

УДК 004.652.4

МОДЕЛЬ ДАННЫХ ОБЪЕКТНО-РЕЛЯЦИОННОГО ТИПА

А.М. Копейкин

ГОУ ВПО «Северо-Западный государственный заочный технический университет», г. Санкт-Петербург

Рецензент В.Ф. Калинин

Ключевые слова и фразы: граф; задача декомпозиции; множество атрибутов; понятия объекта и связи; предметная область; реляционная база.

Аннотация: Рассматриваются реляционные и объектно-реляционные модели представления данных, служащие основой при проектировании баз данных любого уровня. Показан последовательный эволюционный переход от традиционной реляционной модели к объектно-реляционной модели представления данных. Дается формализованная постановка задачи проектирования реляционной и объектно-реляционной моделей данных.

Принципы реляционной модели как алгебраической системы, были сформулированы Коддом в 1970 году [1]. В основе реляционной модели лежит математическое понятие отношения R с заданным набором алгебраических операций. Любая моделируемая предметная область (ПО) представлялась набором отношений $\{R_i\}$, состав и структура которых определялась на основе всевозможных зависимостей F_R (F_d – функциональных, F_m – многозначных и т.д.). Фактически, каждый элемент R_i является проекцией универсального отношения U_R [2] и должен быть свободен от аномалий модификации хранимых данных.

Процесс преобразования отношения U_R в набор отношений $\{R_i\}$ базируется на основе теории *нормализации*, целью которой является уменьшение избыточности и устранение аномалий модификации хранимых данных.

Копейкин А.М. – аспирант II курса кафедры автоматизации производственных процессов (научный руководитель – доктор технических наук, профессор А.А. Сарвин) ГОУ ВПО «Северо-Западный государственный заочный технический университет», г. Санкт-Петербург.

Суть нормализации отношений, сводится к задаче *декомпозиции без потерь* [3], позволяющей разделить U_R на набор отношений (в каждом из которых синтезируется реляционный ключ K_R и каждое отношение находится в более высокой форме, чем U_R), так чтобы операция эквисоединения восстанавливала U_R из полученных проекций.

Общее правило декомпозиции (приведение отношений к более высокой форме) формально записывается следующим образом.

Пусть $F_R = \{F_R(A_R^i, A_R^j)\}$ (или $F_R = \{A_R^i - F_R^k \rightarrow A_R^j\}$), где $F_R^k \in F_R$, $A_R^i, A_R^j \subseteq A_R$ задано на множестве атрибутов A_R отношения U_R . Тогда для любого состояния U_r справедливо (только для $F_d \subseteq F_R$ и $F_m \subseteq F_R$):

$$U_R = R [A_R \setminus A_R^i] \otimes R_{Fr} (A_R^i, A_R^j),$$

где \otimes – знак операции эквисоединения по множеству атрибутов A_R^i ; \setminus – знак операции разности реляционной алгебры; $[]$ – знак операции проекции; R_{Fr} – отношение, образованное F_R .

Задача проектирования реляционной базы данных ставится следующим образом:

$$\sum_{i=1}^n R_i \rightarrow \min$$

при ограничении $U_R = R_1 \otimes R_2 \otimes \dots \otimes R_n$, где R_i – нормализованное отношение, $i = 1, \dots, n$.

Синтаксическое разложение считается оптимальным при n минимальном. Впервые оптимальное синтаксическое разложение было введено в [4].

Под описанием (схемой) отношения с именем R понимается тройка [5]:

$$Sch(R) = \langle A, F_R, P_{dyn} \rangle,$$

где A – множество атрибутов отношения с заданными областями допустимых значений (доменами), среди которых выделен ключ отношения K ; F_R – множество структурных ограничений целостности отношения (функциональные, многозначные зависимости и зависимости соединений); P_{dyn} – множество динамических ограничений целостности.

Реляционная модель базы данных обладает значительными достоинствами, которые к настоящему времени хорошо изучены. Однако исследования показали некоторые недостатки традиционной реляционной модели (ТРМ), заключающиеся, в основном, в неудовлетворительных свойствах отдельных операций реляционной алгебры и в синтаксической жесткости нормализации.

Доказаны [2–7] следующие положения:

- существуют отношения, которые невозможно сконструировать из более простых отношений средствами алгебры ТРМ;
- наличие неполных функциональных зависимостей может привести к потере информации в ТРМ или к несовместимости модели;
- семантическая перегрузка ТРМ, так как она обладает только одной конструкцией для представления данных и связей между данными – отношением;

- сложность при выполнении рекурсивных запросов;
- неадекватность представления сущностей реального мира.

Для устранения проблем, свойственных чисто реляционному подходу, были предложены так называемые семантические модели данных, к разновидности которых относится объектно-реляционная модель (ОРМ) [3, 5–7].

В основе ОРМ лежат понятия *объекта* и *связи*, которые на концептуальном уровне представляются в виде *фрейма* (понимаемого как иерархически упорядоченное представление стандартной ситуации действительности). Фрейм, в свою очередь, представляет собой иерархическое отношение. Введение иерархического отношения позволяет использовать реляционную алгебру в ОРМ.

В ТРМ не выделяется специальное средство, позволяющее уникально идентифицировать кортеж данного объекта, не учитывая отношения, в котором этот объект содержится. Это естественное следствие ориентации на синтаксис исчисления отношений, а не на связь семантики предметной области с этим синтаксисом. В ОРМ уникальная идентификация объектов необходима, и для выделения объекта введено специальное средство – *уникальный идентификатор*, которое принято называть якорем.

Введение понятия уникального идентификатора позволяет выделить объект независимо от того, существуют ли на атрибутах объекта F_R зависимости. Свойство быть уникальным идентификатором является глобальным для всей базы данных и не зависит от модификации отдельных отношений, в отличие от локальности реляционного ключа для тех же отношений.

Базовый набор отношений в ТРМ синтаксически связан с помощью реляционных ключей и удовлетворяет критерию минимального числа отношений в модели. Набор объектов (отношений) в ОРМ синтаксически (структурно) связан, но критерий выбора конфигурации ОРМ определяется суммарной мощностью формируемого набора фреймов.

Можно сформулировать следующие требования к ОРМ:

- выделенный набор фреймов должен иметь минимальную избыточность;
- запросы к системе могут быть разрешены по любому сочетанию значений атрибутов модели;
- сформированная совокупность фреймов должна быть синтаксически связана с помощью якорей, предохраняющих схему модели от аномалий модификации;
- для обеспечения синтаксической связности каждый объект, заданный для хранения в базе данных, должен быть частью (проекцией), по крайней мере, одного фрейма модели;
- удаление (или модификация) любого набора объектов, содержащихся в модели, не должно приводить к потере информации в других (некорректируемых) объектах;
- реляционные ключи, выделенные на основе F_R и бизнес-правил, должны обеспечивать истинность исходного отношения.

Исходными данными для объектно-реляционной модели являются те же данные, что и для традиционной реляционной модели.

Модель объектно-реляционной базы данных описывается следующей тройкой:

$$R_{or} = \langle \mathbf{Q}, \mathbf{E}, t \rangle,$$

где \mathbf{Q} – набор объектов (фреймов); \mathbf{E} – набор связей между объектами (со статусами – предикатными свойствами); t – время фиксации модели.

В свою очередь фрейм $Q \in \mathbf{Q}$ представляет собой тройку:

$$Q = \langle \mathbf{S}, \mathbf{M}, \mathbf{P}_q^s \rangle,$$

где \mathbf{S} – множество частей фрейма; \mathbf{M} – набор ориентированных связей (со статусами) между частями (матрица смежности с указанием предикатных свойств связей); \mathbf{P}_q^s – правила перехода между частями и существования частей. Правило перехода представляет собой тройку <текущая часть, следующая часть, правила существования части >.

В свою очередь часть $S \in \mathbf{S}$ представляет собой:

$$S = \langle \mathbf{A}_s, \text{St}, N_s, P_s \rangle,$$

где \mathbf{A}_s – содержимое части = {<атрибут, домен атрибута, семантический код>} семантический код := простое поле | классификатор | вычисляемое поле | наследуемое поле | предикат ограничения целостности; St – статус части (основная, внутренняя, внешняя и т.п.); N_s – кардинальность части (однозначная, многозначная); P_s – предикат ограничения целостности элементов фрейма.

Основной постулат ОРМ заключается в том, что любую ПО реального мира можно отобразить в виде ориентированного графа (в общем случае ультраграфа) с изменяемой структурой. Вершинами графа являются объекты (части), а ребрами являются связи между объектами (частями). При этом связи между объектами могут нести различную нагрузку (подчиненность объектов, наследование объектов, рекурсивную обработку и т.п.), которая выражается *статусами* и установленными в процессе проектирования ПО правилами.

Задача выбора конфигурации модели может быть сформулирована следующим образом.

Пусть задано универсальное отношение U_R со схемой $Sch(U_R)$. В процессе выделения из исходного отношения набора объектов со связями (приведения к объектно-реляционной форме) необходимо минимизировать число вершин и связей при условии, что обратный процесс соединения полученных объектов и связей должен создавать отношение покрывающее исходное (содержание должно быть одинаковым, но схема может получиться другой).

Исходное U_R преобразуется в $\langle \mathbf{Q} = \{Q_1, \dots, Q_n\}, \mathbf{E} = \{E_1, \dots, E_m\} \rangle$, такое что \mathbf{Q} и \mathbf{E} адекватно отображают предметную область и

$$\sum_{i=1}^n Q_i \sum_{j=1}^m E_j \rightarrow \min$$

при ограничении $U_R = Q_1 \oplus Q_2 \oplus \dots \oplus Q_n \oplus E_1 \oplus E_2 \oplus \dots \oplus E_m$, где \oplus – операция соединения в ОРМ.

Рассмотренная конфигурация модели реализована в системе ORD (www.ord.com.ru).

Список литературы

1. Codd, E. A Relational Model of Data for Latge Shared Data Banks / E. Codd // SACM. – June 1970. – 13, № 6.
 2. Мейер, М. Теория реляционных баз данных / М. Мейер. – М. : Мир, 1987. – 608 с.
 3. Дейт, К. Введение в системы баз данных / К. Дейт. – 8-е изд. – М. : ИД «Вильямс», 2006. – 1328 с.
 4. Неклюдова, Е.А. Синтез логической схемы реляционных баз данных / Е.А. Неклюдова, М.Ш. Цаленко // Программирование. – 1979. – № 6. – С. 58 – 68.
 5. Цаленко, М.Ш. Моделирование семантики в базах данных / М.Ш. Цаленко. – М. : Наука, 1989. – 288 с.
 6. Конноли, Т. Базы данных: проектирование, реализация и сопровождение. Теория и практика / Т. Конноли. – 3-е изд. – М. : ИД «Вильямс», 2003. – 1453 с.
 7. Проектирование объектно-реляционных баз данных / Г.П. Воронин [и др.]. – Л. : Судостроение, 1986. – 179 с.
-

Model of Data of Object-Relation Type

A.M. Kopeikin

North-Western State Part-Time Technical University, St. Petersburg

Key words and phrases: graph; decomposition task; set of attributes; notions of object and link; subject area; relation base.

Abstract: Relation and object-relation models of data presentation are considered; those are used as the basis for designing databases of all types and levels. Gradual evolutionary transfer from conventional relation model of data presentation to that of object-relation is shown. Formalized setting of the task of designing relation and object-relation models of databases is given.

© А.М. Копейкин, 2007