

## АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

### «МЕТОДЫ АНАЛИЗА ДАННЫХ»

<b>Направление подготовки (специальность)</b>	10.03.01 Информационная безопасность
<b>Направленность (профиль) подготовки</b>	Безопасность автоматизированных систем
<b>Цель освоения дисциплины</b>	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.
<b>Общая трудоемкость дисциплины</b>	10 зачетных единиц, 360 часов
<b>Форма промежуточной аттестации</b>	Зачет, экзамен
<b>Краткое содержание дисциплины:</b>	<ul style="list-style-type: none"> <li>• Введение в анализ данных. Проблема обработки данных. Матрица данных.</li> <li>• Классификация данных с использованием детерминированных моделей. Решающие поверхности и дискриминантные функции.</li> <li>• Процедуры обучения с коррекцией ошибок: правило с фиксированным приращением, правило абсолютной коррекции, частично корректирующее правило.</li> <li>• Классификация данных на основе статистических моделей. Функция потерь. Байесовская дискриминантная функция.</li> <li>• Примеры построения статистических дискриминантных функций для различных статистических нескольких моделей данных.</li> <li>• Кластер-анализ. Основные типы задач кластер-анализа. Меры подобия и функции расстояния. Выбор критерия кластеризации.</li> <li>• Методы снижения размерностей данных. Анализ матриц исходных данных.</li> <li>• Системы DATA MINING. в задачах анализа и интерпретации данных</li> <li>• Частотный анализ лексики и ключевые слова. Частотное распределение лексики в языке. Закон Ципфа. Доля <math>n</math> самых legomena.</li> <li>• Автоматическое определение тематики. Векторное представление текста для задач информационного поиска.</li> </ul>

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

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

