

ГЛОССАРИЙ ПО ДИСЦИПЛИНЕ «БИЗНЕС-АНАЛИЗ»

Data Mining — исследование и обнаружение "машиной" (алгоритмами, средствами искусственного интеллекта) в сырых данных скрытых знаний, которые ранее не были известны, нетривиальны, практически полезны, доступны для интерпретации человеком.

ETL (Extract, Transform, Load — дословно «извлечение, преобразование, загрузка») — один из основных процессов в управлении хранилищами данных, который включает в себя: извлечение данных из внешних источников; их трансформацию и очистку, загрузку в хранилище данных.

HOLAP – гибридная схема многомерного представления данных на основе как многомерных, так и реляционных средств.

MOLAP – многомерное представление данных на основе средств многомерных или объектно-ориентированных СУБД.

OLAP (On-Line Analysis Processing) – подсистема информационно-аналитической системы, удовлетворяющая требованиям Кодда или теста FASMI.

OLTP (On-Line Transaction Processing) – автоматизированные системы оперативной обработки транзакций.

ROLAP – многомерное представление данных на основе средств реляционных СУБД.

Агрегация данных – это вычисление обобщенных показателей из детальных данных для поддержки стратегического или тактического управления.

Ассоциативные модели — выявление закономерностей между связанными событиями. Ассоциация – связь, возникающая при определенных условиях между двумя или несколькими событиями.

Витрина данных – это предметно-ориентированное хранилище данных, как правило, предназначенное для использования группой пользователей в рамках конкретного вида деятельности предприятия.

Генетический алгоритм – это алгоритм, основанный на природных механизмах селекции и генетики. Эти алгоритмы обеспечивают выживание сильнейших решений из множества сгенерированных, формируя и изменяя процесс поиска на основе моделирования эволюции исходной популяции решений.

Дерево решений является одним из наиболее популярных подходов к решению задач Data Mining. Они создают иерархическую структуру классифицирующих правил типа «ЕСЛИ ... ТО», имеющую вид дерева. Дерево решений состоит из узлов, ветвей и листьев. Дерево решений всегда начинается с корневого узла и растет вниз, разделяя данные на каждом уровне на новые узлы.

Интеграция данных – это объединение данных, которые изначально вводятся в разные системы.

Интеллектуальный анализ данных – термин, используемый для описания обнаружения и выделения знаний в базах данных, изыскания и исследования данных, обработки образцов данных, очистки и сбора данных.

Искусственная нейронная сеть – обобщенное название группы математических алгоритмов, обладающих способностью обучаться на примерах, «узнавая» впоследствии черты встреченных образов и ситуаций.

Искусственный интеллект – 1) научное направление, в рамках которого ставятся и решаются задачи технического и программного моделирования тех видов человеческой деятельности, которые считаются интеллектуальными (например, рассуждение, обучение, принятие решений и др.); 2) искусственно созданная система, проявляющая свойство интеллекта, т. е. способность выбора и принятия целесообразного решения при большом

многообразии целей на основе ранее полученного опыта или рационального анализа внешних воздействий.

Информационная компетентность - общая способность к эффективному использованию различных источников информации в процессе познавательной деятельности при решении конкретных задач в реальных жизненных ситуациях.

Информационная культура – это область корпоративной культуры, определяющая качество обмена информацией и знаниями в организации.

Информационно-аналитическая система (ИАС) – комплекс аппаратных, программных средств, информационных ресурсов, методик, которые используются для обеспечения автоматизации аналитических работ в целях обоснования принятия управленческих решений и других возможных применений.

Информационное пространство – совокупность информационных объектов, информационно отображающих свойства системы и протекающие в ней процессы.

Информационное хранилище – подсистема ИАС, которая выполняет задачи накопления сведений о деятельности предприятия, партнеров, других информационных ресурсов из различных источников, в том числе баз данных, отображающих отдельные бизнес-процессы, автоматизированных рабочих мест, информационных систем и других источников информации, в том числе из глобальных информационных сетей.

Источники данных – локальные таблицы и реляционные БД автоматизированных систем, в которых хранятся корпоративные данные для анализа.

Классификация – система соподчиненных понятий (классов объектов) какой-либо предметной области, используемая для установления связей между этими понятиями.

Кластеризация – позволяет объединять объекты в группы (кластеры). Отличается от классификации тем, что сами группы заранее не заданы.

Консолидация данных – это сбор в единую базу данных из различных источников предприятия.

Машинное обучение – область искусственного интеллекта, включающая модели, методы и алгоритмы, ориентированные на автоматическое накопление и формирование знаний с использованием процедур анализа обобщения данных.

Метаданные – данные о данных или совокупность атрибутов, характеризующих какую либо группу показателей.

Нейронные вычисления – вычислительный подход к моделированию человеческого мозга, который основывается на соединении большого количества простых процессоров для генерации сложного поведения.

Нейронные сети — математические модели, а также их программные или аппаратные реализации, построенные по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма.

Нечеткая переменная – величина, которая может принимать лингвистические значения. Например, нечеткая переменная «скорость» может принимать такие значения, как «малая», «небольшая», «средняя», «высокая».

Показатель – высказывание с законченным смыслом, включающее как название переменной величины, так и ее конкретное количественное значение со всеми качественными признаками,

Самоорганизующаяся карта Кохонена – мощный механизм кластеризации, позволяющий отобразить результаты в виде компактных и удобных для интерпретации двумерных карт.

Семантическая сеть – метод представления знаний посредством сети узлов, соответствующих понятиям или объектам, связанным дугами, описывающими отношение между дугами.

Система искусственного интеллекта – программная система, имитирующая на компьютере мышление человека.

Схема «звезда» – схема многомерного представления данных в виде факт-таблицы с увязанными с ней таблицами размерностей.

Схема «снежинка» – развитие схемы «звезда». Ее отличает большое количество консольных таблиц, они имеются практически на каждой таблице размерности и могут иметь несколько уровней иерархии.

Схема «созвездие» – объединение нескольких схем «звезда» и/или «снежинка» через общие таблицы размерности или консольные таблицы.

Таблица измерений (размерности) – таблица, наполняющая содержанием факт-таблицу. Такие таблицы содержат постоянные или редко и мало изменяемые данные и должны находиться в отношении «один ко многим» к таблице фактов.

Таблица фактов – таблица, в которой помещаются все данные относительно интересующего пользователя обобщающего показателя, то есть объекта или события, которые интересуют пользователя.

Тест FASMI – тест для определения принадлежности ИС к классу OLAP-систем, комплекс требований к OLAP-системам, которые в русском переводе означают: Быстрый Анализ Разделяемых Многомерных Данных..

Технология Data Mining предназначена для анализа данных с помощью математических моделей, основанных на статистических вероятностных и оптимизационных методах, с целью выявления в них заранее неизвестных закономерностей, зависимостей и извлечение знаний.

Транзакция – неделимая с позиции воздействия на БД последовательность операции манипулирования данными.

Хранилище данных – это предметно-ориентированное, привязанное ко времени и неизменяемое собрание данных, применяемых для поддержки процессов принятия управленческих решений.

Цифровизация— замена аналоговых (физических) систем сбора и обработки данных технологическими системами, которые генерируют, передают и обрабатывают цифровой сигнал о своем состоянии. В широком смысле — процесс переноса в цифровую среду функций и деятельности (бизнес-процессов), ранее выполнявшихся людьми и организациями.