

УДК 681.3:510.22

Н. Н. БУСЛИК, канд. техн. наук

**К ОПРЕДЕЛЕНИЮ РАВЕНСТВА АТРИБУТОВ
РЕЛЯЦИОННОЙ БАЗЫ ДАННЫХ**

Для многих задач проектирования баз данных (БД) и семантического анализа запросов к БД ключевым является вопрос о сходстве атрибутов, принадлежащих одному или разным отно-

нениям. При его решении в научной литературе и на практике используется широкий спектр понятий от интуитивного понятия «одинаковых по смыслу» атрибутов до строгих теоретико-множественных и формально-логических определений [1, 2]. «Широта» спектра связана с двумя очевидными обстоятельствами. Во-первых, сходство атрибутов имеет несколько типов (степеней), среди которых мы, в частности, будем выделять равенство. (Можно назвать также такие типы сходства, как частичное совпадение значений, взаимная функциональная зависимость и т. д.). Во-вторых, поскольку атрибут задается своим именем и множеством значений, каждый тип сходства, в том числе и равенства, можно рассматривать в различных аспектах. Например, можно говорить о равенстве имен атрибутов, равенстве множеств допустимых значений, равенстве множеств значений в каждом допустимом состоянии БД.

В настоящей работе не ставится задача классификации всех аспектов сходства или равенства атрибутов. Вместо этого выбирается один аспект равенства, а именно, равенство множеств значений атрибутов в каждом допустимом состоянии БД. Такой тип равенства будем называть в дальнейшем вполне равенством и обозначать его символом « \equiv ». Рассматриваемый аспект равенства оказывается чрезвычайно важным на этапе построения схемы интегрированной реляционной БД. Так, вполне равенство может играть роль статического ограничения на состояния БД и использоваться для устранения дублирования данных [3, 4].

Несмотря на то, что формальные признаки вполне равенства прямо или косвенно вводятся многими авторами, существует необходимость строгого и однозначного толкования этого понятия. Прежде чем перейти к формулировке определений, уточним место вводимого понятия в системе понятий теории реляционных СД. На наш взгляд, различия в подходах к определению равенства атрибутов связаны с различиями концептуальных уровней моделирования БД. Мы будем исходить из того, что определение вполне равенства строится на нижнем концептуальном уровне, каким является уровень логической схемы. Более высокие уровни моделирования БД, такие, как даталогический, инфологический, будем называть в дальнейшем концептуальными, а уровень логической схемы — примитивным. Вводимое ниже понятие вполне равенства ни в коем случае не может заменить понятий «концептуального» равенства атрибутов. Наше определение лишь помогает уточнить, какое место займет то или иное «концептуальное» равенство при его интерпретации на примитивном уровне моделирования. Так, проводя мысленный эксперимент со всеми возможными «правильными» (т. е. допустимыми в концептуальном понимании) экземплярами отношений БД по сопоставлению атрибутов на предмет вполне равенства и получая положительный результат, можно говорить, что для этих атрибутов выполняется предикат концептуального равенства соответствующего типа.

Ввиду сказанного мы не будем пользоваться строгим понятием допустимого состояния БД. В противном случае возникает очевидное противоречие: определение вполне равенства вводится как совокупность некоторых признаков, касающихся всех допустимых состояний БД, в то же время допустимое состояние БД может даваться с использованием — в качестве ограничения — равенства атрибутов. При разрешении противоречия следует учитывать, что равенство атрибутов как ограничение на состояние БД может объявляться или предполагаться лишь на более высоком концептуальном уровне, поскольку вытекает не из вводимого определения, а из свойств, присущих хранимой (поступающей) в БД информации. Таким образом, объявление ограничения на состояние БД в виде равенства на примитивном уровне является вторичным и может применяться лишь для формализации процедуры контроля поступающей в БД информации (которая в силу ошибок и несанкционированных действий «источника» может не удовлетворять концептуально предполагаемым ограничениям целостности).

Итак, определение допустимого состояния БД является прерогативной концептуального моделирования. В то же время собственно понятие состояния БД, используемое на уровне примитивной модели, хорошо известно в научной литературе [2, 5] (в [2] употребляется термин «реализация БД» и будет использоваться нами в дальнейших определениях без дополнительных пояснений. Допустимое же состояние БД на примитивном уровне понимается как состояние БД, удовлетворяющее концептуальным ограничениям.

Перейдем теперь к определению вполне равенства. Параллельно будет обсуждаться роль вполне равных атрибутов как статических ограничений на состояния БД. Вначале рассмотрим атрибуты одного отношения БД.

О п р е д е л е н и е 1. Два атрибута одного отношения БД вполне равны, если, и только если, для каждого допустимого состояния БД значения этих атрибутов в каждом кортеже отношения совпадают.

Как видим, признак вполне равенства играет в данном определении роль необходимого и достаточного условия, накладываемого на значения атрибутов. Подобный способ определения используется и в дальнейшем.

Из определения 1 однозначно вытекает та роль, которую играет вполне равенство как статическое ограничение на состояние отношения БД. Такое ограничение является более специфическим по сравнению с важнейшим видом ограничений — функциональными зависимостями. Очевидно, из вполне равенства следует взаимная функциональная зависимость атрибутов, обратное в общем случае неверно.

Решение задачи о признаках вполне равенства атрибутов различных отношений проведем в несколько этапов. Трудности определения равенства в этом случае связаны прежде всего с тем, что

такого вида статическое ограничение является заведомо более слабым по сравнению с ограничениями, заданными внутри отношений. Таким образом, некорректная декларация равенства атрибутов различных отношений может войти в противоречие с ограничениями на допустимые состояния каждого из отношений. Кроме того, такой признак, как совпадение множеств значений атрибутов различных отношений, не сопровождаемый правилами о порядке сопоставления значений, может служить лишь признаком «эффемерного» равенства. (Заметим, что в соответствии с определением 1 сопоставляются значения атрибутов в одном кортеже отношения).

Таким образом, наша задача состоит в том, чтобы определить, в каких именно кортежах различных отношений БД следует искать совпадающие значения атрибутов, полагаемых вполне равными. По аналогии с определением 1 нам необходимо найти такое отношение, которое включало бы в себя кортежи обоих отношений, где находятся сравниваемые атрибуты. Для получения искомого отношения из двух исходных используем операцию эквисоединения. Напомним, что операция эквисоединения близка операции естественного соединения с тем отличием, что она сохраняет в результирующем отношении все атрибуты исходных отношений [6].

Очевидно, выбираемое нами эквисоединение должно удовлетворять требованию однозначности восстановления исходных отношений при соответствующей его декомпозиции. Нетрудно показать, что это требование выполняется, если соединение осуществляется по «эффемерно» равным атрибутам*. Действительно, если множества значений атрибутов, по которым выполняется соединение, не совпадают, то в соединении теряются кортежи исходных отношений. Единственным атрибутом отношения, значения которого не зависят от значений других атрибутов этого отношения и, следовательно, могут быть декларируемы как равные значениям атрибута другого отношения, является ключ. Таким образом, прежде, чем определить вполне равенство произвольных атрибутов различных отношений, необходимо ввести строгое понятие равенства ключей.

Поскольку ключ отношения может быть составным, уточним такие понятия, как значение группы атрибутов и равенство значений. Под значением группы атрибутов понимается кортеж, состоящий из значений этих атрибутов в одном кортеже отношения; порядок расположения элементов кортежа задается произвольно и взаимно однозначно соответствует некоторой перестановке атрибутов в группе. Два значения группы атрибутов называются равными, если и только если при одной и той же перестановке атрибутов соответствующие кортежи совпадают.

О п р е д е л е н и е 2. Два ключа двух различных отношений вполне равны, если, и только если, для каждого допустимого со-

* Речь идет об эквисоединении по условию равенства двух атрибутов — по одному из каждого отношения. Аналогичные рассуждения относятся к соединению по нескольким атрибутам.

стояния БД найдется одна и та же перестановка атрибутов одного из ключей, такая, что множества значений этих ключей будут равными.

Для удобства дальнейшего изложения введем несколько вспомогательных понятий.

Компонентами эквисоединения называются группы атрибутов (атрибуты) каждого из соединяемых отношений, по равенству значений которых осуществляется операция эквисоединения.

Каноническим соединением двух отношений называется их эквисоединение, хотя бы одним из компонентов которого является ключ отношения.

Каноническое соединение называется полным если, и только если, множества значений обоих его компонентов равны.

Все последующие определения вполне равенства используют понятие полного канонического соединения и имеют смысл только в случае, когда такое соединение отношений БД существует.

Определение 3. Два атрибута двух отношений вполне равны, если найдется полное каноническое соединение с фиксированными компонентами, в котором эти атрибуты вполне равны (в смысле определения 1).

Фраза «с фиксированными компонентами» в определении означает тот факт, что компоненты соединения в каждом рассматриваемом состоянии БД остаются одинаковыми.

Сформулированный в определении признак вполне равенства соответствует признаку вполне равенства из определения 1 в том смысле, что задается аналогичный порядок сравнения значений двух атрибутов: сравниваются значения из одного кортежа. Естественным является вопрос о единственности сформулированного в определении 3 признака, т. е. о необходимости заданного условия. Нетрудно заметить, например, что атрибуты вполне равных ключей двух отношений попарно вполне равны и, кроме того, не зависят от значений неключевых атрибутов. Можно ли использовать в качестве компонентов канонического соединения отдельные ключевые атрибуты составного ключа? Ответ отрицательный, поскольку значение ключевого атрибута зависит от значения ключа. Поясним это на примере.

Пусть заданы два отношения: ρ_1 с носителем $R_1 = A_1B_1C_1$ и ρ_2 с носителем $R_2 = A_2B_2C_2$. Пусть, далее, в каждом состоянии БД эти отношения равны в том смысле, что для каждого кортежа из ρ_1 найдется равный ему кортеж в ρ_2 , и наоборот. Пусть ключом отношения ρ_1 является группа $K_1 = A_1B_1$, а ключом ρ_2 — группа $K_2 = A_2B_2$. Зададим теперь конкретный экземпляр отношения ρ_1 как множество кортежей $\{ \langle a_1b_1c_1 \rangle, \langle a_1b_2c_1 \rangle, \langle a_2b_2c_2 \rangle, \langle a_2b_2 \times c_1 \rangle \}$. Экземпляр отношения ρ_2 имеет такой же вид. Рассмотрим эквисоединение ρ_1 и ρ_2 по вполне равным ключевым атрибутам B_1 и B_2 . Экземпляр соединения имеет вид:

$\{ \langle a_1b_1c_1a_1b_1c_1 \rangle, \langle a_1b_1c_1a_2b_1c_2 \rangle, \langle a_2b_1c_2a_1b_1c_1 \rangle, \langle a_2b_1c_2a_2b_1c_2 \rangle, \langle a_1b_2c_1a_1b_2c_1 \rangle, \langle a_1b_2c_1a_2b_2c_1 \rangle, \langle a_2b_2c_1a_1b_2c_1 \rangle, \langle a_2b_2c_1a_2b_2c_1 \rangle \}$.

Здесь атрибуты A_1 и A_2 , C_1 и C_2 оказываются неравными, хотя в соответствии с определением 3 $A_1 \equiv A_2$, $C_1 \equiv C_2$.

Тем не менее условие из определения 3 не является необходимым для вполне равенства атрибутов различных отношений БД. Рассматривая вполне равенство как бинарное отношение на множестве атрибутов, легко убедиться, что в соответствии с определением 1 — это отношение эквивалентности. В то же время из определения 3 не следует свойство транзитивности вполне равенства: поскольку одно отношение БД может иметь несколько ключей, возможна ситуация, когда атрибут этого отношения, вполне равный атрибуту другого отношения, равен атрибуту третьего отношения (равенства $A_1 \equiv A_2$, $A_1 \equiv A_3$ обнаруживаются в соединениях по различным ключам первого отношения). Логично установить, что все три атрибута в описанной ситуации являются вполне равными. Для этого используем понятие полного канонического соединения нескольких отношений как последовательности полных канонических соединений пар отношений.

Определение 4. Два атрибута двух различных отношений БД вполне равны, если найдется полное каноническое соединение произвольного числа отношений БД, в котором эти атрибуты вполне равны.

Признак вполне равенства в определении 4 также не является необходимым условием. Поскольку определение вполне равных атрибутов основано на сопоставлении равных ключей, логично расширить это понятие за счет использования функциональных зависимостей атрибутов внутри отношений. А именно, если в отношениях БД заданы функциональные зависимости, то всю БД можно привести в третью нормальную форму (ЗНФ) без потерь информации и с сохранением заданных функциональных зависимостей [6]. Заметим, что в данном случае мы будем использовать не классическое разложение в ЗНФ, а множество всевозможных разложений каждого исходного отношения в ЗНФ, с тем, чтобы получить максимальную возможность сопоставления ключей и выявления вполне равных атрибутов. Заметим также, что получение ЗНФ является пределом разложения отношений для наших целей, так как дальнейшее разложение в нормальную форму Бойса-Кодда не сохраняет заданные функциональные зависимости [6].

Определение 5. Два атрибута двух различных отношений БД вполне равны, если, и только если, в ЗНФ БД найдется полное каноническое соединение, в котором эти атрибуты вполне равны.

Достаточность условия вполне равенства следует из самого определения (фразу «и только если» можно опустить). Интуитивным обоснованием его необходимости служит приведенная выше цепь рассуждений (от определения 1 к определению 5). Для формального обоснования необходимости условия сопоставим его с условием из определения 1, которое служит эталоном вполне равенства.

Теорема. Если в отношении БД атрибуты вполне равны в смысле определения 1, то при любой декомпозиции этого отношения без потерь информации и с сохранением заданных функциональных зависимостей эти атрибуты будут вполне равными в смысле определения 5.

Доказательство очевидно, поскольку какова бы ни была последовательность указанных декомпозиций исходного отношения, всегда можно провести обратные действия как операции полного канонического соединения. При этом гарантируется сохранение информации и заданных функциональных зависимостей.

Список литературы: 1. Хаббард Дж. Автоматизированное проектирование баз данных/Пер. с англ. М., 1984. 296 с. 2. Цикритзис Д., Лоховски Ф. Модели данных/Пер. с англ. М., 1985. 344 с. 3. Мартин Дж. Организация баз данных в вычислительных системах/Пер. с англ. М., 1980. 662 с. 4. Методы проектирования схемы реляционной базы данных/Н. Н. Буслик, Э. А. Дедиков, А. Н. Жадан и др.//Техника средств связи. Сер. ТЭУ. 1985. Вып. 2. С. 49—51. 5. Цаленко М. Ш. Семантические и математические модели баз данных. М., 1985. 208 с. 6. Ульман Дж. Основы систем баз данных/Пер. с англ. М., 1983. 334 с.

Поступила в редколлегию 14.06.89