

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

МЕТОДЫ АНАЛИЗА ДАННЫХ И ЕСТЕСТВЕННО-ЯЗЫКОВЫХ ТЕКСТОВ

(название дисциплины)

10.05.04 «ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИЕ СИСТЕМЫ БЕЗОПАСНОСТИ»

(код направления (специальности) подготовки)

8

(семестр)

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целями освоения дисциплины «Методы анализа данных и естественно-языковых текстов» являются обеспечение подготовки студентов в соответствии с требованиями ФГОС ВО и учебного плана специальности 10.05.04, 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.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

- Данная дисциплина относится к базовой части Блока Б1 (код Б1.Б.15). В учебном плане предусмотрены виды учебной деятельности, обеспечивающие синтез теоретических лекций, лабораторных работ и практических занятий.
- Дисциплина изучается на 4 курсе, требования к «входным» знаниям, умениям и готовностям (пререквизитам) обучающегося определяются требованиями к уровню подготовки по специальности 10.05.04 «Информационно-аналитические системы безопасности», квалификации - специалист по курсам «Информатика», «Математическая логика и теория алгоритмов», «Структуры данных», «Базы данных и экспертные системы», «Моделирование автоматизированных информационных систем», «Методология и организация информационно-аналитической деятельности». Курс тесно взаимосвязан с другими дисциплинами, такими как «Современные платежные системы и их безопасность», «Лингвистическое обеспечение автоматизированных информационных систем», «Формализованные модели и методы решения аналитических задач» и другими.

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ

В результате освоения дисциплины студент должен обладать:

- ПК-2 – способностью применять методы анализа массивов данных и интерпретировать профессиональный смысл получаемых формальных результатов;
- ПСК-1.2 – способностью разрабатывать и применять автоматизированные технологии обработки естественно-языковых текстов и формализованных данных при решении информационно-аналитических задач;
- ПСК-1.3 – способностью решать задачи анализа данных больших объемов.

4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

- 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.
- Методы прогнозирования временных рядов.
- Системы DATA MINING. в задачах анализа и интерпретации данных.
- Автоматическая обработка языка и компьютерная лингвистика.

- Инструменты морфологического анализа для русского языка. Понятия словоформа, лексема, лемма, основа.
- Частотный анализ лексики и ключевые слова. Частотное распределение лексики в языке. Закон Ципфа. Доля \log \log \log .
- Локальные модели контекста. Вероятностные языковые модели
- Автоматическое определение тематики. Векторное представление текста для задач информационного поиска.
- ПО для кластеризации текстов. Пакеты кластеризации для R. gCLUTO. Классификация текстов.
- Извлечение мнений и оценок (Sentiment analysis).
- Извлечение фактов и отношений. Синтаксис и формальные языки.
- Автоматический анализ стиля. Стилometрия.

Составитель:

доцент каф. ИЗИ к.т.н., доцент Монахов Ю.М.

должность, ФИО, подпись

Заведующий кафедрой

ИЗИ

М.Ю. Монахов

ФИО, подпись

Директор института

ИТР

А.А. Галкин

ФИО, подпись

Дата, Печать института (факультета)

