

NGŨ NGHĨA DỮ LIỆU TRONG CƠ SỞ DỮ LIỆU QUAN HỆ CÓ GIÁ TRỊ NULL PHỤ THUỘC NGŨ CẢNH

BÙI THỊ THÚY HIỀN

Abstract. In this paper, we consider semantics of data in relational databases with context dependent null values by giving a formal definition of model for a relation which contains context dependent nulls. Some illustrative examples of data semantics are also given.

1. MỞ ĐẦU

Trong [1], chúng tôi đã giới thiệu về null phụ thuộc ngữ cảnh. Thực chất của null phụ thuộc ngữ cảnh là để biểu diễn những giá trị chưa biết và phụ thuộc vào ngữ cảnh. Đối với mỗi null phụ thuộc ngữ cảnh, người ta xem xét các đặc trưng ngữ cảnh của chúng, kí hiệu chúng bởi các kí tự $\delta_1, \delta_2, \delta_3, \dots$

Để nghiên cứu ngữ nghĩa dữ liệu của null phụ thuộc ngữ cảnh, trước hết hãy xét tình huống sau:

Cho quan hệ người (*tên, tuổi, môn*) lưu thông tin về họ tên, tuổi và môn dạy của các giáo viên thuộc một khoa của một trường đại học; giả sử hệ thống nhận được các thông tin:

- (i) Có một giáo viên tên là A, 45 tuổi.
- (ii) Có một giáo viên tên là B hiện không dạy môn học nào.
- (iii) Có một giáo viên tên là C đang dạy một môn học.

Thông tin (i) nhận được là thông tin "mở" về môn dạy, nghĩa là người giáo viên tên là A có thể dạy một môn học, có thể dạy nhiều môn học, hoặc cũng có thể không dạy môn học nào. Ở trường hợp này nếu chỉ dùng một null phụ thuộc ngữ cảnh δ để biểu diễn cho thông tin về môn dạy, thì khi xét ngữ nghĩa dữ liệu của null phụ thuộc ngữ cảnh người ta đều phải gán cho nó các tình huống để có thể đại diện cho một giá trị xác định, hoặc cho vài giá trị xác định, hoặc là cho một giá trị không tồn tại. Khi đó quá trình xem xét ngữ nghĩa dữ liệu của các quan hệ có chứa giá trị null sẽ phức tạp lên rất nhiều.

Thông tin (ii) cho biết giá trị tại thuộc tính môn dạy của giáo viên B là không tồn tại. Trong trường hợp này, người ta có thể sử dụng null không tồn tại dne thay cho một null phụ thuộc ngữ cảnh δ . Vì giá trị dne là giá trị xác định nên nó mang nhiều thông tin hơn giá trị không xác định δ .

Thông tin (iii) cho biết giáo viên C đang dạy một môn học nên giá trị null tại vị trí môn học sẽ biểu thị là có tồn tại một giá trị xác định nhưng chưa biết.

Từ những tình huống trên cho thấy việc phân loại null phụ thuộc ngữ cảnh cho phù hợp với từng trường hợp thiếu thông tin sẽ hợp lý hơn.

Trường hợp (i) chúng ta sử dụng ý tưởng null "mở" của Gottlob và Zicari [5]:

Cho t là một bộ dữ liệu, A là một thuộc tính, nếu $t[A]$ được gán là null mở (viết $t[A] = \text{open}$) thì thuộc tính A sẽ được xem xét dưới giả thiết thế giới mở (OWA) và vì thế $t[A]$ có thể biểu thị cho một giá trị không tồn tại, có thể biểu thị cho một giá trị xác định hoặc có thể biểu thị vài giá trị xác định, nói cách khác mọi thông tin vẫn còn là mở đối với $t[A]$.

Khái niệm null phụ thuộc ngữ cảnh mở có thể được mở rộng như sau: Một null phụ thuộc ngữ cảnh mở là một null mở và giá trị null đó phụ thuộc vào ngữ cảnh trong CSDL.

Như vậy với trường hợp (i) có thể sử dụng một null phụ thuộc ngữ cảnh mở để biểu diễn thông tin bị thiếu.

Trường hợp (ii) có thể sử dụng null không tồn tại dne.

Trường hợp (iii) có thể sử dụng null phụ thuộc ngữ cảnh chưa biết để biểu thị một giá trị tồn tại nhưng chưa biết và phụ thuộc vào ngữ cảnh.

Về vai trò của null mở có thể tìm thấy trong [5]. Điều muốn được nhấn mạnh ở đây là khả năng biểu diễn thông tin cao khi kết hợp giữa null mở, null chưa biết và null không tồn tại. Có lẽ, không có tình huống thiếu thông tin nào là không biểu diễn chính xác được bởi sự kết hợp giữa ba kiểu null trên. Như vậy cách mà chúng ta lựa chọn khi coi một CSDL null phụ thuộc ngữ cảnh gồm có null phụ thuộc ngữ cảnh chưa biết, null phụ thuộc ngữ cảnh mở và null không tồn tại là hoàn toàn hợp lý.

Để phân biệt với null phụ thuộc ngữ cảnh chưa biết, null phụ thuộc ngữ cảnh mở sẽ được kí hiệu bởi các kí tự như β_1, β_2, \dots

Xét một ví dụ với ba thể hiện r_a, r_b, r_c khác nhau của R

r_a	Tên	Khoa	Điện.thoại	r_b	Tên	Khoa	Điện.thoại	r_c	Tên	Khoa	Điện.thoại
	Vinh	Toán	8543647		Vinh	Toán	8543647		Vinh	Toán	8543647
	Hùng	Toán	δ_1		Vinh	Toán	β_1		β_1	β_2	β_3
	Kiên	Toán	dne								

Trong thể hiện r_a không thuộc tính nào có giá trị mở.

Với thể hiện r_b , có thể suy ra những sự kiện sau:

- Vinh là nhân viên duy nhất.
- Khoa toán là khoa duy nhất.
- Vinh có một số điện thoại: 8543647.
- Vinh có thể không có hoặc còn có thêm một vài số điện thoại nữa.

Ngoài ra, không một sự kiện nào khác là đúng!

Giá trị mở β_1 trong r_b không có nghĩa thuộc tính Điện.thoại có miền giá trị là các tập hợp. Nếu Vinh có thêm vài số điện thoại thì những số điện thoại này sẽ được biểu diễn bởi một tập những bộ, với cùng giá trị Tên và Khoa còn số Điện.thoại thì khác nhau.

Thể hiện r_c mô tả một CSDL dưới giả thiết thể giới mở. Do tất cả những giá trị thuộc tính của bộ thứ hai là mở nên chỉ nói được Vinh làm việc tại khoa Toán, có số điện thoại là 8543647. Ngoài ra có thể còn những nhân viên khác nhưng chưa được biết.

2. NGŨ NGHĨA DỮ LIỆU CỦA QUAN HỆ CÓ CHỨA NULL PHỤ THUỘC NGŨ CẢNH

2.1. Tập khả năng của quan hệ có null thuộc ngữ cảnh

Cho $R(A_1, \dots, A_n)$ là một lược đồ quan hệ được xác định trên một tập thuộc tính A_1, \dots, A_n .

Với mỗi thuộc tính A_i , ta kí hiệu miền giá trị tương ứng là $\text{Dom}(A_i)$. Miền của R là tích Đề các $\text{Dom}(A_1) \times \text{Dom}(A_2) \times \dots \times \text{Dom}(A_n)$ và kí hiệu là $\text{Dom}(R)$.

Chúng ta mở rộng mỗi miền $\text{Dom}(A_i)$ thành $\text{Dom}^*(A_i)$ bằng cách thêm vào một tập hữu hạn các kí hiệu null: $\text{Dom}^*(A_i) = \text{Dom}(A_i) \cup \Delta_{i_1} \cup \Delta_{i_2} \cup \{\text{dne}\}$ trong đó

- Δ_{i_1} là tập null phụ thuộc ngữ cảnh chưa biết và thường được kí hiệu bởi các kí tự Hy Lạp như $\delta_1, \delta_2, \dots$

- Δ_{i_2} là tập null phụ thuộc ngữ cảnh mở và được kí hiệu bởi β_1, β_2, \dots

- dne là kí hiệu cho null không tồn tại.

- $\text{Dom}(A_i), \Delta_{i_1}, \Delta_{i_2}, \{\text{dne}\}$ là các tập không giao nhau.

Tương tự một sự mở rộng của $\text{Dom}(R)$ là $\text{Dom}^*(R) = \text{Dom}^*(A_1) \times \dots \times \text{Dom}^*(A_n)$.

Một thể hiện (hay một quan hệ) có chứa null phụ thuộc ngữ cảnh của một lược đồ R là một tập con của $\text{Dom}^*(R)$. Những quan hệ như vậy được kí hiệu bởi các kí tự thường như r, r_1, \dots và được gọi là các quan hệ một phần.

Tập tất cả các quan hệ một phần trên lược đồ R được kí hiệu là $\text{rel}_1(R)$.

Các quan hệ không chứa null được gọi là quan hệ toàn phần, tập tất cả các quan hệ toàn phần trên lược đồ R được kí hiệu là $rel(R)$.

Một bộ của một quan hệ r là một phần tử của r . Chúng ta kí hiệu những bộ của r bởi những kí tự như t, t', s, s', \dots . Nếu t là bộ của một quan hệ r thì $t[A_i]$ là giá trị của t tại thuộc tính A_i . Nếu $t[A_i]$ khác null ta viết $t[A_i]!$.

Kí hiệu open để chỉ null phụ thuộc ngữ cảnh mở và unk để chỉ null phụ thuộc ngữ cảnh chưa biết.

Tập tất cả các null phụ thuộc ngữ cảnh chưa biết hoặc null phụ thuộc ngữ cảnh mở được gọi chung là các giá trị không xác định, còn những giá trị khác null hoặc dne được gọi là các giá trị xác định.

Do giá trị của null không tồn tại dne được qui định bởi những giá trị đã biết khác trong CSDL nên null không tồn tại dne cũng là null phụ thuộc ngữ cảnh.

Cho t_1 và t_2 là hai bộ (có thể chứa null phụ thuộc ngữ cảnh) trên lược đồ R . A là một thuộc tính, ta kí hiệu $t_1[A] = t_2[A]$ nếu

1. $t_1[A]!$, $t_2[A]!$ và $t_1[A] = t_2[A]$, hoặc
2. $t_1[A] = \delta_i$, $t_2[A] = \delta_j$ và $i = j$, hoặc
3. $t_1[A] = \beta_i$, $t_2[A] = \beta_j$ và $i = j$, hoặc
4. $t_1[A] = dne$ và $t_2[A] = dne$.

Ta viết $t_1[X] = t_2[X]$ nếu $t_1[A] = t_2[A] \forall A \in X$.

Ý nghĩa của phép so sánh $=$ là kiểm tra sự trùng nhau về kí hiệu của các giá trị trong CSDL. Ví dụ $3 = 3$, $\delta_i = \delta_i$, $\beta_j = \beta_j$ và $dne = dne$.

Ngược lại với phép so sánh $=$ là phép so sánh \neq . Ví dụ $3 \neq 4$, $\delta_i \neq \delta_j$, $\delta_i \neq \beta_i$.

Định nghĩa mà chúng ta sẽ phát biểu sau đây đề cập tới tập khả năng của một lược đồ quan hệ. Mỗi khả năng của một quan hệ r sẽ chứa tập tất cả các bộ thuộc r sau khi đã thay thế các giá trị null bởi các giá trị xác định. Ví dụ, xét một quan hệ $r : \{(a, \delta_i, b)\}$ với miền của thuộc tính giữa là $\{1, 2, 3\}$. Khi đó một khả năng của r là $\langle a, 1, b \rangle$ và một khả năng khác sẽ là $\langle a, 2, b \rangle$.

Khi thay thế những giá trị null, do tính chất của các null không tồn tại nên chúng ta sẽ đưa ra kí hiệu \perp để biểu diễn cho dne trong tập khả năng. Ví dụ, xét một quan hệ $r : \langle a, dne, \delta \rangle$, khi đó với miền giá trị của thuộc tính cuối là $\{c, d\}$ thì có hai khả năng cho r là $\langle a, \perp, c \rangle$ và $\langle a, \perp, d \rangle$.

Định nghĩa 1 (khả năng [5]). Cho $R(A_1, \dots, A_n)$ là một lược đồ quan hệ. Khi đó một khả năng của R là một tập con của $Dom^+(R) = Dom^+(A_1) \times \dots \times Dom^+(A_n)$ trong đó $Dom^+(A_i) = Dom(A_i) \cup \{\perp\}$ và $1 \leq i \leq n$.

Tập tất cả các khả năng của R được kí hiệu bằng \hat{R} .

Định nghĩa 2. Cho r là một quan hệ, a là một null phụ thuộc ngữ cảnh ($a \in \Delta_{i_1} \cup \Delta_{i_2} \cup \{dne\}$) xuất hiện tại cột tính A .

i) Một giá trị hoặc một tập giá trị V được gọi là một phép thế có thể của a nếu:

- $V = \emptyset$ hoặc
- $V \neq \emptyset$ và

nếu $a \in \Delta_{i_1}$ thì $V \in Dom(A)$

nếu $a = dne$ thì $V = \perp$

nếu $a \in \Delta_{i_2}$ thì: hoặc là $V = \perp$ hoặc là $V \equiv \{b_1, b_2, \dots, b_m\}$ với $m \geq 1$ và với $1 \leq i \leq m : b_i \in Dom(A)$.

ii) Nếu V là một phép thế có thể của a , ta kí hiệu $r' = S_a^V(r)$ là một quan hệ có được từ r bằng cách:

- nếu $V \neq \emptyset$ thì thay mọi xuất hiện của a trong r bởi V ,
- nếu $V = \emptyset$ thì không thực hiện thay thế a , tức là $r' = r$.

iii) Cho a_1, a_2, \dots, a_k là các null phụ thuộc ngữ cảnh xuất hiện trong r ; V_1, V_2, \dots, V_k tương ứng là các phép thế có thể của a_1, a_2, \dots, a_k . Khi đó $r' = S_{a_1 a_2 \dots a_k}^{V_1 V_2 \dots V_k}(r)$ là quan hệ có được từ r bằng cách với $1 \leq i \leq k$:

- nếu $V_i \neq \emptyset$ thì thay mọi xuất hiện của a_i trong r bởi V_i ,
- nếu $V_i = \emptyset$ thì không thực hiện thay thế a_i .

Ví dụ 1. Xét $r : \{(\beta_1, 1, \delta_1), (\beta_1, 2, 3)\}$ với miền thuộc tính đầu là $\{1, 2, 3\}$, khi đó $r_1 = S_{\beta_1}^{\perp}(r)$ là: $\{(\perp, 1, \delta_1), (\perp, 2, 3,)\}$ và $r_2 = S_{\beta_1}^{\{1,3\}}(r)$ là: $\{(1, 1, \delta_1), (1, 2, 3), (3, 1, \delta_1), (3, 2, 3)\}$.

Mệnh đề 1. Cho r_1, r_2, \dots, r_m là m quan hệ trên cùng một lược đồ; a_1, a_2, \dots, a_k là các null phụ thuộc ngữ cảnh có các khả năng thay thế tương ứng là V_1, V_2, \dots, V_k . Khi đó ta có:

- i) $S_{a_1 a_2 \dots a_k}^{V_1 V_2 \dots V_k}(r_1 \cup \dots \cup r_m) = S_{a_1 a_2 \dots a_k}^{V_1 V_2 \dots V_k}(r_1) \cup \dots \cup S_{a_1 a_2 \dots a_k}^{V_1 V_2 \dots V_k}(r_m)$.
- ii) $S_{a_1 a_2 \dots a_k}^{V_1 V_2 \dots V_k}(r_1 \cap \dots \cap r_m) = S_{a_1 a_2 \dots a_k}^{V_1 V_2 \dots V_k}(r_1) \cap \dots \cap S_{a_1 a_2 \dots a_k}^{V_1 V_2 \dots V_k}(r_m)$.

Chứng minh. Suy trực tiếp từ Định nghĩa 2 và Định nghĩa của các Phép toán quan hệ Hợp, Giao.

2.2. Ngữ nghĩa dữ liệu và mô hình của quan hệ có null phụ thuộc ngữ cảnh

Cho r là một quan hệ trên lược đồ quan hệ $R(A_1, \dots, A_n)$; $f \in \hat{R}$ là một khả năng của r .

Định nghĩa 3. Cho t là một bộ phận của quan hệ r , t_f là một bộ của f . Ta nói t suy ra t_f (kí hiệu $t \triangleright t_f$) nếu và chỉ nếu với $1 \leq i \leq n$ ít nhất một trong bốn điều kiện sau phải được thỏa mãn

1. $t[A_i] = t_f[A_i]$,
2. $t[A_i] = \text{open}$,
3. $(t[A_i] = \text{unk}$ và $t_f[A_i] \neq \perp$),
4. $(t[A_i] = \text{dne}$ và $t_f[A_i] = \perp$).

Ví dụ 2. $t = (1, \text{open}, \text{unk}, \text{dne}, 5) \triangleright (1, 2, 4, \perp, 5)$; $t \triangleright (1, \perp, 1, \perp, 5)$.

Định nghĩa 4. Một khả năng f của r được gọi là một mô hình của quan hệ r nếu và chỉ nếu nó thỏa mãn ba điều kiện sau:

1. Với a_1, a_2, \dots, a_m là tất cả các null phụ thuộc ngữ cảnh xuất hiện trong r , $\exists V_1, V_2, \dots, V_m$ sao cho $f = S_{a_1 a_2 \dots a_m}^{V_1 V_2 \dots V_m}(r)$.
2. $\forall t_f \in f, \exists t'_f \in f : (t_f \neq t'_f \wedge \forall 1 \leq i \leq n : (t_f[A_i] \neq \perp \Rightarrow t'_f[A_i] = t_f[A_i]))$.
3. $(\text{dne}, \text{dne}, \dots, \text{dne}) \in r \Rightarrow f = \emptyset$.

Trước hết, chúng ta đưa ra những giải thích ngắn gọn cho từng điều kiện của Định nghĩa.

Điều kiện 1 bắt buộc mỗi null phụ thuộc ngữ cảnh chưa biết phải được thay thế bởi duy nhất một giá trị xác định, chứ không bởi một tập giá trị. Hơn nữa, nó yêu cầu mỗi bộ t_f trong f phải được suy ra ít nhất từ một bộ của r . Về mặt trực giác, điều này có nghĩa, r phải được diễn giải dưới giả thiết thế giới đóng trừ khi nó có xuất hiện những giá trị null phụ thuộc ngữ cảnh mở nào đó. Điều kiện 1 cũng khẳng định, mỗi bộ của r (trừ ra hai bộ $(\text{open}, \text{open}, \dots, \text{open})$, $(\text{dne}, \text{dne}, \dots, \text{dne})$ đều suy ra ít nhất một bộ trong f . Cụ thể:

- Những bộ mà chỉ chứa dữ liệu xác định của r (không có null) đều xuất hiện trong mô hình f .
- Những bộ có chứa một hoặc nhiều null phụ thuộc ngữ cảnh chưa biết phải suy ra ít nhất một bộ trong f , trong đó những null phụ thuộc ngữ cảnh chưa biết được thay thế bởi những giá trị dữ liệu xác định.

- Những bộ có chứa một hoặc nhiều null phụ thuộc ngữ cảnh mở phải suy ra một hoặc nhiều bộ trong f , trong đó mỗi null phụ thuộc ngữ cảnh mở hoặc thay thế bởi một giá trị \perp , hoặc được thay thế bởi một giá trị xác định hoặc cũng có thể được thay thế bởi nhiều giá trị xác định.

- Cuối cùng thì mỗi null phụ thuộc ngữ cảnh không tồn tại mà xuất hiện trong một bộ nào đó của r sẽ được thay thế bởi giá trị \perp trong một hoặc nhiều bộ tương ứng của f .

Điều kiện 2 biểu diễn ngữ nghĩa mạnh của null phụ thuộc ngữ cảnh không tồn tại. Nó khẳng định nếu tồn tại một bộ t_f trong f mà có vài giá trị \perp thì không thể tồn tại bất kì một bộ nào trong f mà khác với t_f chỉ bởi thay thế một vài giá trị \perp với những giá trị xác định. Ví dụ, xét quan hệ

$r : \langle 3, dne, 4 \rangle$. Điều kiện 1 khẳng định f phải chứa $\langle 3, \perp, 4 \rangle$ và Điều kiện 2 thì không cho phép bộ $\langle 3, x, 4 \rangle$ xuất hiện trong f với x là một giá trị dữ liệu xác định.

Cuối cùng, Điều kiện 3 khẳng định, nếu có bộ $\langle dne, dne, \dots, dne \rangle$ trong r thì f bắt buộc phải là tập rỗng. Điều này xác định thêm ngữ nghĩa cho dne. Đặc biệt, nó nhấn mạnh quan hệ $\{\langle dne, dne, \dots, dne \rangle\}$ và quan hệ rỗng \emptyset đều có chung duy nhất một mô hình, đó là \emptyset .

Nếu f chỉ thỏa mãn Điều kiện 1 thì f được gọi là mô hình yếu của r .

Định nghĩa 5. Cho r là một quan hệ có thể chứa null. Ta nói, ngữ nghĩa của r là tập tất cả các mô hình của r (kí hiệu: $MODELS(r)$).

Định nghĩa 6. Cho r và r' là hai quan hệ trên cùng một lược đồ, r và r' được gọi là tương đương ngữ nghĩa (viết $r \equiv r'$) nếu và chỉ nếu $MODELS(r) = MODELS(r')$.

Ví dụ 3. Với quan hệ r và s như trong các bảng dưới đây thì $MODELS(r) = MODELS(s)$.

1	2	δ_1
2	δ_2	3

r

1	2	δ_3
2	δ_4	3

s

Định nghĩa 7. Cho r và r' là hai quan hệ trên cùng một lược đồ (có thể chứa null hoặc không), t là một bộ của quan hệ r , t' là một bộ của r' . Chúng ta nói t' xác định hơn t (kí hiệu $t' \geq t$) hoặc t ít xác định hơn t' (kí hiệu $t \leq t'$) nếu và chỉ nếu với $1 \leq i \leq n$ ít nhất một trong sáu điều kiện sau phải được thỏa mãn

- $t[A_i] = t'[A_i]$,
- $(t[A_i] = \text{open} \text{ và } t'[A_i] = \perp)$,
- $(t[A_i] = \text{open} \text{ và } t'[A_i] \neq dne)$,
- $(t[A_i] = \text{unk}, t'[A_i] = \perp)$,
- $(t[A_i] = \text{unk}, t'[A_i] = \text{unk})$,
- $(t[A_i] = dne \text{ và } t'[A_i] = \perp)$.

Ví dụ 4.

$$t = \langle 1, \beta_1, 2, \delta_1, dne \rangle \leq \langle 1, \perp, 2, \delta_1, \perp \rangle; t \leq \langle 1, \beta_1, 2, 1, \perp \rangle;$$

$$t \leq \langle 1, \perp, 2, 1, \perp \rangle; t \leq \langle 1, 1, 1, 2, dne \rangle; t \leq \langle 1, \beta_1, 2, \delta_1, \perp \rangle; t \not\leq \langle 1, 1, 2, 3, 1 \rangle.$$

Mệnh đề 2.

- i) Nếu $t_1 \triangleright t_2$ thì $t_1 \leq t_2$ và $t_2 \perp$.
- ii) Nếu $t_1 \leq t_2$ và $t_2 \perp$ thì $t_1 \triangleright t_2$.
- iii) Nếu $t_1 \leq t_2$ và $t_2 \leq t_3$ thì $t_1 \leq t_3$.
- iv) Nếu $t_1 \leq t_2$ và $t_2 \triangleright t_3$ thì $t_1 \triangleright t_3$.

Chứng minh. (i) và (ii) được suy trực tiếp Định nghĩa 3 và Định nghĩa 7.

(iii) được suy trực tiếp từ Định nghĩa 7.

(iv) Do $t_2 \triangleright t_3$, theo (i) $t_1 \leq t_3$ & $t_3 \perp$; theo (ii) và (iii) ta có $t_1 \triangleright t_3$.

Định nghĩa 8. Cho r và s là hai quan hệ trên cùng một lược đồ. r được gọi là xác định hơn s (kí hiệu $r \geq s$) hay s được gọi là ít xác định hơn r (kí hiệu $s \leq r$) nếu với mọi bộ $t_s \in s$ luôn tồn tại bộ $t_r \in r$ sao cho $t_r \geq t_s$.

Nếu $r \geq s$ và $s \geq r$ chúng ta viết $r \cong s$.

Bổ đề 1. Cho r và s là hai quan hệ trên cùng một lược đồ, nếu $MODELS(r) \neq \emptyset$ và $MODELS(r) \subseteq MODELS(s)$ thì

- i) $r \geq s$.

ii) $\forall t_r \in r \exists t_s \in s : t_s \leq t_r$.

Chứng minh. i) Giả sử $r \not\geq s$, khi đó luôn tồn tại bộ $t_s \in s$ để $\forall t_r \in r t_r \not\geq t_s$. Ta xây dựng một mô hình f của r sao cho $\forall t_r \in f : t_r \not\geq t_s$. Rõ ràng $f \in \text{MODELS}(r)$ và $f \notin \text{MODELS}(s)$. Như vậy $\text{MODELS}(r) \not\subseteq \text{MODELS}(s)$, điều này mâu thuẫn với giả thiết. Vì vậy $r \geq s$.

ii) Giả sử $\exists t_r \in r$ để $\forall t_s \in s t_r \not\geq t_s$, khi đó mọi mô hình của r đều chứa bộ được dẫn từ t_r , không thể là mô hình của s , mâu thuẫn với giả thiết, vậy ta có (ii).

Ví dụ 5. Với quan hệ r và s dưới đây thì $\text{MODELS}(r) \subseteq \text{MODELS}(s)$

δ_1	δ_2	1
δ_1	δ_2	2
β_1	2	β_2
4	5	6

s

1	δ_3	1
1	δ_3	2
2	2	β_3
4	5	6

r

Định lý 1. Nếu $r \equiv s$ thì $r \cong s$.

Chứng minh. Suy trực tiếp từ Bổ đề 1.

Định lý 2. Cho r và s là hai quan hệ trên cùng một lược đồ. Nếu tồn tại các null phụ thuộc ngữ cảnh a_1, a_2, \dots, a_k trong s và tồn tại các thay thế V_i của a_i với $1 \leq i \leq k$ sao cho $r = S_{a_1 a_2 \dots a_k}^{V_1 V_2 \dots V_k}(s)$ thì $\text{MODELS}(r) \subseteq \text{MODELS}(s)$.

Chứng minh. Ta phải chứng minh $\forall f$ là mô hình của r thì f cũng là mô hình của s . Do f là mô hình của r , gọi b_1, b_2, \dots, b_n là tất cả các null phụ thuộc ngữ cảnh xuất hiện trong r , khi đó tồn tại các phép thế có thể Y_i của b_i $\forall 1 \leq i \leq n$ sao cho $f = S_{b_1 b_2 \dots b_n}^{Y_1 Y_2 \dots Y_n}(r)$. Theo giả thiết $r = S_{a_1 a_2 \dots a_k}^{V_1 V_2 \dots V_k}(s)$ nên $f = S_{b_1 b_2 \dots b_n}^{Y_1 Y_2 \dots Y_n}(S_{a_1 a_2 \dots a_k}^{V_1 V_2 \dots V_k}(s)) = S_{b_1 b_2 \dots b_n a_1 a_2 \dots a_k}^{Y_1 Y_2 \dots Y_n V_1 V_2 \dots V_k}(s)$. Mặt khác do f là mô hình của r nên f thỏa mãn điều kiện 2 và 3 của Định nghĩa mô hình. Vậy f là mô hình của s .

2.3. Một vài ví dụ

Xét lược đồ quan hệ $R = (X, Y, Z)$ với $\text{Dom}(X) = \{a, b\}$; $\text{Dom}(Y) = \{1, 2\}$ và $\text{Dom}(Z) = \{c, d\}$. Chúng ta sẽ đưa ra những thể hiện khác nhau của R và bàn luận về ngữ nghĩa của chúng.

Rõ ràng, quan hệ mà chỉ chứa những giá trị dữ liệu xác định sẽ có duy nhất một mô hình, mô hình đó chính là gồm tất cả các bộ của r . Ví dụ, xét thể hiện r_1 của R

X	Y	Z
a	1	c
a	2	d

Hiển nhiên, r_1 có duy nhất một mô hình f_1

X	Y	Z
a	1	c
a	2	d

Xét một thể hiện r_2 khác của R

X	Y	Z
a	1	c
a	1	d
a	2	c
a	2	d
β_1	β_2	β_3

Do tính chất của null phụ thuộc cánh mở nên r_2 sẽ có 8 mô hình khác nhau.

a	1	c
a	1	d
a	2	c
a	2	d
b	1	c

f_a

a	1	c
a	1	d
a	2	c
a	2	d
b	2	c

f_b

a	1	c
a	1	d
a	2	c
a	2	d
b	1	d

f_c

a	1	c
a	1	d
a	2	c
a	2	d
b	2	d

f_d

a	1	c
a	1	d
a	2	c
a	2	d
b	\perp	c

f_e

a	1	c
a	1	d
a	2	c
a	2	d
b	\perp	d

f_f

a	1	c
a	1	d
a	2	c
a	2	d
b	1	\perp

f_g

a	1	c
a	1	d
a	2	c
a	2	d
b	2	\perp

f_h

Để thấy có những khả năng của r_2 không là mô hình. Ví dụ:

a	1	c
a	1	d
a	2	c
a	2	d
\perp	\perp	\perp

a	1	c
a	1	d
a	2	c
a	2	d
a	\perp	c

Tiếp theo, xét quan hệ r_3 có chứa null phụ thuộc ngữ cảnh chưa biết:

X	Y	Z
a	δ_1	c
a	δ_1	d
b	δ_2	c

r_3 có bốn mô hình sau:

a	1	c
a	1	d
b	1	c

 f_a

a	1	c
a	1	d
b	2	c

 f_b

a	2	c
a	2	d
b	1	c

 f_c

a	2	c
a	2	d
b	2	c

 f_d

Với quan hệ r_3 nếu chỉ dùng null chưa biết để biểu diễn thông tin không đầy đủ ta sẽ thấy số mô hình của r_3 tăng lên rất nhiều.

X	Y	Z
a	unk	c
a	unk	d
b	unk	c

a	1	c
a	1	d
b	1	c

a	1	c
a	1	d
b	2	c

a	1	c
a	2	d
b	1	c

a	1	c
a	2	d
b	2	c

a	2	c
a	1	d
b	1	c

a	2	c
a	1	d
b	2	c

a	2	c
a	2	d
b	1	c

a	2	c
a	2	d
b	2	c

có tám mô hình.

Như vậy, trong một quan hệ null thuộc ngữ cảnh nếu như số giá trị null có cùng ngữ cảnh càng nhiều thì mối ràng buộc giữa chúng sẽ càng cao và do đó số mô hình mà chúng xác định sẽ ít hơn nhiều so với những trường hợp thông thường khác.

Cuối cùng, chúng ta hãy xét quan hệ r_4 có chứa ba kiểu null khác nhau

X	Y	Z
a	dne	c
a	δ_1	d
b	β_1	d

Trước hết, ta nhận xét rằng, do có bộ $\langle b, \delta_1, d \rangle$ nên giá trị β_1 trong bộ $\langle b, \beta_1, d \rangle$ không thể được thay bằng giá trị \perp trong f (do điều kiện 2). Vì vậy giá trị được thay thế cho β_1 trong quan hệ này chỉ có thể là 1 hoặc 2.

Như vậy r_4 có bốn mô hình sau đây

a	⊥	c
b	1	d

f_a

a	⊥	c
b	1	d
b	2	d

f_b

a	⊥	c
b	2	d

f_c

a	⊥	c
b	2	d
b	1	d

f_d

3. KẾT LUẬN

Nghiên cứu về ngữ nghĩa dữ liệu trong CSDL với thông tin không đầy đủ là việc làm cần thiết để hỗ trợ cho vấn đề mở rộng các phép toán quan hệ cổ điển trong trường hợp CSDL có chứa null. Bài báo này đã trình bày nghiên cứu CSDL null phụ thuộc ngữ cảnh trong sự có mặt của null phụ thuộc ngữ cảnh chưa biết, null phụ thuộc ngữ cảnh mở và null không tồn tại, xác định được ngữ nghĩa cho những quan hệ đó và qua một vài ví dụ về ngữ nghĩa trong bài báo, chúng ta thấy null phụ thuộc ngữ cảnh luôn cung cấp được nhiều thông tin hơn hoặc chí ít là bằng (theo nghĩa số mô hình ít hơn hoặc bằng) so với các trường hợp thông thường khác.

TÀI LIỆU THAM KHẢO

- [1] Bùi Thị Thúy Hiền, Cơ sở dữ liệu quan hệ với giá trị null phụ thuộc vào ngữ cảnh, *Tạp chí Tin học và Điều khiển học* 15 (1) (1999) 1-10.
- [2] N. C. Hồ và L. T. Thăng, Ngữ nghĩa của dữ liệu trong Cơ sở dữ liệu với thông tin không đầy đủ, *Tạp chí Tin học và Điều khiển học*, 11 (2) (1995) 7-15.
- [3] J. Biskup, A formal approach to null values in database relations. In Hervé Gallaire, Jack Minker, and Jean Marie Nicolas, editors, *Advances in Database Theory*, Vol. I, Plenumpress, New York, 1981.
- [4] E. F. Codd, Extending the database relational model to capture more meaning, *ACM TODS* 4 (4) (1979) 397-434.
- [5] N. C. Hồ, A relational model of databases with context dependent null values, *Bull. Pol. Ac. Tech.* 36 (1-2) (1988) 91-105.
- [7] T. Imielinski and W. Lipski Jr., Incomplete information in relational databases, *Journal of the ACM* 31 (4) (1984) 761-791.
- [8] W. Lipski Jr., On semantic issues connected with incomplete information databases, *Journal of the ACM* 3 (3) (1979) 262-296.
- [9] D. Maier, *The Theory of Relational Databases*, Computer Science Press, 1983.
- [10] J. Paredaens et al., *In the Structure of the Relational Database Model*, Chapter 6, pages 157-176, Springer-Verlag, Berlin, Heidelberg, 1989.
- [11] C. Zaniolo, *A Formal Treatment of Nonezistent Values in Database Relations*, Bell Laboratories, unpublished, 1983.
- [12] C. Zaniolo, Database relations with null values, *Journal of Computer and System Science* 28 (1) (1984) 142-166.

Nhận bài ngày 10-3-1999

Nhận lại sau khi sửa ngày 26-5-1999

Khoa Toán - Cơ - Tin học,
Trường Đại học Khoa học tự nhiên, ĐHQG Hà Nội.