LÝ THUYẾT THIẾT KẾ
CƠ SỞ DỮ LIỆU QUAN HỆ

Đặt vấn đề

Các vấn đề nảy sinh với CSDL được thiết kế tồi:

- Dư thừa dữ liệu
- Dị thường khi sửa chữa
- Dị thường khi thêm bộ
- Dị thường khi xoá bộ

Ví dụ: CSDL gồm 1 quan hệ:

MASV	HOTEN	NS	MAHP	TENHP	HK	TC	DIEM
BK10000	Nguyễn Mai Hoa	1-5-85	m1	Đại số	1	4	6
BK10000	Nguyễn Mai Hoa	1-5-85	m2	Giải tích	1	3	7
BK10000	Nguyễn Mai Hoa	1-5-85	m3	Giải tích	2	4	5
BK10001	Trần Mạnh Dũng	2-9-85	m1	Đại số	1	4	8
BK10001	Trần Mạnh Dũng	2-9-85	m2	Giải tích	1	3	4

Khái niệm phụ thuộc hàm

- **Định nghĩa.** Cho tập thuộc tính U hữu hạn khác $\varnothing$. Một phụ thuộc hàm trên U có dạng $X \rightarrow Y$, $X, Y \subseteq U$
- (Y phụ thuộc hàm vào X hay X quyết định Y, X là vế trái và Y là vế phải của phụ thuộc hàm)
- Cho R(U). Nói R thoả $X \rightarrow Y$ nếu $\forall u, v \in R$ mà $u.X = v.X$ thì $u.Y = v.Y$
- Kí hiệu R(f) - quan hệ R thỏa phụ thuộc hàm f
- R thỏa tập phụ thuộc hàm F nếu R(f), với $\forall f \in F$

  kí hiệu R(F).

Lược đồ quan hệ

Định nghĩa. Lược đồ quan hệ α là cặp hai thành phần:

$\boldsymbol{\alpha}$ **= <U, F>**,

trong đó U là tập thuộc tính, F là tập phụ thuộc hàm.

- Quan hệ R(U) gọi là thuộc lược đồ α nếu R(F).
- Ví dụ: Cho lược đồ α = <U, F> với

U = { MASV, HOTEN, NS, MAHP, TENHP, HK, TC, DIEM}

F = {MASV $\rightarrow$ {HOTEN, NS};

MAHP $\rightarrow$ {TENHP, HK, TC};

TENHP $\rightarrow$ MAHP; {MASV, MAHP} $\rightarrow$ DIEM}

Suy dẫn logic

Định nghĩa . Cho lược đồ quan hệ α = <U, F>;

f : phụ thuộc hàm trên U.

Nói f suy dẫn logic được từ F nếu với mọi quan hệ R thuộc α, R(F) thì R(f). Kí hiệu F $\vDash$ f.

Ví dụ. U = ABC; F = {A $\rightarrow$ B, B $\rightarrow$ C}

Chứng minh rằng F$\vDash$ A $\rightarrow$ C

R là một quan hệ bất kì thuộc lược đồ α = <U, F>, ta có R(F).

Ta sẽ chứng minh R(A $\rightarrow$ C):

Giả sử u, v $\in$ R mà u.A = v.A (3.1) và u.C ≠ v.C (3.2).

Vì R(A $\rightarrow$ B), (3.1) ⇨ u.B = v.B ⇨ u.C = v.C (vì R(B $\rightarrow$C))

Mâu thuẫn với (3.2). Vậy u.C = v.C ⇨ R(A $\rightarrow$ C)

Bao đóng của tập phụ thuộc hàm

- **Định nghĩa.** Bao đóng của tập phụ thuộc hàm F là:

$$F^{+} = \{ X \rightarrow Y \mid F \models X \rightarrow Y \}$$

Khi $F = F^{+}$, nói F là họ đầy đủ các phụ thuộc hàm.

- **Tính chất:**

F và G là 2 tập phụ thuộc hàm trên tập thuộc tính U.

1. $F \subseteq F^{+}$
2. Tính đơn điệu

   Nếu $F \subseteq G$ thì $F^{+} \subseteq G^{+}$
3. Tính lũy đẳng

   $F^{++} = F^{+}$
4. $(FG)^{+} \supseteq F^{+} G^{+}$
5. $(FG)^{+} = (F^{+}G)^{+} = (FG^{+})^{+}$

Hệ tiên đề cho các phụ thuộc hàm

Hệ tiên đề Armstrong (A^0)

Cho lược đồ quan hệ $\alpha = <U, F>$, X, Y, Z $\subseteq$ U.

A1) Luật phản xạ (reflexity)

Nếu $Y \subseteq X$ thì $X \to Y$

A2) Luật tăng trưởng (augmentation)

Nếu $X \to Y$ thì $XZ \to YZ$

A3) Luật bắc cầu (transitivity)

Nếu $X \to Y$ và $Y \to Z$ thì $X \to Z$

Kí hiệu $F \vdash f$: phụ thuộc hàm fsuy dẫn được từ F nhờ hệ tiên đề Armstrong (Bằng cách áp dụng các luật của $A^{0)}$

$F^* = \{f \mid F \vdash f\}$

Ví dụ. Chứng minh $\{A \to B, B \to CD\} \vdash AC \to CD$

(1) $A \to B$

(2) $B \to CD$

(3) $A \to CD$ (1, 2, A3)

(4) $AC \to CD$ (3, A2) (||)

*** Các quy tắc khác:**

A4) Luật hợp (union rule)

Nếu $X \to Y$, $X \to Z$ thì $X \to YZ$

A5) Luật tựa bắc cầu (pseudo-transitivity rule)

Nếu $X \to Y$, $WY \to Z$ thì $WX \to Z$

A6) Luật tách (decomposition rule)

Nếu $X \to Y$, $Z \subseteq Y$ thì $X \to Z$

Hệ quả: (Hệ quả quan trọng của luật hợp - tách)

Giả sử $Y = A_1A_2\ldots A_n$ thì:

$X \rightarrow Y \Leftrightarrow X \rightarrow A_i$, $\forall i = 1,2,..n$.

* **Định lý.** *Hệ tiên đề A^0 là đúng và đầy đủ.*

(i) A^0 đúng: $F^* \subseteq F^+$

(ii) A^0 đầy đủ: $F^* = F^+$

<u>Chứng minh (i):</u> Điều (i) có nghĩa:

Với F là một tập phụ thuộc hàm, R là một quan hệ bất kì và R(F), nếu $F \vdash X \rightarrow Y$ thì $F \models X \rightarrow Y$ hay $R(X \rightarrow Y)$

<u>Chứng minh (ii):</u> Do (i) đã có $F^* \subseteq F^+$ nên chỉ cần chứng minh $F^+ \subseteq F^*$ hay $\forall X \rightarrow Y \in F^+$ thì $X \rightarrow Y \in F^*$.

Ví dụ. U = CTHRSG

F = { $C \rightarrow T$, $HT \rightarrow R$, $HS \rightarrow R$, $CS \rightarrow G$, $HR \rightarrow C$}

Chứng minh $CH \rightarrow R$ và $HS \rightarrow U$ được suy dẫn từ F nhờ hệ tiên đề Armstrong.

Bài toán thành viên

Cho tập thuộc tính U, F là tập phụ thuộc hàm trên U, f là một phụ thuộc hàm trên U. Hỏi $X \rightarrow Y \in F^+$ hay không.

Định nghĩa 3.5. Cho F là tập phụ thuộc hàm trên tập thuộc tính U, $X \subseteq U$. Bao đóng của X đối với F là

$$X^+_F = \{A \in U \mid F \vdash X \rightarrow A\}$$

Trong ngữ cảnh cụ thể có thể viết X^+ thay cho X^+_F

Ghi nhớ:

- Các điều sau là tương đương: $F \vDash X \rightarrow Y \Leftrightarrow X \rightarrow Y \in F^+$
$\Leftrightarrow F \vdash X \rightarrow Y \Leftrightarrow Y \subseteq X^+_F$

$$X^+ = \{A \mid F \vdash X \rightarrow A\} = \{A \mid X \rightarrow A \in F^+\}$$

Tính chất

U - tập thuộc tính; F - tập phụ thuộc hàm trên U; X, Y $\subseteq$ U

1. Phản xạ: $X \subseteq X^+$
2. Tính đơn điệu: $X \subseteq Y$ thì $X^+ \subseteq Y^+$
3. Tính luỹ đẳng: $X^+ = (X^+)^+ = X^{++}$
4. $(XY)^+ \supseteq X^+Y^+$
5. $(XY)^+ = (X^+Y)^+ = (XY^+)^+$
6. $F \vdash X \rightarrow Y \Leftrightarrow Y \subseteq X^+$
7. $X \rightarrow Y \Leftrightarrow Y^+ \subseteq X^+$
8. $X \rightarrow X^+$ và $X^+ \rightarrow X$
9. $X^+ = Y^+ \Leftrightarrow X \rightarrow Y$ và $Y \rightarrow X$

Tính bao đóng của tập thuộc tính

U - tập hữu hạn các thuộc tính, $X \subseteq U$, F - tập phụ thuộc hàm trên U.

Phương pháp tính X^+ :

Tính liên tiếp các tập thuộc tính $X^0, X^1, \ldots X^i, \ldots$

1. $X^0 = X$
2. $X^{i+1} = X^i \cup Z^i$, với $Z^i = \cup Y_j$ nếu $X_j \rightarrow Y_j \in F$ và $X_j \subseteq X^i$

Vì $X^0 \subseteq X^1 \subseteq X^2 \subseteq \ldots \subseteq X^i \subseteq \ldots \subseteq U$, dãy đơn điệu tăng và bị chặn trên bởi U do đó $\exists i$: $X^i = X^{i+1}$ (*). Khi đó $X^+ = X^i$, với i là số nguyên nhỏ nhất thỏa mãn (*).

Lưu ý: Một cải thiện của thuật toán: loại bỏ một phụ thuộc hàm sau khi đã dùng nó.

Ví dụ. Cho tập phụ thuộc hàm

$F = \{AC \rightarrow BE, ACE \rightarrow DG, B \rightarrow CE, ACD \rightarrow EGH\}$

a. Tính X^+ với $X = A$

b. Chứng tỏ rằng $ABC \rightarrow EDH$ suy dẫn được từ F nhờ hệ tiên đề A^0

Khoá của lược đồ quan hệ

Định nghiã. Cho lược đồ quan hệ α = <U, F>, K $\subseteq$ U.

K được gọi là khoá (key) của lược đồ α nếu:

i) $K^+ = U$

ii) $\nexists$ K' $\subset$ K mà K' thoả mãn (i)

Nếu K chỉ thoả mãn (i), K là siêu khoá (super key).

Ví dụ. Xét lược đồ quan hệ với

U = {S#, SNAME, P#, PNAME, QTY}

F = {S# $\rightarrow$ SNAME, P# $\rightarrow$ PNAME, {S#, P#} $\rightarrow$ QTY}}

Lược đồ có một khoá là {S#, P#}

Ví dụ 1. Cho lược đồ α = <U, F> với

U = {MASV, HOTEN, NS, MAHP, TENHP, HK, TC, DIEM}

F = {MASV$\rightarrow${HOTEN,NS}; MAHP$\rightarrow$ {TENHP, HK, TC}; TENHP $\rightarrow$ MAHP; {MASV, MAHP} $\rightarrow$ DIEM}

Lược đồ α có 2 khoá là {MASV, MAHP} và {MASV, TENHP}

Ví dụ 2. U = CSZ; F = {CS$\rightarrow$Z, Z$\rightarrow$C}

Khoá là: CS và SZ

*** Nhận xét**:

- Mọi lược đồ quan hệ α đều có ít nhất một khoá.
- Khoá của một lược đồ nói chung không duy nhất
- Các khoá không bao nhau
- Hợp của hai khóa là một siêu khoá, không là một khoá. Giao của hai khoá không là siêu khóa.

Thuật toán tìm khoá của lược đồ quan hệ

Thuật toán. Tìm một khoá của lược đồ quan hệ

Algorithm KEY1

Input U, F

Output Một khoá K của lược đồ $\alpha = \langle U, F \rangle$

Method

1. K := U;
2. For each A in U do

   If $(K \setminus \{A\})^+ = U$ then $K := K \setminus \{A\}$; endif;

   Endfor;
3. Return (K);

End.

Thuật toán. Tìm một khoá của lược đồ quan hệ

Algorithm KEY2

Input $U, F = \{L_i \rightarrow R_i \mid i = 1,2,...m\}$

Output Một khoá K của lược đồ α = <U,F>

Method

1. $K := (U \setminus L_1R_1) \cup L_1$;

/*Giả sử $K= \{A_1, A_2,....,A_n\}$ */

2. For i := 1 to n do

   If $(K \setminus \{A_i\})^+ = U$ then $K := K \setminus \{A_i\}$; endif;

   Endfor;

3. Return (K);

End.

Algorithm **KEY 3**

Input $U, F = \{L_i \rightarrow R_i \mid i = 1,2,...m\}$

Output Một khoá K của lược đồ α = <U, F>

1. Tính giao của mọi khoá

$$I_\alpha = U \setminus \bigcup_{i=1}^{m} (R_i \setminus L_i)$$

If $I_\alpha^+ = U$ then Return (I_α);

2. Tìm tập $P \subseteq N_\alpha$ là tập các thuộc tính không thuộc khoá nào

$$P = \bigcup_{L_i \subseteq I_\alpha} (R_i \setminus L_i)$$

3. $M := (I_\alpha P)^+ \setminus I_\alpha$ /* $P \subseteq M \subseteq N_\alpha$ */
4. $H := U \setminus M \setminus I_\alpha$
5. Xuất phát từ I_α bổ sung dần các thuộc tính trong H cho đến khi tìm được một khoá K.
6. Return (K);

- Nếu H đủ nhỏ có thể tìm được tất cả các khoá của lược đồ bằng thuật toán trên
- Nếu $I_\alpha^+ = U$ thì lược đồ có khoá duy nhất là I_α.
- Bước 3 trong thuật toán dựa vào tính chất sau:

  Nếu $X \subseteq I_\alpha$, $Y \subseteq N_\alpha$, $P = (XY)^+ \setminus X$ thì $Y \subseteq P \subseteq N_\alpha$

Ví dụ. Tìm khoá của lược đồ α = <U, F>, với:

a) U = CSZ, F = {CS → Z, Z → C}

b) U = CTHRSG;

F = {C→T, HT→R, HS→R, CS→G, HR→ C}

c) U = ABCDEGHIK

F = {ACH → BH, BH → ACD, ABCI → DIK,

ADEI → BGC, CGI → AEK, H → BC}

PHỦ CỦA TẬP PHỤ THUỘC HÀM

Tập tối thiểu

Định nghĩa. Tập phụ thuộc hàm F là tập tối thiểu (minimal) nếu:

1) Mỗi vế phải của mỗi phụ thuộc hàm trong F chỉ có 1 thuộc tính.

2) $\nexists X \to A \in F$ sao cho $F^+ = (F \setminus \{X \to A\})^+$.

3) $\nexists X \to A \in F$ mà $F^+ = (F \setminus \{X \to A\} \cup \{Z \to A\})^+$, với $Z \subset X$.

Kiểm tra điều kiện 2)

Với mỗi $X \to A \in F$, $X \to A$ là dư thừa, tức là $F^+ = F1^+$, trong đó

$F1 = F \setminus \{X \to A\} \Leftrightarrow A \in X^+_{F1}$.

Chứng minh:

(i) Nếu $A \in X^+_{F1} \Leftrightarrow X \to A \in F1^+$.

Do $(FG)^+ = (F^+G)^+ = (FG^+)^+$ nên

$(F1 \cup \{X \to A\})^+ = (F1^+ \cup \{X \to A\})^+ = (F1^+)^+$,

hay $F^+ = (F1^+)^+ = F1^+$

⇨ $X \to A$ là dư thừa.

(ii) Nếu $X \to A$ là dư thừa, tức là $F^+ = F1^+$

Do $X \to A \in F$ ⇨ $X \to A \in F1^+$ ⇨ $A \in X^+_{F1}$

Từ (i) và (ii) ⇨ $X \to A$ là dư thừa ⇨ $A \in X^+_{F1}$

Cách kiểm tra điều kiện 3)

Với mỗi $X \to A \in F$, $B \in X$ là dư thừa, tức là

$$F^+ = (F \setminus \{X \to A\} \cup \{Z \to A\})^+, \text{ với } Z = X \setminus \{B\}$$

$$\Leftrightarrow A \in (X \setminus \{B\})^+_F$$

Chứng minh:

(i) Nếu B là dư thừa, từ $F^+ = (F \setminus \{X \to A\} \cup \{Z \to A\})^+$

$$\Rightarrow Z \to A \in F^+ \Rightarrow A \in Z^+_F$$

(ii) Nếu $A \in Z^+_F \Rightarrow Z \to A \in F^+$. Lại có: $X \supset Z \Rightarrow X \to Z$

$$\Rightarrow X \to A \in F^+ \text{ (luật bắc cầu).}$$

Mọi phụ thuộc hàm thuộc F^+ mà suy dẫn được từ $X \to A$ thì cũng suy dẫn được từ $Z \to A$ nên có thể thay $X \to A$ bởi $Z \to A$ hay B là dư thừa.

Từ (i) và (ii) ta có điều phải chứng minh. (||)

PHỦ CỦA TẬP PHỤ THUỘC HÀM

Sự tương đương giữa 2 tập phụ thuộc hàm

- **Định nghĩa.** Cho F, G là 2 tập phụ thuộc hàm trên tập thuộc tính U. F và G là tương đương nếu $F^+ = G^+$.

- **Bổ đề.** *Mọi tập phụ thuộc hàm F đều tương đương với một tập phụ thuộc hàm G mà vế phải của các phụ thuộc hàm trong G bao gồm không quá 1 thuộc tính.*

Chứng minh:

Xây dựng G là tập:

$G = \{X \to A_i\}$, với mỗi $X \to Y \in F$ và $Y = \{A_1, \ldots A_n\}$.

Ta chứng minh $F^+ = G^+$.

(i) Chứng minh $G^+ \subseteq F^+$

Do $X \to Y \vdash X \to A_i$ (luật tách) nên $G \subseteq F^+$

suy ra $G^+ \subseteq F^+$.

(ii) Chứng minh $F^+ \subseteq G^+$

Ta có $\{X \to A_1, \ldots, X \to A_n\} \vdash X \to Y \in F$ (luật hợp)

$\Rightarrow F \subseteq G^+ \Rightarrow F^+ \subseteq G^+$.

Từ (i) và (ii) suy ra $F^+ = G^+$. (||)

Phủ tối thiểu (minimal cover)

Định nghĩa. Cho F là một tập phụ thuộc hàm. Tập phụ thuộc hàm G gọi là phủ tối thiểu của F nếu:

1) G tương đương F,
2) G là tập tối thiểu.

Nhận xét:

- Luôn tìm được ít nhất 1 phủ tối thiểu cho F.
- Một tập phụ thuộc hàm F có thể có nhiều phủ tối thiểu.

Thuật toán. Tìm một phủ tối thiểu

Algorithm Minimal Cover

Input F

Output Tập G là phủ tối thiểu của F

Method

1. G := F;
2. Thay mỗi phụ thuộc hàm $X \rightarrow \{A_1 \ldots A_n\} \in G$ bằng n phụ thuộc hàm $X \rightarrow A_1$, $X \rightarrow A_2, \ldots, X \rightarrow A_n$
3. For each $X \rightarrow A$ in G do

   If $X \rightarrow A$ là dư thừa then $G := G \setminus \{X \rightarrow A\}$;
4. For each $X \rightarrow A$ in G do

   For each B in X do

   If B là dư thừa then

   $G := G \setminus \{X \rightarrow A\} \cup \{X \setminus \{B\} \rightarrow A\}$;
5. Return (G);

Ví dụ 1:

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

Loại 2 phụ thuộc hàm (theo thứ tự): $B \rightarrow A$, $A \rightarrow C$ được phủ tối thiểu $\{A \rightarrow B, B \rightarrow C, C \rightarrow A\}$.

Nếu loại $B \rightarrow C$ được phủ tối thiểu:

$$\{A \rightarrow B, B \rightarrow A, A \rightarrow C, C \rightarrow A\}.$$

Ví dụ 2:

$$F = \{AB \rightarrow C, A \rightarrow B, B \rightarrow A\}$$

Có thể loại A hoặc B từ $AB \rightarrow C$ nhưng không thể loại đồng thời cả hai.

Ví dụ 3:

F = { AB → C, C → A, BC → D, ACD → B, D →E,

D →G, BE →C, CG→B,CG →D, CE →A, CE →G}

Loại các phụ thuộc hàm theo thứ tự

ACD →B, CG →D, CE → A được phủ tối thiểu là:

G = { AB →C, C →A, BC →D, D →E, D →G,

BE →C, CG →B, CE →G}.

Nếu thứ tự loại các phụ thuộc hàm là: CE → A, CG → B và loại A khỏi ACD →B (thay ACD → B bằng CD → B) được kết quả là:

G = { AB →C, C →A, BC→D, D →E, D →G,

BE →C, CD →B, CG →D, CE → G}.

PHÉP TÁCH LƯỢC ĐỒ QUAN HỆ

- **Hình chiếu của tập phụ thuộc hàm trên một tập thuộc tính**

  Cho tập phụ thuộc hàm F trên tập thuộc tính U ,

  $Z \subseteq U$

  $$\Pi_Z(F) = \{X \rightarrow Y \in F^+ | \ XY \subseteq Z\}$$

- **Phép tách lược đồ quan hệ**

  Cho lược đồ quan hệ α = <U, F>. Phép tách lược đồ α là thay α bằng tập các lược đồ quan hệ $\alpha_1, \alpha_2, \ldots, \alpha_k$ sao cho: $\alpha_i = <U_i, F_i>$, $i = 1,2,..,k$, trong đó $U_i \subseteq U$, $F_i = \prod_{Ui}(F)$.

- Ký hiệu phép tách là $\delta = [U_1, U_2, \ldots, U_k]$

- Cho lược đồ α = <U, F> với

U = {MASV, HOTEN, NS, MAHP, TENHP, HK, TC, DIEM}

F = {MSV $\rightarrow$ {HOTEN, NS};

MAHP $\rightarrow$ {TENHP, HK, TC};

TENHP $\rightarrow$ MAHP; {MASV, MAHP} $\rightarrow$ DIEM}

- Xét phép tách $\delta = [U_1, U_2, U_3]$:

U_1 = {MASV, HOTEN, NS}

U_2 = {MAHP, TENHP, HK, TC}

U_3 = {MASV, MAHP, DIEM}

F_1 = {MASV $\rightarrow$ {HOTEN, NS}

F_2 = {MAHP $\rightarrow$ {TENHP, HK, TC};TENHP $\rightarrow$ MAHP}

F_3 = {{MASV, MAHP} $\rightarrow$ DIEM

- Cho lược đồ α = <U, F> với

  U = CSZ; F = $\{CS \to Z, Z \to C\}$

- Xét phép tách ρ = [CS, ZC]

  U_1 = CS; $F_1 = \varnothing$

  U_2 = ZC; $F_2 = \{Z \to C\}$

- Nhận xét:

- Với phép tách ρ, $F_1 \cup F_2 \nvdash F$

– Với phép tách δ, $F_1 \cup F_2 \cup F_3 \vdash F$

- **Phép tách bảo toàn tập phụ thuộc hàm (**(Dependency Preservation)**)**

  Cho lược đồ quan hệ α = <U, F> và một phép tách $\delta = [U_1, \ldots, U_k]$ của α. Khi đó δ là phép tách bảo toàn tập phụ thuộc hàm F nếu: $\cup F_i \vdash F$

Phép tách không mất thông tin

(Lossles - Join Decomposition)

Ví dụ: Quan hệ T trên lược đồ α sau khi tách và kết nối lại có thể nảy sinh những bộ mới:

T (A B C)	T_1 (A B)	T_2 (B C)
$A_1\ b_1\ c_1$	$a_1\ b_1$	$b_1\ c_1$
$a_1\ b_2\ c_1$	$a_1\ b_2$	$b_2\ c_1$
$a_2\ b_2\ c_2$	$a_2\ b_2$	$b_2\ c_2$

$T_1 * T_2 \supset \{(a_1, b_2, c_2), (a_2, b_2, c_1)\}$, trong đó $(a_2, b_2, c_1) \notin T$

Phép tách không mất thông tin đảm bảo 1 quan hệ có thể khôi phục lại từ các phần chiếu của nó (bằng phép kết nối)

- **Định nghĩa**

  Phép tách δ = [U_1, U_2,…,U_k] của lược đồ quan hệ α = <U, F> là không mất thông tin nếu với $\forall$ R thuộc α:

  $$R = R_1 * R_2 * \ldots * R_k,$$

  trong đó $R_i = R[U_i]$, i = 1, 2,…, k

  Ký hiệu $R_1 * R_2 * \ldots * R_k = m_\delta(R)$

- **Bổ đề**. Cho lược đồ quan hệ α = <U, F> và phép tách δ = [U_1, U_2,…,U_k] của lược đồ α. Khi đó, với mọi quan hệ R của lược đồ α ta có:

1. $R \subseteq m_\delta(R)$
2. Nếu $S = m_\delta(R)$ thì với i = 1,2,..,k, ta có: $S[U_i] = R[U_i]$
3. $m_\delta(m_\delta(R)) = m_\delta(R)$

Thuật toán kiểm tra phép tách không mất thông tin

Algorithm Lossles - Join Decomposition

Input Lược đồ quan hệ α = <U, F> ,

$U = \{A_1, \ldots, A_n\}$,

phép tách $\delta = [U_1, U_2, \ldots U_k]$

Output Kết luận δ có là phép tách có kết nối không mất thông tin hay không

<u>Method</u>

1. Xây dựng bảng n cột, k hàng, cột j ứng với thuộc tính Aj $\in$ U, hàng i ứng với lược đồ quan hệ αi. Ở hàng i, cột j ghi kí hiệu:

+ a_j nếu $A_j \in U_i$

+ b_{ij} nếu $A_j \notin U_i$

2. Biến đổi bảng trên theo qui tắc:

- Với mỗi phụ thuộc hàm $X \rightarrow Y \in F$

Xét các hàng, nếu có 2 hàng i, t giống nhau trên X thì làm chúng giống nhau trên Y theo cách:

+ Nếu 1 trong 2 kí hiệu là a_j thì thay kí hiệu kia là a_j

+ Nếu 2 ký hiệu là bij và btj thì thay cả 2 ký hiệu là b_{ij} hoặc b_{tj} đều được.

- Quá trình dừng khi không làm thay đổi bảng được nữa.

3. Xem bảng kết quả

Nếu xuất hiện 1 hàng gồm toàn kí hiệu a_1, a_2,…a_n thì phép tách δ là có kết nối không mất mát thông tin (ngược lại, δ không là phép tách không mất thông tin).

Ví dụ 1. α = <U, F> , U = SAIP, F = {S $\rightarrow$A, SI $\rightarrow$P}

δ = [U1, U2] , U1 = SA , U2 = SIP

Ví dụ 2. U = ABCDE, F = {C $\rightarrow$D, A $\rightarrow$C, B $\rightarrow$C, DE $\rightarrow$C, CE $\rightarrow$A}

δ = [AC, BC , CD, DEC, CEA)

Ví dụ 3. U = ABCDE, F = {C $\rightarrow$D, A $\rightarrow$C, B $\rightarrow$C, , DE $\rightarrow$C, CE $\rightarrow$A}

δ = [AC, BC , CD, DEC, CEA, BE]

- **Định lý.** δ = [U1, U2] là phép tách của α = <U, F>
  δ là phép tách có kết nối không mất thông tin nếu
  $U1 \cap U2 \rightarrow U1 \setminus U2 \in F+$
  (hoặc $U1 \cap U2 \rightarrow U2 \setminus U1 \in F+$)

- Ví dụ 1. α = <U, F> , U = SAIP, F = {S $\rightarrow$A, SI $\rightarrow$P}
  δ = [U1, U2] , U1 = SA , U2 = SIP
  $U1 \cap U2 = \{S\}$; $U1 \setminus U2 = \{A\}$

DẠNG CHUẨN

- Dạng chuẩn 1 (1NF)
- Dạng chuẩn 2 (2NF)
- Dạng chuẩn 3 (3NF)
- Dạng chuẩn Boyce−Codd (BCNF)
- Dạng chuẩn 4
- Dạng chuẩn 5

Dạng chuẩn 1 (1NF-The first normal form)

- Lược đồ quan hệ α =<U, F> ở dạng chuẩn 1NF nếu mỗi thuộc tính $A \in U$ có miền trị chỉ chứa các giá trị nguyên tố.
- Thuộc tính ‘không nguyên tố’ có thể là thuộc tính đa trị (multivalued attribute) hoặc thuộc tính phức/ghép (composite attribute) hoặc kết hợp 2 loại thuộc tính trên

Ví dụ về lược đồ không ở 1NF

Thuộc tính đa trị

- Ví dụ 1

S	(S#,	SNAME,	CITY)
	S1	Smith	{London, Paris}
	S2	Jones	{Rome}

Thuộc tính ghép

- Ví dụ 2

SHHSX	CQD	DCHI	SX(SHMH, SL)
S_1	A	Hà Nội	(H_1, 250)
S_1	A	Hà Nội	(H_2, 300)
S_2	A	Hà Nội	(H_1, 100)
S3	B	Hải Phòng	(H_3, 200)

Đưa quan hệ chưa chuẩn hoá về dạng chuẩn 1

- Ví dụ 1

1. Tách làm 2 quan hệ:

   S (S#, SNAME) có khoá là S#

   S_CITY (S#, CITY) có khoá là {S#, CITY}

2. Quan hệ S trở thành:

   S (S#, CITY, SNAME)

S#	CITY	SNAME
S1	London	Smith
S1	Paris	Smith
S2	Rome	Jones

   - Khoá: {S#, CITY}
   - Nhược điểm: dư thừa dữ liệu

Đưa quan hệ chưa chuẩn hoá về dạng chuẩn 1

3. Biến đổi quan hệ S thành:

S (S#,	CITY1,	CITY2,	CITY3,	SNAME)
S1	London	Paris		Smith
S2	Rome			Jones

Đưa quan hệ chưa chuẩn hoá về dạng chuẩn 1

- Cách 1: Biến đổi thành một quan hệ phẳng (flat realation)

  HSX(SHHSX, CQD, DCHI, SHMH, SL)

  - Nhược điểm: Dư thừa dữ liệu

- Cách 2: Đưa các thuộc tính lồng thành một quan hệ mới và được kết quả CSDL gồm 2 quan hệ sau:

  HSX(SHHSX, CQD, DCHI)

  SX(SHHSX, SHMH, SL)

Dạng chuẩn 1

- Lược đồ α = <U, F>,

  U = {SHHSX, CQD, DCHI, SHMH, SL}

  F = {SHHSX $\rightarrow$ DCHI, DCHI $\rightarrow$CQD, {SHHSX, SHMH}$\rightarrow$SL}

- Nhận xét: - α ở 1NF

  - Khoá duy nhất là {SHHSX, SHMH}

- Các thuộc tính không khoá

  Nα = {CQD, DCHI, SL}

- Dư thừa dữ liệu dẫn đến dị thường cập nhật, loại bỏ, bổ sung.

SHHSX	CQD	DCHI	SHMH	SL
S_1	A	Hà Nội	H_1	250
S_1	A	Hà Nội	H_2	300
S_2	A	Hà Nội	H_1	100
S3	B	Hải Phòng	H_3	200

- F = {SHHSX $\rightarrow$ DCHI, DCHI $\rightarrow$CQD, {SHHSX, SHMH}$\rightarrow$SL}
- Khoá duy nhất: {SHHSX, SHMH}
- Các thuộc tính không khoá: Nα = {CQD, DCHI, SL}
- Dư thừa dữ liệu dẫn đến dị thường cập nhật, loại bỏ, bổ sung.

DẠNG CHUẨN 2 (2NF)

- **Định nghĩa.** Phụ thuộc đầy đủ (full functional dependency)
- Cho phụ thuộc hàm X $\rightarrow$Y, Y được gọi là phụ thuộc đầy đủ vào X (hay X $\rightarrow$Y là phụ thuộc hàm đầy đủ) nếu !$\exists$ Z $\subset$ X mà Z $\rightarrow$Y.
- **Định nghĩa .** Lược đồ α = <U, F> ở 2NF nếu:

  + α ở 1NF

  + Mọi thuộc tính không khoá phải phụ thuộc đầy đủ vào khoá

Ví dụ.

- Lược đồ α = <U, F>,

  U = {SHHSX, CQD, DCHI, SHMH, SL}

  F = {SHHSX $\rightarrow$ DCHI, DCHI $\rightarrow$CQD,

  {SHHSX, SHMH}$\rightarrow$SL}

- Tách thành 2 lược đồ ở 2NF:

  \+ α1 =<U1, F1>,

  U1 = {SHHSX, CQD, DCHI},

  F1 = {SHHSX $\rightarrow$DCHI, DCHI $\rightarrow$ CQD}

  \+ α2 = <U2, F2>,

  U2 = {SHHSX, SHMH, SL}

  F2 = {{SHHSX, SHMH} $\rightarrow$ SL}

- Nhận xét: $\alpha 1$ còn dư thừa dữ liệu: ở những địa chỉ giống nhau thì CQD giống nhau, gây khó khăn khi cập nhật dữ liệu.

SHHSX	**CQD**	**DCHI**
S_1	A	Hà Nội
S_1	A	Hà Nội
S_2	A	Hà Nội
S3	B	Hải Phòng

DẠNG CHUẨN 3 (3NF)

- **Định nghĩa.** (Phụ thuộc bắc cầu)

  Cho α = <U, F>, A $\in$ U, X $\subseteq$ U. Nói A phụ thuộc bắc cầu vào X nếu $\exists$ Y $\subseteq$ U:

  1. X $\rightarrow$Y
  2. Y $\rightarrow$ A
  3. Y $\not\rightarrow$ X
  4. A $\notin$ XY

- **Định nghĩa.** Lược đồ α = <U, F> ở 3NF nếu:

  + α ở 1NF

  + Mọi thuộc tính không khoá không phụ thuộc bắc cầu vào khoá.

$\alpha 1 = \langle U1, F1 \rangle$,

$U1 = \{SHHSX, CQD, DCHI\}$,

$F1 = \{SHHSX \rightarrow DCHI, DCHI \rightarrow CQD\}$

$\alpha 2 = \langle U2, F2 \rangle$,

$U2 = \{SHHSX, SHMH, SL\}$

$F2 = \{\{SHHSX, SHMH\} \rightarrow SL\}$

- $\alpha 2$ ở 3NF.
- $\alpha 1$ không ở 3NF (CQD phụ thuộc bắc cầu vào khoá)
- Tách $\alpha 1$ thành: $\alpha 11$ và $\alpha 12$

$U_{11} = \{SHHSX, DCHI\}$, $U_{12} = \{CQD, DCHI\}$

$F_{11} = \{SHHSX \rightarrow DCHI\}$, $F_{12} = \{DCHI \rightarrow CQD\}$

Lược đồ

$\alpha = <U, F>$

SHHSX	CQD	DCHI	SHMH	SL
S_1	A	Hà Nội	H_1	250
S_1	A	Hà Nội	H_2	300
S_2	A	Hà Nội	H_1	100
S3	B	Hải Phòng	H_3	200

Lược đồ

$\alpha_{11} = <U_{11}, F_{11}>$

SHHSX	DCHI
S_1	Hà Nội
S_2	Hà Nội
S3	Hải Phòng

Lược đồ

$\alpha_{12} = <U_{12}, F_{12}>$

DCHI	CQD
Hà Nội	A
Hải Phòng	B

Lược đồ

$\alpha_2 = <U_2, F_2>$

SHHSX	SHMH	SL
S_1	H_1	250
S_1	H_2	300
S_2	H_1	100
S3	H_3	200

- Xét lược đồ quan hệ α=<U, F>, U =CSZ,

  F ={Z$\rightarrow$C, CS $\rightarrow$Z}

  (trong đó C: City, S: Street, Z: Zipcode)
- Khoá của lược đồ: CS, SZ
- Nα = ϕ
- α ở 2NF và 3NF
- Còn dư thừa dữ liệu, chẳng hạn như:

C	S	Z
c_1	s_1	z_1
c_1	s_2	z_1

Dạng chuẩn Boyce - Codd (BCNF) [1974]

- Lược đồ α = <U, F> ở BCNF nếu có $X \subseteq U$, $A \in U$ và $A \notin X$ mà $X \rightarrow A \in F+$ thì X phải là siêu khoá.

- (Có thể xem 3NF là dạng đơn giản của BCNF vì dạng phát biểu tổng quát của 3NF là: $\forall X \rightarrow A \in F^+$ thì hoặc X là siêu khoá hoặc A là thuộc tính khoá. Trong dạng chuẩn BCNF không cho phép A là thuộc tính khoá).

- Bài tập. Cho lược đồ α = <U, F>,

  U = SIDM, trong đó:

  S: Store I: Item

  D: Department M: Manager

  F = $\{SI \rightarrow D, SD \rightarrow M\}$
- Kiểm tra dạng chuẩn của lược đồ α

- Xét quan hệ R (CTX), trong đó:
- C: Course ; T: Teacher ; T: Text
- Giả thiết: - Mỗi môn học có nhiều giáo viên dạy
  - Mỗi môn học sử dụng nhiều loại SGK
  - Giáo viên và SGK không phụ thuộc nhau

R

C	T	X
CSDL	GS. Hùng	Nhập môn CSDL
CSDL	GS. Hùng	Nguyên lý các hệ CSDL
CSDL	GS. Dũng	Nhập môn CSDL
CSDL	GS. Dũng	Nguyên lý các hệ CSDL

Phụ thuộc đa trị

- **Định nghĩa.** Cho lược đồ quan hệ α = <U, F>, X, Y $\subseteq$ U

  Z = U - XY

- Quan hệ R trên α thoả phụ thuộc đa trị (MVD)

  X $\rightarrow\rightarrow$Y nếu với bất kỳ t1, t2 $\in$ R mà t1.X = t2.X

  thì $\exists$ t3 $\in$ R: t1.Y = t3.Y , t1.X = t3.X, t2.Z = t3.Z.

- Do tính đối xứng của t1, t2, còn $\exists$ t4 $\in$ R: t4.X = t2.X, t4.Y = t2.Y, t4.Z = t1.Z

- Kí hiệu X$\rightarrow\rightarrow$ Y|Z

Ví dụ.

	R	(C	T	H	R	S	G)
t_1		CS101	D	M9	22	K	B
t_4		CS101	D	W9	33	K	B
		CS101	D	F9	22	K	B
t_3		CS101	D	M9	22	L	C
t_2		CS101	D	W9	33	L	C
		CS101	D	F9	22	L	C

Quan hệ R trên thoả C$\rightarrow\rightarrow$HR

t1 = (CS101, D, M9, 22, K, B)
t2 = (CS101, D, W9, 33, L, C)
t3 = (CS101, D, M9, 22, L, C)
t4 = (CS101, D, W9, 33, K, B)

Một số phụ thuộc đa trị khác được thoả trên R: C$\rightarrow\rightarrow$SG, HR $\rightarrow\rightarrow$SG

	R	T	H	R	S	G
	(C					)
t_1	CS101	D	M9	22	K	B
t_4	CS101	D	W9	33	K	B
	CS101	D	F9	22	K	B
t_3	CS101	D	M9	22	L	C
t_2	CS101	D	W9	33	L	C
	CS101	D	F9	22	L	C

Quan hệ R không thoả: C$\rightarrow\rightarrow$H, C$\rightarrow\rightarrow$R

chẳng hạn, nếu thoả C $\rightarrow\rightarrow$H thì với t1, t2 trên phải tìm được
t3 = (CS101, D, M9, 33, L, C) $\in$ R,
nhưng trong CSDL có:
t = (CS101, D, M9, 22, L, C)

Nếu có cả 2 bộ này trong R thì mâu thuẫn với HS$\rightarrow$R

Dạng chuẩn 4 (4NF)

- Lược đồ quan hệ α =<U, D> ở 4NF nếu với mỗi $X \rightarrow\rightarrow Y$, $Y \not\subseteq X$, $XY \neq U$ thì X là siêu khoá của α.

Phụ thuộc kết nối và dạng chuẩn 5

- **Định nghĩa.** Một phụ thuộc kết nối (JD-Join Dependency) kí hiệu là JD(U1,…,Un) của lược đồ quan hệ α trên U xác định một ràng buộc trên các quan hệ R của α sao cho: $\forall$R: R= R[U1]*R[U2]*….*R[Un]

- **Định nghĩa.** Lược đồ α ở dạng chuẩn 5 đối với tập F các FD, MVD, JD, nếu mỗi JD(U1,…,Un) trong F^+, U_i đều là siêu khoá của α.

CHUẨN HOÁ THÀNH BCNF

- **Bổ đề**. Cho lược đồ quan hệ α = <U, F>,

  δ =[U1, U2,…,Uk] là phép tách không mất thông tin của α, δi = [U_{i1}, U_{i2}…, U_{it}] là phép tách không mất thông tin của lược đồ αi = <Ui, Fi>.

- Khi đó phép tách

  δ =[U_1, U_2, …U_{i-1},U_{i1}, U_{i2},…, U_{it}, U_{i+1},…, Uk] của lược đồ α là không mất thông tin.

- **Nhận xét**: Nếu lược đồ α = <U, F> chưa ở BCNF

  $\exists$ X $\rightarrow$A $\in$ F^+ mà X không là khoá.

- Đặt XA =U1 , U \ {A} =U2

  Ta có: U1 $\cap$ U2 = X, U1 \ U2 = A và X $\rightarrow$A $\in$ F+

- Theo định lý 3.2, phép tách δ =[U1, U2] của α là không mất thông tin. Dễ thấy α1 = <U1, F1> ở BCNF.

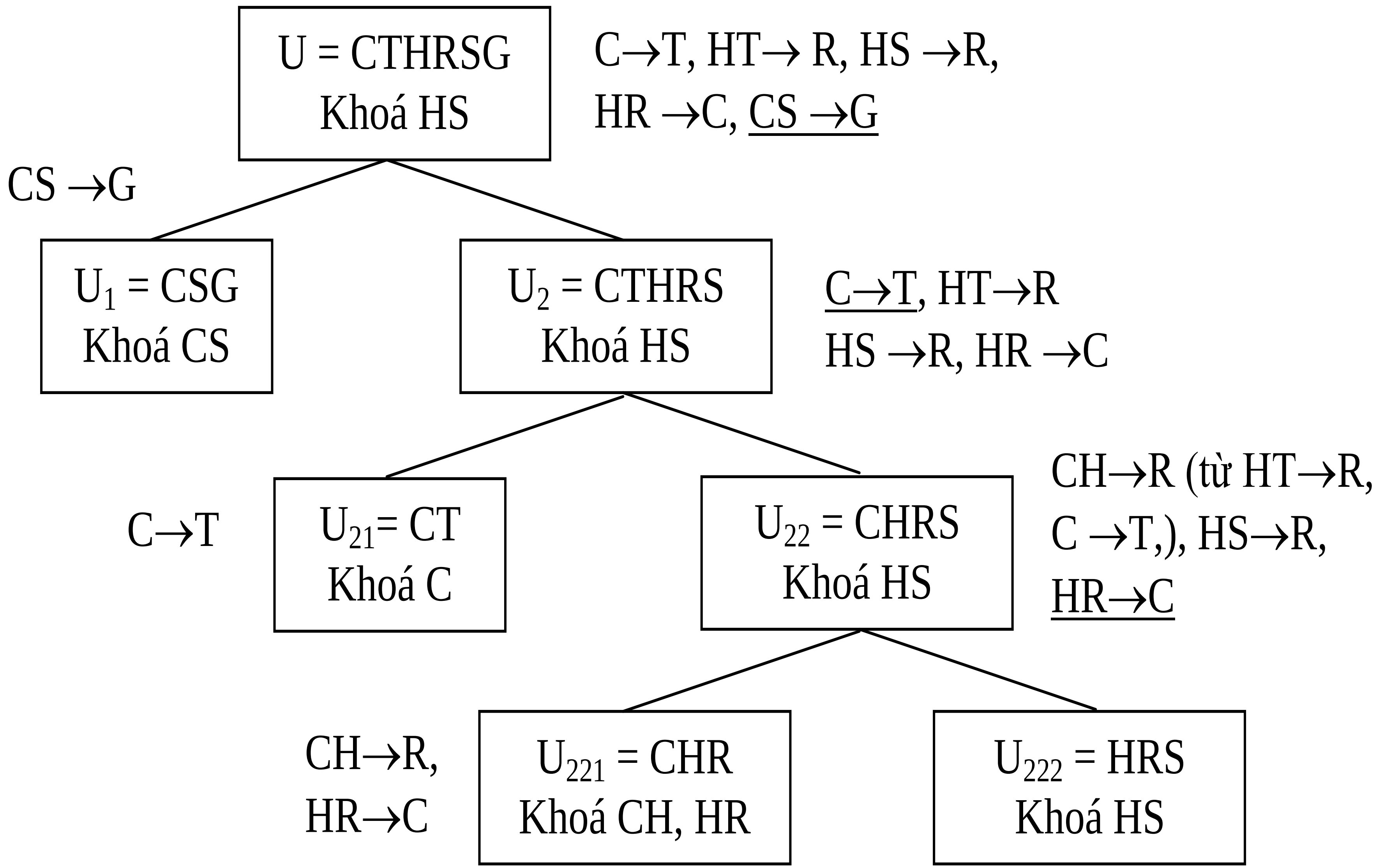
- **Thuật toán 3.6**. Tách lược đồ quan hệ thành các lược đồ con ở BCNF.
- Input Lược đồ quan hệ α = <U, F>
- Output Một phép tách không mất thông tin của lược đồ α sao cho các lược đồ con ở BCNF
- Method

1. Nếu α không ở BCNF thì tách đôi α thành α1, α2 theo cách:

   Chọn $X \rightarrow A \in F$ mà X không là siêu khoá.

   Khi đó phép tách là δ =[U1, U2],

   trong đó U1 = XA, U2 = U \ {A}
2. Tiếp tục quá trình trên đối với α1và α2 cho đến khi các lược đồ con nhận được đều ở BCNF.


U = CTHRSG
Khoá HS
C→T, HT→ R, HS →R,
HR →C, CS →G
CS →G
U1 = CSG
Khoá CS
U2 = CTHRS
Khoá HS
C→T, HT→R
HS →R, HR →C
C→T
U21= CT
Khoá C
U22 = CHRS
Khoá HS
CH→R (từ HT→R,
C →T,), HS→R,
HR→C
CH→R,
HR→C
U221 = CHR
Khoá CH, HR
U222 = HRS
Khoá HS

CHUẨN HOÁ THÀNH 3NF

- Input α = <U, F>
- Output Một phép tách α bảo toàn tập phụ thuộc hàm, có các lược đồ con ở 3NF.
- Method

1. Tính G là phủ tối thiểu của F.
2. Với mỗi X là vế trái của một phụ thuộc hàm trong G, tạo lược đồ quan hệ trên tập thuộc tính $XA_1A_2\ldots A_k$, trong đó $X \rightarrow A_1, \ldots, X \rightarrow A_k \in G$