

НАХОЖДЕНИЕ ВСЕХ ПОТЕНЦИАЛЬНЫХ КЛЮЧЕЙ ОТНОШЕНИЯ ПРИ ЛОГИЧЕСКОМ ПРОЕКТИРОВАНИИ РЕЛЯЦИОННЫХ БАЗ ДАННЫХ

Бутаков Е.А., Шаталова Ю.Г.

Социально-экономический университет,
КАФЕДРА ЭКОНОМИЧЕСКОЙ КИБЕРНЕТИКИ
ул. Героев Севастополя, 13, Севастополь, 340048, Украина,
E-MAIL: *program@msusevastopol.net*

Abstract

A search of all potential keys of the relation is the most important in the problem of database normalizing. The new method of all keys determination is presented. It is based on the matrix representation of boolean functions.

Введение

Важнейшей задачей компьютеризации является разработка баз данных предприятий. База данных должна быть адекватной предметной области и удовлетворять критериям качества; не иметь избыточности, быть свободной от аномалий включения, обновления и удаления. Известно, что база данных в том случае удовлетворяет перечисленным критериям, когда она нормализована и находится, по крайней мере, в третьей нормальной форме (ЗНФ) или в нормальной форме Бойса-Кодда (НФБК) [1]. Какую из этих нормальных форм выбрать решает проектировщик в зависимости от особенностей предметной области, однако НФБК является более строгой, а поэтому предпочтительней. Для приведения схемы базы данных к НФБК необходим анализ функциональных зависимостей (ФЗ) каждого отношения [5], отношение находится в НФБК если детерминант [1] каждой ФЗ этого отношения включает в себя потенциальный ключ [2]. Из этого определения следует *актуальная проблема*, которую необходимо решать при нормализации отношений, а именно — задача поиска всех потенциальных ключей отношения. Статья посвящена *новому методу решения* этой задачи.

1. Методы поиска потенциальных ключей отношения

Существует несколько методов поиска ключей отношения, точнее, методов доказательства того, что анализируемая совокупность атрибутов является одним из потенциальных ключей. Один из методов основан на использовании аксиом Армстронга [2], он заключается в сведении задачи к логическому выводу. Этот метод обладает всеми недостатками задач логического вывода: а) результат зависит от порядка применения аксиом; б) результат зависит от порядка выбора ФЗ, к которым

Заметим, что в (2) y_1, y_2, \dots, y_m — неизвестные, соответствующие конъюнкциям из переменных x_1, x_2, \dots, x_m . Задача нахождения возможных ключей сводится к нахождению этих конъюнкций.

2. Алгоритм поиска всех потенциальных ключей отношения, основанный на матричном представлении булевых функций

Входные данные: схема отношения R ; множество всех нетривиальных функциональных зависимостей F .

Выходные данные: все потенциальные ключи отношения R .

Основные шаги:

1. Сопоставляем атрибутам отношения R булевы переменные.
2. Находим характеристики всех функциональных зависимостей, принадлежащих F .
3. Составляем характеристическую функцию Φ множества F функциональных зависимостей [4].
4. Строим на диаграмме Карно изображение характеристической функции Φ .
5. Строим таблицу 1, в заголовок строк которой выносим символы полупространств с инверсией, а в строки помещаем конъюнкции без инверсий, составляющие в рассматриваемом полупространстве единичный интервал.
6. По определению ключа, конъюнкция из правой части таблицы 1 соответствует ключевому набору атрибутов тогда и только тогда, когда она входит во все полупространства. Для нахождения удовлетворяющих этому условию конъюнкций, составляем список имеющихся в правой части таблицы 1 конъюнкций и строим таблицу, в которой строки соответствуют инверсионным полупространствам, а столбцы — конъюнкциям из списка. Эту таблицу будем называть таблицей включений (В). На пересечении строки, соответствующей некоторому полупространству, и столбца помещаем единицу, если соответствующий конъюнкции интервал является единичным на этом полупространстве.
7. Таблицу включений дополняем, согласно свойству рефлексивности, т.е. помещаем единицы в те строки, символ полупространства которых входит в конъюнкцию, именуемую соответствующий столбец. В остальные элементы таблицы помещаем ноль. В результате получаем дополненную таблицу включений.
8. Совокупности атрибутов, соответствующие конъюнкциям, образующим имена столбцов из единичных элементов, являются потенциальными ключами отношения R .

Достоинством представленного метода является его алгоритмичность: на каждом шаге ясно, что надо делать и как это делать. Метод легко программируется. Другая его особенность заключается в том, что все потенциальные ключи находятся за однократное выполнение алгоритма.

3. Иллюстрация работы на примере

Пусть в отношении $R(A, B, C, D, E, G)$ имеют место следующие функциональные зависимости

$$F = \{ AB \rightarrow C, C \rightarrow A, BC \rightarrow D, ACD \rightarrow B, D \rightarrow EG, BE \rightarrow C, CG \rightarrow BD, CE \rightarrow AG \}.$$

1. Сопоставляем атрибутам A, B, C, D, E, G булевы переменные a, b, c, d, e, g .
2. Находим характеристики функциональных зависимостей:

$$ab\bar{c}, c\bar{a}, b\bar{c}\bar{d}, acd\bar{b}, d\bar{e}, d\bar{g}, be\bar{c}, cg\bar{b}, cg\bar{d}, ce\bar{a}, ce\bar{g}$$

3. Характеристическая функция множества R имеет следующий вид:

$$\Phi = ab\bar{c} \vee c\bar{a} \vee b\bar{c}\bar{d} \vee acd\bar{b} \vee d\bar{e} \vee d\bar{g} \vee be\bar{c} \vee cg\bar{b} \vee cg\bar{d} \vee ce\bar{a} \vee ce\bar{g}$$

4. Строим диаграмму Карно. Диаграмма Карно характеристической функции Φ представлена на рисунке 4.

Рис. 1. Матричное представление множества функциональных зависимостей отношении R

5. В каждом из полупространств $\bar{a}, \bar{b}, \bar{c}, \bar{d}, \bar{e}, \bar{g}$ находим единичные интервалы, соответствующие конъюнкциям, не содержащим переменные с инверсией. Строим таблицу аналогичную таблице 1.
6. Строим таблицу включений (таблица 2).
7. Дополняем таблицу включений (таблица 3).
8. Из таблицы 3 видно, что единичные столбцы именуют конъюнкции: $ab, bc, bd, be, cd, ce, cg$. Следовательно, в отношении $R(A, B, C, D, E, G)$ потенциальными ключами являются следующие совокупности атрибутов $AB, BC, BD, BE, CD, CE, CG$.

Таблица 1.

Полупространство	Интервалы, соответствующие конъюнкциям без инверсий переменных
\bar{a}	bc, bd, be, cd, ce
\bar{b}	cd, ce, cg
\bar{c}	ab, bd, be
\bar{d}	ab, bc, be, ce, cg
\bar{e}	$ab, ad, bc, bd, cd, cg, dg$
\bar{g}	ab, bc, bd, be, cd, ce

Таблица 2. (таблица включений)

Интервалы Полупространство	ab	ad	bc	bd	be	cd	ce	cg	dg
\bar{a}			1	1	1	1	1		
\bar{b}						1	1	1	
\bar{c}	1			1	1				
\bar{d}	1		1		1		1	1	
\bar{e}	1	1	1	1		1		1	1
\bar{g}	1		1	1	1	1	1		

Таблица 3. (дополненная таблица включений)

Интервалы Полупространство	ab	ad	bc	bd	be	cd	ce	cg	dg
\bar{a}	1	1	1	1	1	1	1	1	0
\bar{b}	1	0	1	1	1	1	1	1	0
\bar{c}	1	0	1	1	1	1	1	1	0
\bar{d}	1	1	1	1	1	1	1	1	1
\bar{e}	1	1	1	1	1	1	1	1	1
\bar{g}	1	0	1	1	1	1	1	1	1

СПИСОК ЛИТЕРАТУРЫ

1. Майер Д. Теория реляционных баз данных. // М: Мир. 1987. – 608с.
2. Дейт К.Дж. Введение в системы баз данных. // М: Наука.1998. – 464с.
3. Закревский А.Д. Логика распознавания. // Минск: Наука и техника.1988. – 118с.
4. Шаталова Ю.Г. Эвристический метод оптимизации множества функциональных зависимостей при проектировании реляционных баз данных. – // Искусственный интеллект. – Институт проблем искусственного интеллекта НАНУ. Донецк, 2000. – С. 252-257.

5. Wastl R. Liner Derivations for Keys of a Database Relation Shema // Journal of Universal Computer Science. – 1998. – Vol.4. – Issue12.– P. 883-890.