

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего
образования

НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ

УНИВЕРСИТЕТ им. Р.Е.АЛЕКСЕЕВА

Институт радиоэлектроники и информационных технологий



Конспект

по теме

«Метаданные»

по дисциплине

«Базы данных»

РУКОВОДИТЕЛЬ:

(подпись)

Мисевич П. В.

(фамилия, и.,о.)

СТУДЕНТ:

(подпись)

Сухоруков В.А.

(фамилия, и.,о.)

19-В-2

(шифр группы)

Работа защищена «__» _____

С оценкой _____

Нижний Новгород 2021

Метаданные

Метаданные — информация о другой информации, или данные, относящиеся к дополнительной информации о содержимом или объекте. Метаданные раскрывают сведения о признаках и свойствах, характеризующих какие-либо сущности, позволяющие автоматически искать и управлять ими в больших информационных потоках.

Примеры:

- ✓ Описание типов данных в программах на языках программирования.
- ✓ Описание файлов, встроенных в программы и в спецификациях на языках управления заданиями (IBM JCL).
- ✓ Описание форматов отчётов в языке IBM RPG.
- ✓ Поисковые образы документов в ранних ИПС.
- ✓ Диаграммы потоков данных в CASE-инструментах и др.

Термин «метаданные» появился в начале 1970-х, когда технология БД стала захватывать рынок и проекты стали более массивными. В 80-е гг. термин стал массовым потому, что возникли ЭВМ, которые приблизили данные к пользователю и проекты стали более значимыми.

Примеры метаданных в технологиях баз данных:

- ✓ Концептуальные схемы предметных областей.
- ✓ Схемы баз данных.
- ✓ Описание междууровневых отображений схем в системах баз данных.

Основные свойства метаданных:

- ✓ Относительный характер разделения информационных ресурсов на данные и метаданные
- ✓ Разнообразие областей, в которых используются метаданные, и видов описываемых ресурсов
- ✓ Зависимость свойств метаданных от характера использующей их системы, вида описываемых ресурсов, используемых ИТ, потребностей пользователей систем и т.п.
- ✓ Зависимость состава метаданных от информационной архитектуры системы (примеры в области БД и Веб)
- ✓ Различная степень гранулярности описания ресурса.
- ✓ Метаданные горизонтальной сферы («универсальные») / вертикальной сферы («специализированные»)
- ✓ Автономные (отчужденные от описываемого ресурса)/встроенные
- ✓ Независимые/зависимые от контента описываемых ресурсов
- ✓ Системные/пользовательские метаданные
- ✓ Структурированные/неструктурированные/слабоструктурированные метаданные
- ✓ Статические/динамические (например, схема БД и каталог ЭБ)
- ✓ Формализованные/неформализованные метаданные
- ✓ Явно/неявно представленные (например, HTML-разметка / семантика ссылки в научной публикации)

✓ Многоуровневость метаданных: метаданные – это тоже данные, для них могут быть метаданные.

Функции метаданных:

✓ Метаданные могут использоваться для создания и поддержания функционирования ПО.

✓ Обеспечение интеграции данных из множества источников

✓ Обеспечение интероперабельности и повторного использования ресурсов

✓ Описание предметной области ИС: концептуальная схема, логическая схема, информация об информации в широком смысле слова, онтология

✓ Описание баз данных и других репозиторий структурированных данных, поддержка механизмов управления их ресурсами.

✓ Описание отдельных информационных объектов - таблиц БД, веб-страниц, информационных объектов в ЭБ.

✓ Описание семантики источника информации, отдельного информационного объекта или его фрагмента.

✓ Описание представления данных на разных уровнях информационной архитектуры.

✓ Идентификация описываемых ресурсов.

✓ Обеспечение функций управления данными БД и других источников информационных ресурсов

✓ Поддержка функций поиска информационных ресурсов.

Стандартизация метаданных:

Стандартизация метаданных – основа интероперабельности и повторного использования метаданных и описываемых ресурсов.

Деятельность по стандартизации: официальные органы, промышленные компании и консорциумы, профессиональные сообщества.

Разработано большое число стандартов метаданных «горизонтальной» и «вертикальной» сферы.

Примеры стандартов первой группы («горизонтальная» сфера):

- ✓ дескриптивный подязык языка SQL
- ✓ язык описания объектов ODL консорциума ODMG
- ✓ Open Information Model (OIM) консорциума Metadata Coalition
- ✓ стандарты OMG: UML, CORBA IDL, MOF, Common Warehouse Model (CMW)
- ✓ стандарты W3C: XML, XML Schema, RDF, RDFS, OWL, OWL2, WSDL
- ✓ DCMI, NIST, ISO: Dublin Core (DC)
- ✓ языки описания бизнес-процессов: BPEL, BPML
- ✓ стандарты микроформатов

Во второй группе значительное место принадлежит стандартам научных метаданных, созданным во многих областях исследований.

Применение метаданных в проектах БД и ХД

Одно из основных назначений метаданных — повышение эффективности поиска. Поисковые запросы, использующие метаданные, делают возможным выполнение сложных операций по фильтрации и отбору данных.

Метаданные — высокоуровневые средства отражения информационной модели и описания структуры данных, используемой в ХД. Метаданные должны содержать описание структуры данных хранилища и структуры данных импортируемых источников. Метаданные хранятся отдельно от данных в так называемом репозитории метаданных

Метаданные являются ключевым фактором успеха при разработке и внедрении крупных проектов. Они содержат всю информацию, необходимую для извлечения, преобразования и загрузки данных из различных источников, а также поддерживают навигацию и *поддержку ПО, которое использует БД или ХД*.

Уровни метаданных

Можно выделить два уровня метаданных — технический (административный) и бизнес-уровень.

Технический уровень содержит метаданные, необходимые для обеспечения функционирования хранилища (статистика загрузки данных и их использования, описание модели данных и т. д.).

Бизнес-метаданные обеспечивают пользователю возможность концентрироваться на процессе анализа, а не на технических аспектах работы с хранилищем; они включают бизнес-термины и определения, которыми привык оперировать пользователь.

Метаданные можно использовать тремя способами

1. Пассивно, обеспечивая четкую документацию о структуре, процессе разработки и использовании системы Хранилища данных. Доступная документация необходима всем участникам (т. е. конечным пользователям, системным администраторам, а также разработчикам приложений).

2. Активно, путем хранения конкретных семантических аспектов (например, правил преобразования) в виде метаданных, которые можно интерпретировать и использовать во время исполнения. В этом случае процессы Хранилища данных управляются метаданными.

3. Полуактивно, за счет хранения статической информации (например, определений структур, спецификаций конфигураций), которую будет считывать другой программный компонент во время выполнения.

Цели Создания и управления метаданными

К минимизации работ по разработке и администрированию Хранилища данных
в основном относится:

- 1) Поддержка интеграции систем. Схемы и интеграция данных зависят от метаданных, описывающих структуру и смысл отдельных источников данных и целевых систем;
- 2) Поддержка анализа и проектирования новых приложений. Метаданные повышают контролируемость и надежность процесса разработки приложений, обеспечивая информацию о смысле данных, их структуре и источниках;
- 3) Повышение гибкости системы и возможности повторного использования существующих программных модулей;
- 4) Автоматизация административных процессов. Метаданные управляют запуском различных процессов Хранилища данных (например, загрузки и обновления);
- 5) Усилению механизмов безопасности. Метаданные должны обеспечить правила доступа и пользовательские права для всей системы Хранилища данных.

Вторая цель относится к *эффективному извлечению информации*, а точнее:

- 1) К повышению качества данных. Качество данных определяется следующими характеристиками:
 - ✓ согласованностью (является ли представление данных однородным, нет ли дубликатов, данных с пересекающимися или конфликтующими определениями);
 - ✓ полнотой (все ли данные присутствуют);
 - ✓ точностью (совпадением хранимых и фактических значений);
 - ✓ своевременностью (актуально ли хранимое значение);
- 2) Улучшению взаимодействия внутри системы хранилища данных. Взаимодействие происходит как посредством выполнения простых запросов и отчетных приложений, так и с использованием сложных аналитических инструментов;
- 3) Улучшению анализа данных. Методы анализа данных представлены широко – начиная от простых приложений отчетности и OLAP и заканчивая сложными приложениями Data Mining;
- 4) Применению общей терминологии и языка взаимодействия внутри корпорации.

Группы пользователей метаданных

1. Руководство компании (основные пользователи информационных систем руководителей) хочет знать, что оно может получить от системы, как быстро оно получит ответ на интересующий вопрос, и желательно – в терминах, понятных лицам, которые принимают решения
2. Специалисты организации, например пользователи бухгалтерских систем, хотят, чтобы такая система "разговаривала" на их профессиональном языке, вплоть до того, что разработчики таких систем (как, например, 1С-бухгалтерия) оснащают свои системы специальным, формальным языком, понятным бухгалтерам.
3. Разработчиков приложений интересует информация о модели данных ИС для создания или внедрения дополнительных бизнес-приложений. Они хотят знать, что находится в таблицах БД системы и в каком формате.

Функции метаданных в хранилище данных

Основные функции метаданных в ХД

Представление соответствия данных источников и данных ХД

Управление данными во времени

Поддержка версионности

Интерпретация данных в терминах бизнес - пользователей

Обеспечение открытости

Элементы метаданных ХД

Базовые элементы метаданных ХД

Описание таблиц, их атрибутов, ключей

Описание преобразований данных

История экстрагирования (поступления) данных

Терминологические соответствия понятий

Информация о состояниях и статистики использования данных

Кто отвечает за содержание и актуальность различных источников данных

Алгоритмы агрегации и суммирования данных, критерии выборки из источников, правила преобразования данных источников перед загрузкой в ХД, описание взаимосвязей между объектами ХД, их кардинальность

Типы Метаданных ХД

Типы метаданных ХД

Предметно-ориентированные метаданные

Структурные метаданные

Технические метаданные

Метаданные процесса обработки данных

Концептуальная модель метаданных

