Tạp chí KHOA HỌC TÍNH TOÁN VÀ ĐIỀU KHIỂN TẬP II (1986), Số 3 19-27)

SỰ TƯƠNG ĐƯƠNG CỦA CÁC BIỂU MỞ RỘNG ĐƠN

ĐỖ XUÂN THỌ

I – MỞ ĐẦU

Một trong những vấn đề hiện nay được nhiều người quan tâm là tối ưu hóa việc xử lí các câu hỏi trong cơ sở dữ liệu (CSDL) nhằm làm giảm bớt thời gian thực hiện trên máy tính khi xử lí những câu hỏi đó.

Aho A.V., Sagiv Y., Ullman J.D., [1,2,3] đã sử dụng biểu (tableaux) để tối ưu hóa một lớp các câu hỏi gồm 3 phép tính chọn, chiếu và kết nối tự nhiên). Các thủ tục tối ưu được xây dựng trên cơ sở cực tiểu hóa số dòng của biểu mô tả câu hỏi nhằm loại bỏ được các phép kết nối thừa. Đây là việc làm có ý nghĩa do phép kết nối là « đắt nhất » trong 3 phép tính nêu ra ở trên.

Trong [9, 10] tác giả đã nêu ra những hạn chế của khái niệm biểu, trên cơ sở đó đề xuất dựa vào khái niệm biểu mở rộng. Kết quả cho phép mở rộng lớp các câu hỏi có thể sử dụng biểu để tối ưu hóa.

Do bài toán kiểm tra sự tương đương của 2 biểu là bài toán NP – đầy đủ, vì vậy trong [1,2] các tác giả đã đưa ra khái niệm biểu đơn (simple tableaux) tương ứng lớp con các câu hỏi mà trên thực tế không khác biệt với lớp những câu hỏi xét ở trên là bao. Điều có ý nghĩa quan trọng là thủ tục để kiểm tra sự tương đương của các biểu đơn có độ phức tạp về thời gian là đa thức. Trong bài này nghiên cứu sự tương đương của các biểu mở rộng đơn và điều kiện để một biểu mở rộng đơn có biểu thức quan hệ tương ứng.

II – MỘT SỐ KHÁI NIỆM

Các khái niệm cơ bản và các thuật ngữ về CSDL quan hệ có thể xem chi tiết trong [10]. Ở đây chỉ nhắc lại vài khái niệm sẽ dùng cho các phần sau.

Trên cơ sở đại số quan hệ, câu hỏi của người sử dụng có thể mô tả dưới dạng biểu thức quan hệ với những phép tính chọn, chiếu, kết nối....

Nếu A là thuộc tính của quan hệ r, $\mathcal{C} \subseteq \mathrm{dom}(A)$, phép chọn $\sigma_{A \in \mathcal{C}}(r)$ được định nghĩa [10] :

$$\sigma_{A \in \mathcal{C}}(r) = \{\mu \in r \mid \mu[A] \in \mathcal{C}\}.$$

r_1, r_2 là 2 quan hệ có lược đồ tương ứng R_1 và R_2. Phép kết nối tự nhiên của r_1, và r_2, kí hiệu là $r_1 * r_2$:

$$r_1 * r_2 = \{\mid \mu \mid (\mu \text{ là một bộ của quan hệ có lược đồ } B_1 \cup R_2) \wedge$$
$$((\exists\, \partial_1 \in r_1) \wedge (\exists\, \partial_2 \in r_2) : \partial_1[R_1] = \mu[R_1], \partial_2[R_2] = \mu[R_2])\}.$$

Biểu mở rộng $\mathcal{T}$ là một ma trận 2 chiều có số cột bằng số thuộc tính của quan hệ vũ trụ theo trật tự cố định. $\mathcal{T}$ gồm hàng đầu tiên gọi là giản lược, kí hiệu w_0 và các hàng w_1, w_2,..., w_m. φ là tập các kí hiệu xuất hiện trong biểu.

$$\varphi = \{a_i, b_j, \tilde{\mathcal{C}}_k, \tilde{\mathcal{C}}^*_l, \sqcup\}$$

trong đó a_j là các biến phân biệt, b_j – các biến không phân biệt ; $\tilde{\mathcal{C}}$, $\tilde{\mathcal{C}}^*$ – kí hiệu tập những giá trị $\mathcal{C}$ thuộc miền xác định của các thuộc tính. $\tilde{\mathcal{C}}$ được gọi là kí hiệu tập những giá trị phân biệt. $\tilde{\mathcal{C}}^*$ được gọi là kí hiệu không phân biệt. Trên mỗi cột của biểu $\mathcal{T}$ chỉ xuất hiện nhiều nhất một biến phân biệt a_i hoặc một kí hiệu phân biệt $\tilde{\mathcal{C}}_i$.

Một định giá ρ đối với $\mathscr{T}$ là phép thế mỗi kí hiệu trong φ một giá trị tương ứng và thỏa mãn các điều kiện:

a) Nếu $\tilde{\mathscr{C}}$ là kí hiệu tập giá trị $\mathscr{C}$ thì $\rho(\tilde{\mathscr{C}}) = c$. Trong trường hợp $\mathscr{C} = \phi$ thì $\rho(\tilde{\mathscr{C}}) = \phi$

b) Với w_m $v_1 v_2 \ldots v_n$ là 1 hàng của $\mathscr{T}$ thì

$$\rho(w) = \rho(v_1)\,\rho(v_2)\ldots\,\rho(v_n).$$

Nếu I là một thể hiện của quan hệ vũ trụ, khi đó giá trị của biểu $\mathscr{T}$ ở thể hiện I được xác định như sau:

$$\mathscr{T}(I) = \{\rho(w_o) \mid \rho \text{ là một định giá: } \rho(w_j) \in I,\ j = \overline{1,m}\ \}$$

w_o – giản lược, w_j – các dòng của $\mathscr{T}$.

Trong [9,10] đã nêu qui tắc xây dựng biểu mở rộng $\mathscr{T}$ cho một biểu thức quan hệ thu hẹp E Áp dụng quy tắc đó, chúng ta có thể xây dựng cho mỗi biểu thức quan hệ thu hẹp E một biểu mở rộng tương ứng.

$$V_I(E) = \mathscr{T}(I).$$

$V_I(E)$ – giá trị của biểu thức E với thể hiện I của quan hệ vũ trụ [1,2].

Hai biểu mở rộng $\mathscr{T}_1$ và $\mathscr{T}_2$ là tương đương ($\mathscr{T}_1 \equiv \mathscr{T}_2$), nếu với tất cả thể hiện I của quan hệ vũ trụ:

$$\mathscr{T}_1(I) = \mathscr{T}_2(I)$$

$\mathscr{T}_1$ và $\mathscr{T}_2$ là hai biểu mở rộng. θ là ánh xạ bao hàm nếu ánh xạ các hàng của $\mathscr{T}_1$ vào các hàng của $\mathscr{T}_2$ thỏa những điều kiện sau:

– Nếu w_o là giản lược của $\mathscr{T}_1$ thì $\theta(w_o)$ là giản lược của $\mathscr{T}_2$. θ ánh xạ biến phân biệt sang biến phân biệt, $\tilde{\mathscr{C}}$ sang $\tilde{\mathscr{C}}_1$,

$$\mathscr{C}_1 \subseteq \mathscr{C}$$

– θ ánh xạ kí hiệu $\tilde{\mathscr{C}}^*$ sang $\tilde{\mathscr{C}}_1$ hoặc $\tilde{\mathscr{C}}^*_1$: $\mathscr{C}_1 \subseteq \mathscr{C}$. Nếu trong cột A_k trên 2 hàng i và j của $\mathscr{T}_1$ có cùng $\tilde{\mathscr{C}}^*$ thì trên hàng $\theta(i)$ và $\theta(j)$ của $\mathscr{T}_2$ có cùng kí hiệu.

– Nếu trong cột A_k trên 2 hàng i và j của $\mathscr{T}_1$ có cùng một biến không phân biệt thì trong cột A_k của $\theta(i)$ và $\theta(j)$ có cùng một kí hiệu. Kí hiệu này có thể là biến phân biệt, biến không phân biệt, $\tilde{\mathscr{C}}$ hoặc $\tilde{\mathscr{C}}^*$.

$\mathscr{T}_1$ và $\mathscr{T}_2$ có cùng lược đồ quan hệ đích. Điều kiện cần và đủ để $\mathscr{T}_1 = \mathscr{T}_2$ là tồn tại ánh bao hàm từ $\mathscr{T}_1$ vào $\mathscr{T}_2$ và từ $\mathscr{T}_2$ vào $\mathscr{T}_1$ [9, 10].

III – BIỂU MỞ RỘNG ĐƠN, SỰ TƯƠNG ĐƯƠNG CỦA CÁC BIỂU MỞ RỘNG ĐƠN

Biểu mở rộng đơn là biểu mà nếu trong một cột có biến không phân biệt (b) lặp hoặc kí hiệu tập các giá trị không phân biệt ($\tilde{\mathscr{C}}^*$) lặp thì không có kí hiệu khác lặp trong cột đó Lớp biểu đơn là khá rộng, do đó trong thực tế chúng ta thường ít gặp biểu thức quan hệ tương ứng với biểu là không đơn.

x và w là hai hàng. Nói x phủ w nếu các điều kiện sau đây thỏa:

– x và w có cùng số cột,

– Nếu w có biến a_i (hoặc kí hiệu $\tilde{\mathscr{C}}$) trong một cột thì x cũng có biến a_i (hoặc trong cột đó.

– Nếu w có kí hiệu $\tilde{\mathscr{C}}^*$ trong một cột thì x có kí hiệu

$\tilde{\mathscr{C}}_1$ hoặc $\tilde{\mathscr{C}}^*_1$ trong cột tương ứng: $\mathscr{C}_1 \subseteq \mathscr{C}$.

S là tập các hàng của $\mathscr{T}$.x phủ S nếu x phủ từng hàng trong S.

Từ khái niệm phủ và ánh xạ bao hàm, chúng ta có nhận xét: Nếu $\mathscr{T}_1$ và $\mathscr{T}_2$ là hai biểu mở rộng đơn có giản lược đồng nhất (chỉ khác biệt bằng sự đổi tên các biến phân biệt hoặc k hiệu tập các giá trị phân biệt) và không có biến không phân biệt hoặc kí hiệu tập các giá trị không phân biệt lặp, khi đó $\mathscr{T}_1 \equiv \mathscr{T}_2$ nếu và chỉ nếu mỗi hàng của $\mathscr{T}_1$ phủ một hàng của $\mathscr{T}_2$ và mỗi hàng của $\mathscr{T}_2$ phủ một hàng của $\mathscr{T}_1$.

$\mathscr{T}$ là biểu mở rộng đơn, S – tập các hàng, w – một hàng của biểu $\mathscr{T}$. Bao đóng của tập S đối với w, kí hiệu $CL_w(S)$ là tập tối thiểu hàng chứa S thỏa những điều kiện sau

Nếu $x_1 \in CLw(S)$ và x_2 là một hàng nào đó của $\mathcal{C}$ và x_1, x_2 có cùng biến không phân biệt trong một cột nào đó nhưng trong cột ấy w có kí hiệu khác hoặc x_1 và x_2 có cùng kí hiệu không phân biệt $\tilde{\mathcal{C}}^*$ còn w có $\tilde{\mathcal{C}}_1^*$, hay $\tilde{\mathcal{C}}_1$ với $\mathcal{C}_1 \subseteq \mathcal{C}$ và $\mathcal{C}_1 \neq \mathcal{C}$ thì $x_2 \in CLw(S)$.

Khái niệm bao đóng được đưa vào với mục đích tìm tập các hàng của biểu $\mathcal{C}$ có khả năng ánh xạ vào w. Ta có định lí sau:

Định lí 1.

$\mathcal{C}$ là biểu mở rộng đơn. CLw(S) – bao đóng của tập hàng S đối với w. Ánh xạ θ được xác định như sau:

$$\theta(x) = \begin{cases} w, \text{ nếu } x \in CLw(S) \\ x, \text{ với các hàng còn lại.} \end{cases}$$

Khi đó θ là ánh xạ bao hàm từ $\mathcal{C}$ vào $\mathcal{C}$ nếu và chỉ nếu w phủ CLw(S).

Chứng minh:

Nếu: Giả thiết w phủ CLw(S). Cần chứng minh rằng $\theta(x)$ là ánh xạ bao hàm. Chứng minh theo phương pháp phản chứng. Giả thiết θ không phải là ánh xạ bao hàm, có nghĩa tồn tại 2 hàng y và z của $\mathcal{C}$ có cùng kí hiệu d trong cột B (d có thể là biến không phân biệt hoặc kí hiệu $\tilde{\mathcal{C}}^*$) và có những kí hiệu khác biệt trên các hàm $\theta(y)$ và $\theta(z)$ tương ứng trong B. Xét các trường hợp sau:

a) Cả y và z không thuộc CLw(S). Khi đó $\theta(y)=y$, $\theta(z)=z$. Điều giả thiết mâu thuẫn ở trên không thể xảy ra.

b) $y \in CLw(S)$ và $z \notin CLw(S)$ hoặc ngược lại.

Rõ ràng d không thể là biến không phân biệt hoặc kí hiệu $\tilde{\mathcal{C}}^*$. Thật vậy:

– Nếu $y[B] = z[B] = b$, $w[B] = d' \neq b$. Do $y \in CLw(S)$ suy ra $z \in CLw(S)$ dẫn đến mâu thuẫn với giả thiết.

– Nếu $y[B] = z[B] = \tilde{\mathcal{C}}^*$. Do w phủ CLw(S) và $y \in CLw(S)$ suy ra $w[B] = \tilde{\mathcal{C}}_1$ hoặc $w[B] = \tilde{\mathcal{C}}_1^*$: $\mathcal{C}_1 \subseteq \mathcal{C}$, $\mathcal{C}_1 \neq \mathcal{C}$. Theo quy tắc xây dựng CLw(S) suy ra $\in CLw(S)$, trái với giả thiết.

Như vậy d chỉ có thể là biến phân biệt (u) hoặc kí hiệu tập các giá trị phân biệt ($\tilde{\mathcal{C}}$). Suy ra $z[B] = y[B] = w[B]$.

c) – Cả y và z đều thuộc CLw(S). Khi đó $\theta(y) = \theta(z) = w$.

Trong trường hợp này không thể có giả thiết mâu thuẫn. θ là ánh xạ bao hàm.

Chỉ nếu: Điều cần chứng minh được suy ra trực tiếp từ định nghĩa của ánh xạ bao hàm và quan hệ phủ.

Định lí đã được chứng minh.

w – chuỗi là một dãy các hàng $z_1, z_2, \ldots, z_k$, $k \geq 1$ sao cho $\forall i$, $1 \leq i < k$, nếu z_i và z_{i+1} có cùng biến không phân biệt trong một cột thì w có kí hiệu khác biệt trong cột đó; nếu z_i và z_{i+1} có cùng kí hiệu $\tilde{\mathcal{C}}^*$ thì w có kí hiệu $\tilde{\mathcal{C}}_1^*$ hoặc $\tilde{\mathcal{C}}_1$: $\mathcal{C}_1 \subseteq \mathcal{C}$, $\mathcal{C}_1 \neq \mathcal{C}$.

Theo định nghĩa của bao đóng $z_k \in CLw(S)$ nếu và chỉ nếu tồn tại w–chuỗi $z_1, z_2, \ldots, z_k$ mà $z_1 \in S$.

Định lý 2.

A và B là 2 cột của biểu mở rộng đơn $\mathcal{C}$ với các biến không phân biệt lặp hoặc $\tilde{\mathcal{C}}^*$ lặp trên tập các hàng S_1 và S_2 tương ứng. Giả thiết x phủ $CL_x(S_1)$, θ là ánh xạ bao hàm từ $CLx(S_1)$ vào x và ánh xạ các hàng khác vào chính nó. $\mathcal{C}'$ là biểu nhận được từ $\mathcal{C}$ sau khi loại bỏ các hàng $CLx(S_1) - \{x\}$. $S_3 = S_2 - (CLx(S_1) - \{x\})$. Khi đó nếu S_3 có số hàng ≥ 2 và rong $\mathcal{C}'$ w phủ $CLw(S_3)$ thì trong $\mathcal{C}$ w phủ $CLw(S_2)$.

Chứng minh:

1. Xét trường hợp $x \notin CLw(S_3)$. Cần chứng minh nếu tồn tại w–chuỗi trong $CLw(S_2)$ thì w–chuỗi đó cũng trong $CLw(S_3)$. Suy ra nếu w phủ $CLw(S_3)$ trong $\mathcal{C}'$ thì w cũng phủ $CLw(S_2)$ trong $\mathcal{C}$. Việc chứng minh được tiến hành quy nạp theo độ dài của w – chuỗi rong $\mathcal{C}$ từ s tới y, $z \in S_2$.

Bước cơ sở: Độ dài của w-chuỗi bằng 1. Do y và z có cùng b hoặc $\tilde{\mathcal{C}}^*$ trong cột A nào đó nên $y \in S_2$.

Giả thiết ngược lại $y \notin S_3$ suy ra $y \in CLw(S_1)$. Do S_3 có ít nhất 2 hàng nên ta có thể giả thiết $z \in (S_3 - \{x\})$ vì vậy $z \notin CLx(S_1)$. Khi đó x phải có cùng biến không phân biệt (b) hoặc kí hiệu $\tilde{\mathcal{C}}^*$ trong cột A như trên hàng y và z, vì nếu không $z \in CLx(S_1)$. Như vậy $x \in S_2$ và $x \in S_3$ dẫn tới mâu thuẫn với giả thiết $x \notin CLw(S_3)$. Do đó $y \in S_3$ và $y \in CLw(S_3)$.

Bước quy nạp:

Giả thiết điều cần chứng minh đã đúng với trường hợp w-chuỗi có độ dài k-1. Chứng minh cũng đúng với k. Giả sử w-chuỗi có dạng $z = z_1, z_2, \ldots, z_{k-1}, z_k = y$ trong $\mathcal{T}$ và theo giả thiết quy nạp thì $z_{k-1} \in CLw(S_3)$ trong $\mathcal{T}'$. Do $x \notin CLw(S_3)$ nên $x[A] \neq z_k[A]$ vì trong trường hợp ngược lại $x \in CLw(S_3)$.

$y \notin CLx(S_1)$. Thật vậy, nếu $y \in CLx(S_1)$ thì do $y[A] = z_{k-1}[A]$ và x phủ y nên $z_{k-1} \in CLx(S_1)$. Dẫn đến trái với giả thiết quy nạp là $z_{k-1} \in CLx(S_3)$. Suy ra $y \in CLw(S_3)$ và y là một hàng của $\mathcal{T}'$.

2. Xét trường hợp $x \in CLw(S_3)$ trong $\mathcal{T}'$. θ_2 là ánh xạ bao hàm trong $\mathcal{T}'$; $\theta_2: CLw(S_3) \to w$ và θ_2 ánh xạ các hàng còn lại của $\mathcal{T}'$ vào chính nó. Khi ấy $\theta_2\theta_1$ ánh xạ $CLw(S_2)$ vào w Thật vậy

– Nếu $y \in CLw(S_2)$ và $y \notin CLx(S_1)$, khi đó $\theta_1: y \to x$ và $\theta_2: x \to w$.

– Nếu $y \in CLw(S_2)$ và $y \notin CLx(S_1)$ thì tồn tại w – chuỗi có dạng $z_1, z_2, \ldots, z_n = y$ trong $CLw(S)$. Tương tự như trên bằng phương pháp quy nạp chúng ta chứng minh được $y \in CLw(S_3)$ trong $\mathcal{T}$.

Ta đã chứng minh được $\theta_2\theta$; $CLw(S_2) \to w$ và w phủ $CLw(S_2)$ trong $\mathcal{T}$.
Sử dụng định lí trên có thể giảm số hàng của biểu mở rộng đơn theo quy tắc sau:
Với mỗi hàng i, j của $\mathcal{T}$ $(i \neq j)$, xác định có tồn tại hay không bao đóng CLi (j) bị i phủ. Trong trường hợp tồn tại CLi (j) như vậy, có thể loại bỏ các hàng của CLi (j) từ $\mathcal{T}$. Kết quả nhận được $\mathcal{T}'$ tương đương với $\mathcal{T}$. Bằng cách đó sẽ đưa biểu mở rộng đơn về dạng rút gọn.
Thủ tục COVERS có chức năng kiểm tra hàng i có phủ j không?

Procedure COVERS (i, j, result) /* i có phủ j ? */
for each column A *do*
if row j has distinguished variable of distinguished symbol in column A, and row has a different symbo in column A *then goto* (2) *else*
if row j has nondistinguished symbol $\tilde{\mathcal{C}}^*$ in column A, and row i has no symbol $\tilde{\mathcal{C}}_1$ or $\tilde{\mathcal{C}}_1^*$: $\mathcal{C}_1 \subseteq \mathcal{C}$ *then*
(2) *return* result ← false
return result ← true
and COVERS

Nếu biểu $\mathcal{T}$ có r hàng, c cột và các tập giá trị $\mathcal{C}_1$ được sắp có lực lượng lớn nhất là n thì thủ tục COVERS có độ phức tạp $O(rc + n^2 logn)$.

Thủ tục CLOSE có chức năng tìm bao đóng CLi (j) mà bị hàng i phủ. CLOSE có độ phức tạp $0(r^2c + n^2logn)$.

Procedure CLOSE (i, j, NLi (j), result 1)
k_2 ← j
(2) COVERS (i, k_1, result)
(3) *If* result = « true » *then*
begin
(4) add row k_1 to CLi (j)
for each column A *do*
for each row k *do*
for each row l in CLi (j) *do*

(5) *if* (row k and l have the same nondistinguished variable in A and row i has another symbol in A) or (row k and l have the same nondist — symbol in A and row i has $\tilde{\mathcal{C}}_1$ or $\tilde{\mathcal{C}}_1^*$; $\mathcal{C}_1 \subseteq \mathcal{C}$, $\mathcal{C}_1 \neq \mathcal{C}$) *then*
begin
(6) $k_1 \leftarrow k$; *goto* (2)
end
end
return result 1 ← result
end CLOSE

Thủ tục REDUCE có chức năng giảm các hàng của biểu $\mathcal{C}$ nếu điều đó cho phép. REDUCE có độ phức tạp $0(r^4c + r^2n^2\log n)$.

Procedure REDUCE ($\mathcal{C}$)
for each row i of $\mathcal{C}$ *do*
for each row $j \neq i$ of $\mathcal{C}$ *do*
begin
CLOSE (i, j, CLi (j), result 1)
if result 1 = « true » *then*
remove the rows in CLi (j) from $\mathcal{C}$
end
end REDUCE

Dễ dàng chứng minh định lí sau đây;

Định lí 3. $\mathcal{C}_1$ và $\mathcal{C}_2$ là 2 biểu mở rộng đơn. A là cột với kí hiệu d lặp (d có thể là biến b hoặc kí hiệu $\tilde{\mathcal{C}}^*$) trong tập hàng S_1 của $\mathcal{C}_1$. Trong $\mathcal{C}_1$ không có hàng w nào phủ CLw (S_1). Khi đó: — Nếu $\mathcal{C}_1 \equiv \mathcal{C}_2$ và d là $\tilde{\mathcal{C}}^*$ thì trong cột A tương ứng của $\mathcal{C}_2$ cũng có $\tilde{\mathcal{C}}_1^*$ lặp: $\mathcal{C} = \mathcal{C}_1$. Do đó có thể đồng nhất $\tilde{\mathcal{C}}^*$ với $\tilde{\mathcal{C}}_1^*$.

— Nếu $\mathcal{C}_1 \equiv \mathcal{C}_2$ và d là biến không phân biệt b thì trong cột A của $\mathcal{C}_2$ cũng có biến không phân biệt b_1 lặp. Khi đó có thể thay thế b và b_1 trong $\mathcal{C}_1$, $\mathcal{C}_2$ bằng kí hiệu $\tilde{\mathcal{C}}^*$ mới: $\mathcal{C}$ = dom (A). Sử dụng định lí trên chúng ta có thể xây dựng thủ tục TEST để kiểm tra sự tương đương của 2 biểu đơn $\mathcal{C}_1$ và $\mathcal{C}_2$. Thủ tục TEST có độ phức tạp $o(r^4c + r^2n^{2\log n})$.

Procedure TEST ($\mathcal{C}_1$, $\mathcal{C}_2$)
begin
(1) REDUCE ($\mathcal{C}_1$);
(2) REDUCE ($\mathcal{C}_2$);
(3) *for* each column A of $\mathcal{C}_1$ in which a repeated nondisting. variable b_1 (or nondistinguished symbol $\tilde{\mathcal{C}}_1^*$) *do*
begin
(4) *if* the column A in $\mathcal{C}_2$ has no repeated nondistinguished varible (or nondisting—symbol $\tilde{\mathcal{C}}_2^*$: $\mathcal{C}_1 = \mathcal{C}_2$) *then*
(5) *return* false ; / * $\mathcal{C}_1 \not\equiv \mathcal{C}_2$ * /
(6) let b_2 (or $\mathcal{C}_2^*$) be the repeated nondisting — variable (or repeated nondisting — symbol) in column A of $\mathcal{C}_2$;
(7) make b_1 and b_2 (or $\tilde{\mathcal{C}}_1^*$ and $\tilde{\mathcal{C}}_2^*$) be the same new symbol $\tilde{\mathcal{C}}$; label rows in which has repeated b or $\tilde{\mathcal{C}}^*$
end
(8) *if* every labeled row of $\mathcal{C}_1$ is covered by a labeled row — of $\mathcal{C}_2$ and vice versa *then*

(9) *if* every nolabeled row of $\mathscr{T}_1$ is covered by a nolabeled row of $\mathscr{T}_2$ and vice versa *then*

(10) *return true* ; /* $\mathscr{T}_1 \equiv \mathscr{T}_2$ */

(11) *return false* /* $\mathscr{T}_1 \not\equiv \mathscr{T}_2$ */

end TEST

IV – LỚP CÁC BIỂU MỞ RỘNG ĐƠN CÓ BIỂU THỨC QUAN HỆ TƯƠNG ỨNG

4.1. Một số khái niệm về siêu đồ thị

Giả thiết $X = \{x_1, x_2, \ldots, x_n\}$ tập hữu hạn và $\varepsilon = (E_i \mid i \in \mathcal{J})$ là họ các tập con của X thỏa những điều kiện sau:

1. $E_i \neq \phi$ $(i \in \mathcal{J})$

2. $\bigcup_{i \in \mathcal{J}} E_i = X$

Khi đó cặp $H = (X, \varepsilon)$ được gọi là siêu đồ thị, X là tập các đỉnh, E_i là cạnh của siêu đồ thị.

Giả thiết $A \subseteq X$. $\varepsilon_A = \{E_i \cap A \mid E_i \in \varepsilon ; E_i \cap A \neq \phi$. Khi đó $H_A = (A, \varepsilon_A)$ được gọi là siêu đồ thị con của H.

Cho siêu đồ thị $H = (X, \varepsilon)$; s và t là 2 đỉnh. Đường đi từ s tới t là dãy các cung $E_1, \ldots, E_k$ thỏa những điều kiện sau:

1. $s \in E_1$

2. $t \in E_k$

3. $E_i \cap E_{i+1} \neq \phi$, $1 \leqslant i < k$.

Khi có đường đi từ đỉnh s tới đỉnh t của siêu đồ thị H, ta nói rằng 2 đỉnh s và t là liên thông. Tương tự, ta nói 2 cung E_1 và E_k là liên thông.

$M \subseteq X$, M là tập đỉnh liên thông của siêu đồ thị nếu $\forall$ s,t a $\in M$, $t \in M$ có đường đi từ s tới t.

Thành phần liên thông của siêu đồ thị H là tập đỉnh (hoặc cung) liên thông lớn nhất Giả sử H_A là siêu đồ thị con của H. H_A được gọi là siêu đồ thị con liên thông không tầm thường nếu các đỉnh của nó là liên thông và số đỉnh của siêu đồ thị con > 1.

4.2. Lớp các biểu mở rộng đơn có biểu thức quan hệ tương ứng. Không phải bất kỳ một biểu đơn nào cũng có biểu thức quan hệ tương ứng. Điều kiện để lớp biểu mở rộng đơn có biểu thức quan hệ tương ứng được thể hiện qua định lí dưới đây:

Định lí 4. $\mathscr{T}$ là biểu mở rộng đơn với lược đồ R. $G_{\mathscr{T}}$ (W, ε) là siêu đồ thị gán nhãn được xây dựng theo quy tắc sau: W – tập đỉnh, $W = \{w_i \mid w_i$ hàng của $\mathscr{T}$, $i = \overline{1, n}\}$. ε – tập cạnh, $\varepsilon = \{E_x \mid E_x \subseteq W\}$. E_x là cạnh được gán nhãn X. $E_x = \{u, v, .., w\}$. Trong đó u, v, ..., w, x – các hàng của $\mathscr{T}$; $\exists A \in R$: $u[A] = v[A] = \ldots = w[A] = b$ (hoặc $\mathscr{C}^*$) và $x[A] = a$ hoặc $\mathscr{C}_1^*$. Khi đó biểu $\mathscr{T}$ là tương ứng với một biểu thức quan hệ SPJ, nếu và chỉ nếu không tồn tại siêu đồ thị con H liên thông không tầm thường của $G_{\mathscr{T}}$ mà có cạnh được gán nhãn bởi các nút trong H.

Chứng minh: Chỉ nếu: Giả thiết $\mathscr{T}$ là biểu mở rộng đơn tương ứng với một biểu thức quan hệ SPJ. Cần chứng minh không tồn tại siêu đồ thị con H liên thông không tầm thường. Chứng minh bằng phương pháp phản chứng.

Giả thiết $\exists$ H với các tính chất kể trên và $\exists$ biểu thức quan hệ E: $\mathscr{T}_E = \mathscr{T}$. Xây dựng cây biểu diễn E theo các quy tắc sau: – Dùng những quy tắc biến đổi để gộp dãy phép chọn nhằm thực hiện chúng một lần và đẩy xuống phía lá [6, 7]

$$\sigma_{F_1}(\ldots \sigma_{F_n}(R)\ldots) = \sigma_{F_1 \wedge \ldots \wedge F_n}(R)$$

$$F_i \text{ có dạng } A_i \in \mathscr{C}_i, \ \mathscr{C} \subseteq \mathrm{dom}(A_i)$$

$$\sigma_F(\pi_{A_1 \ldots A_m}(E)) = \pi_{A_1 \ldots A_m}(\sigma_F(E))$$

$\sigma_F(E_1 * E_2) = (\sigma_F(E_1)) * E_2$ nếu những thuộc tính trong F là các thuộc tính của E_1.

$\sigma_F(E_1 * E_2) = \sigma_{F_1}(E_1) * \sigma_{F_2}(E_2)$ nếu $F = F_1 \wedge F_2$ và F_1 chứa những thuộc tính E_1, F_2 chứa thuộc tính E_2.

– Mỗi đỉnh của siêu đồ thị tương ứng với một cây con cơ sở. Mỗi cây con cơ sở chứa một lá tương ứng với lược đồ quan hệ R. Cây con cơ sở – biểu diễn của biểu thức con có dạng tổng quát

$$\Pi X_i(\sigma_{F_{i_1} \wedge \ldots F_{i_n}}.(R))$$

X_i gồm những thuộc tính tương ứng với các cột trên hàng i của biểu $\mathcal{C}$ có kí hiệu lặp, biến a_k hoặc kí hiệu $\tilde{\mathcal{C}}$.

Trong cây biểu diễn biểu thức E, gọi z là nút cha thấp nhất của các cây con cơ sở tương ứng với những nút của siêu đồ thị con H. Giả thiết $y_1, y_2, \ldots, y_k$ – các con trực tiếp của z. $T_1, T_2, \ldots, T_k$ – những cây con có gốc là $y_1, y_2, \ldots, y_k$. Theo cách chọn z $\exists$ T_i và T_j chứa các cây con cơ sở tương ứng với các đỉnh u, v thuộc H : u, v thuộc cung E_W có nhãn w. Điều đó có nghĩa $\exists A \in R$, $u[A] = v[A] = b$ hoặc $u[A] = v[A] = \tilde{\mathcal{C}}^*$ và phép chiếu để tạo ra b hoặc $\tilde{\mathcal{C}}^*$ phải nằm trên z (theo cách chọn z). Mặt khác, $w \in H$ nên trong cây biểu diễn biểuthức E cây con cơ sở tương ứng với w cũng nằm trong T_k nào đó sau z. Vì vậy cây con đó cũng chịu tác động của phép chiếu trên. Nhưng theo giả thiết $w[A] = a$ hoặc $w[A] = \tilde{\mathcal{C}}$ dẫn tới mâu thuẫn.

Nếu: Giả thiết không tồn tại siêu đồ thị con liên thông không tầm thường H với các tính chất kể trên, khi đó tồn tại biểu thức quan hệ E tương ứng với $\mathcal{C}$: $\mathcal{C}_E = \mathcal{C}$. Trước hết ta mô tả thủ tục để xây dựng biểu thức quan hệ E. Sau đó chứng minh $\mathcal{C}_E = \mathcal{C}$. Thủ tục gồm 2 phần:

Phần 1: Xây dựng và thao tác với siêu đồ thị $G_{\mathcal{C}}$

Phần 2: Trên cơ sở kết quả của phần 1 tiến hành xây dựng biểu thức quan hệ tương ứng.

Thủ tục được tiến hành đệ quy.

– Xây dựng siêu đồ thị $G_{\mathcal{C}}$ tương ứng với $\mathcal{C}$ theo quy tắc đã nêu ra ở trên.

– Giả sử $G_1^{(1)}, G_2^{(1)}, \ldots, G_k^{(1)}$, $k \geq 2$ là các siêu đồ thị con liên thông của $G_{\mathcal{C}}$. Những siêu đồ thị con này đều thỏa điều kiện của định lí. $\mathcal{C}_1^{(1)}, \mathcal{C}_2^{(1)}, \ldots, \mathcal{C}_k^{(1)}$ là các biểu con của $\mathcal{C}$ tương ứng với những đồ thị con liên thông $G_1^{(1)}, \ldots, G_k^{(1)}$. Giản lược của $\mathcal{C}_j^{(i)}$ được xác định theo quy tắc sau: Với mỗi cột A_k nếu trong biểu $\mathcal{C}_j^{(i)}$ $\exists$ hàng w: $w[A] = a_k$ thì trên giản lược có biến phân biệt a_k; nếu $w[A] = \tilde{\mathcal{C}}$ thì trên giản lược có kí hiệu $\tilde{\mathcal{C}}$. Trong các trường hợp còn lại, trên giản lược dấu trắng.

– Từ $\mathcal{C}_j^{(i)}$ xây dựng $\mathcal{C}_j'^{(i)}$ bằng cách đổi biến không phân biệt b (hoặc kí hiệu $\tilde{\mathcal{C}}^*$) thành biến phân biệt a (hoặc kí hiệu $\tilde{\mathcal{C}}$) nếu nó đồng thời xuất hiện trong cột A của một $\mathcal{C}_k^{(i)}$ khác. Trên giản lược của $\mathcal{C}_j'^{(i)}$ được ghi thêm biến phân biệt a hoặc $\tilde{\mathcal{C}}$ trong cột A tương ứng. Do $\mathcal{C}$ là biểu mở rộng đơn nên điều đó chỉ xảy ra trên mỗi cột với nhiều nhất là một kí hiệu mà ở cột đó không chứa a hoặc $\tilde{\mathcal{C}}$ (theo cách xây dựng $G_{\mathcal{C}}$).

Rõ ràng là các biểu con $\mathcal{C}_1'^{(1)}, \ldots, \mathcal{C}_k'^{(1)}$ cũng là biểu mở rộng đơn và $G(\mathcal{C}_i'^{(1)})$ là những siêu đồ thị con tương ứng.

Quá trình trên sẽ được lặp với các $\mathcal{C}_i'^{(1)}$ mà có số hàng > 1. Kết quả sẽ nhận được các biểu con $\mathcal{C}_i^{(k)}$ chỉ gồm một hàng.

Phần 3 gồm những thủ tục để xây dựng biểu thức quan hệ $E_j^{(k)}$ tương ứng với biểu đơn $\mathscr{C}_i^{'(k)}$.

Bước cơ sở là thủ tục để biến đổi biểu đơn $\mathscr{C}_j^{(k)}$ có một hàng. Thủ tục làm việc theo quy tắc:

– Duyệt từng cột của biểu theo hàng. Nếu trong cột Ai nào đó xuất hiện kí hiệu $\tilde{\mathscr{C}}$ hoặc $\mathscr{C}^*$ thì thiết lập phép chọn $\sigma_{A_i} \in \mathscr{C}(R)$: R – lược đồ quan hệ vũ trụ. Kết quả nhận được biểu thức chọn có dạng: $\sigma_{F_1 \cap \ldots \cap F_2}(R)$

– $E_j^{(k)} = \Pi_{X_j^{(k)}} (\sigma_{F_1 \cap \ldots \cap F_1}(R)$

$X_j^{(k)}$ – gồm các thuộc tính tương ứng với những cột của biểu $\mathscr{C}_j^{'(k)}$ trên giản lược có kí hiệu khác «⊔».

Theo quy tắc xây dựng trên, dễ dàng chứng minh: $\mathscr{C}_{E_j^{(k)}} = \mathscr{C}_i^{(k)}$.

Giả thiết ở bước thứ (n–1) chúng ta có siêu đồ thị $G_{\mathscr{C}}$ gồm k thành phần liên thông $G_{\mathscr{C}_1}, \ldots, G_{\mathscr{C}_k}$ và đã xây dựng được các biểu thức quan hệ $E'_1, \ldots, E'_k$ tương ứng với $\mathscr{C}'_1, \ldots, \mathscr{C}'_k$:

$$\mathscr{C}_{E'_i} = \mathscr{C}'_i \qquad i = \overline{1, k}$$

Khi đó biểu thức quan hệ E tương ứng với $\mathscr{C}$:

$$E = \Pi_X (*_i E'_i)$$

X là tập thuộc tính tương ứng với các cột trong $\mathscr{C}$ có kí hiệu khác «⊔» trên giản lược của $\mathscr{C}$. Rõ ràng $\mathscr{C}_E = \mathscr{C}$.

Tác giả xin chân thành cảm ơn sự giúp đỡ và những ý kiến góp ý, trao đổi của đồng chí Phó tiến sĩ Lê Tiến Vương.

Ngày nhận bài 1-8-1986

TÀI LIỆU THAM KHẢO

1. Aho A. V., Sagiv Y., Ullman J. D., Equivalences among relational expressions. SIAM J.Comp. 8,2 (1979).

2. Aho A.V., Sagiv Y., Ullman J.D., Efficient optimization of a relational expressions. ACM Trans. Database Syst. 4,4 (Dec, 1979), p. 435 – 454.

3. Sagiv Y. Yannakakis M., Equivalences among relational expressions with the union and difference operators. J.of ACM Vol. 27,N°4, Oct. 1980.

4. Yannakakis M., Papadimitriou C. H., Algebraic Dependencies. J. of Computer and System Sciences 25,2 – 41 (1982).

5. Johnson D. S., Klug A., Testing containment of conjunctive queries under functional and inclusion dependencies.

J. of Comp. and Syst. Sciences 28, 167 — 189 (1984).

6. **Ullman J. D.**, Principles of Database Systems.
Comp. Science Prese 1980.

7. **Talbot S.**, An investigation into logical optimization of relational quary languages
The Computer Journal Vol 27, N°4, 1984.

8. **Berge C.**, Graphs and Hypergraphs 1973.

9. **Đỗ Xuân Thọ**, Biểu mở rộng và sự tương đương một lớp các biểu thức quan hệ.
Báo cáo tại Hội nghị toán học toàn quốc lần thứ 3 Hà-nội tháng 7/1985.

10. ДО СУАН ТХО, Эквивалентность одного класса реляционных выражении.
Acta Matematica Vietnamica N° 1, 1985.

ABSTRACT

Equivalences among simple generalized tableaux

Tableaux are used as a instrument for the optimization of query processing in relational database systems. But it has been proved that the problem of testing the equivalence of two tableaux is a NP - complete one. Therefore the problem which arises is to find such class of tableaux that the procedure for testing the equivalence of which is solvable in polynomial time. The class of simple tableaux possesses such property. The paper investigates the equivalence among simple generalized tableaux and conditions for the existence of corresponding relational expression to a simple generalized tableau.

MỘT CÁCH TIẾP CẬN...

Tiếp theo trang 18)

4. **H. Samet, A. Rosenfeld**, Quadtree structure for region processing. In Proceeding of IJCPR-80, 1980, p. 36 — 41.

5. **L.G. Shapiro**, Data structure for picture processing: A survey. CGIP — Vol 11, 1979, p. 162 — 184.

6. **L.P. Johu, S.S. Iyengar**, Space and Time efficient virtual Quadtree. IEEE Trans. PAMI, Vol PAMI-6, N°2, 1984, p. 244—248.

7. **P.N. Khôi, H. Kiém**, Code conversion in image processing. Institute of Tech. Cybernetics, Bratislava, Rept. N°14, 1984.

8. **P.N. Khôi**, Image processing based on compact representation. Institute of Tech. Cybernetics, Bratislava, Rept., 1985.

9. **P.N. Khôi**, A data structure for quadtree codes and application. Kozlemények, Budapest. N°32, 198 .

ABSTRACT

An approach to represent image objects by shaps

The paper deals with a special data structure - virtual quadtree - for representing and approximating shaps of image objects. Some operator manipulating on this data structure is proposed and followed by their properties. The last section presents the evaluation of memory space efficiency of quadtree code in relation to run length code and chain code for a given image