*Tạp chí Tin học và Điều khiển học, T. 17, S. 1 (2001), 89–96*

# MỘT SỐ VẤN ĐỀ ĐỐI VỚI PHỤ THUỘC KẾT NỐI VÀ DẠNG CHUẨN CHIẾU - KẾT NỐI

PHẠM QUANG TRUNG

**Abstract.** Join dependency (JD) and further normal forms play an important role in the theory of normalize. In this paper we prove new properties on JD implication from a given set of dependencies and present special property of a relational scheme is in project-join normal form (PJNF).

**Tóm tắt.** Phụ thuộc kết nối (join dependency - JD) và dạng chuẩn bậc cao có vai trò quan trọng trong lý thuyết chuẩn hóa. Trong bài báo này sẽ chứng minh một số tính chất về sự suy dẫn các JD từ một tập phụ thuộc cho trước và trình bày tính chất đặc trưng của lược đồ quan hệ ở dạng chuẩn chiếu - kết nối (project-join normal form - PJNF).

## 1. MỞ ĐẦU

**Các ký hiệu**: Quan hệ $R$ trên tập thuộc tính $U$ được kí hiệu là $R(U)$, hợp của hai tập thuộc tính $X, Y$ được viết là $XY$; phép kết nối được kí hiệu bằng dấu *.

Phần này chỉ nêu một số khái niệm và kết quả liên quan, bạn đọc quan tâm chi tiết hơn đề nghị xem [2 - 5].

**Định nghĩa 1.** Cho $R(A_1, A_2, ..., A_n)$ là một lược đồ quan hệ, cho $X$ và $Y$ là các tập con của $\{A_1, A_2, ..., A_n\}$. Chúng ta nói $X \to Y$ (đọc là "*$X$ xác định hàm $Y$*" hay "*$Y$ phụ thuộc hàm vào $X$*") nếu với mọi quan hệ $r$ là thể hiện của $R$, thì trong $r$ không thể có hai bộ trùng nhau trên các thành phần của mọi thuộc tính trong tập $X$ mà lại không trùng nhau trên một hay nhiều hơn các thành phần của các thuộc tính của tập hợp $Y$.

- Quan hệ $r$ *thỏa* phụ thuộc hàm (functional dependency - FD) $X \to Y$, nếu với mọi cặp bộ $\mu, \nu$ trong $r$ sao cho $\mu[X] = \nu[X]$ thì $\mu[Y] = \nu[Y]$ cũng đúng. Nếu $r$ không thỏa $X \to Y$, thì $r$ *vi phạm* phụ thuộc đó.

- Cho $F$ là tập phụ thuộc hàm của lược đồ quan hệ $R$ và cho $X \to Y$ là một phụ thuộc hàm. Chúng ta nói $F$ *suy diễn logic* ra $X \to Y$, viết là $F \models X \to Y$, nếu với mọi quan hệ $r$ của $R$ mà thỏa các phụ thuộc hàm trong $F$ thì cũng thỏa mãn $X \to Y$.

**Định nghĩa 2.** Bao đóng của tập phụ thuộc hàm $F$, ký hiệu là $F^+$, là tập các phụ thuộc hàm được suy diễn logic từ $F$, nghĩa là: $F^+ = \{X \to Y \mid F \models X \to Y\}$.

**Định nghĩa 3.** Cho lược đồ quan hệ $R$ với tập phụ thuộc hàm $F$, cho $X$ là một tập con của $R$, tập $X$ được gọi là *khóa* (key) của lược đồ quan hệ $R$ nếu: 1) $X \to R \in F^+$; 2) Với $\forall Y \subset X$ thì $Y \not\to R$. Tập $X$ nếu chỉ thỏa mãn điều kiện 1) nêu trên được gọi là một *siêu khóa* (superkey). Các khóa (hay siêu khóa) được liệt kê rõ ràng cùng với lược đồ quan hệ được gọi là các *khóa được chỉ định* (designated key).

**Định nghĩa 4.** Hai tập phụ thuộc hàm $F$ và $G$ trên lược đồ $R$ là *tương đương* (equivalent), ký hiệu là $F \equiv G$, nếu $F^+ = G^+$. Nếu $F \equiv G$ thì $F$ là một *phủ* (cover) của $G$.

Phụ thuộc hàm $X \to Y \in F$ là *dư thừa* nếu $F - \{X \to Y\} \models X \to Y$.

**Định nghĩa 5.** Hai tập thuộc tính $X$ và $Y$ là *tương đương* với nhau trên tập phụ thuộc hàm $F$, nếu $F \models X \to Y$ và $F \models Y \to X$ (ký hiệu là $X \leftrightarrow Y$).

**Định nghĩa 6.** *Phụ thuộc hàm phức hợp* (compound functional dependency - CFD) có dạng $(X_1, X_2, ..., X_k) \to Y$, trong đó $X_1, X_2, ..., X_k$ và $Y$ là các tập con khác nhau của lược đồ $R$.

Quan hệ $r(R)$ thỏa phụ thuộc hàm phức hợp $(X_1, X_2, ..., X_k) \to Y$ nếu nó thỏa các phụ thuộc hàm $X_i \to X_j$ và $X_i \to Y$, với $1 \leq i, j \leq k$. Trong phụ thuộc hàm phức hợp này, $(X_1, X_2, ..., X_k)$ được gọi là vế trái, $X_1$, $X_2$, ..., $X_k$ là các tập trái, $Y$ là vế phải.

CFD là cách viết rút gọn hơn tập các phụ thuộc hàm có các vế trái tương đương. Trong trường hợp nếu $Y = \emptyset$, có dạng đặc biệt của $CFD$ là $(X_1, X_2, ..., X_k)$.

**Định nghĩa 7.** Tập $F$ được gọi là *phủ* của $G$ nếu $F \equiv G$, trong đó $F$ và $G$ bao gồm hoặc là tập các phụ thuộc hàm, tập các phụ thuộc hàm phức hợp, hoặc là tập hợp chỉ gồm một loại phụ thuộc.

**Định nghĩa 8.** Tập phụ thuộc hàm $F$ được gọi là *tập đặc trưng* (characteristic set) đối với phụ thuộc hàm phức hợp $(X_1, X_2, ..., X_k) \to Y$, nếu $F \equiv \{(X_1, X_2, ..., X_k) \to Y\}$. Nếu mỗi tập hợp trái của phụ thuộc hàm phức hợp được sử dụng với tư cách là vế trái của phụ thuộc hàm đúng một lần (nghĩa là $F$ có dạng $\{X_1 \to Y_1, X_2 \to Y_2, ..., X_k \to Y_k\}$), thì $F$ được gọi là *tập đặc trưng tự nhiên* (natural characteristic set) đối với phụ thuộc hàm phức hợp đã cho.

**Định nghĩa 9.** Tập phụ thuộc hàm phức hợp $F$ được gọi là *dạng vành* (annular), nếu không có các tập trái $X$ và $Z$ trong các vế trái khác nhau, mà $X \leftrightarrow Z$ trên $F$.

**Định nghĩa 10.** Cho lược đồ quan hệ $R$ với tập phụ thuộc hàm $F$. Cho tập thuộc tính: $X \subseteq R$, thuộc tính $A \in R$. Ta nói thuộc tính $A$ *phụ thuộc bắc cầu* (transitively dependent) vào $X$ trên $R$ nếu tồn tại $Y \subseteq R$ sao cho $X \to Y$ và $Y \to A$ nhưng $Y \not\to X$ với $A \notin XY$.

**Định nghĩa 11.** Một lược đồ quan hệ $R$ với tập phụ thuộc hàm $F$ được gọi là *ở dạng chuẩn thứ ba* (third normal form - 3NF) nếu không có thuộc tính không khóa phụ thuộc bắc cầu vào khóa của $R$. Một lược đồ cơ sở dữ liệu $\mathcal{R}$ là ở dạng chuẩn thứ ba nếu mọi lược đồ quan hệ trong $\mathcal{R}$ là ở $3NF$.

**Chuẩn hóa bằng phép tách** (normalization through decomposition)

Cho lược đồ $R$ và tập phụ thuộc hàm $F$. *Phép tách một lược đồ quan hệ* là việc thay thế một lược đồ $R$ bằng tập các lược đồ con $\rho = \{R_1, R_2, ...R_k\}$ (các $R_i$ không nhất thiết phải rời nhau) sao cho: a) $R_i \subseteq R$, $i = 1, 2, ..., k$; b) $R = R_1 R_2 \ldots R_k$.

- Cho lược đồ quan hệ $R$. Phép tách $\rho$ là *phép tách có kết nối không mất thông tin* (lossless join decomposition) nếu với mọi quan hệ $r$ trên $R$ mà thỏa $F$, ta có: $r = \pi_{R_1}(r) * \pi_{R_2}(r) * ... * \pi_{R_k}(r)$. Tức là quan hệ $r$ là kết nối tự nhiên của các hình chiếu của $r$ trên các $R_i$.

- Phép tách $\rho = \{R_1, R_2, ..., R_k\}$ được gọi là *phép tách bảo toàn (preserve) tập phụ thuộc* $F$, nếu: $G = \pi_{R_1}(F) \cup \pi_{R_2}(F) \cup ... \cup \pi_{R_k}(F)$ suy dẫn ra $F$ (trong đó: $\pi_{R_i} = \{X \to Y \in F^+ \mid X, Y \subseteq R_i\}$).

Có hai kỹ thuật chính để chuẩn hóa lược đồ quan hệ bằng việc tách (decomposition) là *phép phân tích* (analysis) và *phép tổng hợp* (synthesis).

• Chuẩn hóa bằng phép phân tích

Với đặc trưng chính đảm bảo tiêu chuẩn tính kết nối không mất thông tin của các lược đồ thành phần là kỹ thuật thông dụng để chuẩn hóa lược đồ quan hệ với các dạng chuẩn khác nhau. Nếu một lược đồ quan hệ không thỏa dạng chuẩn mong muốn vì một phụ thuộc nào đó thì nó được tách thành hai hoặc một số các lược đồ quan hệ căn cứ vào phụ thuộc này. Mỗi một lược đồ được tách thừa hưởng các ràng buộc thích hợp. Việc tách được lặp lại cho đến khi tất cả các lược đồ đã được chuẩn hóa.

• Chuẩn hóa 3NF bằng phép tổng hợp (normalization through synthesis)

Phần này chỉ giới thiệu phép tổng hợp sử dụng phủ dạng vành.

Quy ước: Ký hiệu $\mathcal{R} = \{R_1, R_2, ..., R_k\}$ là tập lược đồ quan hệ nhận được bởi một thuật toán chuẩn hóa.

**Thuật toán TH-3NF**

VÀO: Tập $U$, tập phụ thuộc hàm $F$ trên $U$.

RA: Tập lược đồ quan hệ ở dạng chuẩn ba, bảo toàn $F$, có kết nối không mất thông tin, có số lượng lược đồ là ít nhất.

PHƯƠNG PHÁP:

1) Bổ sung thuộc hàm $U \to @$ vào tập phụ thuộc hàm $F$ (trong đó @ là tên "thuộc tính giả" không thuộc $U$). Rút gọn về trái của các phụ thuộc hàm. Loại bỏ các phụ thuộc hàm dư thừa. Kết quả của bước này nhận được tập $F'$.

2) Tạo tập phủ dạng vành $G$ đối với $F'$.

3) Tạo tập phụ thuộc hàm đặc trưng tự nhiên $G_1$ tương đương với tập $G$. Gọi $G_2$ là tập $G_1$ đã được rút gọn về phải.

4) Tạo tập phủ dạng vành $G_3$ đối vơi tập $G_2$. Tương ứng với từng phụ thuộc hàm phức hợp trong $G_3$, xây dựng lược đồ quan hệ có tập thuộc tính là tất cả các thuộc tính xuất hiện trong mỗi phụ thuộc hàm phức hợp, tập các khóa chỉ định của mỗi lược đồ tương ứng là bao gồm các tập trái của mỗi phụ thuộc hàm phức hợp.

5) Kết quả là tập lược đồ được xây dựng ở bước 4). Thuộc tính giả @ được loại khỏi lược đồ chứa @.

**Định lý 1.** [4] *Lược đồ cơ sở dữ liệu $\mathcal{R} = (R_1, R_2, ..., R_k)$ được tổng hợp bằng Thuật toán TH-3NF từ tập các phụ thuộc hàm $F$ thỏa mãn các tính chất sau đây:*

1) *Đối với mỗi lược đồ bất kỳ $R_i$ thuộc $\mathcal{R}$, mỗi khóa chỉ định của $R_i$ là một khóa*

2) *Lược đồ cơ sở dữ liệu $\mathcal{R}$ bảo toàn tập thuộc hàm $F$.*

3) *Lược đồ cơ sở dữ liệu $\mathcal{R}$ bao gồm các lược đồ thành phần là ở dạng chuẩn ba.*

4) *Kết nối các lược đồ con của lược đồ cơ sở dữ liệu $\mathcal{R}$ là không mất thông tin.*

5) *Ngoài ra, không tồn tại lược đồ cơ sở dữ liệu nào khác có số lượng lược đồ con ít hơn thỏa mãn các tính chất nêu trên.*

## Phụ thuộc đa trị, phụ thuộc kết nối và dạng chuẩn chiếu - kết nối

**Định nghĩa 12.** Cho $R$ là một lược đồ quan hệ, cho $X$ và $Y$ là các tập con của $R$, và $Z = R - (XY)$. Quan hệ $r(R)$ thỏa *phụ thuộc đa trị* (multivalued dependency - MVD) $X \longrightarrow\!\!\!\!\!\rightarrow Y$ nếu với hai bộ bất kỳ $t_1$ và $t_2$ trong $r$ mà $t_1(X) = t_2(X)$, tồn tại bộ $t_3$ trong $r$ mà $t_3(X) = t_1(X)$, $t_3(Y) = t_1(Y)$ và $t_3(Z) = t_2(Z)$.

Ký hiệu $\Sigma$ là tập các phụ thuộc hàm và phụ thuộc hàm đa trị trên tập thuộc tính $U$.

**Định lý 2.** [2] *Cho $r$ là một quan hệ trên lược đồ $R$, và cho $X$, $Y$ và $Z$ là các tập con của $R$ mà $Z = R - (XY)$. Quan hệ $r$ thỏa phụ thuộc đa trị $X \longrightarrow\!\!\!\!\!\rightarrow Y$ nếu và chỉ nếu $r$ tách có kết không mất thông tin thành các lược đồ quan hệ $R_1 = XY$ và $R_2 = XZ$.*

**Định lý 3.** [5] *Cho $R$ là một lược đồ quan hệ và $\rho = (R_1, R_2)$ là phép tách $R$. Cho $\Sigma$ là tập phụ thuộc hàm và phụ thuộc đa trị trên $R$. Khi đó $\rho$ là phép tách có kết nối không mất thông tin đối với $\Sigma$ nếu và chỉ nếu: $(R \cap R_2) \longrightarrow\!\!\!\!\!\rightarrow (R_1 - R_2)$, hoặc tương đương, nhờ luật bù, $(R_1 \cap R_2) \longrightarrow\!\!\!\!\!\rightarrow (R_2 - R_1)$.*

**Định nghĩa 13.** Cho $\mathcal{R} = \{R_1, R_2, ..., R_p\}$ là một tập lược đồ quan hệ trên $U$. Một quan hệ $r(R)$ thỏa phụ *thuộc kết nối* (JD) $*[R_1, R_2, ..., R_p]$ nếu $r$ được tách có kết nối không mất thông tin thành $R_1, R_2, ..., R_p$. Tức là: $r = \pi_{R_1}(r) * \pi_{R_2}(r) * ... * \pi_{R_p}(r)$.

Ta cũng viết $*[R_1, R - 2, ..., R_p]$ là $*[\mathcal{R}]$.

Điều kiện cần để một quan hệ $r(U)$ thỏa JD $*[R_1, R_2, ..., R_p]$ là $U = R_1R_2...R_p$. MVD là một trường hợp riêng của JD. Một quan hệ $r(R)$ thỏa MVD $X \longrightarrow\!\!\!\!\!\rightarrow Y$ nếu và chỉ nếu $r$ được tách có kết nối không mất thông tin thành $XY$ và $XZ$, với $Z = R - (XY)$. Điều kiện này chính là JD $*[XY, XZ]$.

Một JD $*[R_1, R_2, ..., R_p]$ là *tầm thường* nếu mọi quan hệ $r(R)$ đều thỏa nó. Một JD $*[R_1, R_2, ..., R_p]$ là *áp dụng được* vào lược đồ quan hệ $R$ nếu $R = R_1R_2...R_p$.

**Định nghĩa 14.** Cho $R$ là một lược đồ quan hệ và $\Sigma$ là tập các phụ thuộc hàm và phụ thuộc

kết nối trên $R$. Lược đồ quan hệ $R$ là *ở dạng chuẩn chiếu - kết nối* (PJNF) nếu đối với mỗi JD $*[R_1, R_2, ..., R_p]$ suy dẫn từ $\Sigma$ và áp dụng được vào $R$, thì JD đó là tầm thường hoặc mỗi $R_i$ là một siêu khóa đối với $R$. Một lược đồ cơ sở dữ liệu $\mathcal{R}$ là ở PJNF đối với $\Sigma$ nếu mỗi lược đồ quan hệ $R$ thuộc $\mathcal{R}$ là ở PJNF đối với $\Sigma$.

**Bảng** (tableau)

Một *bảng* là một ma trận gồm tập các dòng. Mỗi cột trong bảng tương ứng với một thuộc tính trong $R$. Mỗi dòng gồm các biến được viết ra từ tập $V$, là hợp phân biệt của hai tập $V_d$ và $V_n$: a) $V_d$ là tập các *biến được đánh dấu* (distinguished variable - dv), một biến ứng với mỗi thuộc tính: nếu $A$ là một thuộc tính được xét, thì $v_A$ là một dv tương ứng. b) $V_n$ là tập các *biến không được đánh dấu* (nondistinguished variable - ndv): ký hiệu là $n_1, n_2, ..., n_k, ...$

Một biến bất kỳ bị hạn chế xuất hiện nhiều nhất trong một cột, một biến được đánh dấu phải xuất hiện trong mỗi cột, và trong một cột chỉ có thể có một biến đánh dấu.

Một *ước lượng* (valuation) là một hàm $\rho$ ánh xạ mỗi biến trong bảng $T$ với một phần tử trong $\text{dom}(A)$, trong đó $A$ là cột mà biến xuất hiện trong đó. Đây là sự mở rộng hàm từ bảng $T$ tới một quan hệ trên $R$ như sau, nếu $\omega = \langle \nu_1, \nu_2, ..., \nu_n \rangle$ là một dòng của $T$, thì $\rho(\omega)$ là bộ $\langle \rho(\nu_1), \rho(\nu_2), ..., \rho(\nu_n) \rangle$ và $\rho(T) = \{\rho(\omega) \mid \omega \text{ là một dòng trong } T\}$.

Cho $\Sigma$ là tập các MVD và FD (một MVD bất kỳ được thể hiện như một JD). *Săn đuổi* (hay *theo dõi* - chase) là kết quả của việc áp dụng các phép biến đổi sau đây vào bảng $T$ cho đến khi không có thể làm biến đổi thêm:

- $F$-qui tắc ($F$-rule): Với mỗi $FD \rightarrow A$ trong $\Sigma$, có một $F$-qui tắc biến đổi bảng như sau. Giả sử bảng $T$ có các dòng $\omega_1$ và $\omega_2$, trong đó $\omega_1[X] = \omega_2[X]$ và cho $\nu_1 = \omega_1[A]$ và $\nu_2 = \omega_2[A]$. Nếu $v_1$ hoặc $v_2$ là biến được đánh dấu và cái kia thì không, thì biến không được đánh dấu được đổi thành biến được đánh dấu. Nếu cả hai là các biến không được đánh dấu, thì biến có chỉ số dưới lớn hơn được thay bằng biến có chỉ số dưới nhỏ hơn.
- $J$-qui tắc ($J$-rule): Cho $*[R_1, R_2, ..., R_p]$ là một $JD$ trong $\Sigma$. Nếu có một dòng $\omega$ sao cho $\omega[R_1] \in T[R_1], ..., \omega[R_p] \in T[R_p]$, $\omega$ được bổ sung vào $T$.

Ký hiệu $\text{chase}_\Sigma(T)$ là bảng kết quả nhận được từ việc áp dụng $F$-qui tắc và $J$-qui tắc đối với mọi phụ thuộc trong $\Sigma$ cho đến khi không có thể thay đổi thêm bảng được nữa. Có thể chứng tỏ rằng [3] chase luôn kết thúc và bảng kết quả là duy nhất, không phụ thuộc vào thứ tự áp dụng các qui tắc để đặt lại tên cho các biến không được đánh dấu.

**Bổ đề 1.** [3] *Cho $T_x$ là bảng được cấu trúc bao gồm hai dòng: một dòng được ký hiệu là $\omega_d$, gồm mọi biến được đánh dấu và dòng kia, được ký hiệu là $\omega_x$, gồm các biến được đánh dấu trong các $X$-cột và các biến không được đánh dấu ở những nơi khác. Nếu $T^* = \text{chase}_\Sigma(T_x)$, thì $FD\ X \rightarrow Y$ là thuộc $\Sigma^+$ nếu và chỉ nếu các $Y$-cột trong $T^*$ chỉ gồm các biến được đánh dấu.*

**Định lý 4.** [1] *Xét lược đồ quan hệ $R(U)$, tập phụ thuộc hàm $F$ trên $U$ và $m$ tập con $U_1, U_2, ..., U_m$ của $U$, với $U_1U_2...U_m = U$. Cho $T$ là một bảng trên $U$, với $m$ dòng $s_1, s_2, ..., s_m$, trong đó với mỗi $i$ $(1 \leq i \leq m)$, và với mỗi $A \in U$, nếu $A \in U_i$, thì $s_i(A)$ là bằng với $dv_A$, và nếu $A \in U - U_i$, thì $s_i(A)$ là một ndv phân biệt. Thế thì, mọi quan hệ trên $R(U)$ mà thỏa $F$ có một phép tách-kết nối không mất thông tin đối với $U_1, U_2, ..., U_m$ khi và chỉ khi bảng $\text{chase}_F(T)$ có một dòng gồm toàn bộ các dv.*

## 2. MỘT SỐ VẤN ĐỀ ĐỐI VỚI JD VÀ PJNF

Lưu ý là ở đây không xét trường hợp các lược đồ quan hệ chỉ có các phụ thuộc hàm tầm thường và các phụ thuộc đa trị,...; và ở mục này khái niệm khóa chỉ định có cùng một ý nghĩa như đối với các lược đồ cơ sở dữ liệu ở $3NF$.

**2.1.** Một vấn đề đặt ra là: có phương pháp nào để suy dẫn các phụ thuộc kết nối từ tập các phụ thuộc cho trước hay không?. Đương nhiên là có thể bằng cách áp dụng Hệ tiên đề cho tập các phụ thuộc [3, 5], nghĩa là phải tính toán bao đóng của tập phụ thuộc đã cho. Lý thuyết về Bảng và Chase

được dùng làm công cụ để kiểm tra một phân tách là có kết nối không mất thông tin hay không, cũng để kiểm tra tính đúng đắn của các dẫn xuất phụ thuộc từ một tập phụ thuộc cho trước, nhưng cũng chỉ được sử dụng để kiểm tra chứ không phải là công cụ để đưa ra các dẫn xuất.

Như đã thấy, việc nghiên cứu vấn đề phân tách lược đồ quan hệ đóng vai trò quan trọng trong lý thuyết chuẩn hóa và là cơ sở hình thành khái niệm phụ thuộc kết nối. Tiếp cận vấn đề này, Mệnh đề 1 và Bổ đề 2 sau đây trình bày phương pháp tạo phụ thuộc kết nối từ kết quả của các thuật toán chuẩn hóa.

**Mệnh đề 1.** *Cho lược đồ quan hệ $R$ với tập phụ thuộc $\Sigma$. Giả sử $\mathcal{R} = \{R_1, R_2, ..., R_k\}$ là lược đồ cơ sở dữ liệu kết quả của việc áp dụng thuật toán chuẩn hóa có tính chất kết nối không mất thông tin. Thì phụ thuộc kết nối: $*[R_1, R_2, ..., R_k]$ là áp dụng được vào lược đồ $R$.*

*Chứng minh.* Với $\mathcal{R}$ là lược đồ cơ sở dữ liệu kết quả của thuật toán chuẩn hóa có tính chất kết nối của các lược đồ quan hệ thành phần $(R_1 * R_2 * ... * R_k)$ là không mất thông tin và $R = R_1R_2...R_k$. Do đó phụ thuộc kết nối $*[R_1, R_2, ..., R_k]$ là áp dụng được vào $R$. □

**Thí dụ 1.** Cho lược đồ quan hệ $R = A\,B\,C\,D\,E\,H\,I$ và tập phụ thuộc $\Sigma = \{A \to B\,C\,H,\ B\,C\,H \to A,\ B\,C\,H\,I \to E,\ E \to B\,H,\ E\,B \to C\}$.

Thực hiện thuật toán tổng hợp đối với các phụ thuộc hàm của $\Sigma$, trường hợp sử dụng phủ dạng vành: $G = \{(A,\ B\,C\,H),\ (B\,C\,H\,I) \to E,\ (E) \to B\,H\}$, và kết hợp với lược đồ thành phần khóa để đảm bảo tínhchất kết nối không mất thông tin: nếu sử dụng lược đồ khóa $A\,D\,I$, thì nhận được lược đồ cơ sở dữ liệu kết quả là $\mathcal{R} = \{A\,B\,C\,H,\ B\,C\,E\,H\,I,\ B\,E\,H,\ A\,D\,I\}$. Theo Mệnh đề 1 phụ thuộc kết nối $*[A\,B\,C\,H,\ B\,C\,E\,H\,I,\ B\,E\,H,\ A\,D\,I]$ là áp dụng được vào $R$; còn nếu kết hợp với lược đồ khóa $C\,D\,E\,I$, thì nhận được lược đồ cơ sở dữ liệu kết quả là $\mathcal{R} = \{A\,B\,C\,H,\ B\,C\,E\,H\,I,\ B\,E\,H, C\,D\,E\,I\}$, và theo Mệnh đề 1 có phụ thuộc kết nối $*[A\,B\,C\,H,\ B\,C\,E\,H\,I,\ B\,E\,H,\ C\,D\,E\,I]$ là áp dụng được vào $R$.

Trường hợp sử dụng phủ dạng vành: $G' = \{(A,\ B\,C\,H),\ (A\,I) \to E,\ (E) \to B\,H\}$ và kết hợp với lược đồ thành phần khóa để đảm bảo tính chất kết nối không mất thông tin thì các phụ thuộc kết nối $*[A\,B\,C\,H,\ B\,C\,E\,H\,I,\ B\,E\,H,\ A\,D\,I]$ và $*[A\,B\,C\,H,\ B\,C\,E\,H\,I,\ B\,E\,H,\ C\,D\,E\,I]$ là áp dụng được vào $R$.

Có thể bằng các phép phân tích-kết nối không mất thông tin liên tiếp đối với tập $\Sigma$ gồm các phụ thuộc hàm và phụ thuộc đa trị để nhận được các phụ thuộc kết nối, nhưng không luôn luôn nhận được mọi phụ thuộc kết nối có thể có đối với lược đồ quan hệ $R$ bất kỳ, có những trường hợp một quan hệ có thể có phép tách-kết nối không mất thông tin không tầm thường (không có lược đồ chiếu trùng với $R$) thành ba lược đồ, mà không có phép tách như vậy thành chỉ một cặp các lược đồ. Thí dụ 2 là một minh họa cụ thể cho điều khẳng định này, phụ thuộc kết nối $*[A\,B,\ A\,C,\ B\,C]$ không thể nhận được bằng cách áp dụng phép phân tích liên tiếp trên lược đồ quan hệ $r(A\,B\,C)$.

**Thí dụ 2.** Quan hệ $r(A\,B\,C)$ trong Hình 1 được tách có kết nối không mất thông tin thành các lược đồ quan hệ $AB, AC$ và $BC$. Các hình chiếu được thể hiện trong Hình 2.

Quan hệ $r$ này không thỏa các phụ thuộc đa trị không tầm thường, nên không có phép tách-kết nối không mất thông tin $r$ thành chỉ một cặp các lược đồ quan hệ $R_1$ và $R_2$ mà $R_1 \neq A\,B\,C$ và $R_2 \neq A\,B\,C$.

$r$

| $A$ | $B$ | $C$ |
|---|---|---|
| $a_1$ | $b_1$ | $c_1$ |
| $a_1$ | $b_2$ | $c_1$ |
| $a_3$ | $b_3$ | $c_3$ |
| $a_4$ | $b_3$ | $c_4$ |
| $a_5$ | $b_5$ | $c_5$ |
| $a_6$ | $b_6$ | $c_5$ |

*Hình 1*

$\pi_{AB}(r)$ = | A | B |
|---|---|
| $a_1$ | $b_1$ |
| $a_1$ | $b_2$ |
| $a_3$ | $b_3$ |
| $a_4$ | $b_3$ |
| $a_5$ | $b_5$ |
| $a_6$ | $b_6$ |

$\pi_{AC}(r)$ = | A | C |
|---|---|
| $a_1$ | $c_1$ |
| $a_1$ | $c_2$ |
| $a_3$ | $c_3$ |
| $a_4$ | $c_4$ |
| $a_5$ | $c_5$ |
| $a_6$ | $c_5$ |

$\pi_{BC}(r)$ = | B | C |
|---|---|
| $b_1$ | $c_1$ |
| $b_2$ | $c_2$ |
| $b_3$ | $c_3$ |
| $b_3$ | $c_4$ |
| $b_5$ | $c_5$ |
| $b_6$ | $c_5$ |

*Hình 2*

**Bổ đề 2.** *Cho lược đồ quan hệ $R$ với tập phụ thuộc $\Sigma$. Nếu áp dụng thuật toán tổng hợp sử dụng phủ dạng vành vào $R$ và nhận được lược đồ cơ sở dữ liệu $\mathcal{R}$ chỉ có duy nhất một lược đồ quan hệ thành phần (ký hiệu $\mathcal{R} = \{R_1'\}$ được hình thành từ phụ thuộc hàm phức hợp duy nhất $(X_1, X_2, ..., X_k) \to Y$. Thì tương ứng với phụ thuộc hàm phức hợp này, các phụ thuộc kết nối có dạng $*[R_1, R_2, ..., R_k]$ là áp dụng được vào $R$, trong đó: ứng với một chỉ số $t$ (với $1 \le t \le k$), thì $R_i = X_t X_j$ (với $1 \le i \le k-1, j \ne t, 1 \le j \le k$) và $R_k = X_t Y$.*

*Chứng minh.* Theo cách tạo phụ thuộc kết nối nêu trong Bổ đề 2: ứng với một chỉ số $t\,(1 \le t \le k)$, thì $R_i = X_t X_j$ (với $1 \le i \le k-1, j \ne t, 1 \le j \le k$) và $R_k = X_t Y$, và có $R = R_1 R_2 ... R_k$. Bởi vì $X_t, X_j$ là khóa của $R$, và cũng là khóa của các $R_i$ (với mọi $i$, mọi $j$), mỗi $R_i$ là một siêu khóa, thì tất cả các hình chiếu của quan hệ $r(R)$ trên các $R_i$ sẽ có cùng số lượng các bộ như $r$. Thêm nữa là, các $R_i$ giống nhau trên khóa $X_t$ nên nếu áp dụng $F$-qui tắc vào bảng $T$ được xây dựng theo Định lý 4, sẽ có một dòng gồm toàn bộ $dv$, do đó kết nối: $r = \pi_{R_1}(r) * \pi_{R_2}(r) * ... * \pi_{R_k}(r)$ là không mất thông tin.

Vì vậy kết luận được rằng, các phụ thuộc kết nối có dạng $*[R_1, R_2, ..., R_k]$ theo cách xây dựng trong Bổ đề 2 là áp dụng được vào $R$. □

**Thí dụ 3.** Cho lược đồ quan hệ gồm tập các thuộc tính $R = A_1\,A_2\,A_3\,A_4\,A_5\,A_6$ và tập phụ thuộc $F = \{A_1 \to A_2\,A_3\,A_6,\ A_2 \to A_3\,A_4,\ A_3 \to A_4\,A_5,\ A_5 \to A_1\,A_4\}$.

Lược đồ cơ sở dữ liệu kết quả của việc áp dụng thuật toán tổng hợp sử dụng phủ dạng vành là $\mathcal{R} = \{A_1\,A_2\,A_3\,A_4\,A_5\,A_6\}$, hình thành từ phụ thuộc hàm phức hợp: $(A_1,\ A_2,\ A_3,\ A_5) \to A_4\,A_6$.

Căn cứ vào Bổ đề 2, các phụ thuộc kết nối áp dụng được vào $R$ là: $*[A_1\,A_2,\ A_1\,A_3,\ A_1\,A_5,\ A_1\,A_4\,A_6]$, $*[A_1\,A_2,\ A_2\,A_3,\ A_2\,A_5,\ A_2\,A_4\,A_6]$, $*[A_1\,A_3,\ A_2\,A_3,\ A_3\,A_5,\ A_3\,A_4\,A_6]$ và $*[A_1\,A_5,\ A_2\,A_5,\ A_3\,A_5,\ A_4\,A_5\,A_6]$.

Hạn chế của việc suy dẫn phụ thuộc kết nối bằng tiếp cận phân tích - kết nối mất thông tin đã được minh họa bởi Thí dụ 2 trên đây. Còn tiếp cận tổng hợp cũng không cho phép trong trường hợp tổng quát có thể suy dẫn ra mọi phụ thuộc kết nối, vì như đã biết, phép tổng hợp chỉ áp dụng trên các phụ thuộc hàm.

Tuy vậy, với Mệnh đề 1 và Bổ đề 2 ta có phương pháp dẫn xuất các phụ thuộc kết nối từ kết quả của việc áp dụng thuật toán chuẩn hóa, là vấn đề khác với Bổ đề 1 và Định lý 4 chỉ cho phép kiểm tra tính đúng đắn của các dẫn xuất.

**2.2.** Khác với các dạng chuẩn: 3NF, BCNF và 4NF, không phải mọi lược đồ quan hệ bất kỳ $R$ với tập phụ thuộc $\Sigma$ đều có thể chuẩn hóa thành PJNF.

**Thí dụ 4.** Cho lược đồ quan hệ $R = A\,B_1\,B_2\,C_1\,C_2\,D\,E\,I_1\,I_2\,I_3\,J$ và tập $\Sigma$:

$$\{\ A \to B_1\,B_2\,C_1\,C_2\,D\,E\,I_1\,I_2\,I_3\,J,$$
$$B_1\,B_2\,C_1 \to A\,C_2\,D\,E\,I_1\,I_2\,I_3\,J, \qquad B_1\,B_2\,C_2 \to A\,C_1\,D\,E\,I_1\,I_2\,I_3\,J,$$
$$E \to I_1\,I_2\,I_3, \qquad C_1\,D \to J, \qquad C_2\,D \to J,$$
$$I_1\,I_2 \to I_3, \qquad I_2\,I_3 \to I_1, \qquad I_1\,I_3 \to I_2, \qquad B_1\,B_2\,I \twoheadrightarrow C_1\,C_2\,D\ \}.$$

Áp dụng thuật toán tổng hợp sử dụng phủ dạng vành, nhận được lược đồ cơ sở dữ liệu kết quả là $\mathcal{R} = \{R_1,\ R_2,\ R_3,\ R_4,\ R_5\}$, trong đó:

$R_1 = A\,B_1\,B_2 C_1\,C_2\,D\,E$; với các khóa chỉ định $\mathcal{K}_1 = \{A,\ B_1\,B_2\,C_1,\ B_1\,B_2\,C_2\}$
$R_2 = E\,I_1\,I_2$; với khóa chỉ định $\mathcal{K}_2 = \{E\}$
$R_3 = C_1\,D\,J$; với khóa chỉ định $\mathcal{K}_3 = \{C_1 D\}$
$R_4 = C_2\,D\,J$; với khóa chỉ định $\mathcal{K}_4 = \{C_2\,D\}$
$R_5 = I_1\,I_2\,I_3$; với các khóa chỉ định $\mathcal{K}_5 = \{I_1\,I_2,\ I_2\,I_3,\ I_1\,I_3\}$

Lược đồ $R$ không là ở PJNF vì theo Mệnh đề 1 thì phụ thuộc kết nối $*[A\,B_1\,B_2\,C_1\,C_2\,D\,E,\ E\,I_1\,I_2,\ C_1\,D\,J,\ C_2\,D\,J,\ I_1\,I_2\,I_3]$ là áp dụng được vào $R$, trong đó có lược đồ thành phần $A\,B_1\,B_2\,C_1\,C_2\,D\,E$ là siêu khóa của $R$, nhưng các lược đồ thành phần $E\,I_1\,I_2$, $C_1\,D\,J$, $C_2\,D\,J$ và $I_1\,I_2\,I_3$ không phải là các siêu khóa của $R$.

Với tiếp cận phương pháp tổng hợp, cụ thể là phép tổng hợp sử dụng phủ dạng vành ta phát hiện một tính chất đặc trưng của lớp lược đồ quan hệ ở PJNF.

**Bổ đề 3.** *Cho lược đồ quan hệ $R$ với tập phụ thuộc $\Sigma$. Nếu lược đồ quan hệ $R$ là ở PJNF thì khi áp dụng thuật toán tổng hợp sử dụng phủ dạng vành vào $R$ và nhận được lược đồ cơ sở dữ liệu $\mathcal{R}$ thì: $\mathcal{R}$ chỉ có duy nhất một lược đồ quan hệ thành phần (ký hiệu $\mathcal{R} = \{R'_1\}$ được hình htành từ phụ thuộc hàm phức hợp duy nhất $(X_1, X_2, ..., X_k) \to Y$.*

*Chứng minh.* Giả sử lược đồ cơ sở dữ liệu $\mathcal{R}$ có hơn một lược đồ quan hệ thành phần, tức $\mathcal{R} = \{R'_1, R'_2, ..., R'_q\}$, với $q \geq 2$.

Theo thuật toán tổng hợp sử dụng phủ dạng vành thì mỗi lược đồ thành phần $i, j$ $(1 \leq i \leq q,\ 1 \leq j \leq q)$ thuộc $\mathcal{R} = \{R'_1, R'_2, ..., R'_q\}$ có thể được ký hiệu như sau:

- Lược đồ thành phần $R'_i = K^i_1\,K^i_2 \ldots K^i_{pi}\,Y^i$; với các khóa chỉ định $\mathcal{K}_i = \{K^i_1, K^i_2, ..., K^i_{pi}\}$, và $Y^i$ là vế trái của phụ thuộc hàm phức hợp thứ $i$.
- Lược đồ thành phần $R'_j = K^j_1\,K^j_2 \ldots K^j_{pj}\,Y^j$; với các khóa chỉ định $\mathcal{K}_j = \{K^j_1, K^j_2, ..., K^j_{pj}\}$, và $Y^j$ là vế trái của phụ thuộc hàm phức hợp thứ $j$.

Vì các $R'_i$ và $R'_j$ là hai lược đồ thành phần được hình thành từ việc phân hoạch tập $\Sigma$, nên không thể có sự tương đương giữa các lược đồ thành phần: $R'_i \leftrightarrow R'_j$ (với mọi $i$, mọi $j$). Bởi vì nếu có sự tương đương như vậy, thì do $K^i_t$ là khóa của $R'_i$ (với mọi $i$, mọi $t$) có $K^i_t \leftrightarrow R'_i$, còn $K^j_h$ là khóa của $R'_j$ (với mọi $j$, mọi $h$) có $K^j_h \leftrightarrow R'_j$, sẽ có sự tương đương giữa các tập trái $K^i_t \leftrightarrow K^j_h$ (với mọi $i, t, j$ và $h$) của các phụ thuộc hàm phức hợp thứ $i$ và $j$. Vì sự tương đương giữa các tập trái, suy ra các tập trái $K^i_t$ và $K^j_h$ (với mọi $i, t, j$ và $h$) phải thuộc cùng vế trái của chỉ một phụ thuộc hàm phức hợp. Tức là các $R'_i$ và $R'_j$ không là hai lược đồ thành phần được hình thành từ việc phân hoạch tập $\Sigma$.

Nhưng nếu không có sự tương đương giữa các lược đồ thành phần: $R'_i \leftrightarrow R'_j$ (với mọi $i$, mọi $j$), thì các $R'_i$ và $R'_j$ không thể cùng là siêu khóa của $R$. Nghĩa là phụ thuộc kết nối $*[R'_1, R'_2, ..., R'_k]$ vi phạm PJNF. Đây là điều mâu thuẫn. □

Bổ đề 3 nêu tính chất đặc trưng của lược đồ quan hệ ở PJNF và là điều kiện cần. Như đã phân tích, lược đồ $R$ trong Thí dụ 4 vi phạm điều kiện nêu trong Bổ đề 3 và không là ở PJNF.

Dễ dàng nhận thấy lược đồ quan hệ $R$ đã cho trong Thí dụ 3, với tập phụ thuộc $\Sigma = \{A_1 \to A_2\,A_3\,A_6,\ A_2 \to A_3\,A_4,\ A_3 \to A_4\,A_5,\ A_5 \to A_1\,A_4,\ *[A_1\,A_2,\ A_1\,A_3,\ A_1\,A_5,\ A_1\,A_4\,A_6],\ *[A_1\,A_2,\ A_2\,A_3,\ A_2\,A_5,\ A_2\,A_4\,A_6],\ *[A_1\,A_3,\ A_2\,A_3,\ A_3\,A_5,\ A_3\,A_4\,A_6],\ *[A_1\,A_5,\ A_2\,A_5,\ A_3\,A_5,\ A_4\,A_5\ A_6]\}$ là ở PJFN, thỏa điều kiện của Bổ đề 3.

Tuy nhiên, cần lưu ý dấu hiệu "lược đồ cơ sở dữ liệu $\mathcal{R}$ chỉ có duy nhất một lược đồ thành phần" không là điều kiện đủ để xác định một lược đồ quan hệ là ở PJNF, như sẽ được chứng tỏ bởi Thí dụ 5 sau đây.

**Thí dụ 5.** Cho lược đồ quan hệ $R = A\,B\,C\,D\,E$ và tập phụ thuộc $\Sigma = \{A \to B\,C\,E,\ B\,C\,E \to A\,D,\ B\,C \twoheadrightarrow A\,E\}$.

Mặc dù lược đồ cơ sở dữ liệu kết quả của việc áp dụng thuật toán tổng hợp sử dụng phủ dạng vành là $\mathcal{R} = \{A\,B\,C\,D\,E\}$, hình thành từ phụ thuộc hàm phức hợp duy nhất: $(A,\ B\,C\,E) \to D$, theo

Bổ đề 2, các phụ thuộc kết nối áp dụng được vào $R$ là: $*[ABCE, AD]$ và $*[ABCE, BCDE]$. Nhưng rõ ràng lược đồ $R$ đã cho không là ở PJNF.

## TÀI LIỆU THAM KHẢO

[1] Atzeni P., De Antonellis V., *Relational Database Theory*, The Benjamin/Cummings Publishing Company, 1993.

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

[3] Maier D., Mendelzon A. O., and Sagiv Y., Testing implications of data dependencies, *ACM Trans. Database Syst.* **4** (4) (1979) 455–469.

[4] Phạm Quang Trung, Nguyễn Xuân Huy, Thuật toán tổng hợp lược đồ cơ sở dữ liệu quan hệ dạng chuẩn ba, *Tạp chí Tin học và Điều khiển học* **16** (2) (2000) 41–50.

[5] Ullman J. D., *Principles of Database Systems*, 2nd edition, Computer Science Press, 1982.

*Nhận bài ngày 12 - 7 - 2000*
*Nhận lại sau khi sửa ngày 19 - 2 - 2001*

*Viện Kiểm sát nhân dân tối cao.*