Tạp chí Tin học và Điều khiển học, T. 14, S. 3 (1998) (52–60)

MÔ HÌNH CƠ SỞ DỮ LIỆU DẠNG KHỐI

NGUYỄN XUÂN HUY[1], TRỊNH ĐÌNH THẮNG[2]

Abstract. At the present time, some types of database models are considered, in which three types of database models are objects of interest: Hierachiral model, Network model and Relational model. This paper introduces a new model for databases, which called a database model of block form. Some basis concepts and block form - relational algebra are introduced for this model. Some of properties related to keys of block schema and block form - relational algebra are claimed and proved. It is shown that the relational database model is special case of the database model of block form in which each relation is a block with the index set, containing only one element.

1. MỞ ĐẦU

Mô hình cơ sở dữ liệu quan hệ do Codd E. F. đưa ra [5] là một họ các quan hệ, trong đó các quan hệ có cấu trúc phẳng (tuyến tính). Do cấu trúc phẳng của quan hệ nên mô hình này chưa đủ đáp ứng đối với các ứng dụng phức tạp, các cơ sở dữ liệu có cấu trúc phi tuyến...

Để mở rộng và khắc phục phần nào những nhược điểm của mô hình quan hệ truyền thống nói trên, bài báo này đưa ra một mô hình cơ sở dữ liệu mới gọi là mô hình cơ sở dữ liệu dạng khối. Mô hình này giúp biểu diễn thế giới thực trong quá trình vận động một cách tự nhiên hơn.

Theo truyền thống của lý thuyết cơ sở dữ liệu ta sẽ sử dụng các kí hiệu sau đây:

Kí hiệu	Ý nghĩa
A, B, C	thuộc tính
X, Y, Z	tập thuộc tính
XY	$X \cup Y$ (hợp của 2 tập thuộc tính X và Y)
ABC	$\{A, B, C\}$ (tập thuộc tính gồm 3 phần tử A, B, C)
dom(A)	miền giá trị của thuộc tính A

Phần còn lại của bài báo này được chia thành 4 phần. Khái niệm về khối, lát cắt và các mệnh đề liên quan với chúng được trình bày trong phần 2. Phần 3 gồm khái niệm về khóa cùng một vài tính chất. Các phép tính trên cơ sở dữ liệu dạng khối được đưa ra ở phần 4, và cuối cùng, phần 5 trình bày đại số quan hệ dạng khối.

2. KHỐI VÀ LÁT CẮT

Khái niệm toán học của mô hình cơ sở dữ liệu dạng khối (gọi tắt là mô hình khối).

Định nghĩa 2.1. Gọi $R = (\mathrm{id}; A_1, A_2, ..., A_n)$ là một bộ hữu hạn các phần tử, trong đó id là tập chỉ số hữu hạn khác rỗng, A_i $(i = \overline{1, n})$ là các thuộc tính. Mỗi thuộc tính A_i $(i = \overline{1, n})$

có miền giá trị tương ứng là dom(A_i). Một khối r trên R, kí hiệu $r(R)$ gồm một số hữu hạn phần tử mà mỗi phần tử là một họ các ánh xạ từ tập chỉ số id đến các miền trị của các thuộc tính A_i, $(i = \overline{1,n})$. Nói cách khác:

$$t \in r(R) \Leftrightarrow t = \{t^i : \text{id} \to \text{dom}(A_i)\},\ i = \overline{1,n}.$$

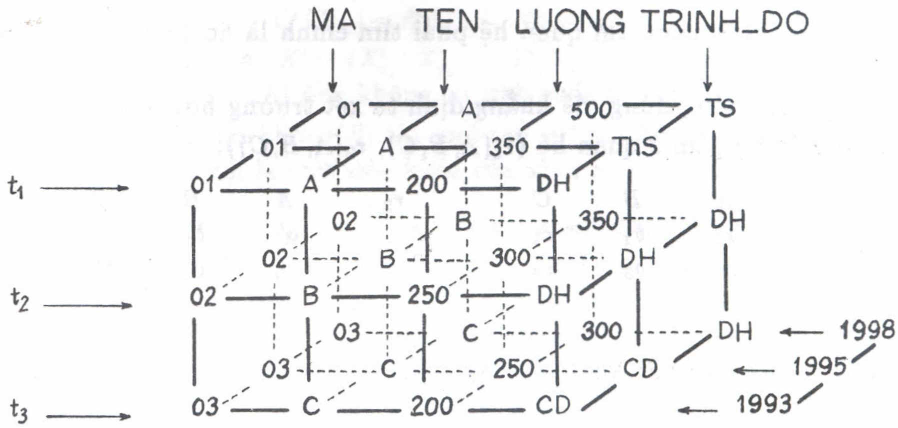
Ta kí hiệu khối đó là $r(R)$ hoặc $r(\text{id}; A_1, A_2, \ldots, A_n)$ đôi khi nếu không sợ nhầm lẫn ta kí hiệu đơn giản là r.

Khi đó khối $r(R)$ được gọi là có lược đồ khối R. Như vậy trên cùng một lược đồ khối R ta có thể xây dựng được nhiều khối khác nhau.

Ví dụ: Cho khối NS(R), ở đây $R = (\text{id}; A_1, A_2, A_3, A_4)$,

trong đó: id = {1993, 1995, 1998}

A_1 = ma, A_2 = ten, A_3 = luong, A_4 = trinh_do



Với khối NS(R) ở trên ta thấy nó gồm 3 phần tử: t_1, t_2, t_3. Khi đó ta có: $t_1(1993, \text{luong}) = 200$, $t_2(1995, \text{luong}) = 300$, $t_3(1998, \text{ma}) = 03$, $t_3(1995, \text{trinh\_do})$ = 'CD' ...

Định nghĩa 2.2. Cho $R = (\text{id}; A_1, A_2, \ldots, A_n)$, $r(R)$ là một khối trên R. Với mỗi $x \in \text{id}$ ta kí hiệu $r(R_x)$ là một khối với $R_x = (\{x\}; A_1, A_2, \ldots, A_n\}$ sao cho:

$$t_x \in r(R_x) \Leftrightarrow t_x = \left\{ t_x^i = t^i \Big|_x \right\}_{i=\overline{1,n}} \text{ với } t \in r(R), \text{ và } t = \left\{t^i : \text{id} \to \text{dom}(A_i)\right\}_{i=\overline{1,n}}.$$

Ở đây $t_x^i(x) = t^i(x)$ với $i = \overline{1,n}$.

Khi đó, $r(R_x)$ được gọi là một lát cắt trên khối $r(R)$ tại điểm x.

Ví dụ: Với khối NS(R) đã cho ở trên, $R = (\text{id}; A_1, A_2, A_3, A_4)$,

trong đó: id = {1993, 1995, 1998}

A_1 = ma, A_2 = ten, A_3 = luong, A_4 = trinh_do

Nếu $x = 1995 \in \text{id}$ thì lát cắt $r(R_{1995})$ có dạng như sau:

MA	TEN	LUONG	TRINH_DO
01	*A*	350	ThS
02	*B*	300	DH
03	*C*	250	CD

Ta có các mệnh đề sau.

Mệnh đề 2.1. *Cho $R = (id;\ A_1, A_2, ..., A_n)$, $r(R)$ là một khối trên R. Với mỗi $x \in id$ thì lát cắt $r(R_x)$ là một quan hệ. Trong trường hợp tập chỉ số id chỉ gồm một phần tử thì $r(R)$ trở thành một quan hệ.*

Như vậy mỗi quan hệ $r(A_1, A_2, ..., A_n)$ là một trường hợp đặc biệt của khối, đó chính là khối $r(R)$ với $R = (\{x\};\ A_1, A_2, ..., A_n)$.

Mệnh đề 2.2. *Cho $R = (id;\ A_1, A_2, ..., A_n)$, $r(R)$ là một khối trên R. Khi đó tồn tại một họ quan hệ duy nhất biểu diễn họ $\{r(R_x)\}_{x \in id}$ các lát cắt của khối $r(R)$. Ngược lại không đúng, nghĩa là với một họ quan hệ cho trước biểu diễn họ các lát cắt của một khối nào đó thì khối tìm được không duy nhất.*

Chứng minh:

1) Với khối $r(R)$ cho trước thì quan hệ phải tìm chính là họ $\{r(R)_x)\}_{x \in id}$ các lát cắt của khối $r(R)$.

2) Điều ngược lại không đúng, để khẳng định ta xét trường hợp sau:

Giả sử ta có một họ gồm 2 quan hệ $\{r_1(A, B, C),\ r_2(A, B, C)\}$:

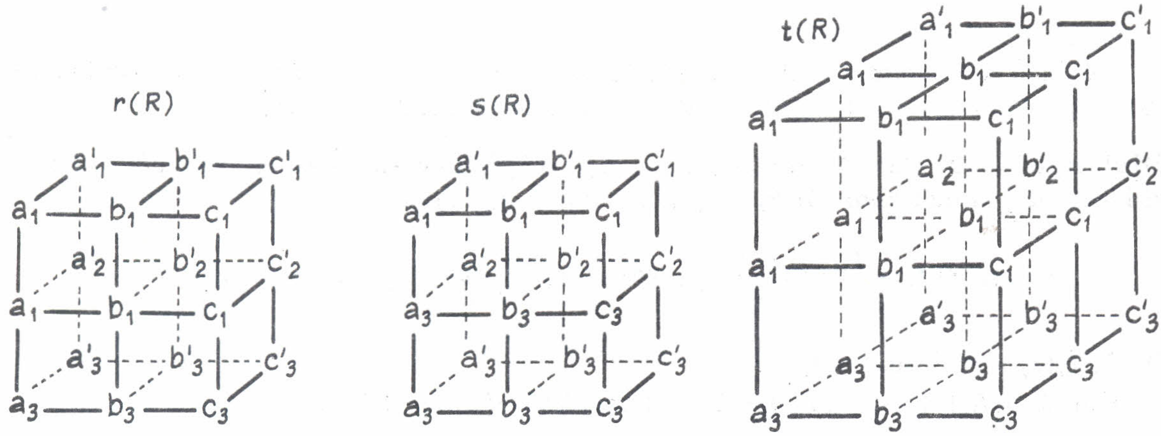
r_1 :

A	B	C
a_1	b_1	c_1
a_3	b_3	c_3

r_2 :

A	B	C
a'_1	b'_1	c'_1
a'_2	b'_2	c'_2
a'_3	b'_3	c'_3

Khi đó ta có chẳng hạn các khối sau đây nhận họ hai quan hệ $\{r_1, r_2\}$ nói trên là họ các lát cắt của nó:



Như vậy sự tồn tại của các khối có họ các lát cắt là họ quan hệ $\{r_1, r_2\}$ nói trên là không duy nhất.

3. KHÓA CỦA KHỐI $r(R)$

Cho $R = (\text{id};\ A_1, A_2, ..., A_n)$, r là một khối trên R. Với mỗi $x \in \text{id}$, $t \in r(R)$, $t = (t^1, t^2, ..., t^n)$, ta kí hiệu $t(x; A_i)$ $(i = 1, .., n)$ là giá trị của phần tử t ở thuộc tính A_i tại chỉ số x.

Để thuận lợi cho việc trình bày, ta đặt $x_i = (x; A_i)$, $x \in$id và như vậy: $t(x_i) = t(x; A_i) = t^i(x)(i = 1,..., n)$. Từ đó, ta kí hiệu:

$$\text{id}_i = \{x_i\}_{x\in \text{id}}\text{, như vậy id}_i = \{(x; A_i)\}_{x\in \text{id}}.$$

Với $X_i \subseteq \text{id}_i$ thì ta kí hiệu: $t(X_i) = \{t(y_i) \mid y_i \in X_i\}$.

Giả sử t_1, $t_2 \in r(R)$ với $t_1 = \{t_1^i : \text{id} \rightarrow \text{dom}(A_i)\}_{i=1,...,n}$, $t_2 = \{t_2^i : \text{id} \rightarrow \text{dom}(A_i)\}_{i=1,...,n}$, khi đó ta định nghĩa khóa của khối $r(R)$ như sau:

Định nghĩa 3.1. Khóa của khối r trên lược đồ khối $R = (\text{id}; A_1, A_2,..., A_n)$ là một tập $K = \{X_{i_1}, X_{i_2},..., X_{i_h}\}$, trong đó $X_{i_k} \neq \emptyset$, $X_{i_k} \subseteq \text{id}_{i_k}$ $(k = 1,..., h)$, thỏa mãn 2 tính chất:

a) Với bất kỳ 2 phần tử t_1, $t_2 \in r$ đều tồn tại một $X_{i_k} \in K$ sao cho:

$$t_1^{i_k}(X_{i_k}) \neq t_2^{i_k}(X_{i_k}).$$

Nói cách khác, không tồn tại 2 phần tử mà:

$$t_1^{i_k}(X_{i_k}) = t_2^{i_k}(X_{i_k}),\ \forall k = 1,..., h.$$

b) Với bất kỳ tập K' nào $K' = \{X'_{i_1}, X'_{i_2},..., X'_{i_h}\}$, với $X'_{i_k} \subseteq X_{i_k}$, $(k = 1,..., h)$ và tồn tại $X'_{i_m} \subset X_{i_m}$, với $m \in \{1, 2,..., h\}$ đều không có tính chất a) nói trên.

Nếu tập K là khóa của khối $r(R)$ thì mọi tập $K'' = \{X''_{i_1}, X''_{i_2},..., X''_{i_n}\}$, trong đó $X_{i_k} \subseteq X''_{i_k}$ $(\forall k = 1,..., h)$, được gọi là một siêu khóa của khối r.

Mệnh đề 3.1. *Cho $R = (id; A_1, A_2,..., A_n)$, $r(R)$ là một khối trên R, $id = \{x\}$. Khi đó $r(R)$ trở thành quan hệ $r(A_1, A_2,..., A_n)$ và mỗi khóa $K = \{X_{i_1}, X_{i2},..., X_{i_h}\}$ trong đó $X_{i_k} \subseteq id_{i_k}$ $(k = 1,..., h)$ của khối $r(R)$ lại trở thành khóa của quan hệ $r(A_1, A_2,..., A_n)$.*

Mệnh đề 3.2. *Cho $R = (id; A_1, A_2,..., A_n)$, $r(R)$ là một khối trên R, khi đó với $x \in id$ mà ta có $\{X_{i_1}, X_{i_2},..., X_{ik}\}$ là khóa khối $r(R)$ thì ta cũng có với mọi $y \in id$, $\{A_{i_1}, A_{i_2},..., A_{i_k}\}$ là khóa của quan hệ $r(R_y)$.*

Mệnh đề 3.3. *Cho $R = (id; A_1, A_2,..., A_n)$, $r(R)$ là một khối trên R. Khi đó nếu với $x \in id$ nào đó mà ta có $\{X_{i_1}, X_{i_2},..., X_{i_k}\}$ là khóa của lát cắt $r(R_x)$ thì $\{id_{i_1}, id_{i_2},..., id_{i_k}\}$ là khóa của khối $r(R)$.*

4. CÁC PHÉP TÍNH TRÊN CƠ SỞ DỮ LIỆU DẠNG KHỐI

Các phép tính cơ bản thường được áp dụng cho một cơ sở dữ liệu là:

- phép chèn (insert)
- phép loại bỏ (delete)
- phép cập nhật (change)

Trong mô hình cơ sở dữ liệu dạng khối thì các phép tính này cũng được áp dụng cho từng phần tử của các khối lưu trữ trong máy. Cụ thể như sau:

a) Phép chèn: Khi chèn thêm một phần tử t vào khối r ta có:

$r = r \cup t$

INSERT $(r; t^1, t^2,..., t^n)$

b) Phép loại bỏ: Phép loại bỏ là phép xóa một phần tử ra khỏi một khối cho trước. Chẳng hạn, phép loại bỏ phần tử t_0 ra khỏi khối r có dạng:

$r = r - t_0$

DEL$(r; t^1 = t_0^1, t^2 = t_0^2,..., t^n = t_0^n)$

c) Phép cập nhật: Cũng giống như phép chèn và phép loại bỏ, phép cập nhật phần tử t_0 thành phần tử t_0' có dạng:

$$r = r - t_0 \cup t_0'$$

$$\mathrm{CH}(r;\ t^1 = t_0^1, t^2 = t_0^2, ..., t^n = t_0^n;\ t^{i_1} = t^{i_1}{}', t^{i_2} = t^{i_2}{}', ..., t^{i_k} = t^{i_k}{}')$$

Đối với phép loại bỏ và phép cập nhật, để xác định một phần tử cần loại bỏ hay cập nhật thì thay vì ta phải dùng bộ ánh xạ $(t^1, t^2, ..., t^n)$ để xác định phần tử $t = (t^1, t^2, ..., t^n)$ như ở trên, ta có thể chỉ dùng các ánh xạ thu gọn mà miền xác định của chúng tạo nên khóa của khối $r(R)$.

Chẳng hạn, nếu khóa của khối $r(R)$ là bộ $K = (X_{m_1}, X_{m_2}, ..., X_{m_k})$ thì phép loại bỏ phần tử t_0 của khối r có dạng:

$$r = r - t_0,$$

$$\mathrm{DEL}(r;\ t^{m_1}(X_{m_1}) = t_0^{m_1}(X_{m_1}), t^{m_2}(X_{m_2}) = t_0^{m_2}(X_{m_2}), ..., t^{m_k}(X_{m_k}) = t_0^{m_k}(X_{m_k}))$$

và phép cập nhật phần tử t_0 thành phần tử t_0^n có dạng

$$r = r - t_0 \cup t_0',$$

$$\mathrm{CH}(r;\ t^{m_1}(X_{m_1}) = t_0^{m_1}(X_{m_1}, t^{m_2}(X_{m_2}) = t_0^{m_2}(X_{m_2}, ..., t^{m_k}(X_{m_k}) = t_0^{m_k}(X_{m_k});$$
$$t^{i_1} = t^{i_1}{}', t^{i_2} = t^{i_2}{}', ..., t^{i_k} = t^{i_k}{}').$$

Đối với các phép chèn, loại bỏ và cập nhật nêu trên thì khối r suy biến thành một quan hệ, nghĩa là khi tập chỉ số id chỉ gồm một phần tử thì chúng lại trở thành các phép chèn, loại bỏ và cập nhật trên một quan hệ trong mô hình cơ sở dữ liệu quan hệ [5].

5. ĐẠI SỐ QUAN HỆ TRÊN KHỐI

Cho r là một khối trên $R = (\mathrm{id}; A_1, A_2, ..., A_n)$, ở đây ta luôn giả thiết rằng r là một khối gồm một tập hữu hạn các phần tử. Cũng tương tự như đại số quan hệ trong mô hình cơ sở dữ liệu quan hệ, ở đây các phép toán của đại số quan hệ lại được áp dụng cho các khối; bên cạnh đó còn có thêm hai phép toán mới được xây dựng là: tích Đề các theo tập chỉ số và phép nối dài.

Đối với các phép hợp, giao và trừ thì hai khối tham gia phải là khả hợp (nghĩa là nó cùng một lược đồ khối).

a. Phép hợp:

Cho 2 khối r và s khả hợp, khi đó hợp của r và s, kí hiệu $r \cup s$ là một khối gồm các phần tử thuộc một trong trong 2 khối r và s đã cho.

Biểu diễn hình thức phép hợp có dạng:

$$r \cup s = \{t \mid t \in r \text{ hoặc } t \in s\}.$$

b. Phép giao:

Cho 2 khối r và s khả hợp, khi đó giao của 2 khối r và s là một khối, kí hiệu $r \cap s$, mà các phần tử của nó thuộc đồng thời cả 2 khối r và s đã cho.

Biểu diễn hình thức phép giao có dạng:

$$r \cap s = \{t \mid t \in r \text{ và } t \in s\}.$$

c. Phép trừ:

Cho 2 khối r và s khả hợp, khi đó hiệu của 2 khối r và s là một khối, kí hiệu $r - s$, mà các phần tử của nó thuộc r nhưng không thuộc s.

Biểu diễn hình thức phép trừ có dạng:

$$r - s = \{t \mid t \in r \text{ và } t \notin s\}.$$

(Ta có mối quan hệ giữa phép giao và phép trừ: $r \cap s = r - (r - s)$).

d. Tích Đề các:

Định nghĩa 1. Cho $R = (\text{id}; A_1, A_2, ..., A_n)$, $S = (\text{id}; B_1, B_2, ..., B_m)$, ở đây $\{A_1, A_2, ..., A_n\} \cap \{B_1, B_2, ..., B_m\} = \emptyset$. Khi đó tích Đề các của hai khối $r(R)$ và $s(S)$ là một khối, kí hiệu $r \times s$, khối này có khung $R \times S = (\text{id}; A_1, A_2, ..., A_n, B_1, B_2, ..., B_m)$, mỗi phần tử thuộc khối này là một bộ gồm $n + m$ ánh xạ, trong đó n ánh xạ đầu có dạng một phần tử thuộc r, còn m ánh xạ sau có dạng một phần tử thuộc s.

Biểu diễn hình thức của tích Đề các có dạng:

$$r \times s = \{t \mid t(R) \in r \text{ và } t(S) \in s\},$$

trong đó $t = (t^1, t^2, ..., t^n, t^{n+1}, ..., t^{n+m})$, $t(R) = (t^1, t^2, ..., t^n)$ và $t(S) = (t^{n+1}, ..., t^{n+m})$.

e. Tính đề các theo tập chỉ số:

Định nghĩa 2. Cho $R = (\text{id}; A_1, A_2, ..., A_n)$, $S = (\text{id}'; A_1, A_2, ..., A_n)$. Khi đó tích Đề các của 2 khối $r(R)$ và $s(S)$ theo tập chỉ số là một khối, kí hiệu $r \times_{\text{id}} s$, khối này có khung $R \times_{\text{id}} S = \{\text{id} \coprod \text{id}'; A_1, A_2, ..., A_n\}$, với $\text{id} \coprod \text{id}'$ là kí hiệu tích rời rạc của hai tập chỉ số id và id'. Mỗi phần tử thuộc khối này là một bộ gồm n ánh xạ $(t^1, t^2, ..., t^n)$ với $t^i : \text{id} \coprod \text{id}' \to A_i$, $i = 1, ..., n$, mỗi ánh xạ này được cảm sinh từ 2 ánh xạ thứ i tương ứng của r và s.

Cụ thể hơn, giả sử có 2 phần tử là $t_r \in r$ và $t_s \in s$: $t_r = (t_r^1, t_r^2, ..., t_r^n)$, $t_s = (t_s^1, t_s^2, ..., t_s^n)$, khi đó ta có ánh xạ cảm sinh của t_r và t_s, phần tử cảm sinh của t_r và t_s kí hiệu là t_{rs}.

Gọi $j_1 : \text{id} \to \text{id} \coprod \text{id}'$, $j_2 : \text{id}' \to \text{id} \coprod \text{id}'$ là các phép nhúng thì ta được:

$$t_{rs} j_1 \in r \text{ và } t_{rs} j_2 \in s.$$

Biểu diễn hình thức của chúng có dạng:

$$r \times_{\text{id}} s = \{t \mid t j_1 \in r \text{ và } t j_2 \in s\}.$$

g. Phép chiếu:

Cho $R = (\text{id}; A_1, A_2, ..., A_n)$, r là một khối trên R.

Khi đó ta gọi $P = (\text{id}'; A_{i_1}, A_{i_2}, ..., A_{i_h})$ là lược đồ con của lược đồ R nếu $\text{id}' \subseteq \text{id}$, $A_{i_j} \in \{A_1, A_2, ..., A_n\}$, $j = 1, ..., h$.

Một phép chiếu của khối r trên lược đồ con P, kí hiệu $\Pi_P(r)$, là một khối có lược đồ khối P và mỗi phần tử thuộc khối này có dạng:

$(t^{i_1}, t^{i_2}, ..., t^{i_h})\Big|_{\text{id}'}$, trong đó: $t^{i_j} \in \{t^1, t^2, ..., t^n\}$, $j = 1, ..., h$, và $(t^1, t^2, ..., t^n) \in r$.

Biểu diễn hình thức của phép chiếu có dạng:

$$\Pi_P(r) = \left\{ (t^{j_1}, t^{j_2}, ..., t^{j_h})\Big|_{\text{id}'} \,\Big|\, t^{i_j} \in \{t^1, t^2, ..., t^n\},\ j = 1, ..., h,\ (t^1, t^2, ..., t^n) \in r \right\}.$$

Ta dễ dàng chứng minh được một số tính chất sau đây của phép chiếu:

Mệnh đề 5.1.

1) $\Pi_P(\Pi_P(r)) = \Pi_P(r)$

2) Nếu $P \subseteq Q$ *thì* $\Pi_P(\Pi_Q(r)) = \Pi_P(r)$

3) $\Pi_P(r \cup s) = \Pi_P(r) \cup \Pi_P(s)$

4) $\Pi_P(r \cap s) = \Pi_P(r) \cap \Pi_P(s)$

5) $\Pi_P(r - s) \supseteq \Pi_P(r) - \Pi_P(s)$

ở đây r và s là các khối khả hợp trên lược đồ R, còn P và Q là các lược đồ con của lược đồ R.

h. Phép chọn:

Cho $R = (\text{id}; A_1, A_2, ..., A_n)$ và khối $r(R)$.

Cho một phép chọn nghĩa là ta xây dựng một tập con các phần tử của khối đã cho thỏa mãn biểu thức F xác định. Biểu thức F được diễn tả bằng một tổ hợp Boolean của các toán hạng, mỗi toán hạng là một phép so sánh đơn giản giữa hai biến là hai giá trị điểm của hai ánh xạ thành phần nào đó, hoặc giữa một biến là giá trị điểm của một ánh xạ thành phần và một hằng.

Các phép so sánh trong F là $<, =, >, \geq, \leq, \neq$, còn các phép toán logic trong F là: $\vee, \wedge, \neg$.

Biểu diễn hình thức của phép chọn có dạng:

$$\sigma_F(r) = \{t \in r \mid F(t)\}$$

trong đó $F(t)$ là giá trị đúng của biểu thức Boolean F tại phần tử $t \in r$.

Mệnh đề 5.2. *Với r và s là các khối trên lược đồ R ta có:*

1) $\sigma_F(r \cup s) = \sigma_F(r) \cup \sigma_F(s)$

2) $\sigma_F(r \cap s) = \sigma_F(r) \cap \sigma_F(s)$

3) $\sigma_F(r - s) = \sigma_F(r) - \sigma_F(s)$

i. Phép kết nối:

Cho $R = (\text{id}; A_1, A_2, ..., A_n)$ và $S = (\text{id}; B_1, B_2, ..., B_m)$, cùng với 2 khối $r(R)$ và $s(S)$ tương ứng.

Gọi $T = (\text{id}; C_1, C_2, ..., C_p)$, trong đó $\{C_1, C_2, ..., C_p\} = \{A_1, A_2, ..., A_n\} \cup \{B_1, B_2, ..., B_m\}$.

Phép kết nối của 2 khối r và s, kí hiệu $r \bowtie s$ là khối $t(T)$ định nghĩa như sau:

$$t(T) = \{t \mid \exists t_r \in r \text{ và } t_s \in s \text{ sao cho } t(R) = t_r,\ t(S) = t_s\}.$$

Phép kết nối này cũng gọi là phép kết nối tự nhiên của hai khối $r(R)$ và $s(S)$, đôi khi sử dụng kí hiệu $r * s$.

Đặc biệt khi các khối $r(R)$ và $s(S)$ có tập chỉ số id trong lược đồ khối của chúng chỉ gồm 1 phần tử thì các khối này trở thành các quan hệ và phép kết nối tự nhiên của hai khối lại trở thành phép kết nối tự nhiên của hai quan hệ trong mô hình cơ sở dữ liệu quan hệ [5].

Nếu $\{A_1, A_2, ..., A_n\} \cap \{B_1, B_2, ..., B_m\} = \emptyset$ thì $r * s$ trở thành tích Đề các của hai khối đã cho.

Ta có thể mở rộng khái niệm kết nối như sau:

Giả sử $A_{i_k} \in \{A_1, A_2, ..., A_n\}$, $B_{i_k}\{B_1, B_2, ..., B_m\}$, và $\text{dom}(A_{i_k} = \text{dom}B_{i_k}$, $1 \leq k \leq h$, (ở đây A_{i_k} và B_{i_k} không nhất thiết phân biệt). Khi đó kết nối của r và s theo $A_{i_1}, A_{i_2}, ..., A_{i_h}$, và $B_{i_1}, B_{i_2}, ..., B_{i_h}$ là khối $t(T)$, khối này được định nghĩa như sau:

$$t(T) = \{t \mid \exists t_r \in r \text{ và } t_s \in s \text{ sao cho } t(R) = t_r,\ t(S) = t_s,\ t_r^{i_k} = t_s^{i_k},\ 1 \leq k \leq h\},$$

ở đây $t_r = (t_r^1, t_r^2, ..., t_r^n)$, $t_s = (t_s^1, t_s^2, ..., t_s^m)$.

Thay cho kí hiệu $r \bowtie s$ ở đây ta kí hiệu rõ hơn:

$$t(T) = r\,[t_r^{i_k} = t_s^{i_k},\ 1 \leq k \leq h]\,s.$$

k. Phép nối dài:

Cho hai khối $r(\text{id}; A_1, A_2, ..., A_n)$ và $s(\text{id}'; A_1, A_2, ..., A_n)$, ở đó nếu $\text{id}_1 \cap \text{id}_2 \neq \emptyset$ mà ta có với $t \in r$ và $k \in s$:

$$t^1\Big|_{\text{id}_1\cap\text{id}_2} = k^1\Big|_{\text{id}_1\cap\text{id}_2}$$

$$t^2\Big|_{\text{id}_1\cap\text{id}_2} = k^2\Big|_{\text{id}_1\cap\text{id}_2}$$

$$.........$$

$$t^n\Big|_{\text{id}_1\cap\text{id}_2} = k^n\Big|_{\text{id}_1\cap\text{id}_2}$$

thì khi đó ta xây dựng được một phần tử mới có dạng:

$$u = (u^1, u^2, ..., u^n) \text{ với } u^h : \text{id} \cup \text{id}' \to A^h$$

sao cho: $u^h\Big|_{\text{id}} = t^h$, $u^h\Big|_{\text{id}'} = k^h$, $\forall h = 1, ..., n$, và kí hiệu: $u^h = t^h *_{\text{id}} k^h$, $\forall h = 1, ..., n$.

Những phần tử $u = (u^1, u^2, ..., u^n)$ này tạo ra một khối mới, được kí hiệu: $r *_{\text{id}} s$, gọi là khối nối dài của 2 khối r và s.

Phép toán được xây dựng ở trên gọi là phép nối dài của 2 khối r và s đã cho.

Biểu diễn hình thức của phép nối dài có dạng:

$$r *_{\text{id}} s = \left\{ u = (u^j)_{j=\overline{1,n}} \mid \left(u^j\Big|_{\text{id}}\right)_{j=\overline{1,n}} \in r \text{ và } \left(u^j\Big|_{\text{id}'}\right)_{j=\overline{1,n}} \in s \right\}.$$

Mệnh đề 5.3. *Giả sử $r(R)$, $r'(R)$, $q(Q)$, $s(S)$ là các khối đã cho, khi đó ta có:*

1) $(q \bowtie r) \bowtie s = q \bowtie (r \bowtie s)$

2) $(r \cup r') \bowtie s = (r \bowtie s) \cup (r' \bowtie s)$

3. $(r \cap r') \bowtie s = (r \bowtie s) \cap (r' \bowtie s)$

4. $(r - r') \bowtie s = (r \bowtie s) - (r' \bowtie s)$

l. Phép chia:

Cho khối $r(\text{id}; A_1, A_2, ..., A_n)$ và khối $s(\text{id}; A_{i_1}, A_{i_2}, ..., A_{i_h})$, trong đó $A_{i_k} \in \{A_1, A_2, ..., A_n)\}$, $\forall k = \overline{1, h}$, khi đó phép chia của khối r cho khối s, kí hiệu $r \div s$, là một khối gồm các phần tử $t = (t^1, t^2, ..., t^{n-h})$ sao cho $\forall u = (u^1, u^2, ..., u^h)$, $u \in s$ thì phần tử $tu \in r$, ở đây phần tử tu có dạng: $tu := (t^1, t^2, ..., t^{n-h}, u^1, u^2, ..., u^h)$.

Biểu diễn hình thức của phép chia có dạng:

$$r \div s = \{t \mid \forall u \in s,\ tu \in r\}.$$

Trong trường hợp các khối r và s có tập chỉ số id chỉ gồm 1 phần tử thì khi đó các khối này trở thành các quan hệ và phép toán chia của hai khối trở thành phép toán chia của hai quan hệ trong mô hình cơ sở dữ liệu quan hệ [5].

Mệnh đề 5.4. *Giả sử* $R = (id;\ A_1, A_2, ..., A_n)$, $S = (id;\ A_{i_1}, A_{i_2}, ..., A_{i_h})$, *trong đó* $A_{i_k} \in \{A_1, A_2, ..., A_n\}$, $\forall k = \overline{1,h}$ *và* $r(R)$, $r'(R)$, $s(S)$, $s'(S)$ *là các khối đã cho, khi đó ta có:*

1) $(r \cup r') \div s \supseteq (r \div s) \cup (r' \div s)$.

2) $(r \cap r') \div s = (r \div s) \cap (r' \div s)$.

3) $(r - r') \div s \subseteq (r \div s) - (r' \div s)$.

4) Nếu $s \subseteq s'$ *thì* $(r \div s) \supseteq (r \div s')$.

TÀI LIỆU THAM KHẢO

[1] Armstrong W. W., Dependency structures of database relationships, Proc. 1974 EFIP Congress, pp. 580-583, North-Holland, Amsterdam.

[2] Beeri C., On the membership problem for functional and multivalued dependencies, *ACM Trans. on Database System* **5** (3) (1980) 241-259.

[3] Demetrovics J., Nguyen Xuan Huy, Closed sets and translation of relation schemes, *Computers Math. Applic.* **21** (1) (1991) 13-23.

[4] Demetrovics J., Nguyen Xuan Huy, Representation of closure for functional, multivalued and join dependencies, *Computers and Artificial Intelligence* **11** (2) (1992) 143-154.

[5] Maier D., *The Theory of Relational Databases*, Computer Science Press, Rockville, Md., 1983.

[6] Ullman J. D., *Principles of Database Systems*, Computer Science Press, Rockville, Md., 1982.

[7] Ullman J. D., *Principles of Database and Knowledge - Base systems*, Vol. 1, 2, Computer Science Press, 1989.

Received: March 10, 1998

(1) Viện Công nghệ thông tin, Trung tâm KHTN và CNQG.

(2) Trường Đại học Sư phạm - Đại học Quốc gia Hà Nội.