Tạp Chí Tin Học và Điều Khiển Học
Tập VIII(1992), Số 2(1-11)

Về các quan hệ tối giản ở dạng không chuẩn một

Trần Thái Sơn
Bộ nội vụ

I. Mở đầu.

Mô hình cơ sở dữ liệu dạng quan hệ do Codd [1] đưa ra đã thu hút được sự chú ý mạnh mẽ của các nhà nghiên cứu cũng như của những người làm thiết kế ứng dụng do tính đơn giản của mô hình, sự độc lập dữ liệu và cơ sở toán học khá cao của nó. Tuy nhiên, việc đòi hỏi mọi quan hệ trong cơ sở dữ liệu phải ở dạng chuẩn một (tức là giá trị tại mỗi thuộc tính của một bộ là giá trị đơn) là đòi hỏi quá chặt chẽ vì trong thực tế nhiều khi cần thiết phải mở rộng miền giá trị của thuộc tính, cho phép nó chứa giá trị - tập hợp hoặc giá trị quan hệ. Thí dụ trong quan hệ PHIEU - NHAN - SU, ở thuộc tính NGOAI - NGU có thể có chứa cả tập giá trị như (Anh, Nga, Pháp). Việc mở rộng mô hình Codd cho phép ta một mặt tránh được dư thừa (như một người chỉ cần biểu diễn bởi một bộ có chứa giá trị (Anh, Nga, Pháp) ở thuộc tính ngoại ngữ thay vì tách ra làm ba bộ), mặt khác cho phép biểu diễn thế giới thực một cách tự nhiên hơn (bằng cách cho phép có thuộc tính quan hệ, tức là cho phép biểu diễn cấu trúc phân cấp), đáp ứng phần nào yêu cầu " hướng đối tượng " của các hệ phần mềm hiện đại. Lần đầu tiên, mô hình dữ liệu quan hệ dạng không chuẩn một (NFR - non - first normal form relation, hay $\neg$1NF, hay Nested Relation Model) được đề cập đến trong [2], trong đó tác giả đề xuất mô hình và nghiên cứu sự chuẩn hoá thông qua các phụ thuộc hàm và phụ thuộc đa trị mở rộng. Từ đó đến nay, hàng loạt các công trình nghiên cứu đã ra đời, chủ yếu xoay quanh việc mở rộng các kết quả đã biết trong mô hình của Codd như chuẩn hoá quan hệ, nghiên cứu các kiểu phụ thuộc dữ liệu, vấn đề cập nhật dữ liệu, tối ưu hoá việc xử lý câu hỏi ... (xem [3], [4], [7], [8]). Ngoài các phép toán quen thuộc của mô hình Codd được mở rộng, trong mô hình NFR không thể thiếu được hai phép toán quan trọng: phép nhóm và phép tách. Trong [5] đã đưa ra các yêu cầu đối với hai phép toán này để có thể thực hiện các phép toán quan hệ như chiếu, kết nối ... được thuận tiện, tuy nhiên chưa trả lời được câu hỏi " với điều kiện gì thì các yêu cầu đó được đáp ứng." Trong [6], các tác giả đã đi xa hơn, nghiên cứu các tính chất của phép toán này dựa trên

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

các khái niệm phụ thuộc đa trị yếu và quan hệ chính quy. Trên cơ sở đó, các tác giả dã đưa ra điều kiện đủ để tập các quan hệ tối giản và tập các quan hệ chính tắc là đồng nhất. Bài báo này tiếp tục hướng nghiên cứu trên, chứng minh được rằng, trong trường hợp quan hệ được rút gọn trên tập gồm hai thuộc tính, điều kiện cần và đủ để tập các quan hệ tối giản và tập các quan hệ chính tắc bằng nhau là không tồn tại đỉnh treo trên đồ thị tương ứng với quan hệ hoặc quan hệ là chính quy trên tập thuộc tính đã nêu.

II. Một số khái niệm và kết quả cơ sở.

Mở đầu chúng tôi đưa ra định nghĩa về NFR. Các khái niệm cơ bản về mô hình của Codd có thể xem trong [9].

Giả sử R là lược đồ quan hệ trên tập thuộc tính U={E1,E2,...,En}, ở đó mỗi miền giá trị của Ei, ký hiệu dom(Ei) là nguyên tố. Với mỗi $a_i \in dom(Ei)$, $(a_1, a_2, ..., a_n)$ được gọi là n - bộ (hoặc bộ), ký hiệu $[E1(a_1), ...En(a_n)]$.

Bộ ghép trên $\{E1, ..., En\}$ là bộ có dạng

$$t = [E1(a_1^1, ..., a_1^{m_1}), ..., En(a_n^1, ..., a_1^{m_n})]$$

Tập các bộ ghép được gọi là quan hệ dạng không chuẩn một. Trừ khi có thể gây hiểu lầm, dom(E) sẽ được ký hiệu đơn giản là E.

Tập tách tương ứng của bộ t là tập các bộ

$$\{[E1(a_1^{i_1}), ..., En(a_n^{i_n})] : i_k = 1, ..., m_k, \forall k\}$$

và ký hiệu là t*. Hợp các tập tách tương ứng của mọi bộ của NFR r được gọi là quan hệ chuẩn một (1NF) tương ứng và ký hiệu là r*.

Cho hai bộ t_1 và t_2 trong NFR r, ở đó

$$t_1 = [E1(e_1), ..., En(e_n)]$$

$$t_2 = [E1(f_1), ..., En(f_n)],$$

e_1, f_1 là các bộ ghép.

Phép nhóm t_1 và t_2 trên E1, ký hiệu là $V(E1, t_1, t_2)$ được định nghĩa bằng $[E1(e_1 + f_1), E2(e_2), ...En(e_n)]$ nếu $e_i = f_i$ với i=2,3,...,n và bằng t_1 ở các trường hợp còn lại, ở đó $e_1 + f_1 = e_1 \cup f_1$. Trong trường hợp thứ nhất t_1 và t_2 gọi là *nhóm được*. Phép *nhóm hoàn toàn của r trên E1 là việc tiến hành mọi phép nhóm có thể trên E1*, ký hiệu V(E1,r). Ta ký hiệu V(E,V(F,r)) bởi V(EF,r).

Tương tự, *phép tách* của t_1 trên E1 cho ta tập

$$\{[E1(a_i), E2(e_2), ..., En(e_n)] : a_i \in e_i\}$$

Cho tập con $E \subseteq U$, U là tập tất cả các thuộc tính của quan hệ r, E={E1,...,Ek }. Phép nhóm r trên E được hiểu là phép nhóm trên Ei ∈ E. NFR r được gọi là *E - chính tắc* của 1NF s* với một hoán vị p của tập chỉ số { 1,...,k } khi và chỉ khi r được tạo nên bởi việc áp dụng liên tiếp phép nhóm hoàn toàn s* theo p:

$$r = V(Ep(1),...,Ep(k),\ s*).$$

r được gọi là *E - tối giản* của s* khi và chỉ khi r được tạo bởi các phép nhóm s* trên và chỉ trên E và không còn một phép nhóm nào trên E có thể thực hiện được nữa. Ký hiệu tập mọi quan hệ E - chính tắc của s* và tập mọi quan hệ E- tối giản tương ứng là Can(E,s*) và Irr(E,s*). Dễ dàng thấy Can(E,s*) có thể hơn một phần tử.

Ví dụ:

$s^* =$

A	B	C
a	b_1	c_1
a	b_1	c_2
a	b_2	c_2

A	B	C
a	b_1	c_1c_2
a	b_2	c_2

V(BC, S*)

A	B	C
a	b_1	c_1
a	b_1b_2	c_2

V(CB, S*)

E=$\{B,C\}$, $Can(E,s*) = \{V(BC,s*), V(CB,s*)\}$.

Ngoài ra có quan hệ là E - tối giản nhưng không phải là E - chính tắc:

A	B
a_1	b_1
a_1	b_2
a_2	b_2
a_2	b_3
a_3	b_3

A	B
a_1	b_1b_2
a_2	b_2
a_2a_3	b_3

Tính chất 2.1 [6]: $Can(E,s*) \subseteq Irr(E,s*)$.

Ưu điểm của các quan hệ chính tắc là được tạo bởi cách nhóm theo một trình tự nhất định. Còn ưu điểm của các quan hệ tối giản là có thể tiến hành cập nhật dữ liệu trên những vùng riêng biệt của quan hệ để tạo nên những quan hệ tối giản. Lớp chung của chúng kết hợp được những ưu điểm của cả hai loại quan hệ. Kết quả chủ yếu trong bài báo này là đưa ra điều kiện cần và đủ để chúng bằng nhau.

Định nghĩa: NFR r thỏa phụ thuộc đa trị yếu

$$U - EF - (W) \rightarrow E|F$$

khi và chỉ khi nếu trong r có ba bộ

$$\begin{matrix} a & b_1 & c_1 \\ a & b_2 & c_2 \\ a & b_1 & c_2 \end{matrix}$$

thì cũng phải có bộ $a \;\; b_2 \;\; c_1$

Tính chất 2.2 [7]:

$$V(EF,r) = V(FE,r) \Longleftrightarrow U - EF - (W) \rightarrow E|F$$

Định nghĩa: NFR r trên U= { E1...En,F1...Fm }, được gọi là *E - chính quy*, ở đó E={E1,...En} nếu và chỉ nếu r* thỏa các phụ thuộc

$$\{U - EiEj - -(W) \rightarrow Ei|Ej, \forall i,j = 1,...,n \; i \neq j\}$$

Định lý 2.1 [6]. *Irr(E,r*) = Can (E,r*) và chúng chỉ chứa một phần tử khi và chỉ khi là E - chính quy.*

Tuy nhiên chính quy không phải là điều kiện cần và đủ để Irr(E,r*)=Can(E,r*) trong trường hợp tổng quát như ví dụ dưới đây chỉ ra

r* =

A	B
1	1
2	1
2	2

r1 =

A	B
(1,2)	1
2	2

r2 =

A	B
1	1
2	(1,2)

Ở đây, Irr(E,r*)=Can(E,r*) và chứa hai phần tử. Phần sau sẽ chỉ ra điều kiện cần và đủ để Irr(E,r*)=Can(E,r*) trong trường hợp E={E1,E2}.

III. Các quan hệ chính tắc và tối giản trên tập hai thuộc tính.

Sau đây ta luôn luôn coi E={E1,E2}.

Định nghĩa: Tập $A_E(t)$ các bộ có liên quan với bộ t trên tập thuộc tính E của NFR r bao gồm:

1) $t \in A_E(t)$,

2) Nếu $u \in A_E(t)$ thì mọi $v \in r$, v có thể nhóm được với u trên Ei ∈ E, cũng thuộc $A_E(t)$, i ∈ {1,2}.

Định nghĩa: G(E,r)=(V,F) là một đồ thị vô hướng ứng với r trong đó:

- V là tập các đỉnh gồm và chỉ gồm các đỉnh ứng với mỗi bộ trong r, ký hiệu đỉnh bằng chính bộ tương ứng.

- F là tập các cạnh. Giữa hai đỉnh t_1 và t_2 có một cạnh nối khi và chỉ khi t_1 nhóm được với t_2 trên E.

Dễ thấy tập các $A_E(t)$ tạo ra một phân hoạch G(E,r) ra thành các thành phần liên thông. Vì các phép nhóm và tách chỉ tác động lên các bộ có liên quan nên từ đây trở về sau, nếu không nói rõ, chúng ta sẽ hiểu r *chỉ bao gồm một thành phần liên thông trên E.*

Định nghĩa: Đỉnh t $\in$ G(E,r) được gọi là *đỉnh treo* trên E nếu tồn tại hai cặp bộ u_1, u_2 và v_1, v_2 sao cho u_1 nhóm được với u_2 trên thuộc tính khác với thuộc tính trên đó u_2 nhóm được với t. Cũng như vậy với v_1, v_2.

Ví dụ có các bộ (a_1,b_1), (a_1,b), (a,b), (a,b_2), (a_2,b_2), trong đó (a,b) là đỉnh treo.

Để đơn giản cách viết mà không ảnh hưởng đến tính tổng quát của bài toán, ta sẽ coi r là quan hệ chỉ gồm hai thuộc tính E1,E2 (vì phần ngoài của $A_E(t)$ là giống nhau).

Định nghĩa: Hai quan hệ r_1, r_2 được gọi là bằng nhau qua phép nhóm, ký hiệu $r_1 \simeq r_2$ nếu mỗi quan hệ có thể nhóm bằng dãy các phép nhóm nào đó (không nhất thiết giống nhau) để cho cùng một quan hệ kết quả.)

Bổ đề 3.1 *Nếu Irr(E,r*) chứa hơn hai phần tử thì trong ba quan hệ bất kỳ thuộc Irr(E,r*) tồn tại một bộ thuộc một trong ba quan hệ đó và không thuộc hai quan hệ còn lại.*

Chứng minh. Giả sử có $r_1, r_1, r_3 \in Irr(E, r*)$ và i,j,k à một hoán vị của tập chỉ số {1,2,3 }.

Ta chứng minh bổ đề bằng phản chứng. Giả sử không tồn tại bộ nào thuộc r_i và không thuộc $r_i \cap r_k$. Khi đó r_i chỉ gồm các bộ đồng thời cũng có mặt trong r_j, hoặc r_k và r_i có thể tách làm ba tập hợp không giao nhau từng đôi một

$$
\begin{aligned}
q &= r_i \cap r_j \cap r_k \\
r_i^1 &= r_i \cap r_j - q \\
r_i^1 &= r_i \cap r_k - q
\end{aligned}
$$

Tương tự

$$
\begin{aligned}
r_j^1 &= r_i \cap r_j - q \\
r_j^2 &= r_j \cap r_k - q
\end{aligned}
$$

và

$$
\begin{aligned}
r_k^1 &= r_i \cap r_k - q \\
r_k^2 &= r_i \cap r_j - q
\end{aligned}
$$

Khi đó $r_k^1 = r_i^2 = r_i - r_i^1 + q$ và $r_k^2 = r_j^2 - r_j^1 + q$. Do $r_i^1 = r_j^1$ và $(r_i)* = (r_j)*$ nên $(r_k^1)* = (r_k^2)*$. Điều này vô lý vì trong $(r_k)*$ không có hai bộ nào trùng nhau. Bổ đề được chứng minh.

Bổ đề 3.2. *Nếu Irr(E,r*) chứa hơn hai phần tử thì tồn tại một đỉnh treo trong đồ thị G(E,r*) tương ứng.*

Chứng minh. Giả sử

$$r_1,\ r_2,\ r_3\ \in Irr(E, r*)$$

Xét trường hợp $r_i \cap r_j \neq \emptyset$.

Vì r* theo quy ước là một thành phần liên thông nên tồn tại hai bộ (a,b) và (a_1, b) sao cho (a,b) $\in$ $(r_1 - r_1 \cap r_2)*$, a, $a_1 \in Ei_1$, $b \in Ei_2$, i_1, i_2 là hoán vị của tập chỉ số { 1,2 }. Rõ ràng (a,b) $\in$ $(r_2 - r_1 \cap r_2)*$ do đó, tồn tại hai bộ khác nhau t, u chứa (a,b), $t \in r_1$, $u \in r_2$. Do $t \neq u$ nên ngoài a,b ra t hoặc u còn chứa ít nhất một giá trị nữa khác a_1. Vì vai trò t,u là đối xứng, chỉ cần xét hai trường hợp $t = ([aa_2], [b])$ và $t = ([a], [b, b_2])$, ở đó [x] ký hiệu tập chứa x, a_2, b_2 không nằm trong u.

Xét trường hợp thứ nhất. Do bộ (a_2, b) $\in$ $r*$ không nhóm vào được với u ở r_2, suy ra trong r* tồn tại bộ (a_2, b_2) (để nhóm với (a_2, b)) hoặc (a, b_2) (để nhóm với u). Ngoài ra, vì bộ (a_i, b) $\in (r_1 \cap r_2)*$ cũng không nhóm được với t nên trong r* tồn tại bộ (a_1, b_1) hoặc đồng thời hai bộ (a, b_1) và (a_2, b_1). Như vậy luôn tồn tại một trong các dãy bộ

$$\begin{array}{l}(a,b),\ (a_1,b),\ (a_2,b),\ (a_2,b_2),\ (a_1,b_1)\\ (a,b),\ (a_1,b),\ (a_2,b),\ (a,b_2),\ (a_1,b_1)\\ (a,b),\ (a_1,b),\ (a_2,b),\ (a_2,b_2),\ (a,b_1)\\ (a,b),\ (a_1,b),\ (a,b_2),\ (a_2,b_1),\ (a,b_1)\end{array}$$

với các đỉnh treo tương ứng là (a, b), (a_2, b), (a_1, b) và (a, b_2).

Trường hợp t=([a],[bb_2]) làm tương tự.

Trường hợp $r_1 \cap r_2 = \emptyset$. Vì r_1, r_2, r_3 có vai trò như nhau, có thể quy về trường hợp r_1, r_2, r_3 rời nhau hoàn toàn. Khi đó lấy một bộ (a,b) $\in$ r* bất kỳ, tồn tại ba bộ khác nhau từng đôi một t, u, v tương ứng thuộc r_1, r_2, r_3 và chứa (a,b). Từ sự khác nhau của t, u, v ta có các trường hợp sau:

$$t = ([aa_1], [b]),\ u = ([aa_2], [b]),\ v = ([a], [b])$$

hoặc

$$t = ([aa_1], [b]),\ u = ([a], [bb_1]),\ v = ([a], [b])$$

và các trường hợp đối xứng khác. Bằng cách chứng minh hoàn toàn tương tự ta thấy trong mọi trường hợp, r* đều có đỉnh treo. Bổ đề được chứng minh.

Định lý 3.1: *Nếu r* không có đỉnh treo trên E thì*

$$Irr(E, r*)\ =\ Can(E, r*)$$

Chứng minh. Từ bổ đề 3.2 suy ra Irr(E,r*) chứa khong quá hai phần tử. Nếu Irr(E,r*) chứa một phần tử, hiển nhiên Irr(E,r*) = Can(E,r*). Nếu Irr(E,r*) chứa hai phần tử, Can(E,r*) cũng chứa hai phần tử, vì nếu V(E1E2, r*) = V(E2E1,r*), tức Can(E,r*) chỉ chứa một phần tử thì theo Tính chất 2.2 có U-E1E2-(w)—E1 và do đó Irr(E,r*) cũng chỉ chứa một phần tử trái giả thiết. Từ đó có điều phải chứng minh.

Bổ đề 3.3: *Nếu $r_1 \simeq r_2$ và $t \in r_1$, $u \in r_2$ sao cho t và u giao nhau không rỗng trên tập E1 và E2 thì trong quan hệ tối giản được tạo nên bởi r_1 và r_2 có chứa bộ hợp cuả t và u.*

Chứng minh. Giả sử $c \in t[E1] \cap u[E1]$, $d \in t[E2] \cap u[E2]$. Giả sử s là bộ được tạo nên từ việc nhóm trong r_1 bộ t với các bộ nào đó, s sẽ chứa c và d trong các cột E1, E2, tương ứng. Như vậy từ r_2, s sẽ được tạo nên từ các phép nhóm có u tham gia vì trong r_2 (c,d) đã nằm trong u. Suy ra s phải chứa u. Vậy s chứa cả u lẫn t.

Định lý 3.2 *Nếu G(E,r*) có đỉnh treo thì*

$$Irr(E, r*) = Can(E, r*)$$

khi và chỉ khi r là E - chính quy.

Chứng minh. Khi r* là E - chính tắc thì

$$Irr(E, r*) = Can(E, r*)$$

theo định lý 2.1.

Bây giờ ta sẽ chứng minh chiều ngược lại.

Giả sử t=(a,b) là đỉnh treo và (x,y) $\in A_E(t)$ bất kỳ. Giả sử $t_1, t_2, ..., t_k$ là đường đi ngắn nhất từ t đến (x,y), ở đó a, x $\in Ei_1$; b, $y \in Ei_2$, i_1, i_2 là một hoán vị cuả tập chỉ số {1,2}. Ta có dãy (a_1, b_1), (a_1, b_2), (a_2, b_2), (a_2, b_3), ..., (a_n, b_n), ở đó $a_1 = a$, $b_1 = b$, $a_n = x$, $b_n = y$. Chú ý rằng vì t_1, t_2, ..., t_k là đường đi ngắn nhất từ t đến (x,y) nên nếu bộ t_{i-1} có giá trị trên E1 trùng với t_i thì t_i có giá trị trên E2 trùng với t_{i+1}.

Với $n \geq 3$, tức nếu đường đi từ (a,b) đến (x,y) có chứa ít nhất một đỉnh khác nào đó, ta sẽ chứng minh bằng quy nạp rằng trong quan hệ kết quả có bộ $([a_1...a_n], [b_1...b_n])$.

Với n=3 có dãy (a_1, b_1), (a_1, b_2), (a_2, b_2), (a_2, b_3) và (a_3, b_3). Ta thấy bước đầu có ba cách nhóm, cho ba quan hệ:

r1 =

E1	E2
a_1	$b_1 b_2$
a_2	$b_2 b_3$
a_3	b_3
..	..

r2 =

E1	E2
a_1	b_1
$a_1 a_2$	b_2
$a_2 a_3$	b_3
..	..

r3 =

E1	E2
a_1	$b_1 b_2$
a_2	b_2
$a_2 a_3$	b_3
..	..

Vì r1 và r2 là cách nhóm chính tắc, mà theo giả thiết thì Irr(E,r*) = Can(E,r*) tức Irr(E,r*) chứa không quá hai phần tử nên có r3≃r1 hoặc r3 ≃ r2. Giả sử có r3 ≃ r2. (Trường hợp r3 ≃ r1 hoàn toàn tương tự). Theo bổ đề 3.3, ta thấy trong quan hệ đích có bộ $([a_1, a_2], [b_1, b_2])$. Từ đó suy ra trong r* có bộ (a_2, b_1). Với sáu bộ đã biết, có thể có ba cách nhóm, cho ba quan hệ:

r4 =

$E1$	$E2$
a_1a_2	b_1b_2
a_2a_3	b_3
..	..
..	..
..	..

r5 =

$E1$	$E2$
a_1	b_1b_2
a_2	$b_1b_2b_3$
a_3	b_3
..	..
..	..

r6 =

$E1$	$E2$
a_1	b_1
a_1a_2	b_2
a_2	b_1b_3
a_3	b_3
..	..

Nếu r6≃r4 thì theo bổ đề dễ dàng thấy trong quan hệ đích phải có bộ $([a_1, a_2, a_3], [b_1, b_2, b_3])$. Nếu r6≃r5 thì có bộ $([a_1, a_2], [b_1, b_2, b_3])$. Trong trường hợp này r* phải có chứa thêm bộ (a_1, b_3). Ta lại có thể nhóm r* theo ba cách được các quan hệ:

r7 =

$E1$	$E2$
a_1a_2	$b_1b_2b_3$
a_3	b_3
..	..
..	..

r8 =

$E1$	$E2$
$a_1a_2a_3$	b_3
a_1a_2	b_1b_2
..	..
..	..

r9 =

$E1$	$E2$
a_1a_3	b_1
a_2	$b_1b_2b_3$
a_1	b_1b_2
..	..

Dễ thấy trong cả hai trường hợpr9≃r7 và r9≃r8 đều có theo bổ đề 3.3, trong quan hệ đích bộ $([a_1, a_2, a_3], [b_1, b_2, b_3])$.

Bây giờ khẳng định đúng với n, chứng minh nó cũng đúng với n+1. Ta có trong bước nhóm đầu, ba quan hệ như sau:

r1 =

$E1$	$E2$
a_1	b_1b_2
a_2	b_2b_3
..	..
a_n	b_nb_{n+1}
a_{n+1}	b_{n+1}
..	..

, r2 =

$E1$	$E2$
a_1	b_1
a_1a_2	b_2
..	..
a_na_{m+1}	b_{n+1}
..	..
..	..

, r3 =

$E1$	$E2$
a_1	b_1b_2
a_2	b_2
a_2a_3	b_3
..	..
a_na_{m+1}	b_{n+1}
..	..

Giả sử r3≃r2 có theo bổ đề 3.3, trong quan hệ đích bộ $([a_1, a_2], [b_1, b_2])$. Như vậy trong r* phải có thêm bộ (a_2, b_1). Từ các bộ đã có và thêm bộ (a_2, b_1) có thể có ba cách nhóm như sau:

$r4 =$

E1	E2
a_1	b_1b_2
a_2	$b_1b_2b_3$
..	..
a_n	b_nb_{n+1}
a_{n+1}	b_{n+1}
..	..

$r4 =$

E1	E2
a_1a_2	b_1b_2
a_2a_3	b_3
..	..
a_na_{n+1}	b_{n+1}
..	..

$r6 =$

E1	E2
a_1	b_1
a_1a_2	b_2
a_2	b_1b_3
a_3	b_3b_4
..	..
a_n	b_nb_{n+1}
a_{n+1}	b_{n+1}
..	..

Nếu r6≃r4 hay r6≃r5 đều có bộ $([a_1a_2], [b_1b_2b_3])$ trong quan hệ kết quả. Như vậy trong r* có thêm bộ (a_1, b_3). Ta sẽ có sau bước nhóm ban đầu một quan hệ có các bộ (a_1a_2, b_1b_2), (a_1a_2, b_3), $(a_3, b_3), \ldots, (a_{n+1}, b_{n+1})$. Nếu đặt $a_1a_2 = a'_2$, $b_1b_2 = b'_2$, ta thấy đường đi có độ dài ngắn đi một. Theo giả thiết quy nạp sẽ có trong quan hệ đích bộ $([a'_2a_3...a_{n+1}], [b'_2b_3...b_{n+1}])$ tức là $([a_1...a_{n+1}], [b_1...b_{n+1}])$.

Trường hợp n < 3 ta có thể có các dãy sau:

(a_1, b_1), (a_1, b_2), (a_2, b_2), (a_2, b_3);

(a_1, b_1), (a_1, b_2), (a_2, b_2);

(a_1, b_1), (a_1, b_2);

(a_1, b_1);

ở đó bộ đầu là (a,b), bộ cuối là (x,y) (trường hợp cuối (a,b) trùng với (x,y)). Trong hai trường hợp đầu, ngoài cặp bộ(a_1, b_2), (a_2, b_2) do (a_1, b_1) là đỉnh treo, trong r* còn tồn tại ít nhất cặp bộ (a_0, b_0) (a_0, b_1) (hoặc (a_0, b_0), (a_1, b_0)). Cả hai trường hợp này ta đều có đường đi từ (a_0, b_0) đến (x,y) có độ dài không nhỏ hơn 3, và theo kết quả phần trên ta có điều phải chứng minh. Trường hợp cuối khi (a,b)=(x,y) ta có ngay kết quả.

Xét trường hợp thứ ba, khi (x,y)=(a_1, b_2). Vì (a,b) là đỉnh treo nên theo định nghĩa tồn tại hai cặp bộ sao cho chúng nhóm được với nhau trên thuộc tính khác với thuộc tính mà trên đó chúng nhóm được với (a,b). Để xác định giả sử có (a_1, b_1), (a_1, b) và (a, b_2), (a_2, b_2), $(x, y) = (a, y)$. Cùng với (a,b) chúng có thể nhóm theo ba cách khác nhau cho ba quan hệ:

$$r1 = \begin{array}{|cc|}\hline E1 & E2 \\ a_1 & bb_1 \\ a & bb_2y \\ a_2 & b_2 \\ .. & .. \\ .. & .. \\ \hline \end{array}, \; r2 = \begin{array}{|cc|}\hline E1 & E2 \\ a_1 & b_1 \\ a_1a & b \\ aa_2 & b_2 \\ a & y \\ .. & .. \\ \hline \end{array}, \; r3 = \begin{array}{|cc|}\hline E1 & E2 \\ a_1 & bb_1 \\ a & by \\ aa_2 & b_2 \\ .. & .. \\ .. & .. \\ \hline \end{array}$$

Nếu, chẳng hạn, r3≃r2 thì có bộ ([a_1a], [by]). Như vậy trong r* có thêm bộ (a_1, y). Ta sẽ có đường đi từ (a,b) đến (a,y) qua các bộ (a, b_2), (a_2, b_2), (x, y) trở lại trường hợp đã chứng minh, khi n $\geq$ 3. Tóm lại, trong mọi trường hợp có $A_E(t)$ chỉ bao gồm một bộ. Từ đó suy ra r là E - chính quy. Định lý được chứng minh.

Kết luận

Trong bài báo, chúng tôi đã giải quyết được vấn đề đưa ra trong [6] là tìm điều kiện cần và đủ để tập Irr(E,r*)=Can(E,r*) cho trường hợp riêng quan trọng E={E1,E2}. Trường hợp E có nhiều hơn hai thuộc tính đang còn để mở. Khả năng giải quyết vấn đề có lẽ gắn liền với việc nghiên cứu các phụ thuộc dữ liệu kiểu mới (chẳng hạn phụ thuộc kết nối yếu WJD).

Tác giả xin chân thành cảm ơn phó tiến sỹ Lê Tiến Vượng và Đinh Thị Ngọc Thanh đã đóng góp nhiều ý kiến quý báu trong quá trình hình thành bài báo.

Tài liệu tham khảo

1. Codd E.F., *A relational model for large shared date bank*, Communs ASS. Comput. Mach. **13**(1970), p. 377-387.
2. Makinouchi A., *A consideration on normal form of not necessarily normalized relations in the relational data model*, Proc. Very large Data Basses (1977),447-453.
3. Mark A. Roth; Hendry F. Roth & Don S. Batory, SQL/NF: *A Query Language For ¬1NF Relational Databasses*, Intor. Systems **12**,(1978), p.99-114.
4. Tansel A.U. & Garnett L.L., *Nested Historical Relations*, SIGMOD Record, **18**(1989), n^0 2.
5. Desal B.C.; Goyal P. & Sadri F., *Non-first Normal Form Universal Relations*, An application to Information Retrieval Systems, Information Systems **12**(1987), p. 49-55.
6. Miura T & Moriya K.K..., *On the Irreducible Non-First Normal Form Relations*, Inform. Systems **12**(1987), n^0 3,p.229-238.
7. Fischer F. & Gucht D Van., *Weak multivalued dependencies*, Proc.ACM SIGACT SIGMOD Symp. on principles of Database Systems, Waterloo ONT,(1984), p. 266-274.
8. Kambayashi Y.; Tanaka K & Takeda K., Synthesis of unnormalized relation incorporating more meaninh, Infor. Sci. **29**(1983),p. 201-247.

9. Ultman J.D., Principies of database and knowledge - base systems, Computer Science Press, Rockville, (1989)

Abstract

On the irreducible Non - First normal form Relations

In this paper the propreties of the nest and unnest operations of Non - First normal form Relations (NFR) are discussed. Some necessary and sufficient conditions, under which a collection of irreducible relations and a collection of canonical relations over a set of attributes $E = \{E1, E2\}$ *are identified, are proposed.*