

УДК 004.652.5

А.Б. Кунгурцев, канд. техн. наук, проф.,
А.С. Неизвестный, магистр,
Одес. нац. политехн. ун-т

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ОБЪЕКТНОГО ПРЕДСТАВЛЕНИЯ РЕЛЯЦИОННОЙ БАЗЫ ДАННЫХ

О.Б. Кунгурцев, О.С. Неизвестный. **Математична модель об'єктного подання реляційної бази даних.** Удосконалено об'єктне подання реляційних баз даних шляхом використання словника предметної області. Запропоновано математичний опис об'єктного представлення реляційних баз даних. Введено поняття ідентифікуючих атрибутів, що дозволяють інтегрувати локальні бази даних на рівні об'єктного подання.

A.B. Kungurtsev, A.S. Neizvestny. **Mathematical model of objective representation of a relational database.** Objective representation of relational databases by use of the dictionary of a subject domain is advanced. The mathematical description of objective representation of relational databases is offered. The concept of the identifying attributes is entered, allowing to integrate local databases at a level of objective representation.

Для хранения и обработки своей информации практически все крупные компании используют реляционные базы данных (РБД). Однако, со временем структура РБД в связи с появлением новых задач и сущностей становится сложной для восприятия человеком из-за возрастающего количества отношений и их полей, множества связей между атрибутами и отношениями, несоответствия названий отношений и полей общепринятой терминологии предметной области, а также наличия неинформативных полей. Следовательно, усложняется процесс извлечения из нее данных.

В крупной РБД возникают проблемы с модернизацией структуры, т.к. при большом количестве отношений и их полей несложно запутаться и создать похожее отношение или нарушить связи. Также из-за несоответствия названий отношений и полей общепринятой терминологии предметной области возникают проблемы с освоением РБД при разработке прикладного программного обеспечения (ПО).

Проблемы терминологии также возникают при модернизации или написании нового прикладного ПО к РБД.

Для простого и корректного представления РБД предлагается использовать объектное представление (ОП) РБД, которое однозначно повторяет структуру РБД, имеет иерархический вид и скрывает неинформативные поля. Само ОП хранится в объектном словаре данных (ОСД) [1].

ОП рекомендуется для понимания и изучения структуры и содержания РБД; выполнения им функций посредника при работе с РБД; интеграции локальных РБД на уровне объектного представления.

В настоящее время средства для представления отношений в объектном виде являются коммерческими, и механизмы их работы недоступны для анализа и модернизации [2]. К недостаткам указанных средств представления отношений РБД в объектной форме можно отнести следующие:

— невозможность использования ОП для других систем управления базы данных (СУБД) кроме той, для которой они разработаны;

— названия отношений из РБД автоматически переносятся в ОП, что не всегда удобно, т.к. довольно часто отношения, представленные аббревиатурами на английском языке, не соответствуют именам соответствующих сущностей из предметной области;

— отсутствуют программные средства, позволяющие скрывать неинформативные поля и отношения;

— фирменное ПО имеют достаточно высокую стоимость.

Недостатки определяют необходимость создания ОП, отображающего структуру РБД и отвечающего следующим требованиям:

- возможность работы в любой СУБД;
- сохранение данных о структуре РБД со всеми отношениями, полями отношений и связями между отношениями;
- предоставление средств контроля за структурными изменениями РБД и способов получения информации из РБД;
- возможность сокрытия неинформативных полей и отношений от пользователя;
- возможность именования объектов и их атрибутов, в соответствии с принятой терминологией предметной области.

Фактически ОСД предоставляет при разработке прикладного ПО интерфейс для работы с РБД, при этом можно не знать ее структуру, наименование полей и таблиц.

Для работы с ОСД предлагается использовать специальный язык — объектный SQL (OSQL), который в основном сохраняет структуру и синтаксис SQL, но позволяет использовать не таблицы и поля, а объекты и их атрибуты [1].

Для составления математической модели РБД отношения, входящие в нее, представим в виде множества R , а поля отношений — в виде множества P .

Множество P , в свою очередь, рекомендуется представить в виде двух: V , содержащем ключевые поля, предназначенные для связи между отношениями, автоинкрементные и индексные, и B , содержащем все оставшиеся поля.

В итоге модель РБД будет иметь вид кортежа

$$M = \langle R, P, B, V \rangle \quad (1)$$

Для построения ОСД, адекватного представлению (1), строится математическая модель.

Необходимо исходить из следующих положений.

В ОП объекты соответствуют реальным отношениям РБД. Представим их множеством O . Также как и в РБД у отношений есть поля, так и у объектов есть атрибуты, которые соответствуют полям отношений. Представим их множеством A_O .

Объекты в ОП непосредственно связаны с отношениями в РБД, поэтому необходимо отобразить связи между отношениями и объектами. Представим эти связи множеством D_O . Атрибуты объектов также непосредственно связаны с полями отношений в РБД. Обозначим эти связи множеством D_{AOAR} . Аналогично связям между отношениями существуют связи между объектами. Представим эти связи множеством D_{OO} , а связи между отношениями в РБД — множеством V_O .

Предлагается ОП на две части: внешнюю и скрытую. Внешняя часть используется для решения прикладных задач. Скрытая — представляет собой связи с отношениями, полями отношений, внешние ключевые поля отношений, скрытые атрибуты объектов (внутренние ключи), внутренние ключи отношений. Множество скрытых атрибутов представим как O_K . Множество внутренних ключей отношений представим как A_K .

Таким образом, внешняя часть ОП представляет собой кортеж $\langle O, A_O, D_{OO} \rangle$, а скрытая часть — кортеж $\langle D_O, D_{AOAR}, V, O_K, A_K \rangle$.

Тогда объект будет представлен в виде

$$Ob = \langle O, A_O, D_O, D_{AOAR}, O_K, A_K \rangle, \quad (2)$$

а объектное представление РБД примет вид:

$$M_O = \langle Ob, D_{OO}, V_O \rangle. \quad (3)$$

Соответствие между моделями (1) и (3) достигается тем, что

$$\{O_1, O_2, \dots, O_N\} \text{ соответствует } \{R_1, R_2, \dots, R_N\},$$

$$A_{O_i} \text{ соответствует } B_i,$$

$$D_{DIOJ} \text{ соответствует } V_{IJ},$$

O_{ik} соответствует P_i .

Следовательно, предложенное ОП позволяет адекватно представлять РБД.

Представлены связи ОСД с РБД, где O_1, O_2, O_3 обозначают объекты, а R_1, R_2, R_3 соответствующие им отношения. $P_{R_1}, P_{R_2}, P_{R_3}$ являются соответственно полями отношений R_1, R_2, R_3 (рис. 1). Этим полям соответствуют атрибуты $A_{O_1}, A_{O_2}, A_{O_3}$ объектов O_1, O_2, O_3 . Связи между отношениями и объектами обозначены D_{OR_1} (связь между O_1 и R_1), D_{OR_2} (связь между O_2 и R_2) и D_{OR_3} (связь между O_3 и R_3). Скрытые внутренние ключи обозначены $V_{K_2} = O_{K_2}$ и $V_{K_3} = O_{K_3}$. Связи между объектами обозначены D_{OO_1} (связь между O_2 и O_1) и D_{OO_2} (связь между O_3 и O_2). Связь между отношениями обозначена V_1 .

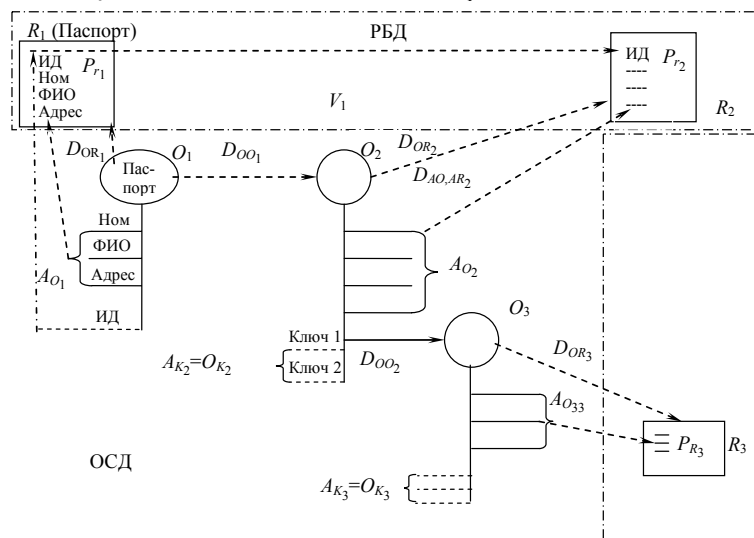


Рис. 1. Схема связей между ОП и РБД

В ОСД отображаются только те объекты и их атрибуты, которые относятся к видимой части ОП, а также связи между объектами (рис. 2).

В РБД и соответственно в ОСД существует проблема корректного именования отношений и их полей. Имена отношений и их поля могут и не соответствовать терминологии, принятой в предметной области, к которой относится РБД, что создает дополнительные трудности при использовании отношений и полей в прикладных задачах.

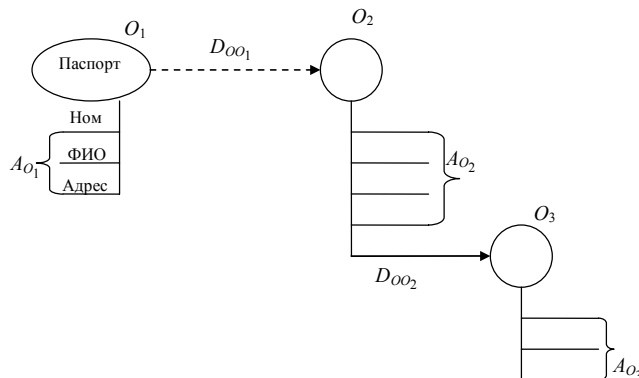


Рис. 2. Видимая часть схемы ОСД

Для корректного именования объектов и их атрибутов в терминах предметной области при построении ОСД предлагается применить словарь предметной области (СПО)[3]. Источником

данных для формирования СПО являются документы, используемые в организации (отчеты, должностные инструкции, справки, приказы, внутренние документы, внешние документы). Качество СПО напрямую зависит от качества и количества выбранных документов.

СПО представляет собой список терминов предметной области с синонимами из морфологически, синтаксически и семантически разобранного текста. Использование синонимов значительно повышает степень узнаваемости объектов и их атрибутов.

Каждый термин в СПО представляется как

$$CPO = \{T, N, S\}, \quad (4)$$

где T — множество имен,

N — множество групп ссылок на синонимы,

S — количество появлений каждого термина.

Каждому $T_i \in T$ соответствует группа ссылок на синонимы $N_i \in N$.

При составлении ОСД имена объектов и их атрибутов должны быть взяты из СПО. Это отражается отношениями $O \in T, A \in T$.

Создание объектного словаря позволяет найти эффективное решение для интеграции локальных БД. Например, пусть требуется объединить базы данных двух филиалов некоторой корпорации. Первый занимается производством некоторых узлов автомобиля, а второй — сборкой автомобиля. Соответствующие БД представляют разные предметные области, которые, однако, имеют общие сущности.

Существует несколько вариантов решения:

— физическая интеграция двух БД в одну;

— создание некоторого прикладного ПО, которое будет использовать две БД.

В первом варианте строится общая БД, в которой будут храниться непосредственно все отношения и их поля интегрируемых БД. Размер объединенной БД равен сумме размеров составляющих ее БД. Кроме этого такое решение требует модификации существующего прикладного ПО не только для использования объединяемой БД, но и для решения локальных задач.

Недостатки второго варианта:

— необходимость изучения структуры каждой БД;

— большие временные и материальные затраты на разработку ПО;

— ПО, объединяющее две БД, будет сложным и не универсальным.

Учитывая указанные недостатки, предлагается выполнить интеграцию БД путем создания центрального ОСД (ЦОСД) [4]. При этом можно работать как с локальным ОСД, так и с ЦОСД. При работе с ЦОСД в распоряжении оказываются объекты из объединенных ОСД.

Интеграция БД из двух предметных областей целесообразна, если пересечение множества имен соответствующих СПО не пустое, т.е.

$$T_1 \cap T_2 \neq \emptyset, \quad (5)$$

где $T_1 \in CPO_1$, а $T_2 \in CPO_2$.

Интеграцию двух БД имеет смысл выполнять, если пересечение имен и связанных с ним синонимов имен объектов из соответствующих ОСД не пустое, т.е.

$$O_1 \cap O_2 \neq \emptyset, \quad (6)$$

где $O_1 \in M_{O_1}$, а $O_2 \in M_{O_2}$.

При интеграции с помощью ОСД возникает проблема соответствия объектов из разных ОСД. В процессе работы интегрированных РБД может возникнуть проблема идентификации экземпляров объектов.

Поскольку в СПО для разных предметных областей возможно несовпадение имен одинаковых понятий, возрастает роль синонимов. Использование синонимов в некоторой степени порождает сомнения в идентичности сопоставляемых понятий.

Для решения проблемы узнавания объектов предлагается ввести понятие “идентифицирующий атрибут” (ИА). ИА может быть любой атрибут, который самостоятельно или в комбинации с другими ИА способен установить сущность некоторого объекта. Например, для объекта “личность” список ИА может быть следующим: номер паспорта, идентификационный код, фамилия, имя, отчество, дата рождения. Для каждого объекта не имеет смысла использовать все ИА. Нужно сформировать группы ИА (ГИА), каждая из которых дает возможность идентифицировать объект и по возможности содержит минимальное количество элементов. Например, для объекта “личность” ГИА будут следующими:

- фамилия, дата рождения;
- фамилия, номер паспорта;
- номер паспорта, дата рождения;
- идентификационный код;
- номер паспорта.

Таким образом, сравнение объектов между собой можно выполнять, сравнивая ГИА.

Если хотя бы одна ГИА является общей, то объекты представляют одну сущность, и их можно объединить.

Дополним структуру объекта (2) в объектном представлении ИА.

$Ob = \langle O, A_O, D_O, D_{AOAR}, O_K, A_K, I \rangle$, где $I = \{I_1, I_2, \dots, I_Q\}$ — множество комбинаций ИА.

Для каждого объекта $I_j = \langle \{A_{O1}, A_{O2}, \dots, A_{OG}\}, \{A_{O3}, A_{O4}, \dots, A_{OG}\}, \dots, \{A_{O6}, A_{O9}, \dots, A_{OG}\} \rangle$.

Для решения проблемы идентификации экземпляра объекта также можно использовать идентифицирующие атрибуты, но требуются дополнительные исследования.

ИА также могут использоваться для идентификации экземпляров объектов. Такая задача возникает при использовании ЦОСД. В общем случае ГИА для этой задачи будут другими. Например, если определены объекты “студент” и “преподаватель”, то выбор конкретного лица можно выполнить только по номеру и серии паспорта.

В заключении можно отметить, что предложенное объектное представление позволяет адекватно отображать РБД. Использование словаря предметной области для построения объектного представления РБД дает возможность установить единую и понятную терминологию для именования объектов и атрибутов. Предложенный механизм идентифицирующих атрибутов позволяет установить соответствие между элементами объектных представлений интегрируемых РБД.

В соответствии с предложенным создано программное обеспечение для построения словаря предметной области и объектного словаря данных, независимое от типов используемых СУБД, продолжаются исследования в направлении идентификации экземпляров объектов и обработки запросов к ЦОСД.

Литература

1. Завалин А. А. Использование словаря данных в информационных системах с логической формой представления данных. Завалин А. А., Кунгурцев А. Б., Блажко А. А. // Тр. Одес. политехн. ун-та. — Одесса, 2002. — Вып. 2 (18). — С. 121 — 126.
2. Объектное представление о реляционной модели. www.alexus.ru/russian/articles/dbms/ooop_rm/index.htm. 18.09.2006
3. Кунгурцев А. Б. Формирование словаря предметной области. Кунгурцев А. Б., Тыхан И.В. // Искусств. интеллект. — Одесса, 2006. — №1. — С. 166 — 172
4. Кунгурцев А.Б. Интеграция баз данных на основе объектного представления. Кунгурцев А.Б., Неизвестный А.С. // Зб. наук. пр. Одес. інституту Сухопут. військ. — Одеса, 2006. — Вип. 12. — С. 63 — 65

Поступила в редакцию 27 марта 2007 г.

