

АННОТАЦИЯ ДИСЦИПЛИНЫ

«Информационные технологии и методы обработки больших объемов данных»

Дисциплина «Информационные технологии и методы обработки больших объемов данных» является частью программы бакалавриата «Компьютерные системы» по направлению «09.03.01 Информатика и вычислительная техника».

Цели и задачи дисциплины

Цель дисциплины - освоение принципов, методов, технологий и инструментов обработки и использования больших данных в информационных системах.

Задачи освоения дисциплины:

- ознакомление студентов с основными классами задач представления и обработки больших данных;
- изучение методов и технологий подготовки, хранения, обработки и анализа больших данных;
- освоение методов аналитической обработки больших объёмов данных в информационных системах;
- приобретение навыков использования технологий и инструментов Big Data при решении практических задач.

Изучаемые объекты дисциплины

Анализ и развитие технологий обработки больших данных в научных исследованиях
Анализ методов и подходов к технологиям обработки больших данных
Архитектура технологий обработки больших данных
Решение практических задач с помощью технологии обработки больших данных

Объем и виды учебной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах
		Номер семестра
		5
1.Проведение учебных занятий (включая проведение текущего контроля успеваемости) в форме: Контактная аудиторная работа, из них:	58	58
- лекции (Л)	18	18
- лабораторные работы (ЛР)	20	20
- практические занятия, семинары и (или) другие виды занятий семинарского типа (ПЗ)	18	18
- контроль самостоятельной работы (КСР)	2	2
- контрольная работа		
1.2. Самостоятельная работа студентов (СРС)	50	50
2. Промежуточная аттестация		
Экзамен	36	36
Дифференцированный зачет		
Зачет		
Курсовой проект (КП)		
Курсовая работа (КР)		
Общая трудоемкость дисциплины	144	144

Краткое содержание дисциплины

Наименование разделов дисциплины с кратким содержанием	Объем аудиторных занятий по видам в часах			Объем внеаудиторных занятий по видам в часах
	Л	ЛР	ПЗ	СРС
5-й семестр				
Тема 1. Анализ системы обработки больших данных	2	0	2	4
Объем генерируемых данных. Многообразие больших данных Скорость генерирования или обработки данных Изменчивость, нестабильность данных во времени. Достоверность, собранных данных Трудоемкость процесса корреляции и построения взаимосвязей между данными.				
Тема 2. Современные методы визуализации информации в приложениях Big Data	2	4	2	4
сложная столбчатая диаграмма хордовая диаграмма солнечная диаграмма поточковый график матричная диаграмма диаграмма рассеяния диаграмма ohlc				
Тема 3. Задачи параллельной обработки данных в научных исследованиях	2	4	2	6
проблемные точки процесса реализации программы параллельные структуры алгоритмов параллелизм на уровне инструкций параллелизм на уровне задач параллелизм на уровне алгоритмов				
Тема 4. Фреймворк Apache Spark как способ обработки больших объёмов данных	2	2	2	6
горизонтальная масштабируемость основная технология распараллеливания кластеры среды hadoop особенности системы хранения главные конкуренты и вычислительная мощность				
Тема 5. Фреймворк mapreduce как способ обработки больших объёмов данных	2	2	2	6
преимущества и недостатки mapreduce возможности восстановления работы после сбоя узлов примеры работы mapreduce				
Тема 6. Анализ массивов и временных рядов	2	2	2	6
Rolling window estimations Экспоненциальное сглаживание, модель Хольта-Винтерса Кросс-валидация на временных рядах, подбор параметров Линейные модели на временных рядах Извлечение признаков (Feature extraction) Линейная регрессия vs XGBoost				

Тема 7. Вопросы хранения больших данных	2	2	2	6
вопросы безопасности больших данных резервное копирование облака собственные алгоритмы шифрования типичные проблемы				
Тема 8. Извлечение, сбор и валидация Big Data	2	2	2	6
Определение проблемы; Извлечение данных; Подготовка данных — очистка данных; Подготовка данных — преобразование данных; Предсказательные модели				
Тема 9. Практические примеры применения Big Data	2	2	2	6
Обзор примеров практического применения, успешные и не успешные варианты				
ИТОГО по 5-му семестру	18	20	18	50
ИТОГО по дисциплине	18	20	18	50

Тематика примерных практических занятий

№ п.п.	Наименование темы практического (семинарского) занятия
1	Обработка данных. Выбор экземпляров (Instance Selection)
2	Обработка данных. Дискретизация для классификации (Discretization)
3	Организация распределённых вычислений
4	Практические задачи с использованием различных инструментов обработки больших данных

Тематика примерных лабораторных работ

№ п.п.	Наименование темы лабораторной работы
1	Знакомство с языком R
2	Стратегии работы с большими массивами данных
3	Подготовка исходных данных
4	Обработка данных. Выбор признаков (Feature Selection)