы данных
5. Конструируемые типы данных
6. . Массивы
7. Записи:
8. . Записи с вариантами;
9. . Множества;
10. Указатели;
11. . Динамическое распределение памяти и списки
12. Абстрактные (определяемые пользователями) типы данных:
13. . Представление типа;
14. Реализация типа;
15. Инкапсуляция;
16. Наследование типов;
17. . Разновидности полиморфизма
18. Типы и структуры данных, применяемые в реляционных базах данных
19. Типы и структуры данных, применяемые в объектно-реляционных базах данных:
20. Строчные типы данных;
21. . Наследование таблиц и семантика включения;
22. Типы коллекций;
23. Объектные типы данных.
24. Описание информационного представления предметной области.
25. ER-диаграмма.
26. Построение концептуальной модели в виде ER-диаграммы.
27. Основные этапы построения.
28. Моделирование локальных представлений.
29. Объединение локальных моделей.
30. Ограничения целостности.
31. Представление концептуальной модели средствами модели данных СУБД.
32. Типовые модели данных СУБД и представление концептуальной модели.
33. Сетевая модель данных.
34. Иерархическая модель данных.
35. Реляционная модель данных.
36. Многомерная модель данных.
37. Восстановление функциональных закономерностей из эмпирических данных.
38. Основы стохастики. Стохастические модели.
39. Алгоритмы и структуры данных.

ОПК-7

40. Средства автоматизированного проектирования концептуальной модели.
41. Формализация реляционной модели.
42. Формализованное описание отношений и схемы отношений.
43. Манипулирование данными в реляционной модели.
44. Операции реляционной алгебры.

Зачет оценивается по двухбалльной шкале: «зачтено»/ «незачтено».

Оценка «Зачёт» ставится, если:

1. полно раскрыто содержание материала билета;
2. материал изложен грамотно, в определенной логической последовательности, точно используется терминология;
3. показано умение иллюстрировать теоретические положения конкретными примерами, применять их в новой ситуации;
4. продемонстрировано усвоение ранее изученных сопутствующих вопросов, сформированность и устойчивость компетенций, умений и навыков;
5. ответ прозвучал самостоятельно, без наводящих вопросов;
6. допущены одна – две неточности при освещении второстепенных вопросов, которые исправляются по замечанию.
7. в изложении допущены небольшие пробелы, не исказившие содержание ответа;

Оценка «Незачёт» ставится, если:

1. не раскрыто основное содержание учебного материала;
2. обнаружено незнание или непонимание большей или наиболее важной части учебного материала;
3. допущены ошибки в определении понятий, при использовании терминологии, которые не исправлены после нескольких наводящих вопросов.
4. не сформированы компетенции, умения и навыки.

Перечень вопросов для подготовки к экзамену:

ОПК-1

45. Понятие типа данных.
46. Встроенные типы данных
47. Уточняемые типы данных
48. Перечисляемые типы данных
49. Конструируемые типы данных
50. . Массивы
51. Записи:
52. . Записи с вариантами;
53. . Множества;
54. Указатели;
55. . Динамическое распределение памяти и списки
56. Абстрактные (определяемые пользователями) типы данных:
57. . Представление типа;
58. Реализация типа;
59. Инкапсуляция;
60. Наследование типов;
61. . Разновидности полиморфизма
62. Типы и структуры данных, применяемые в реляционных базах данных
63. Типы и структуры данных, применяемые в объектно-реляционных базах данных:
64. Строчные типы данных;
65. . Наследование таблиц и семантика включения;
66. Типы коллекций;
67. Объектные типы данных.
68. Описание информационного представления предметной области.
69. ER-диаграмма.
70. Построение концептуальной модели в виде ER-диаграммы.
71. Основные этапы построения.
72. Моделирование локальных представлений.

73. Объединение локальных моделей.
74. Ограничения целостности.
75. Представление концептуальной модели средствами модели данных СУБД.
76. Типовые модели данных СУБД и представление концептуальной модели.
77. Сетевая модель данных.
78. Иерархическая модель данных.
79. Реляционная модель данных.
80. Многомерная модель данных.
81. Восстановление функциональных закономерностей из эмпирических данных.
82. Основы стохастики. Стохастические модели.
83. Алгоритмы и структуры данных.

ОПК-7

84. Средства автоматизированного проектирования концептуальной модели.
85. Формализация реляционной модели.
86. Формализованное описание отношений и схемы отношений.
87. Манипулирование данными в реляционной модели.
88. Операции реляционной алгебры.
89. Описание хранилищ данных.
90. Архитектура хранилища данных.
91. Информационные потоки в хранилище данных.
92. Инструменты и технологии хранилищ данных.
93. Магазины данных.
94. Разработка web-интерфейсов к базам данных.
95. Установка соединения.
96. Запись данных в базу данных.
97. Представление информации из базы данных в виде форм.
98. Проектирование и разработка баз данных для построения динамического web-сайта.

Экзамен оценивается по четырехбалльной шкале: «отлично» / «хорошо» / «удовлетворительно» / «неудовлетворительно».

Оценка **«отлично»** ставится студенту, ответ которого содержит:

- глубокое знание программного материала, а также основного содержания и новаций лекционного курса по сравнению с учебной литературой;
- знание концептуально-понятийного аппарата всего курса, а также свидетельствует о способности:
- самостоятельно критически оценивать основные положения курса;
- увязывать теорию с практикой.

Оценка «отлично» не ставится в случаях систематических пропусков студентом лабораторных и лекционных занятий по неуважительным причинам, а также неправильных ответов на дополнительные вопросы преподавателя.

Оценка «хорошо» ставится студенту, ответ которого свидетельствует о полном знании материала по программе, а также содержит в целом правильное, но не всегда точное и аргументированное изложение материала.

Оценка «хорошо» не ставится в случаях пропусков студентом лабораторных и лекционных занятий по неуважительным причинам.

Оценка **«удовлетворительно»** ставится студенту, ответ которого содержит:

- поверхностные знания важнейших разделов программы и содержания лекционного курса;
- затруднения с использованием научно-понятийного аппарата и терминологии курса;

- стремление логически четко построить ответ, а также свидетельствует о возможности последующего обучения.

Оценка «неудовлетворительно» ставится студенту, имеющему существенные пробелы в знании основного материала по программе, а также допустившему принципиальные ошибки при изложении материала.

7. Методические указания для обучающихся по освоению дисциплины

7.1. Методические указания к занятиям лекционного типа

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. Обозначить вопросы, термины, материал, который вызывает трудности, пометить и попытаться найти ответ в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на консультации, на практическом занятии.

7.2. Методические указания к лабораторным занятиям

Лабораторные занятия

При подготовке к лабораторным занятиям необходимо заранее изучить методические рекомендации по его проведению. Обратит внимание на цель занятия, на основные вопросы для подготовки к занятию, на содержание темы занятия.

Лабораторное занятие проходит в виде выполнения определенного задания на компьютере с использованием специального программного обеспечения. Студент должен сдавать лабораторную работу в виде наглядной демонстрации достигнутых результатов преподавателю.

7.3. Методические указания по организации самостоятельной работы

Материал, законспектированный на лекциях, необходимо регулярно прорабатывать и дополнять сведениями из других источников литературы, представленных не только в программе дисциплины, но и в периодических изданиях.

При изучении дисциплины сначала необходимо по каждой теме прочитать рекомендованную литературу и составить краткий конспект основных положений, терминов, сведений, требующих запоминания и являющихся основополагающими в этой теме для освоения последующих тем курса. Для расширения знания по дисциплине рекомендуется использовать Интернет-ресурсы; проводить поиски в различных системах и использовать материалы сайтов, рекомендованных преподавателем.

При ответе на экзамене необходимо: продумать и четко изложить материал; дать определение основных понятий; дать краткое описание явлений; привести примеры. Ответ следует иллюстрировать схемами, рисунками и графиками.

8. Учебно-методическое и информационное обеспечение дисциплины

8.1. Перечень основной и дополнительной учебной литературы

Основная литература:

ГрасД. Г77 Data Science. Наука о данных с нуля: Пер. с англ. - 2-е изд., перераб. и доп. - СПб.: БХВ-Петербург, 2021. - 416 с.: ил. [Электронный ресурс]. - Режим доступа: https://vk.com/doc44301783_566624925?hash=1pW8jCbZOw9I2hLTyQjLKUS3znW9NJ8HjJ3z zpWQh1o (дата обращения: 03.07.2022).

Дополнительная литература

1. Винстон, У. Л. Excel 2008. Анализ данных и бизнес-моделирование. [Текст] / У. Л. Винстон. - Санкт-Петербург : Русская Редакция ; Санкт-Петербург : БХВ-Петербург, 2008. -

594 с.

2. Халафян, А. А. STATISTICA 6: СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ [Текст] / А. А. Халафян. - 2 изд. - Москва : Бином-Пресс, 2010. - 522 с.

3. Большаков, В. Д. Теория ошибок наблюдений [Текст] : учебник / Большаков В.Д. - 2-е изд., перераб. доп. - Москва : Недра, 1983. - 222 с.

8.3. Перечень программного обеспечения

1. MS Office
2. MySQL
3. Deductor Studio
4. Adobe Dreamweaver

8.4. Перечень информационных справочных систем

Не используется

8.5. Перечень профессиональных баз данных

Электронно-библиотечная система eLibrary

9. Материально-техническое обеспечение дисциплины

Материально-техническое обеспечение программы соответствует действующим санитарно-техническим и противопожарным правилам и нормам и обеспечивает проведение всех видов аудиторных занятий и самостоятельной работы студентов.

Учебная аудитория для проведения занятий лекционного типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации.

Учебная лаборатория прикладных информационных технологий – укомплектована специализированной (учебной) мебелью, компьютерами, служащими для работы с информацией.

Учебная аудитория для текущего контроля и промежуточной аттестации - укомплектована специализированной (учебной) мебелью, техническими средствами обучения, служащими для представления учебной информации.

Помещение для самостоятельной работы – укомплектовано специализированной (учебной) мебелью, оснащено компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечено доступом в электронную информационно-образовательную среду организации.

Помещение для хранения и профилактического обслуживания учебного оборудования.

10. Особенности освоения дисциплины для инвалидов и лиц с ограниченными возможностями здоровья

Обучение обучающихся с ограниченными возможностями здоровья при необходимости осуществляется на основе адаптированной рабочей программы с использованием специальных методов обучения и дидактических материалов, составленных с учетом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья таких обучающихся (обучающегося).

При определении формы проведения занятий с обучающимся-инвалидом учитываются рекомендации, содержащиеся в индивидуальной программе реабилитации инвалида, относительно рекомендованных условий и видов труда.

При необходимости для обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья создаются специальные рабочие места с учетом нарушенных функций и ограничений жизнедеятельности.