TẠP CHÍ KHOA HỌC TÍNH TOÁN VÀ ĐIỀU KHIỂN, Tập I (1985), Số 4 (19–23)

# VẤN ĐỀ MỞ RỘNG ĐẠI SỐ QUAN HỆ

PHẠM HẢI BẮC

## I – MỞ ĐẦU

Để lấy được thông tin cần thiết trong cơ sở dữ liệu người dùng phải biểu diễn yêu cầu của mình trên một ngôn ngữ hỏi đáp nào đó.

Trong ngôn ngữ tự nhiên, một yêu cầu thường có dạng: cho biết các tính chất A của một tập con của một lớp đối tượng nào đó, hãy xác định tập này hoặc tìm ra các tính chất B của các phần tử của tập này.

Trong cơ sở dữ liệu quan hệ, mỗi thuộc tính của một quan hệ đều tương ứng được với một tính chất của một lớp đối tượng nào đó. Điều ngược lại có thể không đúng: có những tính chất của một lớp đối tượng không tương ứng được với một thuộc tính của một quan hệ nào trong cơ sở dữ liệu. Khi một yêu cầu động chạm đến các tính chất như thế, nếu không mở rộng thêm thuộc tính cho quan hệ thì nói chung các ngôn ngữ hỏi đáp không đáp ứng được các nhu cầu của người dùng.

Mở rộng đại số quan hệ để nó khắc phục được tình trạng trên trong trường hợp thuộc tính cần thêm vào có thể tính được qua các thuộc tính khác theo một qui tắc nào đó là vấn đề được giải quyết ở bài báo này.

## II – ĐẠI SỐ QUAN HỆ MỞ RỘNG

### A – Các định nghĩa và ví dụ:

Giả sử $U_R$ là tập các thuộc tính của quan hệ R. Ta gọi $A \in U_R$ là thuộc tính phụ nếu với mọi $r \in R$, $r[A]$ có thể tính được từ $r[\overline{A}]$ và $R[\overline{A}]$ theo một qui tắc xác định ($r[A]$ là giá trị của A trong bộ r, $r[\overline{A}]$ là phần còn lại của bộ r sau khi bỏ đi $r[A]$, $R[\overline{A}]$ là phần còn lại của R sau khi bỏ đi thuộc tính A). Ta gọi $B \in U_R$ là thuộc tính ẩn của quan hệ R nếu B là thuộc tính phụ của quan hệ R' (R' là quan hệ nhận được từ R bằng cách thêm vào nó thuộc tính B).

Ví dụ: Quan hệ CÁN-BỘ với các thuộc tính

HỌ-TÊN, NĂM-SINH, LƯƠNG-CHÍNH, PHỤ-CẤP có các thuộc tính ẩn là TUỔI, TỔNG-LƯƠNG...

Ta nhận thấy rằng điều kiện cần để A là thuộc tính phụ của quan hệ R là A phụ thuộc hàm vào $\overline{A}$ ($\overline{A} = U_R \setminus A$).

Giả sử $W_R = U_R \cup V_R$ trong đó $U_R$ là tập các thuộc tính, $V_R$ là tập các thuộc tính ẩn của quan hệ R. Khi đó $A \in W_R$ được gọi là thuộc tính có thể của R, miền giá trị của A ký hiệu là D(A), giá trị của A ứng với $r \in R$ ký hiệu là $r[A]$.

Ta gọi một ánh xạ là khả dụng nếu miền xác định của nó là $\bigcup_{\alpha \in f} (R_\alpha \times \{R_\alpha\})$, trong đó $)R_\alpha)_{\alpha \in f}$ là một họ quan hệ.

Ánh xạ khả dụng f mà $f(r, R) = r[A]$ với mọi $r \in R$ được gọi là giải thích của A.

Ta nhận thấy rằng mọi thuộc tính có thể của một quan hệ bất kỳ đều tồn tại một giải thích cho nó.

Gọi : $T_1 = \{=, \neq, <, \leqslant, >, \geqslant\}$

$T_2 = \{=, \neq, \subset, \subseteq, \supset, \supseteq, \cap, \overline{\cap}\}$,

Giả sử cho quan hệ R và $\theta_1 \in T_1$, $\theta_2 \in T_2$.

Hai ánh xạ khả dụng f và g được gọi là $\theta_1$ so sánh được trên quan hệ R nếu miền xác định của chúng bao hàm tập $R \times \{R\}$ và miền giá trị của chúng là $\theta_1$ – so sánh được. h được. (Hai tập D và E là $\theta_1$ – so sánh được nếu quan hệ d $\theta_1$ e xác định với mọi $d \in D$ và $e \in E$).

Hai ánh xạ khả dụng f và g được gọi là $\theta_2$ – so sánh được trên quan hệ R nếu miền xác định của chúng bao hàm tập $R \times \{R\}$ và miền giá trị của chúng là tập con của cùng tập $\mathcal{P}(D)$ với D là tập bất kỳ ($\mathcal{P}(.)$ là tập các tập con của .).

**1. Phép chiếu mở rộng:**

Cho quan hệ R và $A = (A_1, A_2, \dots, A_k)$ trong đó $k \geqslant 1$ và $A_i \in W_R$ với $i = \overline{1, k}$. Chiếu quan hệ R lên A là quan hệ, ký hiệu là R [A], được định nghĩa như sau :

$$R[A] = \{(f_1(r, R), f_2(r, R), \dots, f_k(r, R)) \| r \in R\}$$

trong đó $f_i$, với $i = \overline{1, k}$ là giải thích của A.

Nếu $A_i \in U_R$, với $i = \overline{1, k}$ thì phép chiếu nói trên trở thành phép chiếu thông thường.

**2. Phép chọn mở rộng:**

Giả giả các ánh xạ $f_1$ và $g_1$ là $\theta_1$ – so sánh được, $f_2$ và $g_2$ là $\theta_2$ – so sánh được trên quan hệ R. Khi đó $R(f_1\ \theta_1\ g_1)$ và $R(f_2\ \theta_2\ g_2)$ là hai quan hệ được định nghĩa như sau :

$$R(f_1\ \theta_1\ g_1) = \{r \in R \mid f_1(r, R)\ \theta_1\ g_1(r, R)\}$$

$$R(f_2\ \theta_2\ g_2) = \{r \in R \mid f_2(r, R)\ \theta_2\ g_2(r, R)\}$$

Trong đó ký hiệu $f(r, R) \cap g(r, R)$ được hiểu là $f(r, R) \cap g(r, R) \neq \phi$ và ký hiệu $f(r, R)\ \overline{\cap}\ g(r, R)$ được hiểu là $f(r, R) \cap g(r, R) = \phi$. Các ký hiệu khác mang ý nghĩa thông thường.

Nếu f là giải thích của A, g là giải thích của B với $A, B \in U_R$ và D(A), D(B) là $\theta_1$ – so sánh được thì phép chọn $R(f_1\ \theta_1\ g_1)$ trở thành phép chọn thông thường dạng $R(A\ \theta_1\ B)$. Nếu $g_1 = C$ (C là hằng) còn $f_1$ là giải thích của A thì phép chọn nói trên trở thành phép chọn thông thường dạng $R(A\ \theta_1\ C)$.

**3. Phép chia mở rộng:**

Cho hai quan hệ R và S.

Giả sử $A \subseteq U_R$, $B \subseteq U_S$, R[A] và S[B] là khả hợp, $\theta$ là phần tử tùy ý thuộc $T_2$ và $\overline{A} = U_R \setminus A$. Khi đó $R/A\ \theta\ B/S$ là quan hệ được định nghĩa như sau :

$$R/A\ \theta\ B/S = \{x \in R[\overline{A}] \| im_R\ x\ \theta\ S[B]\}.$$

Trong đó $im_R\ x = \{y \in R[A] \mid (x, y) \in R\}$.

Nếu $\theta = \supseteq$ thì phép chia nói trên trở thành phép chia thông thường.

**4. Các phép toán khác:**

Ngoài các phép toán chiếu, chọn và chia đã được trình bày ở trên, các phép toán còn lại bao gồm hợp, giao, trừ, tích Đề – các và nối vẫn được giữ nguyên như trong Đại số quan hệ của Codd :

| Phép toán | Ký hiệu | Định nghĩa |
|---|---|---|
| Hợp | $R \cup S$ | $R \cup S = \{z \mid z \in R \vee z \in S\}$ |
| Giao | $R \cap S$ | $R \cap S = \{z \mid z \in R \wedge z \in S\}$ |
| Trừ | $R \setminus S$ | $R \setminus S = \{z \mid z \in R \wedge z \overline{\in} S\}$ |
| Tích Đề – Các | $R \times S$ | $R \times S = \{(z, t) \mid z \in R \wedge t \in s\}$ |
| Nối | $R * A\ \theta\ B * S$ | $R * A\ \theta\ B * S = \{(z, t) \mid z \in R \wedge t \in S \wedge z[A]\ \theta\ t[B]\}$ |

Trong các phép toán hợp, giao, trừ, hai quan hệ toán hạng phải khả hợp. Trong phép nối $A \in U_R$, $B \in U_S$ và $\theta \in T_1$.

Đại số quan hệ bao gồm các phép toán đã được trình bày ở trên gọi là B đại số.

Dưới đây ta sẽ xét một số ví dụ minh họa khả năng biểu diễn của đại số này.

Trước hết ta giới thiệu một vài ánh xạ khả dụng sẽ dùng đến trong các ví dụ.

Giả sử R là một quan hệ nào đó, A, B $\in U_R$.

Các ánh xạ:

$$f_1 = \Sigma A \text{ THEO } B, \qquad f_2 = \text{MAX } A \text{ THEO } B.$$
$$J_3 = A + B \quad \text{và} \quad f_4 = A$$

được xác định như sau:

$$f_1(r, R) = \sum_{\substack{r^* \in R \\ \wedge r^*[B] = r[B]}} r^*[A]$$

$$f(r, R) = \max_{\substack{r^* \in A \\ \wedge r^*[B] = r[B]}} r^*[A]$$

$f_3(r, R) = r[A] + r[B]$ và $f_4(r, R) = r[A]$ với mọi $r \in R$.

Quan hệ R mà ta xét là quan hệ gồm 5 thuộc tính:

A1 = Số hiệu cán bộ, A2 = Số hiện đơn vị, A3 = Năm sinh, A4 = Lương chính, $A_5$ = Phụ cấp.

*Ví dụ 1*:

Tìm số hiệu, đơn vị công tác, tuổi và tổng lương của mỗi cán bộ.

Ta có quan hệ $R_2 = R_1[A_1, A_2, B_1, B_2]$

Trong đó $B_1$ và $B_2$ là các thuộc tính ẩn.

Giải thích của $B_1$ là 1985 − $A_3$,

Giải thích của $B_2$ là A4+A5.

*Ví dụ 2*: Tìm tổng lương phải trả cho mỗi đơn vị.

Ta có quan hệ $R_3 = R_2[A_2, C_1]$

Trong đó $C_1$ là thuộc tính ẩn.

$C_1$ có giải thích là $B_2$ THEO $A_2$.

*Ví dụ 3*: Tìm những cán bộ có tổng lương cao nhất ở mỗi đơn vị.

Ta có quan hệ

$$R_4 = R_2\,(B_2 = \text{MAX } B_2 \text{ THEO } A_2).$$

**B — Một số tính chất:**

Từ định nghĩa của các phép toán chiếu, chọn và chia của B đại số ta suy ra các tính chất dưới đây:

1) Giả sử f và g là $\theta$−so sánh được trên quan hệ R, trong đó $\theta \in \{=, \neq, >, \geq, <, \leq\}$, f là giải thích của thuộc tính ẩn $B_1$, g là giải thích của thuộc tính ẩn $B_2$. Ta có,

$$R(f\theta g) = ((R[U_R, B_1, B_2])(B_1\theta B_2))[U_R].$$

2) Giả sử $\theta \in \{=, \neq, \subset, \subseteq, \supset, \supseteq, \cap, \overline{\cap}\}$ và $C = S[B]$ Khi đó

$$R/A\ \theta\ B/S = R(T\ A \text{ THEO } \overline{A}\ \theta\ C)[\overline{A}]$$

Ánh xạ r = T A THEO $\overline{A}$ được xác định như sau:

$$f(r, R) = \{r^*[A] \mid r^*[\overline{A}] = r[\overline{A}] \wedge r^* \in R\}$$

trong đó $A \in U_R$, $\overline{A} = U_R \setminus A$ và $r \in R$.

3) $R/A \subseteq B/S = R[\overline{A}] \setminus (R \setminus (R[\overline{A}]\ S[B]))[\overline{A}]$.

Chứng minh:

$$3x \in R/A \subseteq B/S \Leftrightarrow x \in R[\overline{A}] \wedge im_R x \subseteq S[B]$$

$$\Leftrightarrow (x, y) \in R[\overline{A}] \times S[B] \text{ với mọi } y \text{ mà } (x, y) \in R$$

$$\Leftrightarrow (x, y) \in R \setminus (R[\overline{A}] \times S[B]) \text{ với mọi } y$$

$$\Leftrightarrow x \overline{\in} (R \setminus (R[\overline{A}] \times S[B]))[\overline{A}]$$

$$\Leftrightarrow x \in R[\overline{A}] \setminus (R \setminus (R[\overline{A}] \times S[B]))[\overline{A}].$$

4) $R/A \supseteq B/S = R[\overline{A}] \setminus ((R[\overline{A}] \times S[B]) \setminus R))[\overline{A}]$

Chứng minh tương tự như trên.

5) $R/A \cap B/S = (R \cap (R[\overline{A}] \times S[B]))[\overline{A}]$

Chứng minh:

$$x \in R/A \cap B/S \Leftrightarrow x \in R[\overline{A}] \wedge \exists y \in S[B]\,(x, y) \in R$$

$$\Leftrightarrow x \in R[\overline{A}] \wedge \exists y\,(x, y) \in R \cap (R[\overline{A}] \times S[B])$$

$$\Leftrightarrow x \in (R \cap (R[\overline{A}] \times S[B]))[\overline{A}].$$

6) $R/A \overline{\cap} B/S = R[\overline{A}] \setminus (R \cap (R[\overline{A}] \times S[B]))[\overline{A}].$

Chứng minh: dựa vào đẳng thức

$$R/A \overline{\cap} B/S = R[\overline{A}] \setminus R/A \cap B/S.$$

## III – KẾT LUẬN

B – đại số quan hệ có thể sử dụng như một tiêu chuẩn đo sức chọn của các ngôn ngữ hỏi đáp với cơ sở dữ liệu quan hệ. So với đại số quan hệ của Codd, B–đại số có những khả năng mở rộng cần thiết cho thực tế xử lý dữ liệu nhất là về phương diện nâng cao khả năng tính toán trên các quan hệ.

Trên cơ sở của B – đại số có thể thiết kế các ngôn ngữ hỏi đáp khác nhau. Một trong những ngôn ngữ loại đó đã được cài đặt là ngôn ngữ ALQ trong hệ ngân hàng dữ liệu DBl.

Tác giả chân thành cảm ơn giáo sư Nguyễn Lâm, phó tiến sĩ Tô Tuấn và phó tiến sĩ Nguyễn Xuân Huy đã cho những ý kiến bổ ích để đi đến những kết quả đã được trình bày ở trên.

*Nhận ngày 10-10-1985*

### *TÀI LIỆU THAM KHẢO*

1. Codd E.A., Data Base Sublanguage Founded on the Relational Calculus–Proc. 1971 ACM SIGFIDET Worshop on Data Description, Access and Control.

2. Codd E., Relational Completness of Data Base Sublanguages, – Data Base Systems. Courant Computer Science Sumposia Series, 1972, V. 6, Englewood Cliffs, N.V.: Prentice–Hall,

## РЕЗЮМЕ

### ОДИН ПОДХОД К РАСШИРЕНИЮ РЕЛЯЦИОННОЙ АЛГЕБРЫ

Описывается один подход к проблеме расширения реляционной алгебры КОДДа с целью повышения её селективной мощности и удобства пользования.