

Chương 7

Lý thuyết chuẩn hóa cơ sở dữ liệu

PHỤ THUỘC HÀM

Định nghĩa phụ thuộc hàm

Cho lược đồ quan hệ $Q(A_1, A_2, \dots, A_n)$; X, Y là hai tập con khác rỗng của Q^+ .

- Ta nói X xác định Y (hay Y phụ thuộc vào X) nếu với r là một quan hệ nào đó trên Q , $\forall t_1, t_2 \in r$ mà $t_1.X = t_2.X \Rightarrow t_1.Y = t_2.Y$
- Khi đó ta ký hiệu $X \rightarrow Y$ và gọi đó là một *phụ thuộc hàm* (functional dependency).
- Ví dụ: $SOBD \rightarrow HOTEN$ là một phụ thuộc hàm; biết được $SOBD$ thì biết được $HOTEN$.

Phụ thuộc hàm hiển nhiên

- Phụ thuộc hàm $X \rightarrow Y$ với $X \supseteq Y$ gọi là phụ thuộc hàm hiển nhiên.
- Ký hiệu F để chỉ tập các phụ thuộc hàm được định nghĩa trên một lược đồ quan hệ Q ;
- Vì Q hữu hạn nên F cũng hữu hạn, ta có thể đánh số các phụ thuộc hàm của F là f_1, f_2, f_3, \dots

Xác định phụ thuộc hàm

- Để xác định đúng các phụ thuộc hàm cho một lược đồ quan hệ thì ta cần dựa vào một thông tin duy nhất chính là phần diễn giải/mô tả (còn được gọi là tân từ) của lược đồ quan hệ đó.

Một số quy ước

- Chỉ cần mô tả các phụ thuộc hàm không hiển nhiên trong tập F ; còn các phụ thuộc hàm hiển nhiên thì được hiểu là đã có trong tập F .
- Nếu thuộc tính chỉ có một ký tự duy nhất thì việc có ghi dấu phẩy giữa các thuộc tính hay không là không quan trọng.
- Nếu tập phụ thuộc hàm có nhiều phụ thuộc hàm, thì các phụ thuộc hàm được liệt kê cách nhau bởi dấu chấm phẩy – riêng sau phụ thuộc hàm cuối cùng không cần ghi dấu chấm phẩy.
- Tập phụ thuộc hàm, tập các thuộc tính được liệt kê giữa hai dấu ngoặc {}.

Tính chất phụ thuộc hàm

Để có thể xác định thêm các phụ thuộc hàm khác từ tập các phụ thuộc hàm đã được xác định; ta cần dùng các tính chất sau đây (còn gọi là hệ luật dẫn Armstrong).

Với $X, Y, Z, W \subseteq Q^+$

i. Luật phản xạ (reflexivity):

$$X \supseteq Y \Rightarrow X \rightarrow Y$$

ii. Luật tăng trưởng (augmentation):

$$X \rightarrow Y \Rightarrow XZ \rightarrow YZ$$

iii. Luật bắc cầu (transitivity):

$$X \rightarrow Y; Y \rightarrow Z \Rightarrow X \rightarrow Z$$

Tính chất phụ thuộc hàm (...)

Các luật *i,ii,iii* có thể được chứng minh bằng cách sử dụng định nghĩa của phụ thuộc hàm. Từ ba luật trên, ta có thể suy luận thêm các luật sau:

iv. Luật hợp (the union rule):

$$X \rightarrow Y; X \rightarrow Z \Rightarrow X \rightarrow YZ$$

v. Luật bắc cầu giả (the pseudotransitivity rule):

$$X \rightarrow Y; WY \rightarrow Z \Rightarrow XW \rightarrow Z$$

vi. Luật phân rã (the decomposition rule):

$$X \rightarrow Y; Z \subseteq Y \Rightarrow X \rightarrow Z$$

Ví dụ 1.

Cho lược đồ quan hệ $Q(A, B, C, D, E)$, r là quan hệ xác định trên Q được cho như sau:

A	B	C	D	E
a ₁	b ₁	c ₁	d ₁	e ₁
a ₁	b ₂	c ₂	d ₂	e ₁
a ₂	b ₁	c ₃	d ₃	e ₁
a ₂	b ₁	c ₄	d ₃	e ₁
a ₃	b ₂	c ₃	d ₁	e ₁

Thì trong bốn phụ thuộc hàm $A \rightarrow D$; $A, B \rightarrow D$; $E \rightarrow A$; $A \rightarrow E$ có hai phụ thuộc hàm $A, B \rightarrow D$; $A \rightarrow E$ là thỏa mãn r .

BÀI TOÁN BAO ĐÓNG

Định nghĩa bao đóng tập phụ thuộc hàm

Bao đóng (closure) của tập phụ thuộc hàm F (ký hiệu là F^+) là tập tất cả các phụ thuộc hàm có thể suy ra từ F dựa vào các luật dẫn Armstrong.

Sau đây là một số tính chất liên quan tập F^+

- *Tính phản xạ:* $F \subseteq F^+$
- *Tính đơn điệu:* Nếu $F \subseteq G$ thì $F^+ \subseteq G^+$
- *Tính lũy đẳng:* $F^{++} = F^+$

Bao đóng tập thuộc tính

Cho lược đồ quan hệ Q và F là tập các phụ thuộc hàm định nghĩa trên Q , $X \subseteq Q^+$. Bao đóng của tập thuộc tính X đối với tập phụ thuộc hàm F ký hiệu là X^+ (rõ hơn là X_F^+) là tập tất cả các thuộc tính $A \in Q^+$ được suy ra từ X dựa vào các phụ thuộc hàm trong F và hệ luật dẫn Armstrong, nghĩa là:

$$X^+ = \{A: A \in Q^+ \text{ và } X \rightarrow A \in F^+\}$$

Tính chất của bao đóng

$$i. X \subseteq X^+$$

$$ii. X \subseteq Y \Rightarrow X^+ \subseteq Y^+$$

$$iii. X^{++} = X^+$$

$$iv. (XY)^+ \supseteq X^+Y^+$$

$$v. (X^+Y)^+ = (XY^+)^+ = (X^+Y^+)^+$$

$$vi. X \rightarrow Y \in F^+ \Leftrightarrow Y \subseteq X^+$$

$$vii. X \rightarrow Y \Leftrightarrow Y^+ \subseteq X^+$$

Bài toán thành viên

Vấn đề nảy sinh khi nghiên cứu lý thuyết cơ sở dữ liệu là: Cho trước tập các phụ thuộc hàm F và một phụ thuộc hàm f , bài toán kiểm tra phụ thuộc hàm $f: X \rightarrow Y$ có thuộc vào tập F^+ hay không được gọi là *bài toán thành viên*.

Bài toán thành viên(...)

- Bài toán thành viên thuộc loại bài toán khó giải;
- F dù rất nhỏ nhưng F^+ thì có thể không liệt kê hết được; tuy nhiên ta có thể giải bài toán thành viên bằng cách tính X^+ và so sánh X^+ với tập Y .
- Dựa vào tính chất $X \rightarrow Y \in F^+ \Leftrightarrow Y \subseteq X^+$ ta sẽ có ngay câu trả lời $X \rightarrow Y \in F^+$ hay không ?
- Nghĩa là thay vì giải *bài toán thành viên* ta đưa về giải bài toán tìm bao đóng của tập thuộc tính.

Thuật toán tìm bao đóng tập thuộc tính

Dữ liệu vào: Q, F và $X \subseteq Q^+$

Dữ liệu ra: X^+

Bước 1: Đặt $X^+ = X$;

Bước 2: $T = X^+$;

for $\forall f: U \rightarrow V \in F$;

if $(U \subseteq X^+)$

{

$X^+ = X^+ \cup V$;

$F = F - f$;

}

Bước 3: if ($X^+ = T$)
 {
 ‘ X^+ chính là kết quả cần tìm’;
 Dừng thuật toán;
 }

 else
 Quay trở lại bước 2:

Ví dụ 2.

Cho lược đồ quan hệ $Q(A,B,C,D,E,G,H)$ và tập phụ thuộc hàm F

$$F = \{ \begin{array}{ll} f_1: & B \rightarrow A; \\ f_2: & D, A \rightarrow C, E; \\ f_3: & D \rightarrow H; \\ f_4: & G, H \rightarrow C; \\ f_5: & A, C \rightarrow D \end{array} \}$$

Tìm bao đóng của các tập $X = \{A, C\}$ dựa trên F .

Giải Ví dụ 2

$$X^+ = A, C$$

Do f_1, f_2, f_3, f_4 không thoả. f_5 thoả:

$$X^+ = A, C, \underline{D}$$

Lập lại bước 2: f_1 không thoả, f_2 thoả:

$$X^+ = A, C, D, \underline{E}$$

f_3 thoả:

$$X^+ = A, C, D, E, \underline{H}$$

Đến đây thì không có phụ thuộc hàm nào làm thay đổi X^+ nữa, thuật toán dừng lại và kết quả $X^+ = A, C, D, E, H$.

(Bạn đọc hãy tính BC^+ , BCG^+)

KHÓA CỦA LỰC ĐỒ QUAN HỆ

Định nghĩa khoá của quan hệ

Cho quan hệ $Q(A_1, A_2, \dots, A_n)$ được xác định bởi tập thuộc tính Q^+ và tập phụ thuộc hàm F định nghĩa trên Q , cho $K \subseteq Q^+$.

K là một khoá của Q nếu thoả đồng thời cả hai điều kiện sau:

i. $K \rightarrow Q^+ \in F^+$ (hay $K_F^+ = Q^+$)

(K chỉ thoả điều kiện i thì được gọi là siêu khoá)

ii. Không tồn tại $K' \subset K$ sao cho $K'^+ = Q^+$

Thuật toán tìm tất cả các khoá (cơ bản)

Bước 1: Xác định tất cả các tập con của Q^+ .

Để xác định tất cả các tập con của một lược đồ quan hệ $Q(A_1, A_2, \dots, A_n)$ ta lần lượt duyệt tất cả $2^n - 1$ tập hợp con khác rỗng của Q^+ , kết quả tìm được giả sử là tập các thuộc tính: $S = \{X_1, X_2, \dots, X_{2^n - 1}\}$

Bước 2: Tính X_i^+

Bước 3: Nếu $X_i^+ = Q^+$ thì X_i là siêu khoá.

Giả sử sau bước này có m siêu khoá: $S = \{S_1, S_2, \dots, S_m\}$.

Bước 4: Xây dựng tập chứa tất cả các khoá của Q từ tập S .

Xét mọi cặp $S_i, S_j \subset S$ ($i \neq j$), nếu $S_i \subset S_j$ thì loại S_j ($i, j = 1..m$), kết quả còn lại chính là tập các khoá cần tìm.

Thuật toán tìm tất cả các khoá (cải tiến)

Tập nguồn (TN):

Chứa tất cả các thuộc tính có xuất hiện ở vế trái mà không xuất hiện ở vế phải của bất kỳ một phụ thuộc hàm nào. Những thuộc tính không tham gia vào bất kỳ một phụ thuộc hàm nào (cả bên trái và bên phải) thì cũng đưa vào tập nguồn.

Tập trung gian (TG):

Chứa tất cả các thuộc tính vừa tham gia vào vế trái vừa tham gia vào vế phải của bất kỳ một phụ thuộc hàm nào.

Dữ liệu vào: Lược đồ quan hệ Q và tập phụ thuộc hàm F .

Dữ liệu ra: Tất cả các khoá của quan hệ.

Bước 0: Tìm tập TN , tập TG .

if ($TG = \emptyset$)

{

Khoá chính là tập TN ;

Kết thúc;

}

else

Qua bước 1;

- Bước 1:* Tìm tất cả các tập con của tập TG, gọi đó là tập X_i
- Bước 2:* Tính $TN \cup X_i$;
- Bước 3:* Tính $(TN \cup X_i)^+$;
- Bước 4:* Nếu tập $TN \cup X_i = Q^+$ thì $TN \cup X_i$ là một siêu khoá của Q . Giả sử sau bước này, ta có m siêu khoá: $S = \{S_1, S_2, \dots, S_m\}$
- Bước 5:* Xây dựng tập chứa tất cả các khoá của Q từ tập S .
Xét mọi cặp $S_i, S_j \subset S$ ($i \neq j$), nếu $S_i \subset S_j$ thì ta loại S_j ($i, j = 1..m$),
kết quả còn lại chính là tập tất cả các khoá cần tìm.

Lưu ý

Thuật toán tìm khóa cơ bản tìm được tất cả siêu khóa và tất cả khóa, trong khi thuật toán tìm khóa cải tiến thì tìm được tất cả các khóa nhưng không chắc tìm được tất cả các siêu khóa.

Ví dụ 3.

Tìm tất cả các khoá của lược đồ quan hệ Q và tập phụ thuộc hàm F sau:

$Q(A,B,C)$ và $F = \{A \rightarrow B; A \rightarrow C; B \rightarrow A\}$

⊕

X_i	X_i^+	Siêu khóa	khóa
A	Q^+	A	A
B	Q^+	B	B
C	C	-	
A,B	Q^+	A,B	
A,C	Q^+	A,C	
B,C	Q^+	B , C	
A,B,C	Q^+	A,B,C	

Vậy lược đồ quan hệ Q có hai khoá là: $\{A\}$ và $\{B\}$. □

Ví dụ 4.

(Giải lại ví dụ 3)

Áp dụng thuật toán cải tiến trên ta có lời giải như sau:

$$TN = \emptyset; TG = \{A, B\}$$

Gọi X_i là các tập con của tập TG:

X_i	$(TN \cup X_i)$	$(TN \cup X_i)^+$	Siêu khoá	Khoá
\emptyset	\emptyset	\emptyset		
A	A	Q^+	A	A
B	B	Q^+	B	B
A, B	A, B	Q^+	A, B	

Vậy quan hệ trên có hai khoá là: $\{A\}$ và $\{B\}$.

PHỦ TỐI TIỂU

Tập phụ thuộc hàm tương đương

Cho F và G là hai tập phụ thuộc hàm, F và G là hai tập phụ thuộc hàm tương đương (equivalent functional dependency) và ký hiệu là $F^+ = G^+$ nếu và chỉ nếu mỗi phụ thuộc hàm thuộc F đều thuộc G^+ và mỗi phụ thuộc hàm thuộc G đều thuộc F^+ .

Định nghĩa phủ tối thiểu

F được gọi là một tập phụ thuộc hàm tối thiểu nếu F thoả đồng thời ba điều kiện sau:

- i. Vế phải của F chỉ có một thuộc tính.
- ii. Không $\exists f: X \rightarrow A \in F$ và $Z \subset X$ mà
$$F^+ = (F - (X \rightarrow A) \cup (Z \rightarrow A))^+$$
- iii. Không $\exists X \rightarrow A \in F$ mà $F^+ = (F - (X \rightarrow A))^+$

Định nghĩa phủ tối thiểu (...)

- Trong đó vế phải của mỗi phụ thuộc hàm ở điều kiện a) chỉ có một thuộc tính, nên bảo đảm không có thuộc tính nào ở vế phải là dư thừa,
- Điều kiện b) bảo đảm không có một thuộc tính nào tham gia vế trái của phụ thuộc hàm là dư thừa,
- Điều kiện c) bảo đảm cho tập F không có một phụ thuộc hàm nào là dư thừa.
- Chú ý rằng một tập phụ thuộc hàm luôn tìm ra ít nhất một phủ tối thiểu và nếu thứ tự các phụ thuộc hàm trong tập F là khác nhau thì có thể sẽ thu được những phủ tối thiểu khác nhau.

Ví dụ 5.

Cho lược đồ quan hệ $Q(ABCDEFGH)$, thì hai tập phụ thuộc hàm F và G sau (xác định trên Q) là tương đương.

$F = \{B \rightarrow A; DA \rightarrow CE; D \rightarrow H; GH \rightarrow C; AC \rightarrow D; DG \rightarrow C\}$

$G = \{B \rightarrow A; DA \rightarrow CE; D \rightarrow H; GH \rightarrow C; AC \rightarrow D; BC \rightarrow AC;$

$BC \rightarrow D; DA \rightarrow AH; AC \rightarrow DEH\}$

(Bạn đọc hãy kiểm chứng lại ví dụ nhận xét này bằng cách sử dụng định nghĩa về tập phụ thuộc hàm tương đương và tính chất $X \rightarrow Y \in F^+ \Leftrightarrow Y \subseteq X^+$)

Ví dụ 6.

Chẳng hạn hai tập phụ thuộc hàm sau là tương đương:

$Q(A,B,C)$

$F = \{ A \rightarrow B; A \rightarrow C; B \rightarrow A; C \rightarrow A; B \rightarrow C \}$

$G = \{ A \rightarrow B; C \rightarrow A; B \rightarrow C \}$

(việc chứng minh xem như bài tập dành cho bạn đọc)

Ví dụ 7.

Cho lược đồ quan hệ Q và tập phụ thuộc F như sau:

Q(ABCD)

$F = \{$
 $A \rightarrow B;$
 $AB \rightarrow CD;$
 $B \rightarrow C;$
 $C \rightarrow D\}$

Hãy tìm phủ tối thiểu của F.

DẠNG CHUẨN CỦA LỰC ĐỒ QUAN HỆ

Thuộc tính khoá/thuộc tính không khoá

- A là một thuộc tính khoá nếu A có tham gia vào bất kỳ một khoá nào của quan hệ, ngược lại A gọi là thuộc tính không khoá.
- Ví dụ: LĐ QH $Q(A,B,C)$ và $F = \{A \rightarrow B; A \rightarrow C; B \rightarrow A\}$
Thuộc tính khóa: A,B:
Thuộc tính không khóa: C:

Thuộc tính phụ thuộc đầy đủ- phụ thuộc hàm đầy đủ.

- A là một thuộc tính phụ thuộc đầy đủ vào tập thuộc tính X nếu $X \rightarrow A$ là một phụ thuộc hàm đầy đủ (tức là không tồn tại $X' \subset X$ sao cho $X' \rightarrow A \in F^+$).
- Ví dụ: LĐ QH $Q(A,B,C)$ và $F = \{A \rightarrow B; A, B \rightarrow C; A \rightarrow C\}$
 $A, B \rightarrow C$ là phụ thuộc hàm không đầy đủ.

Dạng chuẩn 1

- Lược đồ quan hệ Q được gọi là đạt dạng chuẩn 1 (first normal form-1NF) nếu và chỉ nếu toàn bộ các thuộc tính của Q đều mang giá trị đơn.
- ❖ **Ghi chú:** Mọi lược đồ quan hệ ta xem xét trên máy tính đều xem như đã đạt dạng chuẩn tối thiểu là 1.

Ví dụ 1

Xác định dạng chuẩn của LĐQH sau:

$Q(A,B,C,D)$,

$F = \{ A, B \rightarrow C, D;$

$B \rightarrow D;$

$C \rightarrow A \}$

Dạng chuẩn 2

- Một lược đồ quan hệ Q đạt dạng chuẩn 2 (second normal form) nếu Q đạt dạng chuẩn 1 và tất cả các thuộc tính không khoá của Q đều phụ thuộc đầy đủ vào khóa.
- ❖ **Ghi chú:** Nếu một lược đồ quan hệ không đạt chuẩn 2 thì ta nói nó đạt chuẩn 1.

Ví dụ 2

Xác định dạng chuẩn của LĐQH sau:

$Q(G, M, V, N, H, P),$

$F = \{ \quad G \rightarrow N;$

$\quad G \rightarrow H;$

$\quad G \rightarrow P;$

$\quad M \rightarrow V;$

$\quad N, H, P \rightarrow M \}$

Dạng chuẩn 3

- Một lược đồ quan hệ Q đạt dạng chuẩn 3 (third normal form) nếu với mọi phụ thuộc hàm $X \rightarrow A \in F$ thì một trong 2 điều kiện sau được thỏa: hoặc X là một siêu khoá của Q hoặc A là một thuộc tính khoá của Q .

Ví dụ 3

Xác định dạng chuẩn của LĐQH sau:

$Q(A,B,C,D),$

$F=\{A,B \rightarrow C ;$

$D \rightarrow B;$

$C \rightarrow A,B\}$

DC

DA

Ví dụ 4

Xác định dạng chuẩn của LĐQH sau:

$Q(N, G, P, M),$

$F = \{N, G, P \rightarrow M;$

$M \rightarrow P\}$

Dạng chuẩn BC

- Một lược đồ quan hệ Q ở dạng chuẩn BC (Boyce Codd normal form) nếu với mỗi phụ thuộc hàm $X \rightarrow A \in F$ thì X là một siêu khoá của Q .
- ❖ **Ghi chú:** Một lược đồ quan hệ đạt chuẩn BC thì đạt chuẩn 3.

Ví dụ 5

Xác định dạng chuẩn của LĐQH sau:

$Q(A, C, D, E, I, B),$

$F = \{A, C, D \rightarrow E, B, I;$

$C, E \rightarrow A, D\}$

Ví dụ 6

Cho lược đồ quan hệ $Q(A,B,C,D,E,G)$ và tập phụ thuộc hàm

$F = \{A \rightarrow D;$

$E \rightarrow B;$

$A,E \rightarrow G;$

$B \rightarrow C\}$

a. Tìm tất cả các khóa của lược đồ quan hệ Q .

b. Hãy xác định dạng chuẩn của lược đồ quan hệ Q .

Ví dụ 7

Cho lược đồ quan hệ $p = (Q, F)$ với $Q(A, B, C, D, E, G)$ và tập phụ thuộc hàm

$F = \{C \rightarrow B;$

$A, D \rightarrow C;$

$A, B \rightarrow D;$

$C, E \rightarrow G;$

$A, E \rightarrow D\}$

a. Chứng minh rằng $\{A, E\}$ là một khóa của lược đồ quan hệ p .

b. Tìm tất cả các khóa của lược đồ quan hệ p , hãy cho biết p có bao nhiêu siêu khóa ?

c. Hãy xác định dạng chuẩn của lược đồ quan hệ p .

DẠNG CHUẨN CỦA LƯỢC ĐỒ CSDL

- Dạng chuẩn của một lược đồ cơ sở dữ liệu là dạng chuẩn thấp nhất của các lược đồ quan hệ thành phần

CHUẨN HÓA CƠ SỞ DỮ LIỆU

KIỂM TRA TÍNH BẢO TOÀN THÔNG TIN VÀ TÍNH BẢO TOÀN PHỤ THUỘC HÀM CỦA MỘT PHÂN RÃ

KIỂM TRA TÍNH BẢO TOÀN THÔNG TIN

Định nghĩa

Cho $Q(A_1, A_2, \dots, A_n)$ là một lược đồ quan hệ có n thuộc tính, p là một phép phân rã Q thành các lược đồ quan hệ con $Q_1, Q_2 \dots, Q_k$ và F là tập phụ thuộc hàm xác định trên Q . Ta nói p là phép phân rã bảo toàn thông tin đối với F nếu với mỗi quan hệ r trên Q thoả mãn:

$$Q = \Pi_{Q_1}(r) \bowtie \Pi_{Q_2}(r) \bowtie \dots \bowtie \Pi_{Q_k}(r)$$

tức là r được tạo nên từ phép kết tự nhiên của các hình chiếu của nó trên các Q_i ($i = 1 \dots k$).

Thuật toán kiểm tra phép phân rã bảo toàn thông tin

- *Dữ liệu vào:* Lược đồ quan hệ $Q(A_1, A_2, \dots, A_n)$ và tập các phụ thuộc hàm F , phép phân rã $p = \{Q_1, Q_2, \dots, Q_k\}$.
- *Dữ liệu ra:* p có bảo toàn thông tin không ?

Bước 1:

- Thiết lập bảng với $k + 1$ dòng, $n + 1$ cột (dòng/cột được tính từ 0).
- Cột j ứng với thuộc tính A_j ($j=1..n$), hàng i ứng với lược đồ quan hệ Q_i ($i=1..k$).
- Tại vị trí hàng i , cột j ta điền ký hiệu a_j nếu $A_j \in Q_i$, còn các ô khác được điền các ký hiệu b_t với t bắt đầu từ 1, t tăng dần một đơn vị theo chiều từ trái sang phải và từ trên xuống dưới.

Bước 2:

- Xét lần lượt các phụ thuộc hàm trong F , áp dụng cho bảng vừa mới thành lập ở trên.
- Chẳng hạn xét $X \rightarrow Y \in F$, ta tìm những hàng giống nhau ở tất cả các thuộc tính của X , nếu có những hàng thoả mãn tính chất này thì sẽ làm cho các ký hiệu của hai hàng này bằng nhau ở tất cả các thuộc tính của Y .

Khi làm cho hai ký hiệu của hai hàng này bằng nhau, ta sẽ gặp ba trường hợp sau đây:

- i) Nếu một trong hai ký hiệu là a_j thì cho ký hiệu kia trở thành a_j ,
- ii) Nếu hai ký hiệu là b_k hoặc b_l thì có thể đưa chúng trở thành b_t với $t = \min(k, l)$.
- iii) Nếu cả hai ký hiệu là a_j thì giữ nguyên (lúc đó chỉ số j của các ký hiệu này phải giống nhau).

Chú ý rằng bước này có thể được lặp lại cho đến khi nào ở một lần duyệt qua tất cả các phụ thuộc hàm trong F mà bảng không có sự thay đổi nào.

Bước 3:

- Xét bảng kết quả, nếu có một hàng nào đó chứa toàn giá trị a_j ($j=1..n$) thì kết luận phép phân rã là bảo toàn thông tin, ngược lại thì kết luận phép phân rã là không bảo toàn thông tin.

Định lý

- Nếu $p = \{Q_1, Q_2\}$ là một phân rã của Q , và F là một tập phụ thuộc hàm, thì p là bảo toàn thông tin ứng với F nếu và chỉ nếu:
 - $(Q_1 \cap Q_2) \rightarrow (Q_1 - Q_2)$ hoặc $(Q_1 \cap Q_2) \rightarrow (Q_2 - Q_1)$

Ví dụ 8

- Cho lược đồ quan hệ $Q(S,A,I,P)$, phép phân rã $p = \{Q_1(S,\underline{A}); Q_2(S,I,P)\}$ và tập phụ thuộc hàm $F = \{S \rightarrow A; S,I \rightarrow P\}$.

Hỏi p có bảo toàn thông tin không ?

Ví dụ 9

- Cho lược đồ quan hệ $Q(A,B,C)$ và $F = \{A \rightarrow B\}$.
- Phép phân rã Q thành $p=\{Q_1(A,B); Q_2(A,C)\}$ có bảo toàn thông tin không ?
- Phép phân rã Q thành $p=\{Q_1(A,B); Q_2(B,C)\}$ có bảo toàn thông tin không ?

Ví dụ 10

Cho lược đồ quan hệ $Q(A,B,C,D,E)$, tập phụ thuộc hàm

$F = \{ A \rightarrow C;$

$B \rightarrow C;$

$C \rightarrow D;$

$D, E \rightarrow C;$

$C, E \rightarrow A \}$

và phép phân rã $p = \{R_1, R_2, R_3, R_4, R_5\}$ của Q như sau:

$R_1(A,D); \quad R_2(A,B); \quad R_3(B,E); \quad R_4(C,D,E); \quad R_5(A,E)$

Kiểm tra xem phép phân rã p có bảo toàn thông tin không ?

KIỂM TRA TÍNH BẢO TOÀN PHỤ THUỘC HÀM

Định nghĩa

- Cho phân rã $p = \{Q_1, Q_2, \dots, Q_k\}$ của một lược đồ quan hệ Q và tập phụ thuộc hàm F .
- Hình chiếu của tập F lên tập thuộc tính Z ký hiệu $\Pi_Z(F)$ là tập các phụ thuộc hàm $X \rightarrow Y \in F^+$ sao cho $XY \subseteq Z$.
- Ta nói phép p là phân rã bảo toàn phụ thuộc hàm (dependency preserving decomposition) F nếu hợp của tất cả các phụ thuộc hàm trong $\Pi_{Q_i}(F)$ với $i = 1..k$ suy ra được tất cả các phụ thuộc hàm trong F .

Thuật toán kiểm tra bảo toàn phụ thuộc hàm

Dữ liệu vào: Phân rã $p = \{Q_1, Q_2, \dots, Q_k\}$ và một tập các phụ thuộc hàm $F = \{f_1, f_2, \dots, f_m\}$

Dữ liệu ra: p có bảo toàn phụ thuộc hàm không ?

Gọi G là hợp của các $\Pi_{Q_i}(F)$, để tính xem G có tương đương với F hay không chúng ta xét mỗi phụ thuộc hàm $X \rightarrow Y \in F$ và xác định xem X^+ (được tính ứng với G) có chứa Y hay không ? Nếu có thì phụ thuộc hàm này thuộc G .

- Chúng ta định nghĩa phép toán Q trên tập các thuộc tính Z ứng với một tập phụ thuộc F là phép thế Z bằng $Z \cup ((Z \cap Q)^+ \cap Q)$ trong đó bao đóng luôn được tính ứng với F , phép toán này nối Z với những thuộc tính A sao cho $(Z \cap Q) \rightarrow A \in \Pi_Q(F)$.
- Do đó chúng ta tính X^+ ứng với G bằng cách khởi đầu với X , qua danh sách các Q_i , chúng ta lần lượt thực hiện các phép toán Q_i với mỗi i , nếu tại một lần lặp nào đó không có một phép toán Q_i nào làm thay đổi các tập thuộc tính hiện có thì chúng ta đã thực hiện xong; tập kết quả chính là X^+ .

Ví dụ 11

- Cho lược đồ quan hệ $Q(C,S,Z)$ và tập phụ thuộc hàm $F=\{C,S \rightarrow Z; Z \rightarrow C\}$
- Xét xem phân rã Q thành $Q_1(S,Z)$ và $Q_2(C,Z)$ có bảo toàn thông tin không ? Có bảo toàn phụ thuộc hàm không ?

Ví dụ 12

- Xét tập thuộc tính $Q(A,B,C,D)$ với phân rã $\{\{A,B\},\{B,C\},\{C,D\}\}$ và tập phụ thuộc hàm $F=\{A \rightarrow B; B \rightarrow C; C \rightarrow D; D \rightarrow A\}$.
- Phân rã này có bảo toàn tập phụ thuộc hàm và bảo toàn thông tin không ?

CHUẨN HÓA LƯỢC ĐỒ QUAN HỆ BẰNG PHƯƠNG PHÁP PHÂN RÃ

Phân rã một lược đồ thành các lược đồ con dạng chuẩn 3 bảo toàn phụ thuộc hàm

- *Dữ liệu vào:* Lược đồ quan hệ Q và tập phụ thuộc hàm F (giả sử rằng F đã ở dạng phủ tối thiểu).
- *Dữ liệu ra:* Một phân rã bảo toàn phụ thuộc hàm và mỗi lược đồ quan hệ con đều đạt chuẩn 3NF ứng với hình chiếu của F trên lược đồ đó.

- *Bước 1:*

Nếu có những thuộc tính của Q không nằm trong một phụ thuộc hàm nào của F - dù ở vế phải hay vế trái - thì ta loại chúng ra khỏi Q .

- *Bước 2:*

Nếu có một phụ thuộc hàm nào của F mà liên quan đến tất cả các thuộc tính của Q thì kết quả ra chính là Q (Q không thể phân rã)

- *Bước 3:*

Cứ mỗi phụ thuộc hàm $X \rightarrow A \in F$ thì $\{X, A\}$ là một lược đồ cần tìm.

Phân rã lược đồ thành các lược đồ con dạng chuẩn 3NF vừa bảo toàn phụ thuộc hàm vừa bảo toàn thông tin

- Tìm một phân rã p của Q có dạng 3NF như thuật toán trên, tiếp theo là tìm một khoá X của Q , nếu trong các lược đồ con Q_i có một lược đồ nào đó chứa trọn vẹn khóa X thì p đã bảo toàn thông tin; ngược lại thì ta thêm một lược đồ quan hệ mới là X có tập thuộc tính chính là tập khóa của lược đồ quan hệ Q .

Ví dụ 13

Phân rã lược đồ quan hệ sau thành lược đồ cơ sở dữ liệu đạt dạng chuẩn 3.

$Q(G,H,A,B,C,D),$

$F=\{GH \rightarrow AD;$

$AG \rightarrow B;$

$CD \rightarrow GH;$

$C \rightarrow A;$

$BH \rightarrow C\}$

Ví dụ 14

$Q(A,B,C,D,E,G,H)$

$F = \{ A, B \rightarrow D;$

$E, H \rightarrow G;$

$G \rightarrow C;$

$D \rightarrow C \}$

Hãy phân rã Q theo thuật toán vừa bảo toàn phụ thuộc hàm vừa bảo toàn thông tin.

Phân rã một lược đồ thành các lược đồ con đạt dạng chuẩn BC và bảo toàn thông tin

- **Bổ đề 1**
- Giả sử Q là một lược đồ quan hệ với tập phụ thuộc hàm F , gọi $p = \{Q_1, Q_2, \dots, Q_k\}$ là một phân rã của Q bảo toàn thông tin ứng với Q . Nếu Q_1 được phân rã thành hai lược đồ con (S_1, S_2) bảo toàn thông tin, thì phân rã của Q thành $(S_1, S_2, Q_2, \dots, Q_k)$ cũng bảo toàn thông tin ứng với F .

Bổ đề 2

- a) Mỗi lược đồ có hai thuộc tính đều ở dạng chuẩn BC.
- b) Nếu Q không đạt dạng chuẩn BC thì chúng ta có thể tìm được các thuộc tính A và B trong Q sao cho $(Q - AB)^+ \rightarrow A$. Phụ thuộc $(Q - AB)^+ \rightarrow B$ có thể đúng trong trường hợp này.

Từ đây ta có thể phát biểu thêm mệnh đề đảo cho mệnh đề này như sau:

b') Nếu trong Q không tìm được các thuộc tính A và B sao cho $(Q - AB)^+ \rightarrow A$ và phụ thuộc $(Q - AB)^+ \rightarrow B$ thì Q đã đạt dạng chuẩn BC

Thuật toán

- *Dữ liệu vào*: Cho lược đồ quan hệ phổ quát Q và tập phụ thuộc hàm F
- *Dữ liệu ra*: Lược đồ CSDL tương ứng đạt dạng chuẩn BC.

- Trọng tâm của thuật toán là lấy lược đồ quan hệ Q rồi phân rã nó thành hai lược đồ.
- Một lược đồ có tập các thuộc tính $\{X, A\}$; nó có dạng chuẩn BC và phụ thuộc hàm $X \rightarrow A$ đúng. Lược đồ thứ hai sẽ là $Q - A$, do đó nối của $R - A$ với $\{X, A\}$ là bảo toàn thông tin.
- Thực hiện đệ quy thủ tục phân rã, với $R - A$ ở vị trí của R cho đến khi chúng ta đạt được một lược đồ đáp ứng được các điều kiện của bổ đề 2.b; chúng ta biết rằng lược đồ này có dạng chuẩn BC, thế thì bổ đề 1 bảo đảm rằng lược đồ này hợp với các lược đồ chuẩn BC được tạo ra từ mỗi bước đệ quy có kết nối bảo toàn thông tin.

Chương trình chính của thuật toán Phân rã

do

{

 Phân rã(Q,F);

$p = p \cup Y$;

$Q = Q - A$;

}

while not timAB(Q,F);

$p = p \cup \langle Q, F \rangle$

Để phân rã [Q,F] ta thực hiện theo thủ tục đệ qui như sau:

Phân rã(Q,F)

{

if (nếu không tìm ra cặp A,B trong Q thoả bổ đề 2)

{

Q đã ở dạng chuẩn BC và dừng quá trình phân rã.

exit;

}

Y = Q;

while timAB(Y,F)

{tìm A,B thoả bổ đề 2}

{

Y=Y - B;

}

Q = Q - A;

Phân rã(Q,F) {gọi đệ qui}

}

tìmcapAB(Q,F)

{

G={Tập phụ thuộc hàm được bao trong Q}

$\forall A, B \in Q, A \neq B$

{

if $(Q - (A \cup B)) \rightarrow A \in G$ thì

Exit; {đến thủ tục phân rã}

}

}

Ví dụ 15

Cho lược đồ quan hệ $Q(B, O, S, Q, I, D)$ và tập phụ thuộc hàm F

$$F = \{ S \rightarrow D,$$

$$I \rightarrow B;$$

$$I, S \rightarrow Q;$$

$$B \rightarrow O\}$$

Phân rã Q theo thuật toán trên.