

## DỊCH CHUYỂN TRUY VẤN OQL VÀO CÁC PHÉP TÍNH BAO HÀM TRANSLATING AN OQL QUERY INTO COMPREHENSION CALCULUS

**Trương Ngọc Châu**

*Trường Đại học Bách khoa, Đại học Đà Nẵng*

### TÓM TẮT

Trong bài báo này chúng tôi trình bày một hướng tiếp cận mới cho việc giải bài toán tối ưu hóa câu truy vấn đối tượng OQL (Object Query Language) trên cơ sở dữ liệu hướng đối tượng dựa trên lý thuyết về nhóm. Bài báo tập trung nghiên cứu và xây dựng các bao hàm nhóm dựa trên nhóm các kiểu, và xem các bao hàm nhóm này như là một biểu diễn trung gian cho câu truy vấn OQL, từ dạng biểu diễn trung gian này chúng ta có thể thực hiện các chiến lược tối ưu bằng các quy tắc viết lại, đồng thời làm trung gian cho các quy tắc biến đổi khác, như biến đổi từ bao hàm sang các phép toán đại số đối tượng.

### ABSTRACT

In the article, we introduce a new aspect of seeking the answer to an optimal OQL problem based on monoid theory. The article focuses on researching and building monoid comprehensions based on the types of monoid and the usages of the monoid comprehensions as a intermediate form of OQL. From this intermediary description, we can carry out an optimal strategy by means of rewriting regulations and it can play an intermediary role in different changing rules such as the change from comprehensions to Object-Oriented algebra.

### 1. Đặt vấn đề

Tối ưu hóa truy vấn hướng đối tượng là một lĩnh vực được nhiều nhà tin học quan tâm và nghiên cứu. Để tối ưu truy vấn dựa trên ngôn ngữ truy vấn đối tượng OQL, có nhiều cách tiếp cận khác nhau, như chuyển đổi mô hình cơ sở dữ liệu đối tượng về mô hình quan hệ hay quan hệ nhúng rồi áp dụng các kỹ thuật tối ưu trên quan hệ theo phương pháp truyền thống [6]. Một cách tiếp cận khác là dựa trên các phép biến đổi trực tiếp từ OQL về các biểu thức đại số đối tượng, rồi sau đó áp dụng các quy tắc biến đổi đại số đối tượng để đạt được phương án tối ưu [1, 2, 3, 6]. Nói chung các cách tiếp cận này có những hạn chế sau:

- Khả năng biểu diễn ngữ nghĩa của OQL bằng các phép toán đại số quan hệ cũng như đại số đối tượng là không đầy đủ.
- Khó biểu diễn các câu truy vấn lồng.

Bài báo giới thiệu một cách tiếp cận khác do Fegaras và Maier đề xuất [5]. Các tác giả xem các kiểu dữ liệu trong cơ sở dữ liệu đối tượng như là các nhóm (gồm nhóm sưu tập và nhóm nguyên thủy), sau đó định nghĩa và sử dụng các bao hàm nhóm dựa trên nhóm các kiểu dữ liệu như là một biểu diễn trung gian. Ưu điểm của bao hàm nhóm là biểu diễn đầy đủ các đặt trưng ngữ nghĩa của OQL và làm cơ sở cho việc biểu diễn

các chiến lược tối ưu khác nhau.

## 2. Một số khái niệm liên quan

### 2.1. Cấu trúc nhóm

Trong hệ thống hướng đối tượng, kiểu dữ liệu  $T$  thường được chia thành hai loại: kiểu dữ liệu nguyên thủy (các kiểu số như kiểu byte, int, short, long, float, double, kiểu logic *bool*, kiểu kí tự char,...) và các kiểu tham chiếu các lớp đối tượng. Trên tập kiểu dữ liệu  $T$ , chúng ta xây dựng cấu trúc nhóm như sau.

**Định nghĩa 1.** Một nhóm kiểu  $T$  là một cặp  $(T_{\oplus}, Z_{\oplus})$ , trong đó:  $T_{\oplus}$  biểu diễn nhóm kiểu  $T$ ,  $\oplus$  là một hàm kết hợp hay phép toán từ  $T \times T \rightarrow T$  và  $Z_{\oplus}$  là phần tử đơn vị của hai vế trái và phải của phép toán  $\oplus$ .

Từ định nghĩa trên ta thấy  $\oplus$  là đóng trên  $T$ ,  $Z_{\oplus}$  phần tử đơn vị của nhóm  $(T_{\oplus}, Z_{\oplus})$  và thỏa mãn:  $Z_{\oplus} \oplus x = x \oplus Z_{\oplus} = x, \forall x$ .

Một nhóm  $(T_{\oplus}, Z_{\oplus})$  có thể là nhóm giao hoán, nghĩa là:  $x \oplus y = y \oplus x$ , với  $\forall x, y$  (gọi là  $\oplus$  giao hoán) hay có thể là nhóm lũy đẳng, nghĩa là:  $x \oplus x = x, \forall x$  (gọi là  $\oplus$  lũy đẳng) hay là vừa giao hoán, vừa lũy đẳng.

Vì phép toán  $\oplus$  xác định duy nhất một nhóm. Do đó, ta thường sử dụng tên phép toán như tên nhóm, ví dụ phép toán  $+$  là nhóm kiểu nguyên, *list* nhóm kiểu danh sách, *set* nhóm kiểu tập hợp, *bag* nhóm kiểu túi. Do đó, thay vì viết nhóm  $T_{\oplus}$  ta có thể viết gọn nhóm  $\oplus$ .

Đặt  $\mathcal{T}$  là tập tất cả các nhóm kiểu dữ liệu của hệ thống, ta định nghĩa ánh xạ  $\psi: \mathcal{T} \rightarrow 2^{\{C, I\}}$ . Khi đó,  $C \in \psi(T_{\oplus})$  khi và chỉ khi  $\oplus$  có tính giao hoán và  $I \in \psi(T_{\oplus})$  khi và chỉ khi  $\oplus$  có tính lũy đẳng. Trên  $\mathcal{T}$  xác định thứ tự bộ phận  $\preceq$  giữa hai nhóm  $T_{\otimes}$  và  $T_{\oplus}$  được định nghĩa như sau:  $\otimes \preceq \oplus \Leftrightarrow \psi(T_{\otimes}) \subseteq \psi(T_{\oplus})$

**Ví dụ 1.**  $\text{list} \preceq \text{bag} \preceq \text{set}$ , vì  $\text{set}$  có tính giao hoán và lũy đẳng,  $\text{bag}$  giao hoán, còn  $\text{list}$  không giao hoán và cũng không lũy đẳng.

### 2.2. Đồng cấu nhóm

**Định nghĩa 2.** Một đồng cấu  $\text{hom}^{\oplus \rightarrow \otimes}(f): A \rightarrow B$  từ nhóm sơ tập  $(T_{\oplus}, Z_{\oplus}, U_{\oplus})$  đến nhóm bất kỳ  $(T_{\otimes}, Z_{\otimes})$ , với  $\otimes \preceq \oplus$ , được định nghĩa bằng các phương trình quy nạp sau:

$$\text{hom}^{\oplus \rightarrow \otimes}(f)(Z_{\oplus}) = Z_{\otimes}$$

$$\text{hom}^{\oplus \rightarrow \otimes}(f)(U_{\oplus}(a)) = f(a)$$

$$\text{hom}^{\oplus \rightarrow \otimes}(f)(x \oplus y) = (\text{hom}^{\oplus \rightarrow \otimes}(f)(x)) \otimes (\text{hom}^{\oplus \rightarrow \otimes}(f)(y))$$

Trong đó,  $A$  là một sơ tập kiểu  $T_{\oplus}$  và  $f$  là một hàm từ  $A$  đến  $T_{\otimes}$ . Biểu thức  $\text{hom}^{\oplus \rightarrow \otimes}(f)A$  thay thế  $Z_{\oplus}$  trong  $A$  bằng  $Z_{\otimes}$ ,  $\oplus$  bằng  $\otimes$  và  $U_{\oplus}$  bởi  $f$ .

**Định lý 1.** Đồng cấu nhóm  $\text{hom}^{\oplus \rightarrow \otimes}(f)$  bảo toàn tính chất của nhóm  $\otimes$  [5].

### 2.3. Bao hàm nhóm

Một bao hàm nhóm đối với nhóm  $\oplus$  có dạng  $\oplus\{e \mid q\}$ , trong đó:  $\oplus$  còn được gọi là phép tích lũy của bao hàm, biểu thức  $e$  được gọi là nguồn của bao hàm, mỗi số hạng  $q_i$  trong dãy  $q = q_1, \dots, q_n$ ,  $n \geq 1$  là một lượng từ và nó ở một trong các dạng sau:

Phần tử sinh có dạng  $v \leftarrow e'$ , với  $v$  là biến nhận các giá trị trong  $e'$ ,  $e'$  là một sơ tập hay còn gọi là miền của phần tử sinh hoặc

Một biểu thức logic nào đó.

**Định nghĩa 3.** Bao hàm nhóm đối với nhóm nguyên thủy hay nhóm sơ tập  $\oplus$  được định nghĩa bởi các phương trình sau:

$$\begin{aligned}\oplus\{e \mid \} &= U_{\oplus}(e) \quad \text{cho nhóm sơ tập} \\ e &\quad \text{cho nhóm nguyên thủy}\end{aligned}$$

$$\oplus\{e \mid x \leftarrow u, q\} = \text{hom}^{\otimes \rightarrow \oplus}(\lambda x. \oplus\{e \mid q\})u$$

$$\oplus\{e \mid \text{pred}, q\} = \text{if pred then } \oplus\{e \mid q\} \text{ else } Z_{\oplus}$$

Dựa vào bao hàm nhóm ta có thể dễ dàng xây dựng các phép biên dịch truy vấn OQL vào các bao hàm này như là một biểu diễn trung gian.

**Định lý 2.** Các bao hàm nhóm và các đồng cấu nhóm có cùng khả năng biểu diễn.

*Chứng minh.* Định nghĩa 3 định nghĩa các bao hàm dưới dạng các đồng cấu nhóm. Do đó, ta cần chứng minh bất kỳ đồng cấu nhóm nào đều có thể biểu diễn được dưới dạng bao hàm nhóm. Cụ thể, ta cần chứng minh đẳng thức:  $\text{hom}^{\oplus \rightarrow \otimes}(f)(A) = \otimes\{f(x) \mid x \leftarrow A, y \leftarrow f(x)\}$  cho nhóm sơ tập  $\otimes$  và  $\text{hom}^{\oplus \rightarrow \otimes}(f)(A) = \otimes\{f(x) \mid x \leftarrow A\}$  cho nhóm nguyên thủy  $\otimes$ . Theo tính chất chung của đồng cấu nhóm, ta có mọi nhóm sơ tập  $\otimes$  đều thỏa  $\text{hom}^{\otimes \rightarrow \otimes}(U_{\otimes})x = x$ . Theo Định nghĩa 3, cho một nhóm sơ tập  $\otimes$ :

$$\otimes\{y \mid x \leftarrow A, y \leftarrow f(x)\} = \text{hom}^{\oplus \rightarrow \otimes}(\lambda x. \text{hom}^{\otimes \rightarrow \otimes}(\lambda y. U_{\otimes}(y))(f(x)))(A) = \text{hom}^{\oplus \rightarrow \otimes}(f)(A)$$

đẳng thức này chứng minh cho trường hợp nhóm sơ tập. Bây giờ ta chứng minh cho nhóm nguyên thủy  $\otimes$ , ta có:

$$\otimes\{y \mid x \leftarrow A, y \leftarrow f(x)\} = \text{hom}^{\oplus \rightarrow \otimes}(\lambda x. \text{hom}^{\otimes \rightarrow \otimes}(\lambda y. y)(f(x)))(A) = \text{hom}^{\oplus \rightarrow \otimes}(f)(A), \text{ đẳng thức này chứng minh cho trường hợp nhóm nguyên thủy.}$$

### 2.4. Các phép tính bao hàm nhóm

Các phép tính bao hàm nhóm có cú pháp dựa trên nhóm các kiểu, nó có thể được

sử dụng để xây dựng các thành phần với cấu trúc phức tạp hơn.

**Định nghĩa 4.** Nhóm phép tính bao gồm các dạng cú pháp sau:

|                                           |                                                     |
|-------------------------------------------|-----------------------------------------------------|
| $NULL$                                    | giá trị null                                        |
| $c$                                       | hằng                                                |
| $v$                                       | biến                                                |
| $c.A$                                     | chiếu bảng ghi $c$ lên $A$                          |
| $\langle A_1=e_1, \dots, A_n=e_n \rangle$ | xây dựng bản ghi                                    |
| $if\ e_1\ then\ e_2\ else\ e_3$           | phát biểu if-then-else                              |
| $e_1\ op\ e_2$                            | $op$ là toán tử: $+$ , $=$ , $<$ , $>$ , ...        |
| $\lambda v:t.e$                           | hàm trừu tượng lamda                                |
| $e_1(e_2)$                                | tham chiếu phần tử $e_2$ chứa trong sưu tập $e_1$ . |
| $Z_{\oplus}$                              | phần tử đơn vị                                      |
| $U_{\oplus}(e)$                           | xây dựng sưu tập đơn vị                             |
| $e_1 \oplus e_2$                          | kết hợp                                             |
| $\oplus\{e \mid q_1, \dots, q_n\}$        | bao hàm                                             |

Trong đó,  $e, e_1, \dots, e_n$  là các thành phần trong nhóm phép tính,  $v$  là biến,  $t$  là một kiểu nhóm và  $q_1, \dots, q_n$  là các lượng từ có dạng  $v \leftarrow e$  hoặc  $e$ .

Ta sử dụng quy ước  $x \equiv u$  để biểu diễn ràng buộc biến  $x$  với giá trị  $u$ . Ý nghĩa của ràng buộc này được cho bởi quy tắc rút gọn:  $\oplus\{e \mid r, x \equiv u, s\} \Rightarrow \oplus\{e[u/x] \mid r, s[u/x]\}$ . Trong đó,  $[u/x]$  tương đương với cho  $x = u$ .

### 3. Biên dịch biểu thức OQL vào các phép tính bao hàm

Tất cả các biểu thức OQL đều được biên dịch trực tiếp vào các phép tính bao hàm. Bảng 1 mô tả kết quả biên dịch các cấu trúc chính của OQL vào các phép tính bao hàm.

**Bảng 1.** Biên dịch truy vấn OQL vào các phép tính bao hàm nhóm

| Biểu thức OQL                                                                                                                                      | Nhóm phép tính                                                                                               |
|----------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|
| 1. Cấu trúc: <i>select – from – where</i>                                                                                                          |                                                                                                              |
| <i>select [distinct] e from <math>x_1</math> in <math>e_1, \dots, x_n</math> in <math>e_n</math><br/>where pred</i>                                | $[\text{set}]\text{bag}\{e \mid x_1 \leftarrow e_1, \dots, x_n \leftarrow e_n, \text{pred}\}$                |
| <i>select <math>e(A_1, \dots, A_n, \text{partition})</math><br/>from <math>x_1</math> in <math>u_1, x_2</math> in <math>u_2(x_1), \dots</math></i> | $\text{bag}\{e(A_1, \dots, A_n, \text{partition}) \mid x_1 \leftarrow u_1, x_2 \leftarrow u_2(x_1), \dots\}$ |

|                                                                                                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $x_m \text{ in } u_m(x_1, \dots, x_{m-1})$<br>where $p(x_1, \dots, x_m)$<br>group by $A_1:e_1(x_1, \dots, x_m), \dots,$<br>$A_n:e_n(x_1, \dots, x_m)$<br>having $h(A_1, \dots, A_n, \text{partition})$ | $x_m \leftarrow u_m(x_1, \dots, x_{m-1})\},$<br>$A_1 \equiv e_1(x_1, \dots, x_m), \dots, A_n \equiv e_n(x_1, \dots, x_m),$<br>$\text{partition} \equiv \text{bag}\{ \langle x_1:y_1, \dots, x_m:y_m \rangle \mid$<br>$y_1 \leftarrow u_1,$<br>$y_2 \leftarrow u_2(y_1), \dots,$<br>$y_m \leftarrow u_m(y_1, \dots, y_{m-1}),$<br>$e_1(y_1, \dots, y_m) = A_1, \dots, e_n(y_1, \dots, y_m) = A_n \},$<br>$p(x_1, \dots, x_m), h_1(A_1, \dots, A_m, \text{partition}) \}$ |
| 2. Các toán tử                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| $e_1 \text{ intersect } e_2$                                                                                                                                                                           | $\text{set}\{x \mid x \leftarrow e_1, x \in e_2\}$                                                                                                                                                                                                                                                                                                                                                                                                                      |
| $e_1 \text{ except } e_2$                                                                                                                                                                              | $\{x \mid x \leftarrow e_1, \text{all}\{x \neq y \mid y \leftarrow e_2\}\}$                                                                                                                                                                                                                                                                                                                                                                                             |
| $e_1 \text{ union } e_2$                                                                                                                                                                               | $\text{merge}[\text{set}](e_1, e_2)$                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| 3. Các lượng từ                                                                                                                                                                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| for all $x \text{ in } e : \text{pred}$                                                                                                                                                                | $\wedge\{\text{pred} \mid x \leftarrow e\}$                                                                                                                                                                                                                                                                                                                                                                                                                             |
| exists $x \text{ in } e : \text{pred}$                                                                                                                                                                 | $\vee\{\text{pred} \mid x \leftarrow e\}$                                                                                                                                                                                                                                                                                                                                                                                                                               |
| $e_1 \text{ in } e_2$                                                                                                                                                                                  | $\vee\{x=e_1 \mid x \leftarrow e_2\}$                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| exists( $e$ )                                                                                                                                                                                          | $\vee\{\text{true} \mid x \leftarrow e\}$                                                                                                                                                                                                                                                                                                                                                                                                                               |
| unique( $e$ )                                                                                                                                                                                          | $+\{1 \mid x \leftarrow e\} = 1$                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| 4. Các hàm thống kê                                                                                                                                                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| count( $e$ )                                                                                                                                                                                           | $+\{1 \mid x \leftarrow e\}$                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| sum( $e$ )                                                                                                                                                                                             | $+\{x \mid x \leftarrow e\}$                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| max( $e$ )                                                                                                                                                                                             | $\max\{x \mid x \leftarrow e\}$                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| min( $e$ )                                                                                                                                                                                             | $\min\{x \mid x \leftarrow e\}$                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| 5. Làm phẳng tập                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| flatten( $e$ )                                                                                                                                                                                         | $\text{set}\{x \mid s \leftarrow e, x \leftarrow s\}$                                                                                                                                                                                                                                                                                                                                                                                                                   |

Sự hạn chế trong các phép biên dịch từ OQL sang các phép tính bao hàm, liên quan đến tính lũy đẳng và giao hoán của các nhóm. Ví dụ: xét câu truy vấn “select e from e in GiaoVien where e.luong>100”, ta thấy miền của truy vấn sau mệnh đề from là một tập hợp  $\text{set}(\text{GiaoVien})$  nhưng kết quả của truy vấn trả về dưới dạng túi  $\text{bag}(\text{GiaoVien})$ . Do đó, phép biên dịch truy vấn sau là không hợp lệ:  $\text{bag}\{e \mid e \leftarrow$

GiaoVien, e.luong > 100}, vì không tồn tại đồng cấu từ tập hợp sang túi. Do đó, để hợp lệ, trước hết ta phải biến đổi tập hợp chứa các đối tượng giáo viên thành túi chứa các đối tượng giáo viên nhờ vào hàm bagof(). Khi đó bao hàm trên được viết lại như sau: bag{e | e ← bagof(GiaoVien), e.luong > 100}

#### 4. Chuẩn hóa các phép tính bao hàm

Các phép tính bao hàm có thể được đưa về một dạng tương đương bằng các quy tắc viết lại, tương tự như các quy tắc viết lại trong đại số đối tượng [1,2,3,6,7]. Sự ước lượng các dạng tương đương cho thấy các dạng tương đương thường tạo ra ít cấu trúc dữ liệu trung gian hơn các bao hàm chưa được chuẩn hóa ban đầu. Hơn nữa, trong nhiều trường hợp, quy tắc chuẩn hóa cải tiến hiệu suất thực hiện chương trình. Hình 1 trình bày các quy tắc chuẩn hóa, ví dụ quy tắc R5 được áp dụng khi một bao hàm có phần tử sinh với miền là giá trị  $Z_{\otimes}$  và phần tử sinh này nằm giữa hai hệ sinh khác là  $q$  và  $s$ . Trong trường hợp này bao hàm được chuẩn hóa về giá trị  $Z_{\oplus}$ ; Quy tắc R8 mở lồng bao hàm nhóm lồng, nghĩa là bao hàm có chứa phần tử sinh mà miền của nó là bao hàm khác; Quy tắc R9 mở lồng lượng từ tồn tại; Quy tắc R10 hợp nhất hai bao hàm thành một bao hàm.

|                                                                                         |               |                                                                                                         |                                                |
|-----------------------------------------------------------------------------------------|---------------|---------------------------------------------------------------------------------------------------------|------------------------------------------------|
| $\oplus\{e \mid r, x \equiv u, s\}$                                                     | $\rightarrow$ | $\oplus\{e[u/x] \mid r, s[u/x]\}$                                                                       | R1                                             |
| $(\lambda v.e1)e2$                                                                      | $\rightarrow$ | $e1[e2/v]$                                                                                              | R2                                             |
| $\langle A1=e1, \dots, An=en \rangle.Ai$                                                | $\rightarrow$ | $ei$                                                                                                    | R3                                             |
| $\oplus\{e \mid q, v \leftarrow (\text{if } e1 \text{ then } e2 \text{ else } e3), s\}$ | $\rightarrow$ | $(\oplus\{e \mid q, e1, v \leftarrow e2, s\}) \oplus (\oplus\{e \mid q, \neg e1, v \leftarrow e3, s\})$ | R4<br>cho nhóm giao hoán $\oplus$ hay $q$ rỗng |
| $\oplus\{e \mid q, v \leftarrow Z_{\oplus}, s\}$                                        | $\rightarrow$ | $Z_{\oplus}$                                                                                            | R5                                             |
| $\oplus\{e \mid q, v \leftarrow U_{\oplus}(e'), s\}$                                    | $\rightarrow$ | $\oplus\{e \mid q, v \equiv e', s\}$                                                                    | R6                                             |
| $\oplus\{e \mid q, v \leftarrow (e1 \otimes e2), s\}$                                   | $\rightarrow$ | $(\oplus\{e \mid q, v \leftarrow e1, s\}) \oplus (\oplus\{e \mid q, v \leftarrow e2, s\})$              | R7<br>cho nhóm giao hoán $\oplus$ hay $q$ rỗng |
| $\oplus\{e \mid q, v \leftarrow \otimes\{e' \mid r\}, s\}$                              | $\rightarrow$ | $\oplus\{e \mid q, r, v \equiv e', s\}$                                                                 | R8                                             |
| $\oplus\{e \mid q, v \{ \text{pred} \mid r \}, s\}$                                     | $\rightarrow$ | $\oplus\{e \mid q, r, \text{pred}, s\}$                                                                 | R9<br>cho nhóm lũy đẳng $\oplus$               |
| $\oplus\{\oplus\{e \mid r\} \mid s\}$                                                   | $\rightarrow$ | $\oplus\{e \mid s, r\}$                                                                                 | R10<br>cho nhóm nguyên thủy $\oplus$           |

Hình 1. Các quy tắc chuẩn hóa

**Định lý 3.** Các quy tắc chuẩn hóa trong Hình 1 là bảo toàn ngữ nghĩa.

**Ví dụ 2.** Xét truy vấn lồng bằng ngôn ngữ OQL như sau

select distinct r from r in R where r.B in (select distinct s.D from s in S where r.C=s.C)

câu truy vấn này được biên dịch vào các phép tính bao hàm như sau:

$$\text{set}\{r \mid r \leftarrow R, \forall \{x = r.B \mid x \leftarrow \text{set}\{s.D \mid s \leftarrow S, r.C=s.C\}\}$$

và được chuẩn hóa thành (nhờ các quy tắc R1, R8 và R9)

$$\text{set}\{r \mid r \leftarrow R, x \leftarrow \text{set}\{s.D \mid s \leftarrow S, r.C=s.C\}, x=r.B\}$$

$$\Rightarrow \text{set}\{r \mid r \leftarrow R, s \leftarrow S, r.C=s.C, x \equiv s.D, x=r.B\}$$

$$\Rightarrow \text{set}\{r \mid r \leftarrow R, s \leftarrow S, r.C=s.C, s.D=r.D\}$$

Bao hàm này tương đương với câu truy vấn OQL sau:

$$\text{select distinct } r \text{ from } r \text{ in } R, s \text{ in } S \text{ where } r.C = s.C \text{ and } s.D = r.B.$$

Qua quá trình chuẩn hóa, ta thấy với truy vấn lồng ban đầu ta đã biến đổi nó về dạng truy vấn dạng kết nối.

## 5. Nhận xét

Qua kết quả nghiên cứu và dựa trên quy tắc chuẩn hóa bao hàm chúng tôi rút ra các nhận xét sau:

(i) *Cải tiến kích cỡ và chi phí*: quy tắc chuẩn hóa đã cho ở Hình 1 có khả năng mở lồng và rút gọn các truy vấn OQL. Vấn đề đặt ra ở đây là các quy tắc này có cải tiến được kích cỡ hay chi phí thực hiện hay không. Giả sử rằng, chúng ta biết được kích cỡ trung bình của tất cả các sưu tập trong các đường dẫn của lược đồ cơ sở dữ liệu, ví dụ cho biểu thức  $\text{set}\{e \mid x \leftarrow X, y \leftarrow x.A\}$ , khi đó kích cỡ của  $x.A$  là kích cỡ trung bình của tất cả các tập  $x.A$  ứng với mỗi  $x \in X$ .

Giả sử cho  $\text{size}[[r,s]] = \text{size}[[r]] * \text{size}[[s]]$  và  $\text{cost}[[r,s]] = \text{cost}[[r]] + \text{cost}[[s]]$ . Khi đó kích cỡ và chi phí thực hiện các bao hàm được ước lượng như sau:

$$\text{size}[[\oplus\{e \mid r\}]] = \text{size}[[r]]$$

$$\text{size}[[r, v \leftarrow e]] = \text{size}[[e]] * \text{size}[[r]]$$

$$\text{size}[[r, \text{pred}]] = \text{size}[[r]] * \text{selectivity}[[\text{pred}]]$$

$$\text{size}[[r, v \equiv e]] = \text{size}[[r]]$$

$$\text{size}[[\text{ }]] = 1$$

$$\text{cost}[[\oplus\{e \mid r\}]] = \text{cost}[[e]] + \text{cost}[[r]] + \text{size}[[r]]$$

$$\text{cost}[[r, v \leftarrow e]] = \text{cost}[[e]] + \text{cost}[[r]]$$

$$\text{cost}[[r, \text{pred}]] = \text{cost}[[\text{pred}]] + \text{cost}[[r]]$$

$$\text{cost}[[r, v \equiv e]] = \text{cost}[[e]] + \text{cost}[[r]]$$

$$\text{cost}[[\text{ }]] = 0$$

trong đó,  $\text{size}[[e]]$  là ước lượng kích cỡ của  $e$ ,  $\text{selectivity}[[\text{pred}]]$  là chọn lọc của vị từ

pred, cost[[e]] chỉ phí thực hiện e.

**Ví dụ 3.** phép kết nối  $X \triangleright \triangleleft_{x.A=y.B} Y$  có chi phí  $\text{cost}[[\text{set}\{(x,y) \mid x \leftarrow X, y \leftarrow Y, x.A=y.B\}]]$  và kích cỡ  $\text{size}[[X]] * \text{size}[[Y]] * \text{selectivity}[[x.A=y.B]]$ .

Với các công thức đã cho này, chúng ta có thể chứng minh dễ dàng rằng mỗi quy tắc chuẩn hóa trong Hình 1 đều cải tiến được kích cỡ và chi phí thực hiện. Như về trái của quy tắc R8 trong Hình 1 có chi phí

$$\begin{aligned} & \text{cost}[[\oplus\{e \mid q, v \leftarrow \text{set}\{e' \mid r\}, s\}]] \\ &= \text{cost}[[e]] + \text{cost}[[q, v \leftarrow \text{set}\{e' \mid r\}, s]] + \text{size}[[q, v \leftarrow \text{set}\{e' \mid r\}, s]] \\ &= \text{cost}[[e]] + \text{cost}[[q]] + \text{cost}[[\text{set}\{e' \mid r\}]] + \text{cost}[[s]] + \text{size}[[q]] \times \text{size}[[\text{set}\{e' \mid r\}]] \times \text{size}[[s]] \\ &= \text{cost}[[e]] + \text{cost}[[q]] + \text{cost}[[e']] + \text{cost}[[r]] + \text{size}[[r]] + \text{cost}[[s]] + \text{size}[[q]] \times \text{size}[[r]] \times \text{size}[[s]] \end{aligned}$$

chi phí này thừa một đại lượng  $\text{size}[[r]]$  so với chi phí của vế phải của cùng quy tắc R8 như sau:

$$\begin{aligned} & \text{cost}[[\oplus\{e \mid q, r, v \equiv e', s\}]] \\ &= \text{cost}[[e]] + \text{cost}[[q, r, v \equiv e', s]] + \text{size}[[q, r, v \equiv e', s]] \\ &= \text{cost}[[e]] + \text{cost}[[q]] + \text{cost}[[e']] + \text{cost}[[r]] + \text{cost}[[s]] + \text{size}[[q]] * \text{size}[[r]] * \text{size}[[s]]. \end{aligned}$$

(ii) *Hạn chế*: Mặc dù quy tắc chuẩn hóa có thể mở lồng nhiều dạng truy vấn lồng, nhưng vẫn có một số dạng truy vấn lồng không thể mở lồng được. Ví dụ như dạng truy vấn sau:

$\text{set}\{ \langle E = e, M = \text{set}\{c \mid c \leftarrow e.\text{children}, \wedge \{c.\text{age} > d.\text{age} \mid d \leftarrow e.\text{manager.children}\} \rangle \}$

$e \leftarrow \text{Employee}, e.\text{salary} > \max\{m.\text{salary} \mid m \leftarrow \text{Manager}, e.\text{age} > m.\text{age}\}$ .

Truy vấn con:

$\text{set}\{c \mid c \leftarrow e.\text{children}, \wedge \{c.\text{age} > d.\text{age} \mid d \leftarrow e.\text{manager.children}\}\}$

không thể mở lồng bởi các quy tắc chuẩn hóa bởi vì tập tính toán phải được nhúng trong kết quả của mỗi vòng lặp của bao hàm ngoài. Tương tự, với các lượng từ phổ dụng: all, exist, some,... và hàm gộp: max, min, sum, count, avg,... không thể được mở lồng bởi quy tắc chuẩn hóa.

## 6. Kết luận

Để tối ưu hóa truy vấn một biểu thức OQL bất kỳ trên CSDL hướng đối tượng thì bắt buộc phải dựa vào các phép biến đổi trung gian nào đó. Trong bài báo này chúng tôi đã tập trung nghiên cứu một phương pháp tiếp cận mới, đó là dựa trên lý thuyết về

nhóm trong toán học Đại số; Phương pháp tiếp cận này đơn giản, dễ biểu diễn vì dựa trên cơ sở logic ngữ nghĩa của toán học và đặc biệt là nó biểu diễn được trọn vẹn ngữ nghĩa của câu truy vấn OQL bất kỳ.

### TÀI LIỆU THAM KHẢO

- [1] E. Andonoff, G. Hubert, A. Le Parc, G. Zurfluh, A Query Algebra for Object-Oriented Databases Integrating Version, *IRIT / SIG, IEEE, PP 62-72, 1997.*
- [2] C. Beeri and y. Kornatzky, Algebraic Optimization of Object-Oriented Query Languages. *In Proc. ICDT, Paris, France, 1990.*
- [3] S. Cluetand, c. Delobel, A General Framework for the Optimization of Object-Oriented Queries, *ACM SIGMOD, PP. 382-392, 1992.*
- [4] Elmasri, Navathe, Fundamentals Database Systems, *5<sup>th</sup> Edition*, the United States of America, 2007.
- [5] L. Fegaras and D. Maier. Towards an effective calculus for object oriented query languages. *In Proceedings of the SIGMOD-SIGACT-SIGART international conference on management of data*, pages 47–58, 2000.
- [6] Hoàng Bảo Hùng, Truy vấn và tối ưu hóa truy vấn trong cơ sở dữ liệu hướng đối tượng, *Luận án tiến sĩ toán học*, 2007.