



Федеральное агентство морского и речного транспорта
Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«Государственный университет морского и речного флота
имени адмирала С.О. Макарова»**
Воронежский филиал ФГБОУ ВО «ГУМРФ имени адмирала С.О. Макарова»

Кафедра математики, информационных систем и технологий

АННОТАЦИЯ

дисциплины «*Основы Data Mining*»

Направление подготовки 09.03.02 Информационные системы и технологии

Направленность (профиль) Информационные системы на транспорте

Уровень высшего образования бакалавриат

Промежуточная аттестация зачет

1. Место дисциплины в структуре образовательной программы

Дисциплина «Основы Data Mining» относится к части, формируемой участниками образовательных отношений Блока 1 и изучается на 5 курсе в 9 семестре.

Изучение дисциплины основано на умениях и компетенциях, полученных студентом при изучении дисциплин: «Анализ больших данных», «Автоматизация гидротехнических сооружений и водные пути», «Автоматизация перегрузочного процесса в портах и транспортных терминалах».

Дисциплина необходима в качестве предшествующей для освоения дисциплин: «Методы искусственного интеллекта», «Практика по получению профессиональных умений и опыта профессиональной деятельности», «Технологии интеллектуального анализа данных», «Стандартизация и унификация информационных технологий», «Научно-исследовательская работа», «Преддипломная практика».

2. Планируемые результаты обучения по дисциплине

Код и наименование компетенции	Код индикатора достижения компетенции	Планируемые результаты обучения по дисциплине
ПК-3 Способность проводить научные исследования при разработке, внедрении и сопровождении информационных систем и технологий с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности	ПК-3.1 Анализ исходных данных, оценка качества и эффективности ИС и технологий при разработке, внедрении и сопровождении с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности	Знать: классификацию данных, методы подготовки данных и представления данных. Уметь: проводить анализ исходных данных, оценку качества и эффективности ИС и технологий при разработке, внедрении и сопровождении с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности Владеть: навыками анализа исходных данных, оценки качества и эффективности ИС и технологий при разработке, внедрении и сопровождении с учетом существующего отечественного и зарубежного опыта в профессиональной деятельности

3. Объем дисциплины по видам учебных занятий

Объем дисциплины составляет 3 зачетные единицы, всего 108 часов, из которых по заочной форме 16 часов составляет контактная работа обучающегося с преподавателем (8 часов – занятия лекционного типа, 8 часов – лабораторные работы).

4. Основное содержание дисциплины

12 правил Э. Кодда для реляционной СУБД. OLTP-системы. Механизм транзакций в OLTP-системе. Информационные системы руководства (ИСП, EIS – Executive Information Systems).

Противоречия в требованиях к OLTP-системе и к системе анализа по основным характеристикам.

Свойства ХД и отличие ХД от оперативных источников данных (ОИД). Структура СППР с физическим ХД, её достоинства и недостатки. Структура СППР с виртуальным ХД, её достоинства и недостатки. Основные проблемы создания ХД. СППР с самостоятельными витринами данных (Data Mart). Достоинства и недостатки автономных ВД. СППР с ХД и ВД, достоинства и недостатки СППР с совмещёнными ХД и ВД.

Данные детальные, агрегированные и метаданные. Данные аддитивные, полуаддитивные и неаддитивные. Измерения и факты. Информационные потоки в ХД. Архитектура ХД на уровне потоков информации. ETL-процесс (Extraction Transformation Loading) – извлечение, преобразование, загрузка данных.

Проблемы очистки данных на уровне поля, на уровне записи, на уровне таблицы базы данных, на уровне одиночной базы данных, на уровне множества баз данных. Этапы очистки данных: выявление проблем в данных, определение правил очистки данных, тестирование правил очистки данных, непосредственная очистка данных.

Представление данных в виде гиперкуба. Операции над гиперкубом.

Концептуальное многомерное представление. 18 правил Э. Кодда для определения OLAP-систем (On-Line Analytical Processing). Группы правил (особенности) B, S, R, D. Тест FASMI (Fast Analysis Shared Multidimensional Information).

OLAP-сервер, способы его реализации: MOLAP, ROLAP, HOLAP, DOLAP, JOLAP. Преимущества и недостатки использования многомерных баз данных в OLAP-системах. Схемы реализации многомерного представления данных с помощью реляционных таблиц: «звезда», «снежинка». Наиболее часто встречающиеся типы фактов. Достоинства и недостатки использования реляционных таблиц в OLAP-системах.

Специфика современных требований к переработке «сырых» данных. Характеристика технологии Data Mining, сфера её применения. Уровни знаний, извлекаемых из данных. Некоторые бизнес-приложения Data Mining. Стандартные типы закономерностей, выявляемые методами Data Mining. Классы систем Data Mining: предметно-ориентированные аналитические системы, статистические пакеты, нейронные сети, системы рассуждений на основе аналогичных случаев, деревья решений, эволюционное программирование, генетические алгоритмы, алгоритмы ограниченного перебора, системы для визуализации многомерных данных.

Надстройки интеллектуального анализа данных для Excel. Построитель отчётов. PowerPivot для Excel 2013. Надстройка для Excel Master Data Services.

Продукты Oracle для OLAP и бизнес-анализа. Microsoft SQL Server Analysis Services (SSAS). Компоненты Business Intelligence-решения Microsoft. Унифицированная многомерная модель (UDM). Интеграция SSAS с Microsoft Office.

BI Dev Studio, SQL Server Management Studio (SSMS).

Составитель: к.э.н., доцент Скрипников О.А.

Зав. кафедрой: к.ф.-м. н., доцент Черняева С. Н.