

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

«МЕТОДЫ АНАЛИЗА ДАННЫХ»

Направление подготовки (специальность)	10.03.01 Информационная безопасность
Направленность (профиль) подготовки	Безопасность автоматизированных систем
Цель освоения дисциплины	Обеспечение подготовки студентов в соответствии с требованиями ФГОС ВО и учебного плана бакалавриата 10.03.01 «Информационная безопасность» ознакомление студентов с кругом задач в области автоматической обработки естественного языка (natural language processing) и компьютерной лингвистики (computational linguistics), а также с доступным программным инструментарием для решения прикладных задач обработки текста. В рамках курса рассматриваются основные понятия компьютерной лингвистики, а также существующее программное обеспечение для работы с текстами.
Общая трудоемкость дисциплины	10 зачетных единиц, 360 часов
Форма промежуточной аттестации	Зачет, экзамен
Краткое содержание дисциплины:	<ul style="list-style-type: none">• Введение в анализ данных. Проблема обработки данных. Матрица данных.• Классификация данных с использованием детерминированных моделей. Решающие поверхности и дискриминантные функции.• Процедуры обучения с коррекцией ошибок: правило с фиксированным приращением, правило абсолютной коррекции, частично корректирующее правило.• Классификация данных на основе статистических моделей. Функция потерь. Байесовская дискриминантная функция.• Примеры построения статистических дискриминантных функций для различных статистических нескольких моделей данных.• Кластер-анализ. Основные типы задач кластер-анализа. Меры подобия и функции расстояния. Выбор критерия кластеризации.• Методы снижения размерностей данных. Анализ матриц исходных данных.• Системы DATA MINING. в задачах анализа и интерпретации данных• Частотный анализ лексики и ключевые слова. Частотное распределение лексики в языке. Закон Ципфа. Доля n на n $legomena$.• Автоматическое определение тематики. Векторное представление текста для задач информационного поиска.

	<ul style="list-style-type: none">• ПО для кластеризации текстов. Пакеты кластеризации для R. gCLUTO. Классификация текстов• Извлечение мнений и оценок (Sentiment analysis).• Автоматический анализ стиля. Стилometрия.
--	--

Аннотацию рабочей программы составил доцент кафедры ИЗИ к.т.н. Монахов Ю.М.

