

АННОТАЦИЯ ДИСЦИПЛИНЫ

«Методы и алгоритмы обработки слабоструктурированных данных»

Дисциплина «Методы и алгоритмы обработки слабоструктурированных данных» является частью программы магистратуры «Автоматизированные системы обработки информации и управления» по направлению «09.04.01 Информатика и вычислительная техника».

Цели и задачи дисциплины

Формирование комплекса знаний, умений и навыков в области разработки методов и алгоритмов обработки слабоструктурированных данных и применения современных языков программирования для реализации предложенных методов и алгоритмов в информационных системах..

Изучаемые объекты дисциплины

Слабоструктурированные данные; модели, методы и алгоритмы сбора и обработки слабоструктурированных данных; язык программирования Python; модули и библиотеки анализа данных..

Объем и виды учебной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах			
		Номер семестра			
		4			
1. Проведение учебных занятий (включая проведение текущего контроля успеваемости) в форме:	72	72			
1.1. Контактная аудиторная работа, из них:					
- лекции (Л)				18	18
- лабораторные работы (ЛР)				24	24
- практические занятия, семинары и (или) другие виды занятий семинарского типа (ПЗ)				26	26
- контроль самостоятельной работы (КСР)				4	4
- контрольная работа					
1.2. Самостоятельная работа студентов (СРС)	72	72			
2. Промежуточная аттестация					
Экзамен					
Дифференцированный зачет	9	9			
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
4-й семестр				
Введение в аналитику слабоструктурированных данных	2	0	2	0
Понятие слабоструктурированных данных. Плюсы и минусы использования слабоструктурированных данных. Самоописываемые структуры. Объектный подход к организации данных.				
Форматы слабоструктурированных данных и их обработка	2	4	4	12
Хранилища ключ-значения. Языки разметки. OEM. SOAP. XML, HTML. REST. JSON. YAML. Области применения форматов слабоструктурированных данных. Инструменты сбора и первичной обработки распространенных форматов. СУБД, использующие слабоструктурированные данные.				
Обработка слабоструктурированных данных на естественном языке	2	4	4	12
Сбор данных на естественном языке. Токенизация. Морфологический анализ. Приведение к нормальной форме. Библиотеки анализа естественного языка. Методы и алгоритмы векторного представления данных на естественном языке.				
Применение языка Python для обработки слабоструктурированных данных	4	4	4	12
Обзор возможностей языка Python. Установка и настройка Python для анализа данных. Сравнение Python с другими инструментами обработки слабоструктурированных данных. Библиотеки и фреймворки в экосистеме Python для обработки слабоструктурированных данных.				
Обработка мультимодальных слабоструктурированных данных	4	4	4	12
Понятие мультимодальных данных. Декомпозиция мультимодальных данных. Обработка изображений. Обработка видеоданных. Обработка аудиоданных. Введение в OCR.				

Наименование разделов дисциплины с кратким содержанием	Объем аудиторных занятий по видам в часах			Объем внеаудиторных занятий по видам в часах
	Л	ЛР	ПЗ	СРС
Машинное обучение в обработке слабоструктурированных данных	2	4	4	12
Направления в машинном обучении для обработки слабоструктурированных данных. Использование многомерных пространств для обработки слабоструктурированных данных. Решение задач регрессии, классификации и кластеризации для слабоструктурированных данных. Архитектуры искусственных нейронных сетей для обработки слабоструктурированных данных.				
Преобразование слабоструктурированных данных	2	4	4	12
Задачи, требующие реструктуризации слабоструктурированных данных. DataFrame как промежуточная форма представления данных. Построение схемы данных. Принципы преобразования больших данных. In-memory преобразования. Хранение преобразованных данных как структурированных данных.				
ИТОГО по 4-му семестру	18	24	26	72
ИТОГО по дисциплине	18	24	26	72