Tạp chí Tin học và Điều khiển học, T.26, S.4 (2010), 312–320

PHỦ CỦA TẬP PHỤ THUỘC HÀM VÀ VẤN ĐỀ TỰA CHUẨN HÓA TRONG MÔ HÌNH DỮ LIỆU DẠNG KHỐI

TRỊNH ĐÌNH VINH[1] , VŨ ĐỨC THI[2]

[1]*Khoa Công nghệ thông tin, Trường Đại học Sư phạm Hà Nội 2*

[2]*Viện Công nghệ thông tin, Viện Khoa học và Công nghệ Việt Nam*

Abstract. In recent years, database model of block form is introduced and studied. Some results for block, block scheme, relational algebra over block, keys, functional dependencies, normal forms ... presented in $[1, 2, 3, 5]$...

This paper gives some concepts of the cover of the set of functional dependencies, nonredundant cover, minimalcover, and quatinormization block scheme in the database model of block forms. Based on the concepts, some properties of the cover of the set of functional dependencies and quatinormization algorithms are established.

Tóm tắt. Mô hình dữ liệu dạng khối là một mở rộng của mô hình dữ liệu quan hệ. Trong một vài năm gần đây, mô hình này đã được đề xuất và nghiên cứu. Một số kết quả về khối, lược đồ khối, đại số quan hệ trên khối, khoá, phụ thuộc hàm, các dạng chuẩn ... đã được trình bày trong $[1, 2, 3, 5]$...

Bài báo đưa ra các khái niệm mới về phủ của tập các phụ thuộc hàm, phủ không dư thừa, phủ tối thiểu ... và vấn đề tựa chuẩn hoá một lược đồ khối trong mô hình dữ liệu dạng khối. Từ các khái niệm đưa ra, một số tính chất về phủ của tập phụ thuộc hàm và các thuật toán tựa chuẩn hoá được phát biểu và chứng minh.

1. MÔ HÌNH DỮ LIỆU DẠNG KHỐI

1.1. Khối, lược đồ khối

Định nghĩa 1.1. Gọi $R = (id; A_1, A_2, ..., A_n)$ là một bộ hữu hạn các phần tử, trong đó id là tập chỉ số hữu hạn khác rỗng, A_i $(i = 1, ..., n)$ là các thuộc tính. Mỗi thuộc tính A_i $(i = 1, ..., n)$ có miền giá trị tương ứng là $\mathrm{dom}(A_i)$. 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, $(i = 1, ..., n)$.

Nói một cách khác,

$$t \in r(R) \Longleftrightarrow t = \{t^i : id \to dom(A_i)\}_{i=1,\dots,n}.$$

Ta kí hiệu khối đó là $r(R)$ hoặc $r(id; A_1, A_2, ..., A_n)$, 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.

Định nghĩa 1.2. Cho $R = (id; A_1, A_2, ..., A_n)$, $r(R)$ là một khối trên R. Với mỗi $x \in id$ ta kí hiệu $r(R_x)$ là một khối với $R_x = (\{x\}; A_1, A_2, ..., A_n)$ sao cho

$$t_x \in r(R_x) \Longleftrightarrow t_x = \{t_x^i = t_{|x}^i\}_{i=1,...,n}, t \in r(R)$$

và

$$t = \{t^i : id \to dom(A_i)\}_{i=1,...,n}.$$

ở đây, $t_x^i(x) = t^i(x), i = 1, ..., n$.

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

1.2. Đại số quan hệ trên khối

Cho r là một khối trên $R = (id; A_1, A_2, ..., A_n)$. Tương tự như trong mô hình dữ liệu quan hệ, ở đây các phép toán của đại số quan hệ lại được áp dụng cho các khối như: phép hợp, phép giao, phép trừ, phép chiếu, phép chọn, phép kết nối, phép chia. Ngoài ra còn có một phép toán mới được xây dựng: phép nối dài. Các định nghĩa, các tính chất của các phép toán trên được trình bày trong [2,4].

1.3. Phụ thuộc hàm

Để đơn giản, ta sử dụng các kí hiệu

$$x^{(i)} = (x; A_i); \quad id^{(i)} = \{x^{(i)} | x \in id\}.$$

Gọi $x^{(i)}$, $(x \in id, i = 1, ..., n)$ là các thuộc tính chỉ số của lược đồ khối $R = (id; A_1, A_2, ..., A_n)$.

Định nghĩa 1.3. Cho $R = (id; A_1, A_2, .., A_n)$, $r(R)$ là một khối trên R, $X, Y \subset \cup_{i=1}^n id^{(i)}$, $X \to Y$ là kí hiệu một phụ thuộc hàm. Một khối r thoả $X \to Y$ nếu với mọi $t_1, t_2 \in R$ sao cho $t_1(X) = t_2(X)$ thì $t_1(Y) = t_2(Y)$.

Định nghĩa 1.4. Cho lược đồ khối $R = (id; A_1, A_2, .., A_n)$, F là tập các phụ thuộc hàm trên R. Khi đó bao đóng của F, kí hiệu F^+, được xác định như sau

$$F^+ = \{X \to Y | \ F \Longrightarrow X \to Y\}.$$

Nếu $X = \{x^{(m)}\} \subset id^{(m)}$, $Y = \{y^{(k)}\} \subset id^{(k)}$ thì ta kí hiệu phụ thuộc hàm $X \to Y$ đơn giản là $x^{(m)} \to y^{(k)}$.

Khối $r(R)$ thoả $x^{(m)} \to y^{(k)}$ nếu với mọi $t_1, t_2 \in r$ sao cho $t_1(x^{(m)}) = t_2(x^{(m)})$ thì $t_1(y^{(k)}) = t_2(y^{(k)})$, trong đó

$$t_1(x^{(m)}) = t_1(x; A_m), \quad t_2(x^{(m)}) = t_2(x; A_m),$$

$$t_1(y^{(k)}) = t_1(y; A_k), \quad t_2(y^{(k)}) = t_2(y; A_k).$$

Mệnh đề 1.1. *Cho $R = (id; A_1, ..., A_n)$, $r(R)$ là một khối một khối trên R, $X, Y \subset \cup_{i=1}^n id^{(i)}$, $X \to Y$ là kí hiệu một phụ thuộc hàm. Giả sử $r(R)$ thoả phụ thuộc hàm $X \to Y$. Khi đó nếu $id = \{x\}$ thì*

- $r(R)$ trở thành quan hệ $r(A_1, A_2, ..., A_n)$.

- Phụ thuộc hàm $X \to Y$, $(X, Y \subset \cup_{i=1}^n A_i)$ trở thành phụ thuộc hàm trong mô hình dữ liệu quan hệ.

1.4. Bao đóng của tập thuộc tính chỉ số

Định nghĩa 1.5

Cho lược đồ khối $R = (id; A_1, A_2, .., A_n)$, F là tập các phụ thuộc hàm trên R. Với mỗi , ta định nghĩa bao đóng của X đối với F , kí hiệu X^+, như sau:

$$X^+ = \{x^{(i)}, x \in id, i = 1, ..., n |\ X \to x^{(i)} \in F^+\}.$$

Cho lược đồ khối $R = (id; A_1, A_2, ..., A_n)$, ta kí hiệu tập các phụ thuộc hàm trên R:

$$F_h = \{X \to Y |\ X = \cup_{i \in A} x^{(i)},\ Y = \cup_{j \in B} x^{(j)},\ A, B \subset \{1, 2, ..., n\} \text{ và } x \in id\},$$

$$F_{hx} = F_h \cap \cup_{i=1}^{n} x^{(i)} = \{X \to Y \in F_h |\ X, Y \subset \cup_{i=1}^{n} x^{(i)}\}.$$

1.5. Khoá của lược đồ khối R đối với tập phụ thuộc hàm F trên R

Định nghĩa 1.6. Cho lược đồ khối $R = (id; A_1, A_2, ..., A_n)$, F là tập các phụ thuộc hàm trên R, $K \subset \cup_{i=1}^{n} id^{(i)}$. K gọi là khoá của lược đồ R đối với F nếu thoả 2 điều kiện:

a) $K \to x^{(i)} \in F^+, \forall x \in id, i = 1, ..., n$

b) $\forall\ K' \subset K$ thì K' không có tính chất a)

Nếu K là khoá và $K \subset K''$ thì K'' gọi là siêu khoá của lược đồ khối R đối với F.

1.6. Các dạng chuẩn

Định nghĩa 1.7. Cho lược đồ khối $R = (id; A_1, A_2, ..., A_n)$, F là tập các phụ thuộc hàm trên R. Ta gọi lược đồ khối R thuộc dạng chuẩn 1 nếu và chỉ nếu toàn bộ các miền trị của các thuộc tính $x^{(i)}$, $x \in id$, $i \in \{1, 2, ..., n\}$ đều chỉ chứa các giá trị nguyên tố.

Định nghĩa 1.8. Cho lược đồ khối $R = (id; A_1, A_2, ..., A_n)$, F là tập các phụ thuộc hàm trên R, $X, Y \subset \cup_{i=1}^{n} id^{(i)}$. Ta nói Y là phụ thuộc hàm đầy đủ vào X nếu Y là phụ thuộc hàm vào X nhưng không phụ thuộc hàm vào bất kì một tập hợp con thực sự nào của X.

Định nghĩa 1.9. Cho lược đồ khối $R = (id; A_1, A_2, ..., A_n)$, F là tập các phụ thuộc hàm trên R. Ta gọi lược đồ khối R thuộc dạng chuẩn 2 nếu nó ở dạng chuẩn 1 và mọi thuộc tính không khoá của R là phụ thuộc hàm đầy đủ vào khoá.

Định nghĩa 1.10. Cho lược đồ khối $R = (id; A_1, A_2, ..., A_n)$, F là tập các phụ thuộc hàm trên R. Ta gọi lược đồ khối R thuộc dạng chuẩn 3 nếu nó ở dạng chuẩn 2 và mọi thuộc tính không khoá của R là không phụ thuộc hàm bắc cầu vào khoá.

Định nghĩa 1.11. Cho lược đồ khối $R = (id; A_1, A_2, ..., A_n)$, F là tập các phụ thuộc hàm trên R, $X \subset \cup_{i=1}^{n} id^{(i)}$. Ta gọi lược đồ khối R thuộc dạng chuẩn Boye - Codd nếu $X \to x^{(i)}$ thoả trên R, $x^{(i)} \notin X$, $x \in id$, $i \in \{1, 2, ..., n\}$ thì X là một khoá của R.

Định nghĩa 1.12. Cho lược đồ khối $R = (id; A_1, A_2, ..., A_n)$, F là tập các phụ thuộc hàm trên R. Ta gọi lược đồ khối R thuộc dạng tựa chuẩn 2 (tựa chuẩn 3, tựa chuẩn Boye - Codd) nếu và chỉ nếu $\forall x \in id$ lát cắt R_x thuộc dạng chuẩn 2 (dạng chuẩn 3, dạng chuẩn Boye - Codd).

2. KẾT QUẢ NGHIÊN CỨU

Định nghĩa 2.1. Cho lược đồ khối $R = (id; A_1, A_2, ..., A_n)$, F và G là các tập phụ thuộc hàm trên R. Ta nói F tương đương với G và kí hiệu $F \approx G$ nếu $F^+ = G^+$.

Nếu F và G là tương đương thì ta nói F phủ G hay G phủ F.

Thuật toán kiểm tra sự tương đương của 2 tập phụ thuộc hàm

Thuật toán 2.1 (Derives)

Vào: Hai tập phụ thuộc hàm F và G

Ra: Đúng nếu $F \models G$ và sai nếu ngược lại.

DERIVES (F, G)

Begin

V :=true;

For each $X \to Y \in G$ do $V := V$ and member $(F, X \to Y)$

Return (V);

End.

Thuật toán 2.2 (EQUI)

Vào: Hai tập phụ thuộc hàm F và G.

Ra: Đúng nếu F tương đương với G.

E$QUI(F, G)$

Begin $V :=$ Derives(F, G) and Derives (G, F);

Return (V);

End.

Mệnh đề 2.1. *Các Thuật toán 2.1 và 2.2 sử dụng để xét sự tương đương của hai tập phụ thuộc hàm là đúng.*

Định nghĩa 2.2. Cho lược đồ khối $R = (id; A_1, A_2, ..., A_n)$, tập Fcác phụ thuộc hàm được gọi là không dư thừa nếu không có tập con thực sự F' của F mà $F' \approx F$. Ngược lại, F là dư thừa.

F là phủ không dư thừa cho G nếu F là phủ của G và F không dư thừa.

Thuật toán 2.3 (kiểm tra sự dư thừa)

Vào: Tập phụ thuộc hàm F.

Ra: Đúng nếu F dư thừa, sai nếu ngược lại.

REDUNDANT (F)

Begin

$V :=$ False ; For each $X \to Y \in F$ do

If Member $(F\ \{X \to Y\}, X \to Y)$ then $V :=$ True;

Return (V);

End.

Thuật toán 2.4 (tính phủ không dư thừa)

Vào: Tập phụ thuộc hàm G.

Ra: Phủ không dư thừa F của G.

$NONREDUNDANT(G)$
Begin
$F := G;$
For each $X \to Y \in G$ do
If Member $(F\ \{X \to Y\}, X \to Y)$ then $F := F\ \{X \to Y\};$
Return (F);
End.

Mệnh đề 2.2. *Các Thuật toán 2.3, 2.4 kiểm tra sự dư thừa và tính phủ không dư thừa của tập phụ thuộc hàm là đúng.*

Định nghĩa 2.3. Cho lược đồ khối $R = (id; A_1, A_2, ..., A_n)$, tập F các phụ thuộc hàm, $X \to Y \in F$, $X \to Y$ là một phụ thuộc hàm đầy đủ trong F nếu không tồn tại $X' \subset X$ sao cho $F \approx (F \setminus \{X \to Y\} \cup \{X' \to Y\})$. Khi đó Y gọi là phụ thuộc đầy đủ vào X.

Định nghĩa 2.4. Cho lược đồ khối $R = (id; A_1, A_2, ..., A_n)$. Tập F các phụ thuộc hàm là tối thiểu nếu:

• Vế phải của các phụ thuộc hàm trong F chỉ có một thuộc tính (không có thuộc tính nào ở vế phải là dư thừa).

• Không tồn tại $X \to Y \in F$ và $X' \subset X$ sao cho

$$F \approx (F \setminus \{X \to Y\} \cup \{X' \to Y\})$$

(không có thuộc tính nào ở vế trái là dư thừa).

• Không tồn tại $X \to Y \in F$ sao cho $F \approx (F \setminus \{X \to Y\})$ (không có phụ thuộc hàm nào là dư thừa).

Mệnh đề 2.3. *Cho lược đồ khối $R = (id; A_1, A_2, ..., A_n)$, mọi tập F các phụ thuộc hàm đều có một phủ tối thiểu F' tương đương.*

Thuật toán 2.5 (tìm phủ tối thiểu)
Vào: Tập phụ thuộc hàm F.
Ra: G là phủ tối thiểu của F.
MINIMALCOVER (F,G)
Begin $G := F;$
Thay thế từng phụ thuộc hàm $X \to \{A_1, A_2, ..., A_n\}$ trong G bằng các phụ thuộc hàm $X \to A_1$, $X \to A_2$, ... , $X \to A_n$.
For each $X \to A$ trong G do
For each $B \in X$ do
If $(G \setminus \{X \to A\}) \cup ((X \setminus \{B\}) \to A)$ tương đương với G then thay $X \to A$ bằng $(X \setminus \{B\}) \to A$ trong G;
For each $X \to A$ trong G do
If $(G \setminus \{X \to A\})$ tương đương với G then loại $X \to A$ ra khỏi G;
Return(G);
End.

Mệnh đề 2.4. *Thuật toán 2.5 (Minimalcover) tính phủ tối thiểu là đúng.*

Mệnh đề 2.5. *Cho lược đồ khối $R = (id; A_1, A_2, .., A_n)$, F_h và $F_{hx} = F_h \cap \cup_{i=1}^{n} x^{(i)}$ là các tập phụ thuộc hàm trên R, R_x tương ứng, F'_h là phủ tối thiểu của F_h. Khi đó $F'_{hx} = F'_h \cap \cup_{i=1}^{n} x^{(i)}$ là phủ tối thiểu của tập F_{hx} .*

Chứng minh. Giả sử $F'_{hx} = F'_h \cap \cup_{i=1}^{n} x^{(i)}$ không là phủ tối thiểu của tập F_{hx} , khi đó tồn tại một phụ thuộc hàm thuộc $X \to Y \in F'_{hx}$ sao cho $F''_{hx} = F'_{hx}\ (X \to Y) \approx F'_{hx}$. Mà ta có $F'_h = \cup_{x \in A} F'_{hx}$ nên F'_h không tối thiểu, điều này trái với giả thiết ban đầu là F'_h tối thiểu. Do đó F'_{hx} là phủ tối thiểu của tập F_{hx}.

Các điều kiện phủ tối thiểu còn lại chứng minh tương tự. ■

Mệnh đề 2.6. *Cho lược đồ khối $R = (id; A_1, A_2, .., A_n)$, F_h và F_{hx} là các tập phụ thuộc hàm trên R, R_x tương ứng, $F_h = \cup_{x \in A} F_{hx}$, $A \subset id$, F'_{hx} là phủ tối thiểu của F_{hx}. Khi đó $F'_h = \cup_{x \in A} F'_{hx}$ là phủ tối thiểu của tập F_h .*

Chứng minh. Giả sử $F'_h = \cup_{x \in A} F'_{hx}$ không là phủ tối thiểu của F_h, khi đó tồn tại một phụ thuộc hàm thuộc $X \to Y \in F'_h$ sao cho $F"_h = F'_h \setminus (X \to Y) \approx F'_h$. Vì $F'_h = \cup_{x \in A} F'_{hx}$ nên $X \to Y \in F'_{hx}$ nào đó mà F_{hx} là tập phụ thuộc hàm dư thừa. Điều này trái với giả thiết F'_{hx} là tối thiểu. Vậy F'_h là tối thiểu.

Các điều kiện phủ tối thiểu còn lại chứng minh tương tự. ■

Từ các Mệnh đề 2.5 và 2.6 ta rút ra điều kiện cần và đủ của một phủ tối thiểu trong mô hình dữ liệu dạng khối như sau.

Mệnh đề 2.7. *Cho lược đồ khối $R = (id; A_1, A_2, .., A_n)$, F_h và $F_{hx} = F_h \cap \cup_{i=1}^{n} x^{(i)}$ là các tập phụ thuộc hàm trên R, R_x tương ứng, F'_h là phủ tối thiểu của F_h khi và chỉ khi $F'_{hx} = F'_h \cap \cup_{i=1}^{n} x^{(i)}$ là phủ tối thiểu của tập F_{hx} .*

Định nghĩa 2.5. Cho lược đồ khối $R = (id; A_1, A_2, .., A_n)$, F là tập các phụ thuộc hàm trên R. Phép tách lược đồ khối R là thay thế nó bởi một tập các lược đồ khối $\rho = \{R_1, R_2, ..., R_k\}$, ở đây R_i, $i = 1, 2, ..., k$ là các lược đồ khối con của R có dạng

$$R_i = (id, A_{i1}, A_{i2}, ..., A_{ij})$$

với $\{A_{i1}, A_{i2}, ..., A_{ij}\} \subset \{A_1, A_2, ..., A_n\}$ và $\cup_{im}\{A_{im}\} = \{A_1, A_2, ..., A_n\}$.

Phép tách không mất mát thông tin (Lossless join)

Cho lược đồ khối $R = (id; A_1, A_2, .., A_n)$, F là tập các phụ thuộc hàm trên R, $\rho = \{R_1, R_2, ..., R_k\}$ là một phép tách của R. Khi đó ta nói phép tách này là không mất mát thông tin đối với F nếu mọi khối r trên R thoả F thì

$$r = \prod_{R_1}(r) * \prod_{R_2}(r) * ... * \prod_{R_k}(r) \tag{2.1}$$

Nói cách khác, khối r là kết nối tự nhiên của các khối là phép chiếu của nó trên mỗi R_i .

Ta kí hiệu $m_\rho(r)$ là vế phải của (1), khi đó điều kiện để một phép tách không mất mát thông tin là $r = m_\rho(r)$.

Mệnh đề 2.8. *Cho lược đồ khối $R = (id; A_1, A_2, .., A_n)$, r là một khối trên R, $r_i = \prod_{Ri}(r)$. Khi đó ta có:*

a) $r \subset m_\rho(r)$

b) Nếu $s = m_\rho(r)$ *thì* $\prod_{Ri}(s) = r_i$

c) $m_\rho(m_\rho(r)) = m_\rho(r)$.

Kiểm tra phép tách không mất mát thông tin

Định nghĩa 2.6. Cho lược đồ khối $R = (id; A_1, A_2, .., A_n)$, F_h là tập các phụ thuộc hàm trên R. F_h được gọi là tập đầy đủ các phụ thuộc hàm nếu $F_{hx} = F_h \cap \cup_{i=1}^{n} x^{(i)}$ là như nhau $\forall x \in id$.

Mệnh đề 2.9. *Cho lược đồ khối $R = (id; A_1, A_2, .., A_n)$, F_h là tập các phụ thuộc hàm trên R. Khi đó luôn tồn tại một tập đầy đủ các phụ thuộc hàm F_h^* thoả mãn*

$$F_h \subset F_h^* \subset F_h^+.$$

Thuật toán 2.6 (khẳng định phép tách có mất thông tin hay không)

Vào: Lược đồ khối $R = (id; A_1, A_2, .., A_n)$, tập đầy đủ các phụ thuộc hàm F_h trên R, phép tách $\rho = \{R_1, R_2, ..., R_k\}$.

Ra: Khẳng định phép tách có mất mát thông tin hay không?

Phương pháp

- Xây dựng bảng gồm n cột, k hàng: hàng i tương ứng với lược đồ R_i , cột j tương ứng với thuộc tính A_j . Tại vị trí có toạ độ (i, j) nếu A_j thuộc R_j thì ta ghi a_j , ngược lại ta ghi b_{ij} .
- Ta xét F_{hx} với một x nào đó, với mỗi phụ thuộc hàm $X \to Y \in F_{hx}$ nếu tồn tại hai hàng mà tất cả các cột tương ứng với các thuộc tính của X có giá trị như nhau thì ta làm cho các cột ứng với các thuộc tính của Y cũng có giá trị như nhau trong hai hàng này theo nguyên tắc: nếu có một kí hiệu a_j trong các cột ứng với các thuộc tính của Y thì đồng nhất các kí hiệu là a_j , ngược lại đồng nhất bằng một trong các kí hiệu b_{ij} . Tiếp tục áp dụng các phụ thuộc hàm cho bảng (kể cả việc lặp lại các phụ thuộc hàm đã áp dụng) cho tới khi không thể thay đổi được giá trị nào trong bảng nữa.
- Nếu trong bảng có một hàng gồm các kí hiệu $a_1, ..., a_n$ thì phép tách là không mất mát thông tin, ngược lại phép tách không bảo toàn thông tin.

Mệnh đề 2.10. *Thuật toán 2.6 là đúng.*

Mệnh đề 2.11. *Cho lược đồ khối $R = (id; A_1, A_2, .., A_n)$, F_h là tập đầy đủ các phụ thuộc hàm trên R, $\rho = \{R_1, R_2\}$ là một phép tách trên R, $R_1 = (id, A_{11}, A_{12}, ..., A_{1i})$, $R_2 = (id, A_{21}, A_{22}, ..., A_{2j})$. Phép tách ρ là không mất mát thông tin nếu:*

$$\{x^{(11)}, x^{(12)}, ..., x^{(1i)}\} \cap \{x^{(21)}, x^{(22)}, ..., x^{(2j)}\} \to \{x^{(11)}, x^{(12)}, ..., x^{(1i)}\} \setminus \{x^{(21)}, x^{(22)}, ..., x^{(2j)}\}$$

hoặc

$$\{x^{(11)}, x^{(12)}, ..., x^{(1i)}\} \cap \{x^{(21)}, x^{(22)}, ..., x^{(2j)}\} \to \{x^{(21)}, x^{(22)}, ..., x^{(2j)}\} \setminus \{x^{(11)}, x^{(12)}, ..., x^{(1i)}\}$$

với $x \in id$.

Nhận xét

Cho lược đồ khối $R = (id; A_1, A_2, .., A_n)$, F_h là tập đầy đủ các phụ thuộc hàm trên R và $X, Y \subset \{A_1, A_2, ..., A_n\}$. Nếu $X \to Y$ thuộc F_h^+ thì phép tách $\rho = \{R_1, R_2\}$ với $R_1 = (id, XY)$, $R_2 = (id, XZ)$, trong đó $Z = \{A_1, A_2, ..., A_n\} \setminus XY$ là phép tách không mất mát thông tin.

Phép tách bảo toàn tập phụ thuộc hàm

Cho lược đồ khối $R = (id; A_1, A_2, .., A_n)$, F_h là tập đầy đủ các phụ thuộc hàm trên R, $\rho = \{R_1, R_2, ..., R_k\}$ là một phép tách của R, F_{hi} là hình chiếu của F_h lên R_i và kí hiệu là $\prod_{R_i}(F_h)$ là tập tất cả các phụ thuộc hàm $X \to Y \in F_h^+$ sao cho $XY \subset R_i$. Khi đó ta nói phép tách này bảo toàn tập phụ thuộc hàm F_h nếu hợp của tất cả các phụ thuộc hàm trong $\prod_{R_i}(F)$ với $i = 1, 2, ..., k$ suy diễn logic ra tất cả các phụ thuộc hàm trong F_h.

Mệnh đề 2.12. *Cho lược đồ khối $R = (id; A_1, A_2, .., A_n)$, F_h là tập đầy đủ các phụ thuộc hàm trên R, $\rho = \{R_1, R_2, ..., R_k\}$ là một phép tách không mất thông tin của R đối với F_h . Với mỗi $i = 1, ...k$, ta gọi F_{hi} là hình chiếu của F_h lên R_i và đặt $\delta = (S_1, S_2, ..., S_m)$ là một phép tách không mất mát thông tin của R_i đối với F_{hi} . Khi đó phép tách R thành $(R_1, ..., R_{i-1}, S_1, ..., S_m, R_{i+1}, ..., R_n)$ là không mất thông tin đối với F_h.*

Giả sử có R, F_h , ρ như đã nói ở trên, $\tau = (R_1, R_2, ..., R_k, R_{k+1}, ..., R_n)$ là một phép tách của R chứa các lược đồ của ρ thì τ là một phép tách không mất thông tin.

Thuật toán 2.7 (tách không mất thông tin về dạng tựa chuẩn Boye-Codd)

Vào: Lược đồ khối $R = (id; A_1, A_2, .., A_n)$, tập các phụ thuộc hàm F_h trên R.

Ra: Phép tách $\rho = \{R_1, R_2, ..., R_k\}$ không mất mát thông tin, R_i với mọi $i = 1, 2, ..., k$ đều ở dạng tựa chuẩn Boye-Codd với các phụ thuộc hàm là hình chiếu của F_h lên các R_i.

Phương pháp

- Xây dựng phép tách ρ đối với R theo phương pháp lặp, sau mỗi lần lặp ρ sẽ được tách tiếp với một phép tách không mất mát thông tin đối với F_h.
- Ban đầu phép tách ρ chỉ gồm R. Tiếp theo, nếu S là một lược đồ khối trong ρ không ở dạng tựa chuẩn Boye-Codd, ta xét một phụ thuộc hàm $X \to A$ của S, với điều kiện X không chứa khoá của một lát cắt nào của S và $A \notin X$. Khi đó ta thay thế S bởi S_1, S_2 với $S_1 = (id, A, X)$, $S_2 = (id, Y)$ trong đó Y là các thuộc tính còn lại của S sau khi loại bỏ A.
- Tiếp tục quá trình cho đến khi mọi lược đồ khối con đều ở dạng tựa chuẩn Boye-Codd. Như vậy ta xây dựng được phép tách không mất mát thông tin tựa chuẩn hoá R về dạng tựa chuẩn Boye-Codd.

3. KẾT LUẬN

Những kết quả về phủ của tập các phụ thuộc hàm và vấn đề về tựa chuẩn hoá một lược đồ khối trong mô hình dữ liệu dạng khối được nghiên cứu ở trên chỉ là những kết quả bước đầu. Đó cũng chỉ là các kết quả đối với trường hợp riêng của tập các phụ thuộc hàm F trong lược đồ khối $R = (id; A_1, A_2, .., A_n)$. Những kết quả này cho ta thấy việc tựa chuẩn hoá một lược đồ khối được đưa về việc chuẩn hoá các lát cắt của chúng, đó chính là các quan hệ trong mô hình dữ liệu quan hệ. Từ các kết quả này ta có thể triển khai tiếp quá trình chuẩn hoá và tựa chuẩn hoá trong các trường hợp riêng (đối với tập các phụ thuộc hàm dạng F_h) ..., góp phần làm hoàn chỉnh thêm lí thuyết thiết kế mô hình cơ sở dữ liệu dạng khối.

TÀI LIỆU THAM KHẢO

[1]. Nguyễn Xuân Huy, Trịnh Đình Thắng (1997), Mô hình cơ sở dữ liệu dạng khối , *Kỷ yếu các báo cáo khoa học của Hội thảo một số vấn đề chọn lọc của Công nghệ Thông tin*, Đại lải, 8/1997 (14–19).

[2]. Nguyễn Xuân Huy, Trịnh Đình Thắng, Mô hình cơ sở dữ liệu dạng khối, *Tạp chí Tin học và Điều khiển học* **14** (3) (1998) 52–60.

[3]. Nguyễn Xuân Huy, Trịnh Đình Thắng, Một số kết quả về khoá trong mô hình cơ sở dữ liệu dạng khối, *Kỷ yếu Hội thảo quốc gia về Tin học ứng dụng*, Quy Nhơn, 8/1998 (36–41).

[4]. Nguyễn Xuân Huy, Trịnh Đình Thắng, Một vài thuật toán cài đặt các phép toán của đại số quan hệ trong mô hình dữ liệu dạng khối, *Tạp chí Tin học và Điều khiển học* **15** (3) (1999) 8–17.

[5]. Trịnh Đình Thắng, Một số kết quả về bao đóng, khoá và phụ thuộc hàm trong mô hình dữ liệu dạng khối, *Hội thảo quốc gia lần thứ 4, Một số vấn đề chọn lọc của Công nghệ Thông tin*, Hải Phòng, 2001 (245–251).

[6]. Trịnh Đình Thắng, Trịnh Đình Vinh, Phụ thuộc đa trị trong mô hình dữ liệu dạng khối, *Kỷ yếu Hội thảo quốc gia Một số vấn đề chọn lọc của Công nghệ Thông tin và Truyền thông*, Huế , 12-13/06/2008 (321–328).

Nhận bài ngày 16 - 9 - 2010