

ПОДДЕРЖКА ЦЕЛОСТНОСТИ РЕЛЯЦИОННОЙ МОДЕЛИ ДАННЫХ ПРИ МНОГОЗНАЧНЫХ ФУНКЦИОНАЛЬНЫХ ЗАВИСИМОСТЯХ ДАННЫХ

Введение

В основе всех современных информационных и локальных, и распределенных систем находятся базы данных. Опыт разработки, внедрения и эксплуатации таких систем показал, что наиболее эффективной структурой, удовлетворяющей требованиям как разработчика, так и пользователя, является реляционная модель данных. Теоретические основы проектирования реляционных баз данных, предложенные Коддом [1], с 1970 года по настоящее время практически не изменились.

Неоднократно предпринимались попытки ревизии этой модели, замены ее на различные вариации с объектно-ориентированным подходом, однако опыт разработчиков и время показали – по совокупности выразительных средств и математической основы реляционная модель данных еще долго будет основной моделью при проектировании как локальных, так и распределенных баз данных.

Рассматриваемая статья посвящена исследованию методов поддержки целостности данных в реляционных системах.

1. Проектирование модели базы данных информационной системы

При проектировании информационной системы необходимо провести анализ целей этой системы и выявить требования к ней отдельных пользователей. База данных – это некоторая целевая модель предметной области, в ней находят отражение только те факты о предметной области, которые необходимы для функционирования информационной системы. Предметная область определена, если известны существующие в ней объекты, их свойства и отношения. Проектирование модели базы данных начинается с предварительной структуризации предметной области: объекты реального мира подвергаются классификации, фиксируется совокупность подлежащих отображению в модели типов объектов.

Для каждого типа объектов фиксируется совокупность свойств, посредством которых будут описываться конкретные объекты этого типа, виды отношений и виды взаимосвязей между ними. Сущность инфологического подхода к проектированию информационных систем заключается в установлении соответствия между состоянием предметной области, его восприятием и представлением в базе данных. Согласно инфологическому подходу при проектировании выделяют основные составляющие: объект, свойство, связь (или объектное отношение), время.

Объект в инфологическом подходе – это то, о чем должна накапливаться информация в информационной системе. Выбор объектов производится в соответствии с целевым назначением информационной системы. Объекты могут рассматриваться как атомарные или как составные, причем один и тот же объект в одном приложении может рассматриваться как атомарный, а в другом – как составной. Для составного объекта должны быть определены его внутренние составляющие, внутренняя структура, в соответствии с которой определяется порядок композиции составляющих. Свойства объекта могут не зависеть от его связей (объектных отношений) с другими объектами, т. е. являться локальными [2].

При выборе той или иной модели данных проектировщик определяет способы и формы представления данных на внешнем, пользовательском уровне. Следует подчеркнуть, что современные системы управления базами данных (СУБД) ориентированы в основном на обработку так называемой фактографической информации, представляемой упорядоченной совокупностью фактов (значений в виде чисел, строк символов, логических констант и т.п.). Эти значения, как правило, объединяются в сравнительно небольшие последовательности – записи, которые соответствуют отдельным объектам (явлениям) прикладной предметной области. Совокупности однородных записей объединяются в таблицы (такая форма представления информации, как известно, наиболее широко применяется как в «ручных», так и автоматизированных информационных системах). Таким образом, модель данных задает правила структуризации таблиц, их логического связывания в базу данных, а также правила выполнения операций над элементами «логической» структуры данных.

Несмотря на то, что существует множество моделей представления табличных данных, в настоящее время доминирующее положение на рынке коммерческих СУБД занимают реляционные системы, т.е. СУБД, поддерживающие реляционную модель данных.

2. Реляционная модель данных в задачах интеллектуального анализа

Рассмотрим общий подход к проектированию реляционной базы данных.

Для построения структурной схемы баз данных используются традиционные средства спецификации реляционной модели данных. Основной структурной единицей данных в реляционной

модели является n -арное отношение, представляющее собой конечное подмножество декартова произведения доменов, т.е. множеств атомарных значений элементов данных – атрибутов отношения.

Пусть R – конечное множество имен отношений базы данных; $D = D_1, \dots, D_n$ – множество доменов, где всякий домен D_i есть именованное множество атомарных значений элементов данных; A – конечное множество имен атрибутов отношений; dom – отображение из A в D , определяющее из какого домена выбираются значения атрибутов.

Пару $\langle A_i, \text{dom}A_i \rangle$, где $A_i \in A$, называют атрибутом. Структурную схему S_i отношения R_i $\mathbb{R}_i \in \mathbb{R}$ можно представить в виде $R_i(A_1, \dots, A_n)$, в котором все A_i различны. Отношение r_i можно определить как расширение схемы $S_i : r_i \subseteq \text{dom}A_1 \times \dots \times \text{dom}A_n$.

Перестановка атрибутов в схеме не порождает нового расширения, и множество A_1, \dots, A_n атрибутов отношения R_i задает тип отношения. Для спецификации состава носителя используется выражение $R_i = A_1 \dots A_n$. Структурная схема U реляционной базы данных – это спецификация вида $\mathbb{R}_1, \dots, \mathbb{R}_p$, где $R_i \in \mathbb{R}$ и все R_i различны.

Концептуально реляционная база является информационно-логической моделью некоторой предметной области, такой, что каждое из расширений соответствует некоторому состоянию данной области в определенный момент текущего времени. Каждое состояние моделируется упорядоченной совокупностью значений элементов данных, соответствующих значениям свойств объектов предметной области.

Объекту определенного типа соответствует кортеж отношения определенного типа. Заметим, что реляционная модель данных предполагает сильную типизацию объектов, использование вполне определенных категорий, таких как тип объекта, атрибут (свойство) объекта, домен и отнесение каждого значения и упорядоченной совокупности значений к одной из этих категорий. Объекты определенного типа обладают определенным набором свойств, что задается в реляционной модели схемой отношения.

3. Ограничения целостности в реляционной модели данных

Одним из основополагающих понятий в технологии баз данных является понятие целостности. В общем случае это понятие связано с тем, что база данных отражает в информационном виде некоторый объект реального мира или совокупность взаимосвязанных объектов реального мира. Под целостностью будем понимать соответствие информационной модели предметной области, хранимой в базе данных, объектам реального мира и их взаимосвязям в каждый момент времени.

Целостность (от англ. integrity – неприкосновенность, сохранность) – понимается как правильность данных в любой момент времени. Эта цель может быть достигнута лишь в определенных пределах: СУБД не может контролировать правильность каждого отдельного значения, вводимого в базу данных. Например, нельзя обнаружить, что вводимое значение 2, представляющее номер дня недели, в действительности должно быть равно 3. С другой стороны, значение 9 явно будет ошибочным и СУБД должна его отвергнуть. Однако для этого ей следует сообщить, что номера дней недели должны принадлежать набору (1, 2, 3, 4, 5, 6, 7).

Поддержание целостности базы данных может рассматриваться как защита данных от неверных изменений. Современные СУБД имеют ряд средств для обеспечения поддержания целостности так же, как и средств обеспечения поддержания безопасности.

Выделяют три группы правил целостности: целостность по сущностям, целостность по ссылкам, целостность, определяемая пользователем. Рассмотрим мотивировку двух правил целостности, общих для любых реляционных баз данных.

1. Не допускается, чтобы какой-либо атрибут, участвующий в первичном ключе, принимал неопределенное значение.
2. Значение внешнего ключа должно либо быть равным значению первичного ключа цели, либо полностью неопределенным.
3. Для любой конкретной базы данных существует ряд дополнительных специфических правил, которые относятся к ней одной и определяются разработчиком. Чаще всего контролируется уникальность тех или иных атрибутов, диапазон значений (экзаменационная оценка от 2 до 5), принадлежность набору значений (пол "М" или "Ж").

В реляционной модели объекты реального мира представлены в виде совокупности взаимосвязанных отношений. Любое изменение в предметной области, значимое для построенной модели, должно отражаться в базе данных, и при этом должна сохраняться однозначная интерпретация информационной модели в терминах предметной области.

Мы отметили, что только существенные или значимые изменения предметной области должны отслеживаться в информационной модели. Действительно, модель всегда представляет собой некоторое

упрощение реального объекта, в модели мы отражаем только то, что важно для решения конкретного набора задач.

Рассмотрим фрагмент информационной системы «Библиотека» и реляционную модель базы данных, представленную на рис. 1.

В качестве основных сущностей предметной области «Библиотека» можно выделить следующие:

- СТУДЕНТ(Номер зачетной книжки, Фамилия, Пол, Группа, Дата рождения);

- ЛИТЕРАТУРА(Код книги, Название, Автор, Издательство, Год, Количество страниц, Тип литературы).

Очевидно, в качестве основного логического (или семантического) ограничения целостности будет следующее: любой студент в один день посещения библиотеки не может взять на руки две одинаковые книги или методички, два одинаковых справочника и т.д. Какими средствами реляционной модели поддерживается это ограничение? Решить эту задачу довольно просто. Отношение «СТУДЕНТ» и отношение «ЛИТЕРАТУРА» средствами реляционной модели будут связаны между собой промежуточным отношением «СТУДЕНТЫ_ЛИТЕРАТУРА». Поддержка целостности при этом будет возложена на СУБД реляционного типа, а в качестве механизма поддержки целостности будут выступать свойства составного ключевого атрибута отношения:

СТУДЕНТЫ_ЛИТЕРАТУРА(Номер зачетной книжки, Код книги, Дата выдачи, Дата возврата).

СТУДЕНТЫ				ЛИТЕРАТУРА		СТУДЕНТЫ ЛИТЕРАТУРА		
Номер зачетки	Фамилия	Группа	Пол	Код книги	Автор	Номер зачетки	Код книги	Дата
1	Иванов	ПМ-08-1	М	Д.01.001	Мартин Д	1	Д.01.001	05.02.11
2	Петрова	ПМ-08-1	Ж	Д.01.002	Кодд Э	1	Д.01.001	12.02.11
3	Сидоров	ПМ-08-2	М			1	Д.01.001	15.02.11
						2	Д.01.001	05.02.11
						3	Д.01.002	05.02.11

Рис. 1. Фрагменты отношений информационной системы «Библиотека»

Рассмотренный пример наглядно показывает эффективность реляционного подхода: теоретически правильно спроектированная модель базы данных, в данном случае для информационной системы «Библиотека», автоматически защищает себя на уровне ограничений целостности данных на протяжении всего периода эксплуатации. СУБД реляционного типа, никогда не позволит нарушить заданные и реализованные на уровне модели ограничения целостности. Таким образом, можно сделать следующий вывод: основным средством поддержки логических ограничений целостности в реляционной модели данных является управление ключевыми атрибутами отношений. Приведем несколько примеров в подтверждение изложенного выше утверждения.

Пример 1.

Ограничения целостности. Студент учебного заведения один предмет может изучать на протяжении нескольких семестров. Структура отношения представлена на рис.2.

СТУДЕНТЫ СЕМЕСТРЫ ПРЕДМЕТЫ ПРЕПОДАВАТЕЛИ						
Код студента	Фамилия	Семестр	Код предмета	Предмет	Оценка	Преподаватель
1	Иванов	4	221	Физика	В	Сидоров В.А.
1	Иванов	5	221	Физика	С	Куркин П.А.
2	Петров	4	221	Физика	В	Сидоров В.А.
2	Петров	5	221	Физика	В	Куркин П.А.
2	Петров	6	231	Химия	А	Уваров А.О.

Рис. 2. Фрагменты отношения информационной системы «Деканат».

В отношении СТУДЕНТЫ_СЕМЕСТРЫ_ПРЕДМЕТЫ_ПРЕПОДАВАТЕЛИ ключевые атрибуты – Код студента, Семестр, Код предмета реализуют технологию поддержки ограничений целостности.

Пример 2.

Предметная область и ограничения целостности. В информационной системе «Склад» при формировании налоговых накладных следует учитывать ограничения целостности следующего вида: каждый день нумерация накладных начинается с первого номера. Таким образом, накладная с номером «один» будет встречаться каждый день, если склад в этот день работал. Структура отношения, поддерживающая на уровне ключевых атрибутов, представлена на рис.3.

НАКЛАДНЫЕ ПОСТАВЩИКИ ПОТРЕБИТЕЛИ						
Номер накладной	Дата накладной	Код поставщика	Код потребителя	Поставщик	Потребитель	Сумма накладной
1	05.02.11	1	1	СМИТ	ФОКСМАРТ	3800
2	05.02.11	1	2	СМИТ	АВС	5600
1	06.02.11	1	1	СМИТ	ФОКСМАРТ	12700
2	06.02.11	1	2	СМИТ	АВС	46700
3	06.02.11	1	3	СМИТ	МКС	14540

Рис. 3. Фрагмент отношения информационной системы «Склад»

В отношении НАКЛАДНЫЕ ПОСТАВЩИКИ ПОТРЕБИТЕЛИ ключевые атрибуты – **Номер накладной**, **Дата накладной** реализуют технологию поддержки целостности.

Однако существует класс задач, для которых выполнить логические ограничения целостности при помощи технологии составных ключевых атрибутов, не всегда удастся. К таким задачам относятся задачи различного рода расписаний. Рассмотрим подробно общую постановку такого класса задач.

Пусть в распоряжении разработчика информационной системы «Расписание занятий» имеются приведенные к третьей нормальной форме (ЗНФ) реляционные отношения: ПРЕПОДАВАТЕЛИ, АУДИТОРИИ, ПАРЫ:

ПРЕПОДАВАТЕЛИ		АУДИТОРИИ		ПАРЫ	
Код преп.	Фамилия	Код ауд.	Аудитория	Код пары	Пара
1	Иванов	1	101	1	первая
2	Петров	2	102	2	вторая
3	Мухин	3	103	3	третья

Рис. 4. Основные отношения информационной системы «Расписание занятий»

Ограничения целостности сформулируем в следующем виде:

- преподаватель может проводить занятия только один раз на одной паре в любой аудитории;
- одна аудитория на одной паре может быть занята только один раз.

На предварительном этапе синтеза общей модели данных логично объединить попарно отношения ПРЕПОДАВАТЕЛИ и ПАРЫ через промежуточное отношение ПРЕПОДАВАТЕЛИ_ПАРЫ. В этом случае двойной составной ключ **Код преподавателя** и **Код пары** позволит удовлетворить ограничениям целостности данных. Аналогично сформируем отношение АУДИТОРИИ_ПАРЫ(**Код аудитории** и **Код пары**). И это отношение решает задачу поддержки целостности. Однако остается открытым вопрос, какая будет общая реляционная модель базы данных, позволяющая хранить данные трех отношений: ПРЕПОДАВАТЕЛИ, АУДИТОРИИ, ПАРЫ. Сформируем, как следует из теории синтеза реляционных моделей, универсальное отношение и все атрибуты определим как ключевые:

ПРЕПОДАВАТЕЛИ_ПАРЫ_АУДИТОРИИ

Преподаватель	Пара	Аудитория
Иванов	первая	101
Иванов	первая	102
Иванов	вторая	101
Петров	вторая	101

Рис. 5. Отношение «ПРЕПОДАВАТЕЛИ_ПАРЫ_АУДИТОРИИ»

Анализ приведенного отношения позволяет сделать следующий вывод: если в отношении выделены атрибуты и объединены в двойной составной ключ (Преподаватель-Пара), тогда выполняются ограничения целостности – все кортежи уникальны и один преподаватель может быть добавлен в отношение только в сочетании с новой парой занятий. Это ограничение одного четкого правила. Но, как только в отношение добавляется новый атрибут, в нашем примере Аудитория, и этот атрибут добавляется в состав составного ключа, *предыдущее ограничение поддержки целостности двойного составного ключа автоматически нарушается*. На рис.5 все атрибуты ключевые, и ни о какой поддержке целостности на уровне Преподаватель-Пара уже речь идти не может. «Иванов» на «первой» паре может быть как в аудитории «101», так и в аудитории «102». С точки зрения реляционной модели все корректно – все кортежи различны, а ограничения целостности Преподаватель-Пара и Аудитория-Пара не выполняются.

В реляционной теории баз данных, рассмотренный выше случай относится к исследованию многозначных функциональных зависимостей общей теории нормализации отношений [3].

4. Декомпозиция универсального отношения при многозначных зависимостях атрибутов

Технология декомпозиции универсального отношения при многозначных зависимостях атрибутов позволит автоматически поддерживать ограничения целостности непосредственно при вводе данных.

Пусть r – отношение со схемой R . X, Y, Z – подмножества из R , такие что $Z = R - (X, Y)$.

Определение. Отношение r удовлетворяет множественным функциональным зависимостям тогда и только тогда, когда r разделяется без потерь на отношения $R_1 = XY$ и $R_2 = XZ$.

Следствие из определения. Если отношение r делимо на два отношения $R_1 = XY$ и отношение $R_2 = XZ$ без потерь, то отношение r может быть замещено эквивалентным отношением r' , удовлетворяющим ограничениям целостности, как для отношения R_1 , так и для отношения R_2 .

Декомпозиция отношений: $R_1(X, Y) \rightarrow R_{11}(X), R_{12}(Y); R_2(X, Z) \rightarrow R_{21}(Z); r'(X, Y, Z)$.

Методику преобразования отношения содержащего многозначные зависимости рассмотрим на примере отношения «ПРЕПОДАВАТЕЛИ_ПАРЫ_АУДИТОРИИ», представленного на рис.5. Исходное отношение $r = \langle \text{ПРЕПОДАВАТЕЛИ_ПАРЫ_АУДИТОРИИ} \rangle$. Отношения, на которые без потерь разделяется отношение r : $R_1 = \langle \text{ПРЕПОДАВАТЕЛЬ_ПАРА} \rangle$ $R_2 = \langle \text{АУДИТОРИЯ_ПАРА} \rangle$.

Декомпозиция отношений:

$R_1 = \langle \text{ПРЕПОДАВАТЕЛЬ_ПАРА} \rangle \rightarrow R_{11} = \langle \text{ПРЕПОДАВАТЕЛИ} \rangle; R_{12} = \langle \text{ПАРЫ} \rangle;$

$R_2 = \langle \text{АУДИТОРИИ_ПАРЫ} \rangle \rightarrow R_{21} = \langle \text{АУДИТОРИИ} \rangle;$

$r' = \langle \text{ПРЕПОДАВАТЕЛИ_ПАРЫ_АУДИТОРИИ} \rangle$

Нижнее подчеркивание обозначает ключ отношения.

Рассмотренный в статье метод поддержки целостности данных реляционных моделей в случае многозначных функциональных зависимостей позволит повысить надежность и эффективность информационных систем, основанных на технологии баз данных.

Выводы

На основе анализа особенностей проектирования реляционной модели данных рассмотрены основные проблемы поддержки целостности в случае многозначных функциональных зависимостей. В статье исследованы многозначные функциональные зависимости атрибутов и предложен метод поддержки целостности средствами реляционной модели. Приведен ряд примеров, поясняющих общую проблему поддержки целостности реляционных моделей данных, а также технология специфических модельных ограничений и метод решения поставленной задачи.

ЛИТЕРАТУРА:

1. Саймон, А.Р. Стратегические технологии баз данных: менеджмент на 2000 год : пер. с англ. // Под ред. и с предисл. М.Р. Коголовского. – М. : Финансы и статистика, 1999. – 479 с.
2. Конноли, Т., Бегг, К. Базы данных: проектирование, реализация и сопровождение. Теория и практика, 2-е изд. : пер с англ. – М. : «Вильямс», 2000. – 1120 с.
3. Мейер, Д. Теория реляционных баз данных : пер. с англ. – М. : Мир, 1987. – 608 с., ил.

Филатов Валентин Александрович

Доктор технических наук, профессор кафедры Искусственного интеллекта Харьковского национального университета радиотехники

Filatov_val@ukr.net

Дзюба Нина Степановна

Инженер 1 кат. Научно-исследовательской лаборатории системных технологий Харьковского национального университета радиотехники

Костина Зоя Леонидовна

аспірант кафедри Искусственного интеллекта Харьковского национального университета радиоэлектроники

ПІДТРИМКА ЦІЛІСНОСТІ РЕЛЯЦІЙНОЇ МОДЕЛІ ДАНИХ ПРИ БАГАТОЗНАЧНИХ ФУНКЦІОНАЛЬНИХ ЗАЛЕЖНОСТЯХ ДАНИХ

В.О. Філатов, Н.С. Дзюба, З.Л. Костіна

Стаття присвячена дослідженню багатозначних функціональних залежностей в реляційних моделях даних. Розглянуто основні проблеми підтримки цілісності засобами систем управління базами даних (СКБД) при наявності обмежень розглянутого типу. Запропоновано метод підтримки цілісності засобами реляційної моделі. Наведено ряд прикладів, пояснюючих загальну проблему підтримки цілісності реляційних моделей даних, а також технологія специфічних модельних обмежень і метод вирішення поставленого завдання.

Ключеві слова: реляційна модель, база даних, багатозначні функціональні залежності.

ПОДДЕРЖКА ЦЕЛОСТНОСТИ РЕЛЯЦИОННОЙ МОДЕЛИ ДАННЫХ ПРИ МНОГОЗНАЧНЫХ ФУНКЦИОНАЛЬНЫХ ЗАВИСИМОСТЯХ ДАННЫХ

В.А. Филатов, Н.С. Дзюба, З.Л. Костина

Статья посвящена исследованию многозначных функциональных зависимостей в реляционных моделях данных. Рассмотрены основные проблемы поддержки целостности средствами систем управления базами данных (СУБД) при наличии ограничений рассматриваемого типа. Предложен метод поддержки целостности средствами реляционной модели. Приведены ряд примеров, поясняющих общую проблему поддержки целостности реляционных моделей данных, а также технология специфических модельных ограничений и метод решения поставленной задачи.

Ключевые слова: реляционная модель, база данных, многозначные функциональные зависимости.

Support the integrity Relational Data Model FOR multivalued dependency

V. Filatov, N. Dzuba, Z. Kostina

The article is devoted to the multivalued dependencies in relational data models. The main problems support the integrity by the means of database management systems (DBMS) under the constraints of this type. A method to support the integrity with the means of the relational model proposed. A number of examples illustrating the general problem of supporting the integrity of relational data models, technology-specific restrictions on the model and the method of solving this problem are given.

Key words: relational model, database, multivalued dependency.