

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Костромской государственный университет»
(ФГБОУ ВО КГУ)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

ТЕХНОЛОГИИ ОБРАБОТКИ ДАННЫХ

Направление подготовки *«09.03.02 Информационные системы и
технологии»*

Все направленности

Квалификация выпускника: бакалавр

**Кострома
2020**

Рабочая программа дисциплины **Технологии обработки данных** разработана в соответствии с Федеральным государственным образовательным стандартом 09.03.02 Информационные системы и технологии, утв. приказом Министерства образования и науки РФ от 19 сентября 2017 г. № 926

Разработал: Красавина М.С., доцент каф. ИСТ, к.т.н.

Рецензент: Панин И. Г., проф. каф. ИВТ, д.т.н., доцент

УТВЕРЖДЕНО:

Заведующая кафедрой информационных систем и технологий

Киприна Л.Ю., к.т.н., доцент

1. Цели и задачи освоения дисциплины

Цель дисциплины - формирование систематизированного представления о концепциях, моделях и принципах технологий обработки информации. Ознакомление с принципами организации информационного обмена и консолидации информации, ее поиска и извлечения. Получение представления о трансформации данных и способах их визуализации.

Задачи дисциплины:

- Изучение технологий обработки информации, получение представления о консолидации, трансформации данных и способах их визуализации
- Развитие умений применять технологии обработки информации с применением инструментальных средств
- Получение навыка и опыта обработки информации

2. Перечень планируемых результатов обучения по дисциплине

Освоить компетенции:

ПКрек-3 Способен выполнять работы и управлять работами по созданию ИС, автоматизирующих задачи организационного управления и бизнес-процессы

Код и содержание индикаторов компетенции:

ПКрек-3.1.Разработка модели и документирование бизнес-процессов заказчика

ПКрек-3.2 Разработка архитектуры, прототипов и баз данных ИС

В результате освоения дисциплины обучающийся должен:

Знать:

- основные понятия консолидации информации;
- этапы ETL процесса, основные понятия о хранилищах данных;
- методы трансформации, визуализации данных;
- методы оценки качества и методы очистки данных;
- современные инструментальные средства обработки информации;

Уметь:

- применять полученные теоретические знания для построения оптимального алгоритма обработки информации;
- различать структуру хранилищ данных; создавать, наполнять хранилища данных, извлекать из них информацию, трансформировать структуру хранилищ;
- пользоваться методами трансформации данных;
- подбирать визуализаторы для максимально полного отражения поведение данных, содержащейся в них информации, тенденций, закономерностей;
- производить оценку качества данных и строить алгоритм очистки данных;
- осуществлять выбор инструментальных средств для обработки информации
- использовать инструментальные средства обработки информации;

Владеть:

- терминологией в области обработки информации;
- практическими навыками для построения алгоритма обработки информации;
- способами визуализации данных;
- методами работы с данными хранилищ;
- методами оценки качества данных и методами очистки информации;
- типовыми инструментальными средствами обработки информации;

3. Место дисциплины в структуре ОП ВО

Дисциплина относится к вариативной части учебного плана. Изучается в 3 семестре обучения.

Изучение дисциплины основывается на ранее освоенных дисциплинах/практиках: основы программирования и алгоритмизации, математическое обеспечение ИС.

Изучение дисциплины является основой для освоения последующих дисциплин/практик: управление данными, выполнение ВКР, научных исследований.

4. Объем дисциплины (модуля)

4.1. Объем дисциплины в зачетных единицах с указанием академических (астрономических) часов и виды учебной работы

Виды учебной работы,	Очная форма		Очно-заочная	Заочная
Общая трудоемкость в зачетных единицах	6			
Общая трудоемкость в часах	216			
Аудиторные занятия в часах, в том числе:	50			
Лекции	16			
Практические занятия				
Лабораторные занятия	34			
Самостоятельная работа в часах	127,65			
ИКР	2,35			
Практическая подготовка	22			
Форма промежуточной аттестации	Экзамен (36)			

4.2. Объем контактной работы на 1 обучающегося

Виды учебных занятий	Очная форма		Очно-заочная	Заочная
Лекции	16			
Практические занятия				
Лабораторные занятий	34			
Консультации				
Зачет/зачеты				
Экзамен/экзамены	2,35			
Курсовые работы				
Курсовые проекты				
Всего	52,35			

5. Содержание дисциплины (модуля), структурированное по темам (разделам), с указанием количества часов и видов занятий

5.1 Тематический план учебной дисциплины

№	Название раздела, темы	Всего з.е/час	Аудиторные занятия			Самостоятельная работа
			Лекц.	Практ	Лаб.	
1	Введение в обработку информации	14/0,38	2		2	10
2	Информационный обмен и консолидация информации.	34/0,96	4		6	24
3	Трансформация данных	66/1,83	4		12	47,65
4	Визуализация информации	34/0,96	2		6	26
5	Очистка и предобработка информации	32/0,88	4		8	20
	ИКР	0,35				
	Практическая подготовка				22	
	Контроль: экзамен	36/1				
	Итого:	216/6	16		34 (22)	127,65

5.2. Содержание:

Раздел 1 Введение. Определение информации с точки зрения теории информации, мера информации, формы адекватности информации, данные как вид информации в удобной для обработки форме. Классификация информационных объектов. Методы кодирования. Современные технологии анализа данных, Базовая терминология анализа данных, понятие модели и моделирования. Машинное обучение и классы задач Data Mining. Классификация программных продуктов для создания аналитических решений. Характеристики аналитических платформ. Языки визуального моделирования в аналитических платформах.

Раздел 2 Информационный обмен и консолидация информации. Системы и сети информационного обмена. Обобщенная схема процесса консолидации. Предпосылки появления ХД, Основные требования к ХД, Задачи, решаемые ХД, Детализированные и агрегированные данные, метаданные, Многомерное представление данных и многомерный куб, MOLAP; измерения и факты; операции с многомерным кубом, ROLAP, схемы "звезда" и снежинка", HОLAP, преимущества и недостатки гибридной архитектуры ХД. Концепция виртуальных хранилищ данных. Процесс ETL, его основные цели и задачи, Выбор используемых источников данных, Организация процесса извлечения данных, Уровни очистки данных, Классификация проблем в "грязных" данных. Преобразование структур данных: агрегирование, перевод значений и пр., Организация процесса загрузки в ХД, Многопоточная загрузка и постзагрузочные операции. Преимущества и недостатки отказа от создания ХД.

Раздел 3 Трансформация данных. Понятие трансформации. Цели трансформации и ее роль в процессе ETL. Основные методы трансформации. Трансформация временных рядов: скользящее окно, интервал и горизонт прогноза, глубина погружения. Преобразование даты и времени, группировка и разгруппировка данных. Объединение данных. Внутреннее и внешнее соединение. Цели квантования, выбор числа интервалов

квантования, методы квантования.

Раздел 4 Визуализация информации. Цели и задачи визуализации, группы методов визуализации. Общие визуализаторы: графики, диаграммы, гистограммы, статистика, OLAP-анализ, Манипуляции с OLAP-кубами. Матрицы классификации, диаграммы рассеяния, ретропрогноз, коэффициенты регрессии, визуализация контроля обучения моделей. Древовидные визуализаторы, визуализаторы связей, двумерные карты.

Раздел 5 Очистка и предобработка информации. Концепция управления качеством информации. Уровни качества данных, оценка пригодности данных к анализу. Оценка качества данных по их происхождению, профайлинг данных. Визуальная оценка качества данных. Выявление трудно формализуемых ошибок, Предобработка данных и ее отличие от очистки. Типичный набор инструментов предобработки в аналитическом приложении. Фильтрация данных. Обобщенная модель дубликатов и противоречий. Обработка дубликатов и противоречий. Происхождение пропусков в данных. Методы восстановления пропущенных значений. Редактирование аномальных значений. Виды аномальных значений. Этапы обработки аномальных значений. Методы корректировки аномальных значений. Снижение размерности входных данных: Этапы задачи снижения размерности данных и режимы сокращения данных. Устранение незначущих факторов.

5.3. Практическая подготовка

Код компетенции	Индикатор компетенции	Содержание задания на практическую подготовку по выбранному виду деятельности	Число часов практической подготовки			
			Всего	Лекции	Курсовой проект	Лаб.раб
ПКрек-3	ПКрек-3.2	Создание, наполнение хранилища данных, извлечение информации, построение OLAP-кубов	10			10
ПКрек-3	ПКрек-3.2	Модификация структуры хранилища данных	4			4
ПКрек-3	ПКрек-3.1	Создание сценариев для обработки данных для решения аналитических задач в аналитической системе и/или с использованием языка python	6			6
ПКрек-3	ПКрек-3.1	Изучение способов визуализации информации. Создание отчетов	2			2

6. Методические материалы для обучающихся по освоению дисциплины

6.1. Самостоятельная работа обучающихся по дисциплине (модулю)

№ п/п	Раздел (тема) дисциплины	Задание	Часы	Методические рекомендации по выполнению задания	Форма контроля
1	Введение в обработку информации	1. Повторение пройденного материала. 2. Подготовка к лабораторной работе. 3. Создание отчета по лабораторной работе.	10	См. список литературы	Устная Отчет
2	Информационный обмен и консолидация информации.		24		
3	Трансформация данных		47,65		
4	Визуализация информации		26		
5	Очистка и предобработка информации		20		
6	Подготовка к экзамену			См. список литературы	Экзамен

6.2. Тематика и задания для практических занятий (при наличии)

6.3. Тематика и задания для лабораторных занятий

1. Базовые навыки работы в аналитической системе. Основные методы обработки.
2. Создание и наполнение хранилища данных в аналитической системе
3. Извлечение информации из хранилища данных, построение OLAP-кубов
4. Модификация структуры хранилища данных
5. Манипуляция с упорядоченными данными, групповые операции с данными в аналитической системе и/или с использованием языка python
6. Соединение и разбиение наборов данных в аналитической системе и/или с использованием языка python
7. Изучение способов визуализации информации. Создание отчетов
8. Обработка дубликатов и противоречий в аналитической системе и/или с использованием языка python
9. Предобработка данных, выявление аномалий и шумов, сглаживание рядов в аналитической системе и/или с использованием языка python
10. Решение аналитической прикладной задачи в аналитической системе и/или с использованием языка python

6.4. Методические рекомендации для выполнения курсовых работ (проектов) при наличии

7. Перечень основной и дополнительной литературы, необходимой для освоения дисциплины (модуля)

а) основная:

1. Информационные технологии : учебное пособие / Министерство образования и науки Российской Федерации, Федеральное государственное автономное образовательное учреждение высшего профессионального образования «Северо-Кавказский федеральный университет» ; сост. К.А. Катков, И.П. Хвостова и др. - Ставрополь : СКФУ, 2014. - Ч. 1. - 254 с.: ил. - Библиогр. в кн. ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=457340>

2. Структуры и алгоритмы обработки данных: Учебное пособие / Колдаев В.Д. - М.:ИЦ РИОР, НИЦ ИНФРА-М, 2014. - 296 с.: 60x90 1/16. - (Высшее образование: Бакалавриат) (Переплёт 7БЦ) ISBN 978-5-369-01264-2 <http://znanium.com/catalog/product/418290>

3. Информационные аналитические системы [Электронный ресурс] : учебник / Т. В. Алексеева, Ю. В. Амириди, В. В. Дик и др.; под ред. В. В. Дика. - М.: МФПУ Синергия, 2013. - 384 с. - (Университетская серия). - ISBN 978-5-4257-0092-6 <http://znanium.com/bookread2.php?book=451186>

б) дополнительная:

1. Популярное введение в современный анализ данных в системе STATISTICA: Учебное пособие для вузов / В.П. Боровиков. - М.: Гор. линия-Телеком, 2013. - 288 с.: ил.; 70x100 1/16 + CD-ROM. (обложка, cd rom) ISBN 978-5-9912-0326-5 <http://znanium.com/catalog/product/425084>

2. Статистический анализ данных в MS Excel: Учебное пособие / А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов. - М.: ИНФРА-М, 2012. - 320 с.: 60x90 1/16. - (Высшее образование). (переплет) ISBN 978-5-16-004579-5 <http://znanium.com/catalog/product/238654>

8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

Информационно-образовательные ресурсы:

1. Федеральный портал «Российское образование»;
2. Официальный сайт министерства образования и науки Российской Федерации.

Электронные библиотечные системы:

1. ЭБС «Лань».
2. ЭБС «Университетская библиотека online».
3. ЭБС «Znanium».
4. ЭБС «ИНТУИТ».

9. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Для проведения всех видов занятий по дисциплине необходимо следующее материально-техническое обеспечение:

№ п/п	Специализированные аудитории и классы	Номер аудитории
1	Лекционная аудитория, оборудованная мультимедиа	Е-326
2	Компьютерные классы	Е-319

Учебное оборудование		
	Персональные компьютеры, объединенные в локальную сеть, с выходом в Интернет	
№ п/п	Программное обеспечение	
1	MS Windows (Dream Spark Premium)	Поставщик ООО Форвард Софт Бизнес Договор № 6-ЭА-2014 от 31.10.2014
2	MS Office Std	Поставщик ЗАО Софт Лайн Трейд Договор № 50156/ЯР4393 от 11.12.2014