

# VỀ MỘT SỐ DẠNG CHUẨN TRONG CƠ SỞ DỮ LIỆU MỜ CHỨA DỮ LIỆU NGÔN NGỮ<sup>1</sup>

LÊ XUÂN VINH, TRẦN THIÊN THÀNH, LÊ XUÂN VIỆT

*Trường Đại học Quy Nhơn*

lexuanvinh@qnu.edu.vn

**Tóm tắt.** Trên quan điểm xem miền trị của mỗi thuộc tính trong cơ sở dữ liệu mờ chứa dữ liệu ngôn ngữ tương ứng với một đại số gia tử thích hợp phụ thuộc hàm mờ và phụ thuộc đa trị mờ đã được nghiên cứu [4, 5, 7]. Trong bài báo này, chúng tôi tiếp tục nghiên cứu một số dạng chuẩn mờ. Các khái niệm khóa mức  $\mathfrak{K}$ , siêu khóa mức  $\mathfrak{K}$  và dạng chuẩn F1NF,  $\mathfrak{RF}2NF$ ,  $\mathfrak{RF}3NF$ ,  $\mathfrak{RFBCNF}$  trong cơ sở dữ liệu mờ chứa dữ liệu ngôn ngữ được định nghĩa. Phép tách một lược đồ cơ sở dữ liệu về dạng chuẩn  $\mathfrak{RF}3NF$  bảo toàn các phụ thuộc hàm mờ cũng như tách về dạng chuẩn  $\mathfrak{RFBCNF}$  bảo toàn thông tin được quan tâm nghiên cứu.

*Từ khóa:* cơ sở dữ liệu mờ, dữ liệu ngôn ngữ, dạng chuẩn mờ, phép tách.

**Abstract.** On the viewpoint that each attribute domain of the fuzzy databases with linguistic data corresponds to an appropriate hedge algebras, the fuzzy functional dependencies and the fuzzy multivalued dependencies are studied [4, 5, 7]. In this paper, we continue to study a number of fuzzy normal forms. Such concepts as  $\mathfrak{K}$ -keys, super  $\mathfrak{K}$ -keys and the normal forms F1NF,  $\mathfrak{RF}2NF$ ,  $\mathfrak{RF}3NF$ ,  $\mathfrak{RFBCNF}$  are defined in fuzzy databases with linguistic data. The paper focuses on researching the fuzzy dependency preserving decomposition into  $\mathfrak{RF}3NF$  and the lossless join decomposition into  $\mathfrak{RFBCNF}$ .

*Keywords:* fuzzy databases, linguistic data, fuzzy normal forms, decomposition.

## 1. GIỚI THIỆU

Sự xuất hiện thông tin không chắc chắn, không chính xác, thông tin mơ hồ mà chúng ta sẽ gọi chung là thông tin mờ trong một mô hình cơ sở dữ liệu (CSDL) nào đó là hết sức tự nhiên. Chẳng hạn khi điều tra, khảo sát thông tin về những công ty, doanh nghiệp cung cấp một số mặt hàng trên thị trường ở một địa bàn trong những năm gần đây thì có nhiều thông tin không hoàn toàn chính xác hoặc theo kiểu định tính như: quy mô doanh nghiệp là nhỏ, rất nhỏ, vừa, lớn,...; giá cả có thể là cao, rất cao,...; thời gian cung cấp hàng là nhanh, chậm,...Việc biểu diễn, xử lý các loại thông tin này một cách trực tiếp là vấn đề nhận được sự quan tâm của nhiều nhà nghiên cứu và mô hình chứa thông tin mờ được gọi chung là CSDL mờ.

---

<sup>1</sup>This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.01- 2011.06

Trong các bài báo gần đây [4, 5, 7], các tác giả đã nghiên cứu một số vấn đề trên mô hình CSDL mờ theo cách tiếp cận đại số giá trị. Ngoài những kiểu dữ liệu đã được đề cập như giá trị rõ, giá trị khoảng, tập hợp các giá trị rõ, chúng tôi đặc biệt quan tâm đến các giá trị của biến ngôn ngữ tương ứng với các thuộc tính (quy mô, giá, thời gian cung cấp). Những giá trị của biến ngôn ngữ ở đây cũng chính là những phần tử, những giá trị ngôn ngữ của một đại số giá trị tương ứng và những CSDL chứa các giá trị ngôn ngữ như vậy chúng ta sẽ gọi là CSDL mờ chứa dữ liệu ngôn ngữ hay cho gọn là CSDL ngôn ngữ.

Phân hoạch miền trị của thuộc tính bằng một quan hệ tương đương phụ thuộc vào một số dương  $k$  cố định trước, có ý nghĩa là độ dài của một từ chuẩn được chọn, để xây dựng quan hệ bằng nhau mờ giữa các dữ liệu trong miền trị. Trên cơ sở này có thể thực hiện các chuyển đổi truy vấn mờ sang truy vấn rõ [4], định nghĩa phụ thuộc hàm mờ, phụ thuộc đa trị mờ, xây dựng tập luật suy diễn đúng đắn và đầy đủ [7]. Một số tác giả đã dựa trên những kết quả này để đề xuất phương pháp biểu diễn và cài đặt cơ sở dữ liệu mờ chứa dữ liệu ngôn ngữ trên XML.

Tương tự trong cơ sở dữ liệu kinh điển, việc thiết kế "tốt" CSDL mờ trên mô hình này sẽ tránh được những hiện tượng "dư thừa", "dị thường" khi thực hiện các thao tác cập nhật dữ liệu. Vì vậy, một cơ sở lý thuyết với các khái niệm khóa, siêu khóa và các dạng chuẩn mờ, phép tách lược đồ quan hệ theo quan điểm thống nhất với những kết quả trong [4, 5, 7] cần được nghiên cứu. Đây chính là mục tiêu của bài báo.

Bài báo được tổ chức như sau: Sau phần giới thiệu, Mục 2 dành cho việc trình bày những kết quả cơ bản nhất về phụ thuộc hàm mờ trong cơ sở dữ liệu ngôn ngữ. Mục 3 trình bày các khái niệm, định nghĩa về khóa, siêu khóa mức  $\mathfrak{K}$  và các dạng chuẩn F1NF,  $\mathfrak{R}F2NF$ ,  $\mathfrak{R}F3NF$ ,  $\mathfrak{R}FBCNF$ . Mục 4 dành cho việc nghiên cứu phép tách lược đồ cơ sở dữ liệu về dạng chuẩn  $\mathfrak{R}F3NF$  bảo toàn các phụ thuộc hàm mờ. Mục 5 trình bày về phép tách lược đồ cơ sở dữ liệu về dạng chuẩn  $\mathfrak{R}FBCNF$  bảo toàn thông tin. Cuối cùng, một số nhận xét và kết luận về những vấn đề đã được nghiên cứu trong bài báo sẽ được trình bày trong Mục 6.

## 2. PHỤ THUỘC HÀM MỜ TRONG CƠ SỞ DỮ LIỆU NGÔN NGỮ

CSDL ngôn ngữ, đã được giới thiệu trong [4, 7], là một tập  $DB = \{U, R_1, \dots, R_m; Const\}$ , ở đây  $U = \{A_1, A_2, \dots, A_n\}$  là không gian các thuộc tính, mỗi  $R_i$  là một lược đồ quan hệ,  $Const$  là tập các ràng buộc dữ liệu. Miền trị  $Dom(A)$  của thuộc tính  $A$  (trong số các  $A_1, A_2, \dots, A_n$ ) có thể chứa giá trị rõ, giá trị khoảng, các kiểu dữ liệu khác đặc biệt là giá trị của biến ngôn ngữ, chúng là những phần tử của đại số giá trị tương ứng với biến ngôn ngữ. Với mỗi  $Dom(A)$  ta chọn tùy ý một số nguyên dương  $k \geq 0$ . Ghép các khoảng mờ thích hợp của tất cả các từ có độ dài  $k + 1$  (chi tiết trong [4]), ta được một phân hoạch trên  $[0, 1]$  và qua một phép tỉ lệ đơn giản ta sẽ thu được một phân hoạch trên  $Dom(A)$ . Phân hoạch này xác định một quan hệ tương đương  $\mathcal{S}_k$ . Những giá trị cùng thuộc một lớp tương đương được cho là bằng nhau mức  $k$  hay là tương tự mức  $k$ . Số  $k$  càng lớn phân hoạch càng mịn, hai giá trị bằng nhau mức  $k$  càng phải gần nhau. Vì vậy, đối với các thuộc tính không chứa các thông tin mờ như tên doanh nghiệp, mặt hàng mức  $k$  đối với các thuộc tính này bằng  $\infty$  với ý nghĩa hai giá trị bằng nhau khi đồng nhất với nhau về kí hiệu. Bộ các số nguyên dương  $\mathfrak{K} = (k_1, k_2, \dots, k_n)$  được chọn tương ứng với các thuộc tính trong  $U$  được gọi là một mức tương tự trên lược đồ quan hệ  $R$ .

Để tiện theo dõi, chúng ta nhắc lại một số thuật ngữ và kí hiệu trong [4]. Nếu  $X \subseteq U$  thì

kí hiệu  $\mathfrak{K}_X$  là một bộ gồm  $|X|$  thành phần thu hẹp của  $\mathfrak{K}$  từ  $U$  xuống  $X$ . Cho  $\mathfrak{K}_X$  và  $\mathfrak{K}'_X$  là hai mức tương tự trên  $X$ , ta nói  $\mathfrak{K}_X \geq \mathfrak{K}'_X$  hoặc  $\mathfrak{K}'_X \leq \mathfrak{K}_X$  nếu  $k_A \geq k'_A$  với mọi  $A \in X$ . Cho  $\mathfrak{K}_X, \mathfrak{K}'_Y$  là hai mức tương tự trên  $X, Y \subseteq U$  tương ứng. Mức tương tự mở rộng trên  $X \cup Y$ , kí hiệu  $\mathfrak{K}_X \vee \mathfrak{K}'_Y$ , được xác định là  $\mathfrak{K}_X \vee \mathfrak{K}'_Y = \mathfrak{K}_{X-Y} \cup \mathfrak{K}'_{Y-X} \cup \mathfrak{K}''_Z$ , trong đó  $Z = X \cap Y$  và  $\mathfrak{K}''_Z = \{k_A \vee k'_A | A \in Z\}$ , ở đây  $k_A \vee k'_A = \max(k_A, k'_A)$ . Nếu  $k_Z = k'_Z$  thì  $\mathfrak{K}_X \vee \mathfrak{K}'_Y$  được kí hiệu là  $\mathfrak{K}_X \cup \mathfrak{K}'_Y$ .

Theo cách tiếp cận đại số gia tử, với  $X, Y \subseteq U$ ,  $Y$  phụ thuộc hàm mờ vào  $X$  trong CSDL ngôn ngữ nếu hai bộ bằng nhau mức  $\mathfrak{K}_X$  trên  $X$  thì cũng bằng nhau mức  $\mathfrak{K}_Y$  trên  $Y$ . Ta có thể dùng mức  $\mathfrak{K}$  để chỉ cho các mức hạn chế  $\mathfrak{K}_X, \mathfrak{K}_Y$  cho gọn nếu không gây nhầm lẫn.

**Định nghĩa 2.1.** [5] Cho DB là một CSDL ngôn ngữ và  $R$  là một lược đồ quan hệ của DB có tập thuộc tính  $U$ . Với  $X, Y \subseteq U$  và  $\mathfrak{K}$  là một mức tương tự trên  $U$ , ta nói  $Y$  phụ thuộc hàm mờ vào  $X$  mức  $\mathfrak{K}$  trên lược đồ quan hệ  $R$ , kí hiệu  $f = X \rightarrow_{\mathfrak{K}} Y$ , nếu với mọi quan hệ  $r \in R$ :

$$(\forall t, s \in r)(t[X] =_{\mathfrak{K}} s[X] \Rightarrow t[Y] =_{\mathfrak{K}} s[Y])$$

**Ví dụ 2.1** Một quan hệ  $R$  của CSDL ngôn ngữ (mô hình cho các ví dụ còn lại)

Tên doanh nghiệp	Quy mô	Mặt hàng	Giá	Thời gian cung cấp
a1	lớn	c1	9.5	chậm
a1	lớn	c1	rất-cao	4
a2	nhỏ	c1	5.8	[2, 4]
a2	khá nhỏ	c1	ít-cao	3
a1	lớn	c2	thấp	nhanh
a3	vừa	c2	thấp	rất-nhanh

Giả sử chọn  $\mathfrak{K} = (\infty, 1, \infty, 1, 2)$ . Các số 1,1,2 trong  $\mathfrak{K}$  là ba mức được chọn với ba thuộc tính Quy mô, Giá và Thời gian cung cấp tương ứng. Nếu mức chọn là  $k$  thì mỗi cụm trong phân hoạch của miền trị sẽ là hợp của các khoảng mờ của một số giá trị ngôn ngữ độ dài  $k + 1$ .

Để minh họa, xét miền trị của thuộc tính "Giá" chứa các giá trị ngôn ngữ cao, thấp, rất cao, ít cao,...và các giá trị số. Giả sử miền trị tham chiếu của "Giá" là  $[0,10]$ . Chúng ta sẽ xây dựng ĐSGT tuyến tính  $\underline{AX} = (X, G, C, H, \leq)$  với  $G = \{thấp, cao\}$ ,  $C = \{0, W, 1\}$ ,  $H = H^- \cup H^+ = \{không, ít\} \cup \{khá, rất\}$ , trong đó,  $ít > không$  và  $rất > khá$ . Để ý rằng,  $không$  là gia tử phủ định địa phương, không phải là toán tử phủ định logic [3]. Các tham số mờ  $fm(cao) = fm(thấp) = W = 0.5$ ,  $\mu(không) = \mu(ít) = \mu(khá) = \mu(rất) = 0.25$ . Mức tương tự của thuộc tính "Giá" là 1.

Sử dụng hàm định lượng ngữ nghĩa [4], tính giá trị định lượng của các giá trị ngôn ngữ và chuyển đổi lên miền trị tham chiếu với hệ số  $r = 10$ . Ta có:

$$v_r(cao) = [W + 0.5 \times fm(cao)] \times 10 = [0.5 + 0.5 \times 0.5] \times 10 = 7.5$$

$$v_r(ít-cao) = [W + 0.5 \times fm(ít-cao)] \times 10 = [0.5 + 0.5 \times 0.25 \times 0.5] \times 10 = 5.625$$

$$v_r(không-cao) = [W + fm(ít-cao) + 0.5 \times fm(không-cao)] \times 10 = [0.5 + 0.25 \times 0.5 + 0.5 \times 0.25 \times 0.5] \times 10 = 6.875$$

$$v_r(khá-cao) = v_r(cao) + 0.5 \times fm(khá-cao) \times 10 = 7.5 + [0.5 \times 0.25 \times 0.5] \times 10 = 8.125$$

$$v_r(rất-cao) = 10 - 0.5 \times fm(rất-cao) \times 10 = 10 - 0.5 \times 0.25 \times 0.5 \times 10 = 9.375$$

Tiếp tục, chúng ta tính các khoảng tương tự  $\mathcal{S}_{1,r}(\cdot)$  trên  $[0,10]$  tạo thành bởi các khoảng mờ  $\mathfrak{I}_r(\cdot)$  của các giá trị ngôn ngữ có độ dài 2:

$$\mathcal{S}_{1,r}(cao) = \mathfrak{I}_r(không-cao) \cup \mathfrak{I}_r(khá-cao) = (7.5 - 0.25 \times 0.5 \times 10, 7.5 + 0.25 \times 0.5 \times 10) = (6.25, 8.75]$$

$$\mathcal{S}_{1,r}(\text{rất-cao}) = (10 - 0.25 \times 0.5 \times 10, 10] = (8.75, 10]$$

$$\mathcal{S}_{1,r}(\text{ít-cao}) = (0.5, 0.5 \times 10 + 0.25 \times 0.5 \times 10] = (5, 6.25]$$

Như vậy, nếu chọn  $k = 1$  thì quan hệ tương tự  $\mathcal{S}_1$  phân hoạch nửa trên miền trị tham chiếu của thuộc tính "Giá" thành các khoảng  $(5, 6.25]$ ,  $(6.25, 8.75]$ ,  $(8.75, 10]$ . Giá trị *ít-cao* nằm trong khoảng  $(5, 6.25]$ , *rất-cao* nằm trong  $(8.75, 10]$ , các giá trị còn lại *không-cao*, *cao*, *khá-cao* nằm trong  $(6.25, 8.75]$ . Tương tự, chúng ta có thể phân hoạch miền trị của các thuộc tính khác có tồn tại giá trị ngôn ngữ như thuộc tính Quy mô, Thời gian cung cấp.

Cuối cùng, trong Ví dụ 2.1, ta có  $9.5 =_1$  *rất-cao*,  $5.8 =_1$  *ít-cao* và *khá-nhỏ*  $=_1$  *nhỏ* (có thể suy ra nhờ tính phổ dụng của ĐSGT).

Vì vậy,  $R$  thỏa các phụ thuộc hàm mờ Tên Doanh nghiệp  $\rightarrow_{\mathfrak{R}}$  Quy mô và (Tên Doanh nghiệp, Mặt hàng)  $\rightarrow_{\mathfrak{R}}$  Giá với  $\mathfrak{R} = (\infty, 1, \infty, 1, 2)$  nhưng rõ ràng không có các phụ thuộc hàm tương ứng theo kiểu kinh điển.

Để định nghĩa và chứng minh một số tính chất về dạng chuẩn, chúng ta sẽ sử dụng một số kết quả chính về phụ thuộc hàm mờ trong [4, 5, 7] như sau:

Giả sử  $F$  là một tập phụ thuộc hàm mờ. Bao đóng của  $F$ , kí hiệu là  $F^+$ , là tập tất cả các phụ thuộc hàm mờ được suy diễn từ  $F$  bởi các quy tắc suy diễn (K1) – (K4):

(K1) *Phản xạ*:  $X \rightarrow_{\mathfrak{R}} Y$  nếu  $Y \subseteq X$ ;

(K2) *Mở rộng*:  $X \rightarrow_{\mathfrak{R}} Y \Rightarrow XZ \rightarrow_{\mathfrak{R}} YZ$ , với mọi  $Z \subseteq U$ ;

(K3) *Giảm mức*:  $X \rightarrow_{\mathfrak{R}} Y \Rightarrow X \rightarrow_{\mathfrak{R}'} Y$  nếu  $\mathfrak{R}'_X = \mathfrak{R}_X$  và  $\mathfrak{R}'_Y \leq \mathfrak{R}_Y$ ;

(K4) *Bắc cầu*:  $X \rightarrow_{\mathfrak{R}} Y$  và  $Y \rightarrow_{\mathfrak{R}'} Z \Rightarrow X \rightarrow_{\mathfrak{R} \vee \mathfrak{R}'} Z$  nếu  $\mathfrak{R}'_Y \leq \mathfrak{R}_Y$

Tập các quy tắc suy diễn (K1) – (K4) đã được chứng minh đúng đắn và đầy đủ [5], tức là  $F^+ = F^*$  (ở đây  $F^*$  là tập các hệ quả logic của  $F$ ), đồng thời ta cũng có:

**Định lý 2.1.** [5] *Giả sử  $F$  là một tập phụ thuộc hàm mờ. Khi đó  $F^+$  có những tính chất sau:*

(i)  $X \rightarrow_{\mathfrak{R}} Y \in F^+ \Rightarrow X \rightarrow_{\mathfrak{R}} A \in F^+, \forall A \in Y$ .

(ii)  $(X \rightarrow_{\mathfrak{R}} Y \in F^+, X \rightarrow_{\mathfrak{R}'} Z \in F^+) \Rightarrow X \rightarrow_{\mathfrak{R} \cup \mathfrak{R}'} YZ \in F^+$ , với điều kiện  $\mathfrak{R}_X = \mathfrak{R}'_X$ .

(iii)  $(X \rightarrow_{\mathfrak{R}} Y \in F^+, V \rightarrow_{\mathfrak{R}'} W \in F^+) \Rightarrow XV \rightarrow_{\mathfrak{R} \vee \mathfrak{R}'} YW \in F^+$ , với  $\mathfrak{R}'_{Y \cap VW} \leq \mathfrak{R}_{Y \cap VW}$ .

(iv) Đặt  $G = \{X \rightarrow_{\mathfrak{R}} A \mid X \rightarrow_{\mathfrak{R}} Y \in F \text{ và } A \in Y\}$ . Khi đó,  $G^+ = F^+$ .

Một số quy tắc khác được chứng minh là đúng đắn bằng cách dùng (K1)-(K4) [7]. Đặc biệt chúng ta quan tâm đến các quy tắc trong mệnh đề sau:

**Mệnh đề 2.1.** [7] *Các quy tắc suy diễn sau đây là đúng đắn*

*Quy tắc hợp*:  $\{X \rightarrow_{\mathfrak{R}} Y, X \rightarrow_{\mathfrak{R}'} Z\} \models X \rightarrow_{\mathfrak{R} \cup \mathfrak{R}'} YZ$  nếu  $\mathfrak{R}_X = \mathfrak{R}'_X$ .

*Quy tắc giả bắc cầu*:  $\{X \rightarrow_{\mathfrak{R}} Y, WY \rightarrow_{\mathfrak{R}'} Z\} \models WX \rightarrow_{\mathfrak{R} \vee \mathfrak{R}'} Z$  nếu  $\mathfrak{R}'_Y \leq \mathfrak{R}_Y$ .

*Quy tắc tách*: Nếu  $X \rightarrow_{\mathfrak{R}} Y$  và  $Z \subseteq Y$  thì  $X \rightarrow_{\mathfrak{R}} Z$ .

### 3. MỘT SỐ DẠNG CHUẨN MỜ TRONG CSDL NGÔN NGỮ

Thiết kế "tốt" một CSDL ngôn ngữ nhằm ngăn chặn hiện tượng "dư thừa", "dị thường" xảy ra khi cập nhật dữ liệu. Chúng ta sẽ mở rộng một số định nghĩa từ CSDL kinh điển để phù hợp với việc sử dụng quan hệ bằng nhau mức  $\mathfrak{R}$  theo cách tiếp cận đại số gia tử.

**Định nghĩa 3.1.** (*Phụ thuộc toàn phần, phụ thuộc bộ phận*) Cho  $X, Y \subseteq U$  với  $U$  là tập thuộc tính của lược đồ CSDL ngôn ngữ  $R$ .  $Y$  được gọi là phụ thuộc hàm mờ toàn phần mức  $\mathfrak{R}$  đối với  $X$  nếu và chỉ nếu  $X \rightarrow_{\mathfrak{R}} Y$  và không tồn tại  $X' \subset X$ ,  $X' \neq \emptyset$  sao cho  $X' \rightarrow_{\mathfrak{R}} Y$ .  $Y$

được gọi là phụ thuộc hàm mờ bộ phận mức  $\mathfrak{K}$  đối với  $X$  khi và chỉ khi  $X \rightarrow_{\mathfrak{K}} Y$  và tồn tại  $X' \subset X, X' \neq \emptyset$  sao cho  $X' \rightarrow_{\mathfrak{K}} Y$ .

Bây giờ chúng ta sẽ định nghĩa khóa, siêu khóa mức  $\mathfrak{K}$  của lược đồ CSDL ngôn ngữ.

**Định nghĩa 3.2.** (*Khóa mức  $\mathfrak{K}$ , siêu khóa mức  $\mathfrak{K}$* ) Cho  $K \subseteq U$  và  $F$  là tập các phụ thuộc hàm mờ trong  $R$  và  $\mathfrak{K}$  là một mức tương tự cho trước. Khi đó,  $K$  được gọi là một khóa mức  $\mathfrak{K}$  khi và chỉ khi  $K \rightarrow_{\mathfrak{K}} U \in F^+$  và  $K \rightarrow_{\mathfrak{K}} U$  là phụ thuộc hàm mờ toàn phần mức  $\mathfrak{K}$ . Tập  $S \subseteq U$  được gọi là một siêu khóa mức  $\mathfrak{K}$  khi và chỉ khi  $S$  chứa một khóa mức  $\mathfrak{K}$ .

Nếu  $R$  có nhiều hơn một khóa mức  $\mathfrak{K}$ , ta chọn một khóa  $K$  bất kỳ làm khóa chính mức  $\mathfrak{K}$ . Những thuộc tính  $A \in K$  được gọi là thuộc tính khóa mức  $\mathfrak{K}$ , những thuộc tính  $A \notin K$  được gọi là thuộc tính không khóa mức  $\mathfrak{K}$ .

Miền trị của mỗi thuộc tính trong CSDL ngôn ngữ chấp nhận các kiểu dữ liệu khác nhau từ kiểu (1)-(7) bao gồm: giá trị rõ, khoảng, tập hợp, giá trị ngôn ngữ và 3 kiểu dữ liệu Null khác "unknown", "applicable", "undefine" và quan hệ bằng nhau mức  $k$  giữa các kiểu dữ liệu cũng đã được định nghĩa (xem chi tiết trong [4]). Vì vậy, dạng chuẩn 1 được định nghĩa như sau.

**Định nghĩa 3.3.** (*F1NF*) Lược đồ quan hệ mờ  $R$  được gọi là ở dạng chuẩn 1 (F1NF) khi và chỉ khi với mọi  $r \in R$  miền trị của các thuộc tính chứa dữ liệu thuộc kiểu (1)-(7) [4].

Ngoại trừ dạng chuẩn F1NF, các dạng chuẩn khác được định nghĩa trên lược đồ CSDL ngôn ngữ dựa trên cơ sở tập phụ thuộc hàm mờ và khóa mức  $\mathfrak{K}$ . Vì vậy, trong các định nghĩa sau đây chúng ta sẽ gọi là dạng chuẩn mờ mức  $\mathfrak{K}$ .

**Định nghĩa 3.4.** ( *$\mathfrak{K}F2NF$* ) Cho  $F$  là tập phụ thuộc hàm mờ của lược đồ quan hệ ngôn ngữ  $R$  và  $K$  là một khóa mức  $\mathfrak{K}$ . Khi đó,  $R$  được gọi là ở dạng chuẩn 2 mức  $\mathfrak{K}$  ( $\mathfrak{K}F2NF$ ) khi và chỉ khi  $R$  ở dạng chuẩn F1NF và với mọi thuộc tính không khóa  $A \notin K$  ta có  $K \rightarrow_{\mathfrak{K}'} A \in F^+$  và  $K \rightarrow_{\mathfrak{K}'} A$  thỏa điều kiện  $\mathfrak{K}'_K = \mathfrak{K}_K$  và  $\mathfrak{K}'_A \geq \mathfrak{K}_A$  là phụ thuộc hàm mờ toàn phần.

**Định nghĩa 3.5.** ( *$\mathfrak{K}F3NF$* ) Cho  $F$  là tập phụ thuộc hàm mờ của lược đồ quan hệ ngôn ngữ  $R$  và  $K$  là một khóa mức  $\mathfrak{K}$ . Khi đó,  $R$  được gọi là ở dạng chuẩn 3 mức  $\mathfrak{K}$  ( $\mathfrak{K}F3NF$ ) khi và chỉ khi  $R$  ở dạng chuẩn F1NF và với mọi phụ thuộc hàm mờ  $X \rightarrow_{\mathfrak{K}'} A \in F^+, A \notin X$  thì  $A$  là thuộc tính khóa mức  $\mathfrak{K}$  của  $R$  hoặc  $X$  là một siêu khóa mức  $\mathfrak{K}$  của  $R$ .

**Định nghĩa 3.6.** ( *$\mathfrak{K}FBCNF$* ) Cho  $F$  là tập phụ thuộc hàm mờ của lược đồ quan hệ ngôn ngữ  $R$  và  $K$  là một khóa mức  $\mathfrak{K}$ . Khi đó,  $R$  được gọi là ở dạng chuẩn mờ Boyce-Codd mức  $\mathfrak{K}$  ( $\mathfrak{K}FBCNF$ ) khi và chỉ khi  $R$  ở dạng chuẩn F1NF và nếu phụ thuộc hàm mờ  $X \rightarrow_{\mathfrak{K}'} A \in F^+, A \notin X$  thì  $X$  là một siêu khóa mức  $\mathfrak{K}$ .

**Ví dụ 3.1.**  $R$  là lược đồ quan hệ ngôn ngữ có tập thuộc tính  $U=ABCDE$ . Mức tương tự  $\mathfrak{K} = (\infty, 2, \infty, 1, 2)$ . Tập phụ thuộc hàm mờ  $F = \{A \rightarrow_{\mathfrak{K}AB} B, AC \rightarrow_{\mathfrak{K}'ACD} D\}$ , ở đây  $\mathfrak{K}' = (\infty, 2, \infty, 2, 2)$ .

$ACE$  là một khóa mức  $\mathfrak{K}$  của  $R$ , bởi vì: Từ  $AC \rightarrow_{\mathfrak{K}'ACD} D$  ta suy ra  $AC \rightarrow_{\mathfrak{K}ACD} D$  theo (K3) do  $\mathfrak{K}_{AC} = \mathfrak{K}'_{AC}$  và  $\mathfrak{K}_D \leq \mathfrak{K}'_D$ ; kết hợp với  $A \rightarrow_{\mathfrak{K}AB} B$  và các phụ thuộc tầm thường  $A \rightarrow_{\mathfrak{K}A} A, C \rightarrow_{\mathfrak{K}C} C, E \rightarrow_{\mathfrak{K}E} E$ , theo Quy tắc hợp, ta được  $ACE \rightarrow_{\mathfrak{K}} ABCDE$ .

$R$  không ở dạng chuẩn  $\mathfrak{K}FBCNF$  vì tồn tại  $A \rightarrow_{\mathfrak{K}AB} B$  mà  $A$  không là siêu khóa mức  $\mathfrak{K}$  của  $R$ .  $R$  cũng không ở dạng chuẩn  $\mathfrak{K}F3NF$ , bởi vì tồn tại  $AC \rightarrow_{\mathfrak{K}'ACD} D \in F^+$  mà  $D$  không là thuộc tính khóa mức  $\mathfrak{K}$  và  $AC$  không là siêu khóa mức  $\mathfrak{K}$  của  $R$ . ■

Tiếp theo, chúng ta sẽ nghiên cứu việc tách lược đồ quan hệ như trong ví dụ 3.1 về các dạng chuẩn  $\mathfrak{R}F3NF$  và  $\mathfrak{R}FBCNF$  bảo toàn một số tính chất. Trước tiên, chúng ta quan tâm đến khái niệm phép chiếu một tập phụ thuộc hàm mờ sau đây.

Giả sử  $R$  là một lược đồ quan hệ của CSDL ngôn ngữ với tập thuộc tính  $U$  và  $F$  là một tập phụ thuộc hàm mờ đối với  $R$ ,  $V$  là tập con của  $U$ . Phép chiếu  $F$  lên  $V$  cho kết quả là một tập phụ thuộc hàm mờ, kí hiệu là  $\Pi_V(F)$ , với

$$\Pi_V(F) = \{X \rightarrow_{\mathfrak{R}'} A \mid X \rightarrow_{\mathfrak{R}'} A \in F^+, XA \subseteq V\}$$

Nếu chúng ta quan tâm một mức  $\mathfrak{R}$  xác định trên  $U$  thì chỉ những phụ thuộc hàm mờ trong tập  $\{X \rightarrow_{\mathfrak{R}'} A \mid X \rightarrow_{\mathfrak{R}'} A \in F^+, XA \subseteq U \text{ và } \mathfrak{R}'_X = \mathfrak{R}_X, \mathfrak{R}'_A \geq \mathfrak{R}_A\}$  mới có ý nghĩa và đáng được quan tâm. Vì vậy, đối với  $V \subseteq U$ , chúng ta định nghĩa

$$\Pi_V(F)|_{\mathfrak{R}} = \{X \rightarrow_{\mathfrak{R}'} A \mid X \rightarrow_{\mathfrak{R}'} A \in F^+, XA \subseteq V \text{ và } \mathfrak{R}'_X = \mathfrak{R}_X, \mathfrak{R}'_A \geq \mathfrak{R}_A\}$$

Rõ ràng,  $\Pi_V(F)|_{\mathfrak{R}} \subseteq \Pi_V(F)$ . Bây giờ, chúng ta sẽ nghiên cứu việc tách lược đồ quan hệ.

#### 4. TÁCH VỀ DẠNG CHUẨN $\mathfrak{R}F3NF$ BẢO TOÀN PHỤ THUỘC HÀM MỜ

Trong CSDL kinh điển, đối với lược đồ quan hệ  $R$  có tập thuộc tính  $U$  và tập phụ thuộc hàm  $D$  bao giờ cũng tồn tại một phép tách  $\rho = \{R_1, \dots, R_m\}$  sao cho mỗi  $R_i$ ,  $i = 1, \dots, m$  đều ở dạng chuẩn 3NF với tập phụ thuộc hàm  $D_i = \Pi_{U_i}(D) = \{X \rightarrow Y \mid X \rightarrow Y \in D^+, XY \subseteq U_i\}$ . Trong đó  $U_i$  là tập các thuộc tính của  $R_i$ . Vấn đề này sẽ xảy ra thế nào đối với CSDL ngôn ngữ và tập phụ thuộc hàm mờ. Trước tiên chúng ta trình bày các định nghĩa mở rộng về tập phụ thuộc hàm mờ tương đương và tập phụ thuộc hàm mờ tối tiểu.

**Định nghĩa 4.1.** Cho  $R$  là một lược đồ quan hệ của CSDL ngôn ngữ. Tập phụ thuộc hàm mờ  $F$  và  $G$  được gọi là tương đương nhau đối với  $R$  nếu và chỉ nếu  $F^+ = G^+$ .

**Định nghĩa 4.2.** Tập phụ thuộc hàm mờ  $F$  được gọi là tối tiểu nếu nó thoả mãn các điều kiện sau đây:

1. Vế phải của mỗi phụ thuộc hàm mờ trong  $F$  đều chứa đúng một thuộc tính;
2. Không có phụ thuộc hàm  $X \rightarrow_{\mathfrak{R}} A \in F$  mà  $F - \{X \rightarrow_{\mathfrak{R}} A\}$  tương đương với  $F$ ;
3. Không có phụ thuộc hàm  $X \rightarrow_{\mathfrak{R}} A \in F$  mà tồn tại  $Z \subset X$  sao cho  $(F - \{X \rightarrow_{\mathfrak{R}} A\}) \cup \{Z \rightarrow_{\mathfrak{R}} A\}$  tương đương với  $F$ .

Bây giờ, chúng ta trình bày thuật toán tách lược đồ quan hệ được mở rộng từ một thuật toán trong [8] như sau:

##### Thuật toán 1

+ Đầu vào: Lược đồ quan hệ  $R$  và tập phụ thuộc hàm mờ tối tiểu  $F$

+ Đầu ra:  $\rho = \{R_1, \dots, R_m\}$  sao cho mỗi  $R_i$  ở dạng chuẩn  $\mathfrak{R}_{(i)}F3NF$  và  $R_i$  thoả các phụ thuộc hàm trong  $\Pi_{U_i}(F)|_{\mathfrak{R}_{(i)}}$

+ Phương pháp:

1. Nếu tồn tại thuộc tính không thuộc bất kỳ vế trái hoặc vế phải của một phụ thuộc hàm mờ nào trong  $F$  thì loại bỏ thuộc tính đó khỏi  $R$

2. Nếu tồn tại phụ thuộc hàm mờ trong  $F$  bao hàm tất cả các thuộc tính của  $R$  thì đặt  $\rho = \{R\}$ , ngoài ra:

3. Với mỗi  $X \rightarrow_{\mathfrak{K}} A \in F$ , xác lập  $R_i$  với tập thuộc tính  $U_i = XA$  là một phần tử của phép tách  $\rho$ .

Trong thuật toán trên, mỗi phụ thuộc hàm mờ  $X \rightarrow_{\mathfrak{K}} A \in F$  có một mức  $\mathfrak{K}$  riêng và phép tách bảo toàn phụ thuộc hàm trong  $\Pi_{U_i}(F)|_{\mathfrak{K}}$  ứng với mức  $\mathfrak{K}$  đó.

**Mệnh đề 4.1.** *Giả sử  $R$  là một lược đồ quan hệ của CSDL ngôn ngữ,  $F$  là một tập phụ thuộc hàm tối thiểu trong  $R$  và  $\rho = \{R_1, \dots, R_m\}$  là một kết quả tách từ  $R$  bằng Thuật toán 1. Khi đó, mỗi  $R_i$  thỏa các phụ thuộc hàm trong  $\Pi_{U_i}(F)|_{\mathfrak{K}}$  và  $R_i$  ở dạng chuẩn  $\mathfrak{K}F3NF$  với  $i = 1, \dots, m$ .*

*Chứng minh.* Rõ ràng  $R_i$  thỏa các phụ thuộc hàm trong  $\Pi_{U_i}(F)|_{\mathfrak{K}}$  bởi vì  $\Pi_{U_i}(F)|_{\mathfrak{K}} \subseteq \Pi_{U_i}(F)$ . Bây giờ, chúng ta sẽ chứng minh  $R_i$  ở dạng chuẩn  $\mathfrak{K}F3NF$  đối với  $\Pi_{U_i}(F)|_{\mathfrak{K}}$ . Ở đây  $U_i = XA$  tương ứng với  $X \rightarrow_{\mathfrak{K}} A$  được xử lý trong bước 3 của Thuật toán 1. Để chứng minh phản chứng, giả sử rằng  $R_i$  không ở dạng chuẩn  $\mathfrak{K}F3NF$  đối với  $\Pi_{U_i}(F)|_{\mathfrak{K}}$ . Khi đó, tồn tại  $Y \rightarrow_{\mathfrak{K}'} B \in \Pi_{U_i}(F)|_{\mathfrak{K}}$  ( $B$  không thuộc  $Y$ ) mà  $Y$  không là siêu khóa mức  $\mathfrak{K}$  và  $B$  không là thuộc tính khóa mức  $\mathfrak{K}$  trong  $R_i$ . Để ý rằng  $A$  là thuộc tính đơn và  $YB \subseteq XA$ . Xét hai trường hợp sau đây:

Trường hợp 1:  $A = B$ . Khi đó  $YB \subseteq XB$ . Vì  $B$  không thuộc  $Y$  nên ta suy ra  $Y \subseteq X$ . Hơn nữa, vì  $X$  là siêu khóa mức  $\mathfrak{K}$  và  $Y$  không là siêu khóa mức  $\mathfrak{K}$  trong  $XA$  nên  $Y \subset X$ . Ta có  $Y \rightarrow_{\mathfrak{K}'} B \in \Pi_{U_i}(F)|_{\mathfrak{K}}$ , tức là  $Y \rightarrow_{\mathfrak{K}'} A \in \Pi_{U_i}(F)|_{\mathfrak{K}}$ . Theo định nghĩa của  $\Pi_{U_i}(F)|_{\mathfrak{K}}$  thì  $\mathfrak{K}'_Y = \mathfrak{K}'_Y, \mathfrak{K}'_A \geq \mathfrak{K}'_A$ . Theo (K3),  $Y \rightarrow_{\mathfrak{K}} A \in \Pi_{U_i}(F)|_{\mathfrak{K}}$  và do đó  $Y \rightarrow_{\mathfrak{K}} A \in F^+$ . Vì vậy, trong  $F$  ta có thể thay  $X \rightarrow_{\mathfrak{K}} A$  bằng  $Y \rightarrow_{\mathfrak{K}} A$  mà bao đóng  $F^+$  không thay đổi. Điều này mâu thuẫn với tính chất tối thiểu của  $F$ .

Trường hợp 2:  $A \neq B$ . Vì  $X \rightarrow_{\mathfrak{K}} A$  nên  $X$  là một siêu khóa mức  $\mathfrak{K}$  trong  $XA$  và tồn tại  $Z \subseteq X$  sao cho  $Z$  là một khóa mức  $\mathfrak{K}$  trong  $XA$ . Do  $A \neq B$  và  $B \subseteq XA$  nên tồn tại  $B'$  trong  $B$  và  $B' \in X$ . Hơn nữa,  $B'$  không là thuộc tính khóa mức  $\mathfrak{K}$  do  $B$  không chứa thuộc tính khóa mức  $\mathfrak{K}$  theo giả thiết phản chứng. Vì vậy  $B' \notin Z$ , ta suy ra  $Z \subset X$ . Như vậy ta có thể thay  $Z \rightarrow_{\mathfrak{K}} A$  cho  $X \rightarrow_{\mathfrak{K}} A$  trong  $F$  mà  $F^+$  không thay đổi. Điều này mâu thuẫn với tính chất tối thiểu của  $F$ .

Cả hai trường hợp đều xảy ra mâu thuẫn, mệnh đề đã được chứng minh. ■

**Ví dụ 4.1.**  $R$  là lược đồ quan hệ của CSDL ngôn ngữ, có tập thuộc tính  $U=ABCD$ . Tập phụ thuộc hàm mờ  $F = \{A \rightarrow_{\mathfrak{K}AB} B, AC \rightarrow_{\mathfrak{K}'ACD} D, A \rightarrow_{\mathfrak{K}AD} D\}$ , trong đó mức tương tự  $\mathfrak{K} = (\infty, 2, \infty, 1)$ ,  $\mathfrak{K}' = (\infty, 2, \infty, 2)$ .  $R$  không ở dạng chuẩn  $\mathfrak{K}F2NF$  vì tồn tại  $D$ , không là thuộc tính khóa mức  $\mathfrak{K}$ , phụ thuộc hàm bộ phận mức  $\mathfrak{K}$  vào  $AC$  do tồn tại  $AC \rightarrow_{\mathfrak{K}ACD} D$  và  $A \rightarrow_{\mathfrak{K}AD} D$ . Trong đó,  $AC \rightarrow_{\mathfrak{K}ACD} D \in F^+$  được suy ra từ  $AC \rightarrow_{\mathfrak{K}'ACD} D$  bởi (K3). Sử dụng Thuật toán 1, tách  $R$  thành  $\rho = (AB, ACD, AD)$ . Rõ ràng,  $\rho$  bảo toàn các phụ thuộc hàm mờ.  $AB$  ở dạng chuẩn  $\mathfrak{K}F3NF$  đối với  $\Pi_{AB}(F)|_{\mathfrak{K}}$ ,  $ACD$  ở dạng chuẩn  $\mathfrak{K}'F3NF$  đối với  $\Pi_{ACD}(F)|_{\mathfrak{K}'}$  và  $AD$  ở dạng chuẩn  $\mathfrak{K}F3NF$  đối với  $\Pi_{AD}(F)|_{\mathfrak{K}}$ . Để ý rằng  $A \rightarrow_{\mathfrak{K}AD} D$  không thuộc  $\Pi_{ACD}(F)|_{\mathfrak{K}'}$  do  $\mathfrak{K}_D \leq \mathfrak{K}'_D$ .

## 5. TÁCH VỀ DẠNG CHUẨN $\mathfrak{K}FBCNF$ BẢO TOÀN THÔNG TIN

### 5.1. Định nghĩa phép kết nối tự nhiên trong CSDL ngôn ngữ

Trong CSDL kinh điển, phép tách lược đồ quan hệ  $R$  với tập phụ thuộc hàm  $F$  thành  $\rho = (R_1, \dots, R_m)$  được gọi là phép tách bảo toàn thông tin (hay tách không tổn thất) nếu

như kết nối tự nhiên các hình chiếu của  $R$  lên các thành phần  $R_i$  có tập thuộc tính  $U_i$  tương ứng ta thu được chính quan hệ  $R$ . Điều này nghĩa là đối với mỗi quan hệ  $r \in R$  thỏa  $F$ , ta có  $r = r_1 * \dots * r_m$ , trong đó  $r_i \in R_i$  và  $r_i = \Pi_{U_i}(r), i = 1, \dots, m$ . Khi thực hiện phép kết nối tự nhiên giữa hai quan hệ, kết quả thu được phụ thuộc vào sự bằng nhau của dữ liệu trên các thuộc tính chung của hai quan hệ. Khi quan hệ bằng nhau thông thường được thay bởi quan hệ bằng nhau mức  $\mathfrak{K}$  trong CSDL ngôn ngữ một số tình huống xảy ra trong ví dụ sau.

**Ví dụ 5.1.** Cho  $R = ABC$ ,  $R_1 = \Pi_{AB}(R)$ ,  $R_2 = \Pi_{BC}(R)$  tương ứng với các quan hệ:

$$r = \{a_1b_1c_1, a_1b_1c_2, a_2b_2c_2\}, r_1 = \{a_1b_1, a_2b_2\}, r_2 = \{b_1c_1, b_1c_2, b_2c_2\}$$

Nếu xét quan hệ bằng nhau thông thường thì  $r_1 * r_2 = r$  nhưng nếu dùng quan hệ bằng nhau mức  $\mathfrak{K}$ , giả sử  $b_1 =_k b_2$  và đặt  $b'$  là giá trị bằng nhau mức  $k$  đối với  $b_1$  và  $b_2$  thì  $r_1 * r_2 \neq r$ , bởi vì  $r_1 * r_2 = \{a_1b_1c_1, a_1b_1c_2, a_1b'c_2, a_2b'c_1, a_2b'c_2, a_2b_2c_2\}$ .

Để hạn chế các bộ sinh ra từ các cặp có giá trị bằng nhau mức  $\mathfrak{K}$  trên những thuộc tính chung, khi xử lý trên CSDL ngôn ngữ, chúng ta sẽ giới hạn phép kết nối tự nhiên chỉ thực hiện đối với những cặp có giá trị bằng nhau thông thường, không sử dụng quan hệ bằng nhau mức  $\mathfrak{K}$  trên những thuộc tính chung. Giả sử  $U = \{A_1, \dots, A_n\}$ ,  $Dom(R), Dom(S) \subseteq U$ , một cách hình thức:

$$r * s = \{(r_1, \dots, r_{k-1}, r_k, \dots, r_m, s_{m+1}, \dots, s_n) | r_i = s_i, i = k, \dots, m\}, \forall r \in R, \forall s \in S$$

Với phép kết nối tự nhiên như vậy, chúng ta định nghĩa phép tách bảo toàn thông tin trong CSDL ngôn ngữ.

**Định nghĩa 5.1.** Cho  $R$  là lược đồ quan hệ CSDL ngôn ngữ có tập thuộc tính  $U$ ,  $F$  là tập phụ thuộc hàm mờ đối với  $R$ . Phép tách  $\rho = (R_1, \dots, R_m)$  được gọi là bảo toàn thông tin đối với  $R$  khi và chỉ khi với mọi  $r \in R$ ,  $r$  thỏa  $F$  ta có  $r = \Pi_{U_1}(r) * \dots * \Pi_{U_m}(r)$ , trong đó  $U_i$  là tập thuộc tính của  $R_i$  tương ứng.

**Bổ đề 5.1.** Cho  $R$  là một lược đồ CSDL ngôn ngữ với tập phụ thuộc hàm mờ  $F$  trên tập thuộc tính  $U$  và  $\rho = (R_1, \dots, R_m)$  là một phép tách của  $R$ . Khi đó,

a) Nếu  $\rho$  là phép tách bảo toàn thông tin và thay thế  $R_i$  bởi phép tách  $\sigma = (S_1, \dots, S_n)$  bảo toàn thông tin thì phép tách  $\tau = (R_1, \dots, R_{i-1}, S_1, \dots, S_n, R_{i+1}, \dots, R_m)$  bảo toàn thông tin đối với  $F$ .

b) Nếu  $\rho = (R_1, \dots, R_m)$  là phép tách bảo toàn thông tin thì bổ sung vào  $\rho$  một số lược đồ quan hệ  $(R_{m+1}, \dots, R_{m+k})$  trên  $U$  ta được phép tách  $\tau = (R_1, \dots, R_m, R_{m+1}, \dots, R_{m+k})$  bảo toàn thông tin đối với  $F$ .

*Chứng minh.* Chứng minh tương tự như trong [8].

## 5.2. Tách bảo toàn thông tin về các dạng chuẩn $\mathfrak{K}$ FBCNF

Phép tách bảo toàn thông tin được đặt ra đối với các lược đồ quan hệ  $R$  không ở dạng chuẩn  $\mathfrak{K}$ FBCNF. Trong CSDL kinh điển, một quan hệ  $R$  với tập phụ thuộc hàm  $F$  không ở dạng chuẩn BCNF đều có thể tách bảo toàn thông tin thành các quan hệ con ở dạng chuẩn BCNF. Trên mô hình CSDL ngôn ngữ, ngữ nghĩa của phụ thuộc hàm mờ mức  $\mathfrak{K}$  và phép chiếu tập phụ thuộc hàm mờ đã định nghĩa, chúng ta cũng có một kết quả tương tự bằng việc mở rộng Thuật toán 5.3 trong [8]



**Thuật toán 2.** Tách bảo toàn thông tin về các dạng chuẩn  $\mathfrak{R}FBCNF$

- + Đầu vào: Lược đồ quan hệ  $R$ , tập phụ thuộc hàm mờ  $F$  và một mức  $\mathfrak{K}$
- + Đầu ra:  $\rho$  là một tách của  $R$
- + Phương pháp:
  1. Đặt  $\rho = \{R\}$
  2. While (tồn tại  $S \in \rho$  không ở dạng chuẩn  $\mathfrak{R}FBCNF$ ) do
    - Chọn phụ thuộc hàm mờ  $X \rightarrow_{\mathfrak{K}'} A \in \Pi_S(F)|_{\mathfrak{K}}$  vi phạm chuẩn  $\mathfrak{R}FBCNF$  trong  $S$ :
      - a. Đặt  $S_1 = XA$ ;
      - b. Đặt  $S_2 = S - A$ .
      - c.  $\rho = \rho - \{S\} \cup \{S_1, S_2\}$

**Mệnh đề 5.1.** Giả sử  $R$  là một lược đồ quan hệ của cơ sở dữ liệu ngôn ngữ,  $F$  là một tập phụ thuộc hàm mờ trong  $R$ ,  $\mathfrak{K}$  là một mức tương tự cho trước và  $\rho = \{R_1, \dots, R_m\}$  là kết quả tách từ  $R$  bằng Thuật toán 2. Khi đó mọi  $R_i, i = 1, \dots, m$  đều ở dạng chuẩn  $\mathfrak{R}FBCNF$  và  $\rho$  bảo toàn thông tin.

*Chứng minh.* Trước tiên, chúng ta chứng minh  $R_i, i = 1, \dots, m$  đều ở dạng chuẩn  $\mathfrak{R}FBCNF$ . Thuật toán 2, tách mỗi lược đồ quan hệ không ở dạng chuẩn  $\mathfrak{R}FBCNF$  thành các lược đồ quan hệ có số thuộc tính ít hơn và Thuật toán dừng sau hữu hạn bước. Khi đó, ta thu được kết quả là các lược đồ quan hệ ở dạng chuẩn  $\mathfrak{R}FBCNF$ . Thật vậy:

Giả sử, tồn tại  $S \in \rho$  không ở dạng chuẩn  $\mathfrak{R}FBCNF$  vì có phụ thuộc hàm mờ  $X \rightarrow_{\mathfrak{K}'} A \in \Pi_S(F)|_{\mathfrak{K}}$  với  $A \notin X$  mà  $X$  không là siêu khóa mức  $\mathfrak{K}$  đối với  $S$ . Thuật toán 2 chia  $S$  thành hai lược đồ quan hệ  $S_1 = XA$  và  $S_2 = S - A$ . Rõ ràng,  $S_2 \subset S$ . Chúng ta sẽ chứng minh  $S_1$  cũng là tập con thực sự của  $S$ . Giả sử ngược lại  $S_1 = S$  tức  $XA = S$ . Theo giả thiết  $X \rightarrow_{\mathfrak{K}'} A \in \Pi_S(F)|_{\mathfrak{K}}$  thỏa  $S$  tức là  $X \rightarrow_{\mathfrak{K}'} A$  với  $\mathfrak{K}'_X = \mathfrak{K}_X$  và  $\mathfrak{K}'_A \geq \mathfrak{K}_A$ . Do đó  $X \rightarrow_{\mathfrak{K}} A$  theo (K3). Hơn nữa, theo (K1)  $X \rightarrow_{\mathfrak{K}} X$ . Theo quy tắc hợp, ta suy ra  $X \rightarrow_{\mathfrak{K}} XA$  tức là  $X \rightarrow_{\mathfrak{K}} S$ , mâu thuẫn với  $X$  không là siêu khóa mức  $\mathfrak{K}$  của  $S$ . Như vậy, số thuộc tính trong  $S_1$  cũng như  $S_2$  luôn ít hơn trong  $S$ .

Thuật toán dừng sau hữu hạn bước vì số thuộc tính trong mỗi lược đồ quan hệ ít dần. Nếu  $S$  chỉ có một thuộc tính thì  $S$  hiển nhiên ở dạng chuẩn  $\mathfrak{R}FBCNF$ . Nếu  $S$  có đúng hai thuộc tính chẳng hạn  $S = AB$  và tồn tại một phụ thuộc hàm mờ, chẳng hạn  $A \rightarrow_{\mathfrak{K}'} B \in \Pi_{AB}(F)|_{\mathfrak{K}}$  thì chứng minh tương tự như trên, ta được  $A \rightarrow_{\mathfrak{K}} AB$ , tức  $A$  là một siêu khóa mức  $\mathfrak{K}$  của  $S$ . Do đó  $S$  ở dạng chuẩn  $\mathfrak{R}FBCNF$ .

Bây giờ ta sẽ chứng minh phép tách  $\rho$  bảo toàn thông tin. Thật vậy, khi khởi tạo  $\rho = \{R\}$ . Tại một thời điểm nào đó, giả sử một lược đồ quan hệ  $S \in \rho$  có  $X \rightarrow_{\mathfrak{K}'} A \in \Pi_S(F)|_{\mathfrak{K}}$  không thỏa điều kiện của  $\mathfrak{R}FBCNF$ . Theo Thuật toán 2,  $S$  được tách thành hai lược đồ quan hệ  $S_1 = XA$  và  $S_2 = S - A$ . Như vậy  $S_1 \cap S_2 = X$ ,  $S_1 - S_2 = A$  và theo giả thiết  $X \rightarrow_{\mathfrak{K}'} A \in \Pi_S(F)|_{\mathfrak{K}}$ , tức là  $S_1 \cap S_2 \rightarrow_{\mathfrak{K}'} S_1 - S_2$ . Tương tự phép chứng minh Định lý 5.5 [8], phép tách  $S$  thành  $(S_1, S_2)$  bảo toàn thông tin. Theo Bổ đề 5.1,  $\rho$  bảo toàn thông tin.

**Ví dụ 5.2.**  $R$  là lược đồ quan hệ của CSDL ngôn ngữ, có tập thuộc tính  $U=ABCDE$ . Mức tương tự  $\mathfrak{K} = (\infty, 2, \infty, 1, 2)$ . Tập phụ thuộc hàm mờ  $F = \{A \rightarrow_{\mathfrak{K}_{AB}} B, AC \rightarrow_{\mathfrak{K}'_{ACD}} D\}$ , ở đây  $\mathfrak{K}' = (\infty, 2, \infty, 2, 2)$ .  $R$  không ở dạng chuẩn  $\mathfrak{R}FBCNF$  vì tồn tại  $A \rightarrow_{\mathfrak{K}_{AB}} B$  mà  $A$  không là siêu khóa mức  $\mathfrak{K}$  của  $R$ . Sử dụng Thuật toán 2 tách  $R$  thành  $\rho = \{AB, ACDE\}$ .  $AB$  ở dạng chuẩn  $\mathfrak{R}FBCNF$  đối với tập phụ thuộc hàm mờ  $\Pi_{AB}(F)|_{\mathfrak{K}} = \{A \rightarrow_{\mathfrak{K}_{AB}} B\}$ .  $ACDE$  không ở dạng chuẩn  $\mathfrak{R}FBCNF$  đối với tập phụ thuộc hàm mờ là  $\Pi_{ACDE}(F)|_{\mathfrak{K}} = \{AC \rightarrow_{\mathfrak{K}'_{ACDE}} D\}$

vì trong lược đồ ACDE, AC không là siêu khóa mức  $\mathfrak{K}$ . Tách ACDE thành  $\{ACD, ACE\}$ . Lúc này, ACD ở dạng chuẩn  $\mathfrak{K}$ FBCNF (do AC là khóa mức  $\mathfrak{K}$  trong ACD, điều này được suy ra từ  $AC \rightarrow_{\mathfrak{K}ACD} D$  bởi vì  $\mathfrak{K}'_{AC} = \mathfrak{K}_{AC}$  và  $\mathfrak{K}'_D \geq \mathfrak{K}_D$ ). ACE ở dạng chuẩn  $\mathfrak{K}$ FBCNF là tầm thường. Như vậy,  $R$  được tách thành  $\rho = \{AB, ACD, ACE\}$ .

## 6. KẾT LUẬN

Trong cơ sở dữ liệu ngôn ngữ, phụ thuộc hàm mờ được nghiên cứu dựa trên quan hệ tương tự để phân hoạch miền trị của các thuộc tính theo mức  $\mathfrak{K}$ . Như một tham số, mức  $\mathfrak{K}$  có thể thay đổi theo ý nghĩa của từng ứng dụng. Vì vậy, mở rộng của các dạng chuẩn từ cơ sở dữ liệu kinh điển sang cơ sở dữ liệu ngôn ngữ là không tầm thường. Khóa mức  $\mathfrak{K}$  và một số dạng chuẩn mờ mức  $\mathfrak{K}$  được định nghĩa. Giới hạn các phụ thuộc hàm mờ hạn chế theo mức  $\mathfrak{K}$  cho phép tách một lược đồ cơ sở dữ liệu ngôn ngữ thành các lược đồ ở dạng  $\mathfrak{K}$ 3NF thỏa các phụ thuộc hàm mờ trong  $\Pi_{U_i}(F)|_{\mathfrak{K}}$  cũng như tách một lược đồ cơ sở dữ liệu ngôn ngữ về dạng chuẩn  $\mathfrak{K}$ FBCNF bảo toàn thông tin đóng góp một phần trong định hướng thiết kế các cơ sở dữ liệu ngôn ngữ. Các dạng chuẩn mờ liên quan đến phụ thuộc đa trị trong cơ sở dữ liệu ngôn ngữ [7] là những vấn đề cần nghiên cứu tiếp theo.

## TÀI LIỆU THAM KHẢO

- [1] G. Chen, E. E. Kerre, J. Vandenbulcke, Fuzzy normal forms and a dependency-preserving decomposition into  $\varphi$ -F3NF, *Proceedings 3rd, IEEE International Conference on Fuzzy Systems*, IEEE Computer Society Press, Los Alamitos, 1994 (156–161).
- [2] J. Mishra and S. Ghosh, Normalization in a fuzzy relational database model, *International Journal of Computer Engineering - Technology* **3** (2) (2012) 506–517.
- [3] N. C. Ho, L. X. Vinh, On some properties of ordering relation in non-homogeneous hedge algebras, *Tạp chí Tin học và Điều khiển học* **19** (4) (2003) 373–381.
- [4] N. C. Hồ, L. X. Vinh, N. C. Hào, Thống nhất dữ liệu và xây dựng quan hệ tương tự trong cơ sở dữ liệu ngôn ngữ bằng đại số gia tử, *Tạp chí Tin học và Điều khiển học* **25** (4) (2009) 314–332.
- [5] Nguyễn Cát Hồ, Vũ Minh Lộc, Hoàng Tùng, Nguyễn Tân Ân, Phụ thuộc hàm mờ trong CSDL quan hệ với dữ liệu ngôn ngữ, *Tạp chí Khoa học và Công nghệ* **51** (2) (2013) 137–151.
- [6] S. Sheno, A. Melton and L. T. Fan, Functional dependencies and normal forms in the fuzzy relational database model, *Information Sciences* **60** (1992) 1–28.
- [7] L. X. Vinh, T. T. Thành, Một số vấn đề về phụ thuộc đa trị trong cơ sở dữ liệu mờ chứa dữ liệu ngôn ngữ, *Tạp chí Tin học và Điều khiển học* **28** (1) (2012) 88–102.
- [8] J. D. Ullman, *Principles of Database Systems*, Computer Sciences Press Inc, 1982.

Ngày nhận bài 04 - 6 - 2013

Nhận lại sau sửa ngày 15 - 10 - 2013