

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ ИНТЕЛЛЕКТУАЛЬНЫЙ ИНФОРМАЦИОННЫЙ ПОИСК

Автор: Стефановский Дмитрий Владимирович

Код и наименование направления подготовки, профиля: 38.04.01 Экономика («Системы больших данных в экономике»)

Квалификация (степень) выпускника: Магистр

Форма обучения: очная

Цель освоения дисциплины:

Сформировать компетенции в сфере аналитической работы в области экономики и финансов, обработки и анализов данных.

План курса:

Тема 1. Системы поддержки принятия решений.

Задачи систем поддержки принятия решений. Базы данных — основа СППР. Неэффективность использования OLTP-систем для анализа данных.

Тема 2. Хранилище данных.

Концепция хранилища данных. Организация ХД. Очистка данных. Концепция хранилища данных и анализ.

Тема 3. OLAP-системы.

Многомерная модель данных. Определение OLAP-систем. Концептуальное многомерное представление: двенадцать правил Кодда; дополнительные правила Кодда; тест FASMI. Архитектура OLAP-систем: MOLAP, ROLAP, HOLAP.

Тема 4. Интеллектуальный анализ данных

Задачи Data Mining: классификация задач; задача классификации и регрессии; задача поиска ассоциативных правил; задача кластеризации. Практическое применение Data Mining: интернет-технологии, торговля, телекоммуникации, промышленное производство, медицина, банковское дело, страховой бизнес и другие области применения. Модели Data Mining: предсказательные модели, описательные модели. Методы Data Mining: базовые методы; нечеткая логика; генетические алгоритмы; нейронные сети. Процесс обнаружения знаний: основные этапы анализа, подготовка исходных данных. Управление знаниями (Knowledge Management). Средства Data Mining.

Тема 5. Визуальный анализ данных — Visual Mining

Выполнение визуального анализа данных. Характеристики средств визуализации данных. Методы визуализации: методы геометрических преобразований; отображение иконок; методы, ориентированные на пиксели; иерархические образы.

Тема 6. Анализ текстовой информации — Text Mining

Задача анализа текстов: этапы анализа текстов; предварительная обработка текста; задачи Text Mining. Извлечение ключевых понятий из текста: общее описание процесса извлечения понятий из текста; стадия локального анализа; стадия интеграции и вывода понятий. Классификация текстовых документов: описание задачи классификации текстов; методы классификации текстовых документов. Методы кластеризации текстовых документов: представление текстовых документов; иерархические методы кластеризации текстов; бинарные методы кластеризации текстов. Задача аннотирования текстов: выполнение аннотирования текстов; методы извлечения фрагментов для аннотации. Средства анализа текстовой информации: средства Oracle — Oracle Text; средства от IBM

— Intelligent Miner for Text; средства SAS Institute — Text Miner; средства Мегасуперкомпьютера Интеллидженс — TextAnalyst.

Тема 7. Стандарты Data Mining

Кратко о стандартах. Стандарт CWM: назначение, структура и состав CWM, пакет Data Mining. Стандарт CRISP: появление, структура, фазы и задачи. Стандарт PMML. Другие стандарты Data Mining: стандарт SQL/MM; стандарт Microsoft Data Mining eXtensions (DMX); стандарт Java Data Mining.

Тема 8. Стандартные диаграммы для признакового описания объектов. Кросс-таблицы. Транзакционная и иерархическая модели данных.

Серии диаграмм. Визуализация наборов точек из многомерного пространства. Иконографический подход.

Стандартные диаграммы для многомерной модели данных. Понятие о семантическом слое. Системы отчётности.

Формальные контексты. Частично-упорядоченные множества. Диаграммы Хассе, TreeMap, диаграммы Эйлера-Венна. Карты, различные виды картограмм.

Понятие о содержательных и эстетических требованиях в визуализации. Первые этапы разработки нового метода визуализации.

Разработка математических процедур для нового метода визуализации. Разбор полного цикла создания нового метода визуализации на примере карты науки.

Распределение фундаментальных задач ИАД и основных инструментов статистики по моделям данных: в разрезе исходных данных, в разрезе результатов.

Модели источника (source models)

Аудиторные часы: 180

Формы текущего контроля и промежуточной аттестации: опросы, домашнее задание, контрольная работа, экзамен.

Основная литература:

1. Барсегян, А. А. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. — 3-е изд., перераб. и доп. — СПб.: БХВ-Петербург, 2009. — 512 с.: ил. + CD-ROM — (Учебная литература для вузов) ISBN 978-5-9775-0368-6.
2. Эдвард Тафти. Визуальное представление больших объемов информации.