MỘT VÀI KẾT QUẢ VỀ CÁC RÀNG BUỘC TRÊN CÁC GIÁ TRỊ KHÔNG TRONG CSDL QUAN HỆ

TRẦN DOÃN TUẤN

1. MỞ ĐẦU

Trong CSDL thế giới thực, có những chỗ có thể mang giá trị đặc biệt, theo nghĩa rằng chúng không mang giá trị từ tập các giá trị trong miền xác định của thuộc tính đó. Những giá trị đặc biệt có thể có nhiều nghĩa «giá trị hiện thời là chưa biết», «thuộc tính là không thể áp dụng»... Các giá trị đặc biệt đó còn được gọi là những giá trị không. Những giá trị không đó nảy sinh trong hầu hết các ứng dụng, và sẽ là không hợp lý nếu như ta loại bỏ chúng ra khỏi CSDL một cách triệt để. Tuy nhiên có thể có những thành phần (những thuộc tính hoặc những trường) trong CSDL mà ở đó không cho phép xuất hiện giá trị không. Vì vậy trong [1], D. Maier đã đưa ra khái niệm các EC (và dạng tổng quát hơn của nó là các DEC) như một công cụ để điều khiển việc sử dụng các giá trị không trong CSDL, đó chính là các ràng buộc lên CSDL nhằm cho phép những thành phần nào được phép xuất hiện giá trị không và thành phần nào không được phép. Ví dụ ta có thể mong muốn các bộ trong một CSLD là hoàn toàn xác định trên một trong số vài tập các thuộc tính là khóa.

Nhiều tác giả khác (như E.F.Codd [3], W.Lipski [4], J.Biskup [5],...) đã đề cập đến các giá trị không khi chúng xuất hiện trong CSDL. Vì thế hệ QTCSDL phải cho phép thực hiện những phép toán của đại số quan hệ một cách khác nhau phụ thuộc vào loại giá trị không nào xuất hiện. Tuy nhiên bài báo của chúng ta đề cập đến vấn đề giá trị không được phép xuất hiện ở đâu chứ hhông phải là các phép toán trên giá trị không đó. Vì thế ta sẽ không phân biệt cụ thể loại giá trị không nào trong bài báo này.

Trong phần 2 ta sẽ nhắc lại một số định nghĩa và kết quả quen thuộc. Phần 3 liên quan đến những trường hợp mà hệ ba qui tắc suy diễn tạo thành một hệ đầy đủ. Phần 4 đề cập đến kết quả chính của bài báo này trong việc tính bao đóng của một tập các thuộc tính liên quan đến một tập các ràng buộc dec cho trước.

2. MỘT SỐ KHÁI NIỆM CƠ BẢN VÀ NHỮNG KẾT QUẢ QUEN THUỘC

Trước hết, chúng ta nhắc lại một số định nghĩa và kết quả cơ bản sẽ cần tới về sau.

Giả sử R là một lược đồ quan hệ, r là một quan hệ trên R và t là một bộ trong r. A là một thuộc tính của R, t(A) là giá trị của bộ t trên thuộc tính A. Tương tự nếu X là một tập các thuộc tính của R, ta kí ihệu t(X) là tập các giá trị của bộ t trên tập thuộc tính X. Nếu t(A) là một giá trị xác định trong miền xác định của thuộc tính A, ta sẽ kí hiệu là t(A)!

Định nghĩa. Giả sử t là một bộ trong quan hệ r(R). Ta gọi tập xác định của t, kí hiệu là def(t), là tập tất cả các thuộc tính $A \in R$ sao cho t(A) !

Định nghĩa. Giả sử X, Y1,..., Yn là các tập con của R. Quan hệ r(R) thỏa mãn dec $X \vdash$ (Y1,...,Yn) nếu với mọi $t \in r$, t(X) ! thì sẽ tồn tại $i \in \overline{1, n}$ sao cho t(Yi) !. Ta nói quan hệ r thỏa tập các dec đã cho nếu nó thỏa mọi dec trong tập này.

Một cách cụ thể, một dec (viết tắt của disjunctive existence constraint) là một ràng buộc trên lược đồ CSDL và nó có dạng $X \vdash$ (Y1, Y2,...,Yn). Ràng buộc này nói lên rằng nếu một bộ trong CSDL là xác định trên tập thuộc tính X thì cũng phải xác định trên các thuộc tính của ít nhất một trong các tập Yi.

Vế phải của một dec được gọi là một tập tuyển (disjunctset), còn mỗi phần tử trong tập này được gọi là một hạng tuyển (disjunct). Trường hợp n=1 thì dec này được gọi là một ràng buộc về sự tồn tại (existence constraint) và được kí hiệu là ec. Ràng buộc này được nghiên cứu nhiều trong công trình của D.Maier [1]

Cho trước một tập H các ràng buộc dec, giả sử R là tập các thuộc tính xuất hiện trong các dec của H, khi đó ta nói rằng H là tập các ràng buộc dec trên R.

Định nghĩa. Tập H các dec được gọi là suy dẫn logic ra một dec d khác nếu mọi quan hệ thỏa H cũng thỏa d.

Định nghĩa này nói chung là không thuận lợi về mặt thực hành vì trong quan hệ trên lược đồ R, nếu có một thuộc tính với miền xác định là vô hạn thì ta không thể liệt kê ra tất cả những quan hệ trên R mà thỏa một tập các dec trên R đã cho. Vì thế ta đưa ra các qui tắc suy diễn cho phép ta dẫn ra các dec khác từ tập các dec đã cho bằng cách áp dụng những qui tắc này.

Bổ đề 2. 1. Cho trước lược đồ R ; X, Y, Z là những tập con của R. Các qui tắc sau là xác đáng (sound) :

D1) Phản xạ

Với $X \subseteq Y$ thì $\mid - (X)$

D2) Tăng danh sách

$X \mid - (Y_1, \ldots, Y_n) \Rightarrow X \mid - (Y_1, \ldots, Y_n, Z)$ với $Z \subseteq R$

D3) Nhân danh sách

$X \mid - (Y_1, \ldots, Y_n)$ và $X \mid - (Z_1, \ldots, Z_m)$

$\Rightarrow X \mid - (Y_1Z_1, \ldots, Y_1Z_m, \ldots, Y_nZ_m)$

D4) Tăng trưởng

$X \mid - (Y_1, \ldots Y_n) \Rightarrow XA \mid - (Y_1, \ldots Y_n)$ với mỗi $A \in R$

D5) Bắc cầu

$X \mid - (Y_1, \ldots, Y_n)$ và với $1 \leqslant i \leqslant n$, $Y_i \{(Z_{i1}, \ldots Z_{ik_i}\}$

$\Rightarrow X \mid - (Z_{11}, \ldots, Z_{1k_1}, \ldots Z_{n1}, \ldots, Z_{nk_n})$

D6) Chiếu

$X \mid - (Y_1Z_1, \ldots, Y_nZ_n \Rightarrow X \mid - (Y_i, \ldots, Y_n)$

D7) Giả bắc cầu

$X \mid - (Y_1, \ldots, Y_n)$ và i, $1 \leqslant i \leqslant in$, $Y_iZ \mid - \{W_{i1}, \ldots, W_{ik_i}\}$

$\Rightarrow XZ \mid - (W_{11}, \ldots, W_{1k_1}, \ldots W_{n1}, \ldots, W_{nk_n})$

D8) Nén

$X \mid - (Y_1, \ldots, Y_n)$ và với i nào đó, $1 \leqslant i \leqslant n$

$Y_i \mid - (Z_1, \ldots Z_k)$ $(Y_i .. Y_{i-1}, Y_{i+1}, \ldots Y_n)$

$\Rightarrow X \mid - (Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots Y_n)$

D9) Trích |

$X \mid - \{Y_1Z, \ldots, Y_nZ\} \Rightarrow X \mid - (Z)$

Dễ kiểm tra rằng chỉ cần với 4 luật D1, D2, D3, D5 ta cũng có thể suy ra được các luật khác. Ta gọi 4 luật này là các qui tắc tồn tại (existence rule). Trong [2], thay vì D1, B.S. Goldstein đã dùng qui tắc $X \mid - (X)$, nhưng chúng tôi thấy phải thay bằng qui tắc D1 mạnh hơn như vậy mới suy ra được các qui tắc khác.

Định nghĩa. Bao đóng của tập H các dec, viết tắt là H*, là tập nhỏ nhất các dec chứa H sao cho khi áp dụng các qui tắc tồn tại tới H* sẽ không suy ra một dec nào không nằm trong H*.

Định nghĩa. Giả sử H là tập các dec trên R và $X \subseteq R$. Tập tuyển max (X, H) là vế phải của dec thu được bởi việc áp dụng qui tắc D3 cho mọi dec trong H* với vế trái là X.

Ví dụ $H = (A \mid - (B))$ là tập các dec trên AB. Khi đó

$H^* = (A \mid - (A), A \mid - (A,B), A \mid - (AB), A \mid - (A,AB), A \mid - (B,AB, A \mid - (AI,BAB)$

$B \mid - (B), B \mid - (B,A), B \mid - (B,AB), B \mid - (A,B,AB)$

$AB \mid - (A), AB \mid - (B), AB \mid - (AB), AB \mid - (A,B), AB \mid - (A,AB),$

$AB \mid - (B,AB), AB \mid - (A,B,AB))$. Khi đó

max (A, H) = max (AB, H) = {AB} và max (B, H) = {B, AB}.

Định nghĩa. H, R, X, vẫn giả thiết như trên. Tập tuyển $\{Y_1, \ldots, Y_n\}$ được gọi là tập tuyển con cực đại của X trong H (sub−maximal disjunct set) nếu $X \mid - \{Y_1, \ldots, Y_n\} \in H^*$ và mỗi Y_i là một hạng tuyển trong max (X, H). Tập tất cả các tập tuyển con cực đại của X trong H được kí hiệu bởi SMD (X, H).

Định nghĩa. H, X, R vẫn giả thiết như trên. Tập tuyển trong SMD (X, H) với số hạng tuyển ít nhất được gọi là bao đóng của X đối với H. Tập này được kí hiệu bởi X*, hoặc X * /H khi H cần phải chỉ rõ.

Để định nghĩa này được chặt chẽ, trong [2], tác giả đã chứng minh rằng tập tuyển có tính chất như vậy là tồn tại và duy nhất.

Bổ đề 2.2 Các qui tắc tồn tại tạo thành một hệ qui tắc đầy đủ.

Đây là một kết quả khá đẹp và đã được chứng minh trong [2]. Xem xét lại, ta thấy qui tắc tăng danh sách D2 tuy là tự nhiên về mặt ý nghĩa, nhưng thực chất nó không cho ta thêm được một chút thông tin nào vì việc thêm vào một tập thuộc tính Z khá mơ hồ.

3. HỆ QUI TẮC ĐẦY ĐỦ THU GỌN

Trong mục này, ta xét xem với những trường hợp nào tập ba qui tắc D1, D3, D5 tạo thành một hệ qui tắc đầy đủ.

Bổ đề 3.1. Giả sử H là một tập các dec trên R, $X \subseteq R$ và giả sử $\max(X, H) = \{Z_1,\dots,Z_m\}$. Khi đó với một dec bất kỳ $X \vdash \{Y_1,\dots,Y_n\} \in H^*$ (dec này với vế trái là X) ta có:

1) $X \vdash \{Z_1,\dots,Z_m\} \in H^*$

2) $\forall Z_i \exists Y_j$ sao cho $Y_j \subseteq Z_i$

3) $\forall Y_j \exists Z_i$ sao cho $Y_j \subseteq Z_i$

Bổ đề 3.2. H, X, R vẫn giả thiết như trên. Giả sử $X^* = \{Z_1,\dots,Z_m\}$. Khi đó với mỗi Z_i trong X^*:

1) $X \subseteq Z_i$

2) $\forall X \vdash \{Y_1,\dots,Y_n\} \in H^* \exists Y_j$ sao cho $Y_j \subseteq Z_i$

Bổ đề 3.3. $X^* = \{Z_1,\dots,Z_m\}$ kéo theo $Z_i^* = \{Z_i\}$, $1 \leqslant i \leqslant m$

Bổ đề 3.4. Với các ec thì hệ ba qui tắc D1, D3, D5 tạo thành một hệ qui tắc đầy đủ.

Chứng minh: Ta phải chứng minh rằng nếu có một ec có dạng $e = X \vdash \{Y\}$ được suy logic từ tập H của các ec thì nó phải được dẫn ra từ H bằng việc áp dụng 3 qui tắc này. Lấy $0 = \{V : V^* = \{V\}\}$, giả sử $0 = \{V_1,\dots,V_h\}$. Ta xây dựng quan hệ r như sau:

$$r = \{t_1,\dots,t_n\} \text{ với } t_i(A_j) = \begin{cases} a_j \text{ nếu } A_j \in V_i. \text{ Ở đây } a_j \in DOM(A_j) \\ \perp \text{ (không xác định) nếu ngược lại.} \end{cases}$$

Ta phải chứng minh hai điều:

1) $\forall e \in H^*$ thì r thỏa e và

2) Nếu e được suy dẫn logic ra từ H thì e cũng được dẫn ra từ H bằng cách áp dụng D1, D3, D5. Theo [2] ta có X^* chỉ gồm một hạng tuyển và việc tính X^* chỉ dùng 3 qui tắc D1, D3, D5. Giả sử $X^* = \{Z\}$.

1) Giả sử $e = X \vdash \{Y\} \in H^*$. Ta phải chứng minh r thỏa e. Với mọi bộ $t \in r$ mà $t(X)$ là không xác định sẽ thỏa e một cách tầm thường. Ta chứng minh rằng những bộ với $t(X)!$ cũng thỏa e.

Theo bổ đề 3.2 ta có $Y \subseteq Z$. Giả sử t_i là bộ sao cho $t_i(X)!$ Theo định nghĩa của r ta có $X \subseteq V_i$, suy ra $V_i \vdash \{X\} \in H^*$ và theo D5 ta có $V_i \vdash \{Z\} \in H^*$. Theo định nghĩa ta có $V_i^* = \{V_i\}$ nên theo bổ đề 3.2 ta có $Z \subseteq V_i$. Từ đó suy ra $Y \subseteq V_i \Rightarrow t_i(Y)! \Rightarrow t_i$ thỏa $e \Rightarrow$ quan hệ r cũng thỏa e.

2) Giả sử $e = X \vdash \{Y\}$ được suy dẫn logic từ H. Khi đó ta có r sẽ thỏa e. Ta có $X^* = \{Z\}$ và theo bổ đề 3.3 ta có $Z^* = \{Z\}$. Lấy bộ $t_z \in r$ mà $t_z(Z)!$. Theo bổ đề 3.2 ta có $X \subseteq Z$ nên $t_z(X)!$. Từ $X \vdash \{Y\}$ nên ta suy ra $t_z(Y)!$. Theo cách xây dựng r ta có $Y \subseteq Z$ suy ra $Z \vdash \{Y\}$ theo D1. Như vậy việc suy dẫn $X \vdash X^* = \{Z\}$ và $Z \vdash \{Y\}$ đều chỉ dùng ba qui tắc D1, D3, D5 nên ta có thể dẫn ra $X \vdash \{Y\}$ mà chỉ dùng D1, D3, D5

Trước khi trình bày kết quả tiếp theo, chúng ta đưa ra định nghĩa sau:

Định nghĩa. Cho H là một tập các dec trên R. Ta nói một ràng buộc $d = X \models \{Y_1,\dots,Y_n\}$ là được kéo theo chặt từ H nếu quan hệ r thỏa H đều thỏa d và $\forall i = \overline{1, n}$, $\exists$ quan hệ r thỏa H và không thỏa dec: $X \vdash \{Y_1,\dots,Y_{i-1}, Y_{i+1},\dots,Y_n\}$.

Nói một cách khác, một dec gọi là được kéo theo chặt từ tập các dec khác nếu như nó là được suy diễn logic theo nghĩa thông thường và không thể bỏ đi một hạng tuyển nào từ tập tuyển của dec đó.

Bổ đề 3.5. Với tập các dec được kéo theo chặt, hệ ba qui tắc D1, D3, D5 tạo thành một hệ đầy đủ.

Chứng minh Giả sử $d = X \mid\!- \{Y1,...,Yn\}$ là dec được kéo theo chặt từ H. Ta phải chứng minh rằng nó có thể được suy dẫn từ chỉ bằng cách áp dụng ba qui tắc trên. Ta xây dựng quan hệ r như trong chứng minh bổ đề 3.4. Lấy ri như trong định nghĩa kéo theo chặt. Vì r thỏa H (chứng minh tương tự như ở 3.4) và ri thỏa H nên $r \cup \left(\bigcup_{i=1}^{n} ri \right)$ thỏa H.

Thay vì chọn quan hệ như ở bổ đề 3.4, ta chọn quan hệ $r \cup \left(\bigcup_{i=1}^{n} ri \right)$. Giả sử $X^* = \{Z1,..., Zm\}$ theo bổ đề 3.3 ta có $Zi^* = \{Zi\}$ nên ta lấy $t_{Zi} \in r$ sao cho $t_{Zi}(Zi)!$ vì $X \subseteq Zi$ nên $t_{Zi}(X)!$ có nghĩa $\exists Yzi$ để $t_{Zi}(Yzi)!$. Suy ra $Yzi \subseteq Zi$ theo định nghĩa của r. Từ đó $Zi \mid\!- \{Yzi\}$ theo D1. Vậy $X \mid\!- X^* = \{Z1, ..., Zm\}$ được suy ra từ H chỉ dùng ba qui tắc D1, D3, D5 (theo nhận xét trong [2] nên $X \mid\!- \{Yz1,...,Yzm\}$ là có thể được suy dẫn từ H chỉ dùng ba qui tắc trên. Ta có $\{Yzi,...,Yzm\} \subseteq \{Y1,..., Yn\}$, nhưng theo định nghĩa kéo theo chặt, tập tuyển của d là không thể bỏ đi một hạng tuyển nào nên ta có: $\{Yz1,...,Yzm\} = \{Y1,...,Yn\}$.

Vậy $X \mid\!- \{Y1,...,Yn\}$ có thể được suy dẫn từ H mà chỉ dùng ba qui tắc D1, D3, D5

4. TÍNH X *

Khi làm việc với các dec, một vấn đề rất quan trọng được đặt ra là tính X^*, bao đóng của một tập các thuộc tính X cho trước đối với một tập các dec H đã cho. Nói một cách khác, tính X^* là tìm một tập đầy đủ và gọn nhất của các thuộc tính có liên quan chặt chẽ và thiết yếu đến tập thuộc tính X đã cho. Việc tính được X^* cho phép ta đưa ra được biểu diễn chính tắc của một tập H các dec đã cho (đó chính là tập $\{X \mid\!- X^*, V \subseteq R\}$) và cũng cho ta một phương pháp để quyết định xem hai tập các dec đã cho có tương đương với nhau hay không? (Hai tập H và I của các dec trên R được gọi là tương đương với nhau nếu $H^* = I^*$). Việc tính được X^* kết hợp với kết quả của bổ đề 4.2 cho phép ta quyết định xem một dec d đã cho có thể được suy dẫn từ tập H các dec đã cho hay không, hay nói một cách khác, d có thuộc H^* hay không?. Một ý nghĩa quan trọng nữa trong việc tính X^* là cho phép ta đưa ra được tập các ràng buộc đối tượng (object) OBJ (R, H) từ một tập H của các dec đã cho. Vì vậy những cố gắng trong phần này tập trung vào việc tính bao đóng X^* của tập thuộc tính X đã cho.

Ta có một số kết quả liên quan đến X^*. Ở đây X, Y, Z luôn kí hiệu là tập các thuộc tính $\subseteq R$.

Bổ đề 4.1. $X^* = \{X\} \Leftrightarrow \exists\, Y$ sao cho $X \in Y^*$.

Chứng minh: Nếu $\exists\, Y$ để $X \in Y^*$ thì theo bổ đề 3.3 ta có $X^* = \{X\}$. Ngược lại, nếu $X^* = \{X\}$ thì ta lấy ngay tập Y chính là X.

Kết quả sau đây cho phép ta quyết định xem một dec bất kì dạng $X \mid\!- \{Z1,..,Zm\}$ là có thuộc H^* hay không?.

Bổ đề 4.2 Giả sử $X^* = \{Y1,.., Yn\}$. Khi đó

$$X \mid\!- \{Z1,..., Zm\} \in H^* \Leftrightarrow \forall Yi\ \exists Zj \text{ sao cho } Zj \subseteq Yi.$$

Chứng minh: Nếu $X \mid\!- \{Z1,...,Zm\} \in H^*$ thì theo bổ đề 3.2 ta có ngay điều cần chứng minh.

Ngược lại, nếu $\forall Yi\ \exists Zji$ sao cho $Zji \subseteq Yi$. Khi đó $Yi \mid\!- \{Zji\} \in H^*$. Áp dụng qui tắc bắc cầu ta có $X \mid\!- X^* = \{Y1,...,Yn\}$ nên $X \mid\!- \{Zj1,...,Zjn\} \in H^*$. Nhưng $\{Zj1, ..,Zjn\} \subseteq \{Z1,...,Zm\}$ nên theo qui tắc tăng danh sách ta có: $X \mid\!- \{Z1,...,Zm\} \in H^*$.

Từ đây ta có hệ quả sau:

Bổ đề 4.3. Giả sử $X^* = X$. Khi đó $X \mid\!- \{Z1,...,Zm\} \in H^* \Leftrightarrow \exists Zj : Zj \subseteq X$.

Bổ đề 4.4. Giả sử $X^* = \{Y1,...,Yn\}$. Khi đó với mọi tập $S \in SMD(X, H)$, $S = \{Z1,...,Zm\}$, ta có $X^* \subseteq S$.

Chứng minh: Ta chứng minh điều này bằng phản chứng. Giả sử ngược lại, Yi không bằng một phần tử nào ở trong S. Vì $X \mid\!- S \in H^*$ theo giả thiết, nên theo bổ đề 3.2 ta có $\exists Zj$ sao cho $Zj \subset Yi$. Mặt khác vì $Zj \in max(X.H)$ và $X \mid\!- X^* \in H^*$ nên theo bổ đề 3.1 ta có $\exists Y_k$ sao cho $Y_k \subseteq Z_j$. Từ đó ta có $Y_k \subset Y_i$ nên ta suy ra $Y_i \models \{Y_k\} \in H^*$. Áp dụng luật bắc cầu đối với $X \models X^* = \{Y_1,\ldots,Y_n\}$ ta được $X \models \{Y_1,\ldots, Y_{i-1}, Y_{i+1},\ldots, Y_n\} \in H^*$. Điều này mâu thuẫn với định nghĩa của X^* là có số hạng tuyển là ít nhất.

Ta thấy nếu W là một phần tử $\in \max(X, H)$ mà nằm ngoài X^* thì luôn tồn tại một hạng tuyển $Y_i \in X^*$ sao cho $Y_i \subset W$. Điều này kết hợp với bổ đề 4.4 cho phép ta tính được X^* bằng cách tính một tập $S \in SMD(X, H)$ (bằng một phương pháp nào đó), sau đó loại bỏ đi từ S những phần tử có chứa một phần tử khác. Các phần tử còn lại sẽ chính là X^*.

Vấn đề còn lại của chúng ta là tìm cách tính một tập S nào đó $\in SMD(X,H)$.

Bổ đề 4.5. Giả sử $X \models \{Z_1,\ldots, Z_m\} \in H^*$. Khi đó $Z_i \in \max(X, H) \Leftrightarrow \forall$ dec có dạng $X \models \{W_1,\ldots, W_p\} \in H^*$, $\exists W_j$ sao cho $W_j \subseteq Z_i$.

Chứng minh. Giả sử mọi dec có dạng $X \models \{W_1,\ldots, W_p\} \in H^*$ đều $\exists$ W_j sao cho $W_j \subseteq Z_i$. Khi đó theo cách tính $\max(X, H)$, khi ta nhân mọi vế phải của các dec này với vế phải của dec $X \models \{Z_1,\ldots, Z_m\} \in H^*$ ở trên, ta có ngay $Z_i \in \max(X, H)$.

Phần ngược lại là kết quả của bổ đề 3.1.

Sau đây ta sẽ đưa ra thuật toán để tính một tập $S \in SMD(X, H)$.

Giả sử tập H của các dec đã cho có dạng :

$$\frac{X_1 \models \{Y_{11},\ldots, Y_{1m_1}\}}{X_n \models \{Y_{n1},\ldots, Y_{nm_n}\}}$$

Đối với một tập thuộc tính $X \subseteq R$ sẽ xảy ra hai khả năng sau :

1) $\forall i = \overline{1,n}$ không tồn tại X_i nào mà $X_i \subseteq X$. Khi đó ta có ngay $X^* = \{X\}$ vì các dec trong tập H đã cho không giúp ta suy dẫn thêm được gì cả. (Các qui tắc tồn tại không cho phép ta thay đổi vế trái của dec).

2) $\exists X_1,\ldots, X_k$ sao cho $X_i \subseteq X$ $\forall i = \overline{1,k}$.

Thuật toán tính tập S :

Khởi đầu, ta lấy $X^{(0)} = \{X\}$ với $X_1 \subseteq X$ ta sẽ thác triển $X^{(0)}$ thành :

$X^{(1)} = \{XY_{11},\ldots, XY_{1m_1}\}$ và ta có $X \models X^{(1)} \in H^*$. Ta cứ tiếp tục quá trình thác triển như vậy. Nếu như có một hạng tuyển nào chứa một tập X_i, chẳng hạn $X_i \subseteq XY_{11}$ thì ta sẽ thác triển XY_{11} thành $\{XY_{11}Y_{i1},\ldots, XY_{11}Y_{im_i}\}$, Khi đó $X^{(1)}$ sẽ trở thành $X^{(2)} = \{XY_{11}Y_{i1},\ldots, XY_{i1}Y_{im_i}, XY_{12},\ldots, XY_{1m_1}\}$ và $X \models X^{(2)} \in H^*$... Ta cứ tiếp tục thác triển cho đến khi không thể thác triển thêm được nữa. Quá trình thác triển như vậy luôn chấm dứt vì số các dec trong tập H cho trước là hữu hạn, hơn nữa các hạng tuyển nằm ở bên phải luôn luôn thỏa mãn điều kiện là phải chứa X và là tập con của R. Vì thế số những hạng tuyển như vậy là hữu hạn. Giả sử tập cuối cùng ta thu được là $S = \{Z_1,\ldots, Z_m\}$.

Theo quá trình thác triển, ta luôn có $X \models X^{(i)} \in H^*$ nên ta có $X \models \{Z_1,\ldots, Z_m\} \in H^*$. Để chứng minh tập $S = \{Z_i,\ldots, Z_m\} \in SMD(X, H)$ $\forall i = \overline{1,m}$ ta chỉ cần chứng minh $Z_i \in \max\{X, H\}$ $\forall i = \overline{1,m}$. Muốn vậy, theo bổ đề 4.5 ta chỉ cần chứng minh rằng dec có dạng $X \models \{W_1,\ldots, W_p\} \in H^*$ luôn tồn tại W_j sao cho $W_j \subseteq Z_i$. Ta chứng minh điều này bằng kết quả của bổ đề sau :

Bổ đề 4.6 Vẫn giả thiết Z_i như trên. Giả sử $Y \subseteq Z_i$. Khi đó $\forall$ dec có dạng $d = Y \models \{U_1,\ldots, U_q\} \in H^*$ luôn tồn tại U_j sao cho $U_j \subseteq Z_i$.

Chứng minh : Giả sử d được suy dẫn từ bằng một suy dẫn có độ dài h. Ta chứng minh bổ đề 4.6 bằng qui nạp theo độ dài h của phép suy dẫn này.

— Với h = 1, khi đó một dec có độ dài 1 chỉ có thể có dạng $Y \models \{W\}$ với $W \subseteq Y$ hoặc $Y \models \{W_1,\ldots, W_r\} \in H$. Trong trường hợp đầu ta có $W \subseteq Y \subseteq Z_i$. Trong trường hợp sau, theo cách ta thác triển để có được Z_i, ta thấy rằng $\exists$ W_j để cho $W_j \subseteq Z_i$.

Giả thiết là khẳng định đúng dec thu được bằng một suy dẫn có độ dài $\leqslant k$. Ta chứng minh nó đúng cho dec thu được bởi suy dẫn có độ dài $k + 1$. Muốn vậy ta lần lượt kiểm tra lại khẳng định cần chứng minh khi bước thứ $k + 1$ là một trong 4 qui tắc tồn tại;

— Nếu bước thứ $k + 1$ là qui tắc phản xạ dạng $Y \mid- \{W\}$ với $W \subseteq Y$, ta có ngay $W \subseteq Y \subseteq Zi$,

— Nếu nó là phép nhân danh sách của hai dec sau :

$Y \mid- \{V1, \ldots, Vt\}$ (với số bước suy dẫn $\leqslant k$)

$Y \mid- \{W1, \ldots, Ws\}$ (với số bước suy dẫn $\leqslant k$)

Theo giả thiết qui nạp ta có $\exists V1$ và $\exists Wj$ sao cho $V1 \subseteq Zi$, $Wj \subseteq Zi$. Sau khi áp dụng phép nhân danh sách ta có :

$Y \mid- \{V1W1, \ldots, V1Ws, \ldots, VtWs\}$ thì sẽ tồn tại hạng tuyển $V1Wj \subseteq Zi$.

— Phép bắc cầu : giả sử bước thứ $k+1$ ta áp dụng qui tắc bắc cầu từ hai dec sau :

$Y \mid- \{V1, \ldots, Vt\}$ (với số bước suy dẫn $\leqslant k$)

$Vj \mid- \{U1, \ldots, Um\}$ (với số bước suy dẫn $\leqslant k$)

Sau khi áp dụng qui tắc bắc cầu ta thu được :

$Y \mid- \{V1, \ldots, Vj - 1, U1, \ldots, Um, Vj+1, \ldots, Vt\}$

Theo giả thiết qui nạp ta có $\exists Vr$ để $Vr \subseteq Zi$. Nếu $r \neq j$ ta sẽ có ngay điều cần chứng minh. Nếu $r=j$ thì $Vj \subseteq Zi$ nên theo giả thiết qui nạp ta có $\exists Uk$ để $Uk \subseteq Zi$ $(k \in \overline{1, m})$. Vậy ta có điều cần chứng minh.

— Phép tăng danh sách : khẳng định là hiển nhiên,

Áp dụng kết quả của bổ đề này, vì $X \subseteq Zi$ nên ta có ngay điều cần có là $Zi \in \max(X, H)$, $\forall i = \overline{1, m}$.

Từ các kết quả trên đây ta có một thuật toán để tính $X*$ như ở hình trang sau.

5. KẾT LUẬN

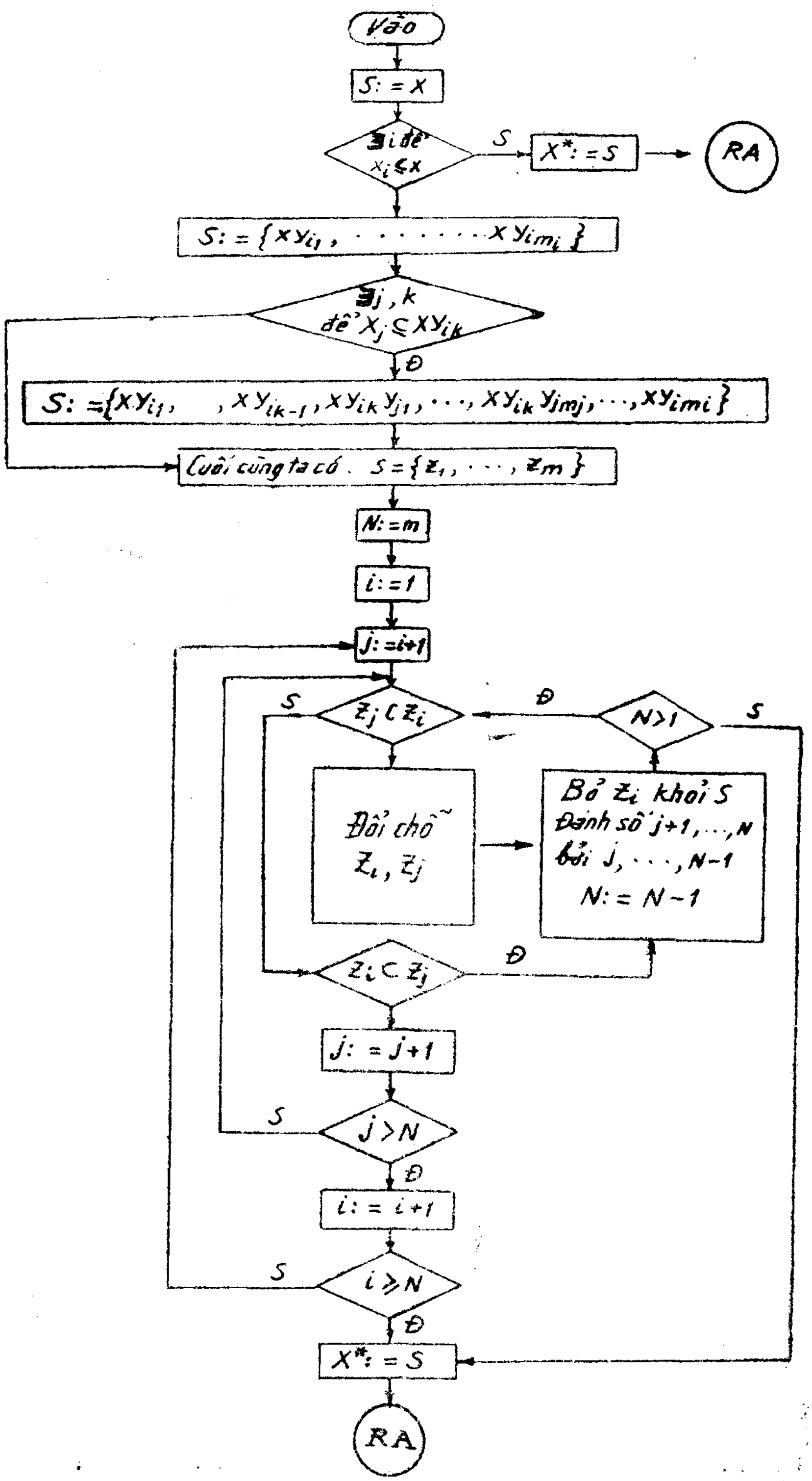
Bài báo này liên quan đến việc hạn chế sự có mặt của các giá trị không trong CSDL. Trong những vấn đề đã được đề cập ở trên, ta không phân biệt giá trị không đó là thuộc loại nào [hoặc là giá trị hiện thời chưa biết, hoặc là không nhất quán...].

Những kết quả ban đầu của chúng ta tập trung vào việc xét xem trong những trường hợp nào, hệ ba qui tắc D1, D3, D5 tạo thành một hệ đầy đủ. Các kết quả tiếp theo tập trung vào việc tính bao đóng $X*$ của một tập thuộc tính X đối với tập các ràng buộc H đã cho. Đây là một vấn đề mà trong [2], tác giả còn để mở. Ý nghĩa của vấn đề này đã được đề cập đến ở đầu phần 4. Thuật toán để tính $X*$ của chúng ta khá giản dị và sáng sủa về mặt thực hiện.

Việc tính toán bao đóng $H*$ của một tập các ràng buộc H cho trước cũng có thể tiến hành một cách tương tự bằng phương pháp thác triển dần dần như chúng ta đã đề cập. Tuy nhiên việc tính $H*$ là khá cồng kềnh (như chúng ta có thể thấy qua một ví dụ đơn giản trong bài này) và có phần không cần thiết về mặt thực tế. Vì thế chúng tôi đã không đề cập đến vấn đề đó ở trong bài báo này.

Có một vấn đề còn mở là sự tương tác giữa ràng buộc giá trị không và các phụ thuộc dữ liệu (như phụ thuộc hàm, phụ thuộc đa trị,.) Cả hai loại ràng buộc này đều biểu thị một phần nào đó về ngữ nghĩa của CSDL. Nếu như cả hai loại ràng buộc này đều xuất hiện trong CSDL thì chúng sẽ tác động lên nhau như thế nào ? Không giống như vấn đề được nói đến trong bài này, việc khảo sát như vậy phụ thuộc vào loại giá trị không nào mà các ràng buộc đó thể hiện.

Nhận ngày 1 — 1 — 1989


Vào
S: = X
∃ i để $x_i \subseteq x$
S
$X^*: = S$
RA
$S: = \{xy_{i_1}, \cdots\cdots xy_{im_i}\}$
∃ j, k để $X_j \subseteq xy_{ik}$
Đ
$S: = \{xy_{i_1}, \ , xy_{ik-1}, xy_{ik}y_{j1}, \cdots, xy_{ik}y_{jmj}, \cdots, xy_{imi}\}$
Cuối cùng ta có. $S = \{z_1, \cdots, z_m\}$
$N: = m$
$i: = 1$
$j: = i+1$
S
$z_j \subset z_i$
Đ
$N > 1$
S
Đổi chỗ z_i, z_j
Bỏ z_i khỏi S
Đánh số $j+1, \ldots, N$ bởi $j, \cdots, N-1$
$N: = N-1$
$z_i \subset z_j$
Đ
$j: = j+1$
S
$j > N$
Đ
$i: = i+1$
S
$i \geqslant N$
Đ
$X^*: = S$
RA

TÀI LIỆU THAM KHẢO

1. Maier, D., Discarding the universal instance assumption : Preliminary results (Tech. repo. 3/1980,).

2. Goldstein, B. S., Formal properties of constraints on Null values in Relational Database (Tech. repo. 7/1981).

3. Cood, E.F., Extending the Database relational model to capture more meaning. ACM. Trans. on Database system 4,4 (1979), pp.397–434.

4. Lipski, W., On semantic issues connected with incomplete Information Database, ACM. Trans. on Database system 4,3 (1979), pp. 262–296.

5. Biskup, J,, A formal approach to null values in database relation : Advances in Database Theory, Vol. 1, Plenum, NewYock, 1981, pp. 299—341.

6. Le Tien Vuong, Ho Thuan, A relational database extended by application of fuzzy set theory anh linguistic variables. Computer and Artificial Intelligence, Vol, 8,1989, N°2, 153 – 168.

ABSTRACT

Some results about cenotraints on null values in relational databases

In [2] B.S.Goldstein introduced the concept of «disjunctive existence constraint» together with a sound and complete set of inference rules (4 rules), as a means for using null values.

In this paper, we investigate cases in which the system of three inference rules is complete,

Finally, an algorithm for computing the closure X^* of X is also given.