Tạp chí Tin học và Điều khiển học, T.25, S.3 (2009), 201–213

TỐI ƯU HÓA CÁC TRUY VẤN ĐỐI TƯỢNG CÓ CHỨA CÁC BIỂU THỨC ĐƯỜNG DẪN

TRƯƠNG NGỌC CHÂU[1], ĐOÀN VĂN BAN[2]

[1] *Trường Đại học Bách khoa, Đại học Đà Nẵng*
[2] *Viện Công nghệ thông tin, Viện Khoa học và Công nghệ Việt Nam*

Abstract. Path query is an independent feature in the system of Oject-Oriented DataBase. So, determining one stategy of optimizing evaluation of querying path is an important issue in the optimization problem of object queries. The article focuses on the researching and the developing the method of optimizing the path expressions, based on semantic analysis of selection and join operations in the path expressions, to apply suitable rewriting rules. Then building a query graph from rewriting expressions in the form of many kinds of selections, joins and finding the maximum common sub path expressions. Therefore, we make out the steps in order to fullfill general query.

Tóm tắt. Truy vấn đường dẫn là một đặc trưng độc lập trong các hệ thống cơ sở dữ liệu hướng đối tượng. Do đó, việc xác định một chiến lược ước lượng tối ưu các truy vấn đường dẫn là một vấn đề quan trọng trong bài toán tối ưu truy vấn đối tượng tổng quát. Bài báo tập trung nghiên cứu và mở rộng phương pháp tối ưu truy vấn các biểu thức đường dẫn, dựa trên cơ sở phân tích ngữ nghĩa của các phép chọn và kết nối trong các biểu thức đường dẫn, để từ đó áp dụng các quy tắt viết lại thích hợp. Sau đó, xây dựng đồ thị truy vấn từ các biểu thức sau khi viết lại với các kiểu khác nhau của phép chọn, kết nối và tìm các biểu thức đường dẫn con chung lớn nhất, để từ đây tạo ra các lược đồ thực thi truy vấn tổng quát.

1. GIỚI THIỆU

Mục đích tối ưu hóa truy vấn đối tượng là để tìm ra một lược đồ thực thi với chi phí thấp nhất có thể. Các bước tối ưu hóa truy vấn có thể liên quan đến hai mức:

- Tối ưu hóa truy vấn ở mức lôgic: viết lại truy vấn bằng cách sử dụng các tính chất ngữ nghĩa của ngôn ngữ để tìm ra các biểu thức tương đương có chi phí thấp hơn truy vấn ban đầu.
- Tối ưu hóa truy vấn ở mức vật lý: dựa trên mô hình chi phí cơ sở để tìm ra thuật toán ước lượng truy vấn tốt nhất.

Các truy vấn đường dẫn, nghĩa là ước lượng các đường dẫn bắt đầu từ một tập các đối tượng và ước lượng một vị từ dựa trên giá trị của biểu thức đường dẫn là một lớp bài toán quan trọng của các truy vấn dạng khai báo. Các truy vấn đường dẫn là một đặt trưng độc lập trong các hệ thống cơ sở dữ liệu hướng đối tượng. Do đó, việc xác định một chiến lược ước lượng truy vấn đường dẫn tối ưu là một vấn đề quan trọng trong bài toán tối ưu truy vấn đối tượng tổng quát.

Có nhiều tiếp cận khác nhau để tối ưu truy vấn biểu thức đường dẫn [1-9], và các tiếp cận này chủ yếu tập trung tối ưu ở mức vật lý dựa trên mô hình chi phí cơ sở [3]. Để tăng hiệu quả ước lượng các truy vấn đường dẫn, các tác giả [4][7][8] đã tận dụng các cơ chế chỉ mục để tìm ra các chiến lược tối ưu biểu thức đường dẫn sao cho hiệu quả nhất. Với các biểu thức đường dẫn bội, Cetin Ozkan [2] đã xây dựng cơ sở heuristic để tối ưu các truy vấn dựa vào chi phí và độ chọn lọc của mỗi biểu thức đường dẫn, bằng cách xem xét một biểu thức đường dẫn như một đơn vị xử lý. Một cách tiếp cận khác của [5] là đề xuất thuật toán xử lý song song để tính toán các biểu thức đường dẫn, thuật toán PCSJ (Parallel Cascade Semi-join) do nhóm tác giả này đề xuất cải tiến một cách hiệu quả hiệu suất thực thi khi tính toán các biểu thức đường dẫn với các vị từ tương ứng.

Tận dụng lý thuyết tối ưu truy vấn trên cơ sở dữ liệu quan hệ, các tác giả trong [1] đã tiến hành biên dịch các biểu thức đường dẫn về các kết nối ngoài dựa trên các bảng (thực thể) của cơ sở dữ liệu quan hệ. Mục đích của phương pháp này là biên dịch các biểu thức đường dẫn thành các bảng tương ứng trong cơ sở dữ liệu quan hệ, sau đó biên dịch truy vấn đối tượng thành các truy vấn quan hệ, rồi áp dụng các chiến lược tối ưu trên truy vấn quan hệ.

Hướng tiếp cận khác do Zhaohui Xie [9] đề xuất nhằm giảm thời gian cần thiết để ước lượng một truy vấn và làm giảm kích thước bộ nhớ tạm thời trong một số giai đoạn ước lượng. Phương pháp tối ưu các biểu thức đường dẫn trong [9] là ở mức lôgic, sử dụng các quy tắc viết lại dựa vào ngữ nghĩa của các phép chọn và kết nối trên các biểu thức đường dẫn. Tương tự như các heuristic tối ưu trong các truy vấn quan hệ [6], vị từ trong phép chọn biểu thức đường dẫn được đơn giản hóa và được đẩy vào thuộc tính cuối cùng của biểu thức đường dẫn nếu có thể, nghĩa là nó được thực hiện trước khi thực hiện các phép kết nối ẩn giữa các lớp kết hợp với biểu thức đường dẫn và được thực hiện sớm như có thể. Tương tự, phép kết nối hiển giữa hai biểu thức đường dẫn có thể được ước lượng trước khi thực hiện các kết nối ẩn giữa các lớp kết hợp với hai biểu thức đường dẫn. Tiếp cận của Zhaohui Xie chỉ xét các truy vấn đường dẫn độc lập mà chưa xét trường hợp các truy vấn đường dẫn ở dạng hội và tuyển. Nghĩa là, các truy vấn đường dẫn được kết hợp với nhau bởi các toán tử lôgic **and** hoặc **or**.

Trong bài báo này, chúng tôi mở rộng phương pháp tối ưu truy vấn đường dẫn dạng hội và tuyển của Zhaohui Xie. Ưu điểm của phương pháp này là đối với các biểu thức đường dẫn trong các truy vấn dạng hội hay tuyển có biểu thức đường dẫn con chung thì biểu thức đường dẫn con chung này chỉ được ước lượng duy nhất một lần trong quá trình ước lượng, do vậy nâng cao được hiệu quả truy vấn.

2. MỘT SỐ KHÁI NIỆM CƠ SỞ

Trong phần này, chúng tôi nhắc lại một số định nghĩa, tính chất và các định lý đã được đề xuất ở [9] làm cơ sở để mở rộng quy tắc tối ưu truy vấn đường dẫn dạng hội và tuyển ở phần nghiên cứu tiếp theo.

Ví dụ 1. Để thống nhất trong trình bày các ví dụ minh họa, chúng ta xét cơ sở dữ liệu được minh hoạ như sau

```
define class NHANSU:
    type tuple( hoTen: String;
        gioiTinh: String;
        namSinh: String;
        )
end NHANSU
define class SINHVIEN inherits NHANSU:
    type tuple( thuocLop: LOP;
        )
end SINHVIEN

define class KHOA:
    type tuple( tenKhoa: String;
        truongKhoa: NHANSU;
        dsNganh : set(String);
        dsLop: set(LOP);
        )
end KHOA
define class LOP:
    type tuple( tenLop: String;
        chuNhiem: NHANSU;
        banCanSu: set(NHANSU);
        khoaHoc: String;
        dsSinhVien: set(SINHVIEN);
        thuocKhoa: KHOA;
        )
end LOP
```

Một biểu thức đường dẫn biểu diễn một điều hướng từ một đối tượng trong một lớp đến các đối tượng trong các lớp khác qua các quan hệ thuộc tính trên một phân cấp hợp thành lớp.

Định nghĩa 1. $o_0.A_1...A_n$ là một biểu thức đường dẫn kết hợp với các lớp $O_0, O_1, ..., O_n$ nếu o_0 là đối tượng thuộc lớp O_0 và A_i là thuộc tính của lớp O_{i-1} biến thiên trên lớp O_i hay tập các đối tượng của O_i, với $i = 1, 2, ..., n$.

Ví dụ, biểu thức đường dẫn sv.thuocLop.khoaHoc cho biết khóa học mà sinh viên sv theo học; sv.thuocLop.banCanSu.gioiTinh cho biết giới tính các cán sự lớp của sinh viên *sv*.

Thuộc tính A_n có thể là nguyên thủy, như các thuộc tính tenLop, khoaHoc hay không nguyên thủy như chuNhiem, dsSinhVien và thuocKhoa.

Định nghĩa 2. $(o_0, o_1, ..., o_n)$ là một đường dẫn đối tượng thỏa mãn $o_0.A_1...A_n$ nếu cho $i = 1, 2, ..., n$ thì $o_i = o_{i-1}.A_i$ khi A_i là một thuộc tính đơn trị (giá trị của thuộc tính A_i có duy nhất một giá trị) và $o_i \in o_{i-1}.A_i$ khi A_i là một thuộc tính đa trị (giá trị của thuộc tính A_i có thể là một sưu tập các giá trị).

Ví dụ, nếu sinh viên sv thuộc lớp '05T1', học khóa '2005-2010' thì (sv, 05t1, 0510) là một đường dẫn đối tượng thỏa sv.thuocLop.khoaHoc, trong đó, 05t1 và 0510 là các định danh đối tượng (OID) của 05T1 và '2005-2010' tương ứng.

Định nghĩa 3. Một biểu thức đường dẫn không có bất kỳ thuộc tính đa trị nào được định nghĩa là một biểu thức đơn trị, ngược lại nó là một biểu thức đa trị.

Ví dụ, sv.thuocLop.khoaHoc là một biểu thức đơn trị, vì mọi thuộc tính trong biểu thức đường dẫn điều là thuộc tính đơn trị và sv.thuocLop.banCanSu.gioiTinh là một biểu thức đường dẫn đa trị, vì có thuộc tính banCanSu là một thuộc tính đa trị. Do đó, một sinh viên có thể có nhiều cán sự lớp (như lớp trưởng, lớp phó, ...), vì vậy, có thể có nhiều hơn một đường dẫn đối tượng thỏa mãn sv.thuocLop.banCanSu.gioiTinh.

Định nghĩa 4. Nếu $o_0.A_1...A_n$ là một biểu thức đường dẫn đơn trị thì giá trị của biểu thức đường dẫn $o_0.A_1...A_n$ là $o_{n-1}.A_n$, trong đó, $(o_0, o_1, ..., o_n)$ là một đường dẫn đối tượng thỏa mãn $o_0.A_1...A_n$. Nếu $o_0.A_1...A_n$ là một biểu thức đường dẫn đa trị thì trị của biểu thức đường dẫn $o_0.A_1...A_n$ là $\{o_n|(o_0, o_1, ..., o_n)$ thỏa $o_0.A_1...A_n$ và $o_n = o_{n-1}.A_n$ nếu A_n là thuộc tính đơn trị hoặc $o_n \in o_{n-1}.A_n$ nếu A_n là thuộc tính đa trị$\}$.

Ví dụ, nếu sinh viên sv học khóa 2005-2010 thì giá trị của sv.thuocLop.khoaHoc là 2005-2010; nếu sinh viên sv có ban cán sự lớp là các bạn Trần Thuận và Lê Thị Vân thì giá trị của sv.thuocLop.banCanSu.hoTen là {Trần Thuận, Lê Thị Vân}.

Để ước lượng biểu thức đường dẫn $o_0.A_1...A_n$ tương đương với việc ước lượng các kết nối ẩn: $O_0 \bowtie O_1 \bowtie ... \bowtie O_{n-1}$, trong đó các thể hiện là các đường dẫn đối tượng thỏa mãn $o_0.A_1...A_n$, với $o_0 \in O_0$. Vì giá trị của các biểu thức đường dẫn có thể là các tập giá trị (đa trị), nên các lượng từ $\exists$ và $\forall$ phải được đưa vào các thao tác chứa biểu thức đường dẫn.

Định nghĩa 5. Các phép chọn trên biểu thức đường dẫn có dạng như sau:

$$s_{q_1}\theta_{q_2}C$$

trong đó, s là biểu thức đường dẫn, C là một hằng hay tập các hằng, $\theta \in \{=, \neq, >, <, \geq, \leq\}$ và $q_1, q_2 \in \{\forall, \exists\}$. Nếu s là biểu thức đường dẫn đơn trị thì q_1 có thể là $\exists$ hay $\forall$. Nếu C là một hằng thì q_2 có thể là $\exists$ hay $\forall$. Trong cả hai trường hợp, q_1 và q_2 luôn chọn mặc định là $\exists$. Phép chọn biểu thức đường dẫn được định nghĩa theo kiểu $\theta(q_1)$.

Ví dụ, cho phép chọn biểu thức đường dẫn $o_{lop}.dsSinhVien.namSinh_{\forall} >_{\exists} \{1985, 1990\}$, ta có kết quả của phép chọn biểu thức đường dẫn này gồm tập các đối tượng $\{o_{lop}|o_{lop} \in LOP\}$ chỉ gồm các sinh viên có năm sinh sau 1985 hoặc 1990, và nó là phép chọn biểu thức đường dẫn có kiểu $>$ ($\forall$).

Tương tự như các heuristic tối ưu trong các truy vấn quan hệ [6], phép chọn trên biểu thức đường dẫn có thể được đẩy vào trước các kết nối ẩn và có thể được thực hiện sớm như có thể. Vấn đề đặt ra ở đây là: *khi nào và làm cách nào để đẩy phép chọn vào trước các phép kết nối ẩn.* Định lý 1 sau chỉ ra các chiến lược tối ưu khác nhau, được áp dụng đến các kiểu khác nhau của các phép chọn.

Định lý 1. ([9])

Nếu phép chọn biểu thức đường dẫn $o.A_1...A_nq_1\theta q_2C$ *có kiểu* $\theta(\exists)$ *thì* $\{o|o.A_1...A_nq_1\theta q_2C, o \in O_0\} = \pi_{(O_0)}(\sigma_{(o.A_1...A_nq_1\theta q_2C)}(O_0 \bowtie ... \bowtie O_n)) = \pi_{(O_0)}(O_0 \bowtie ... \bowtie \sigma_{(o_nq_1\theta q_2C)}(O_n))$.
Nếu phép chọn biểu thức đường dẫn $o.A_1...A_nq_1\theta q_2C$ *có kiểu* $\theta(\forall)$ *thì* $\{o|o.A_1...A_nq_1\theta q_2C, o \in O_0\} = \pi_{(O_0)}(\sigma_{(o.A_1...A_nq_1\theta q_2C)}(O_0 \bowtie ... \bowtie O_n))$.
trong đó, $o.A_1...A_n$ *là một biểu thức đường dẫn được kết hợp với các lớp* $O_0, ..., O_n$; $o_n \in O_n$; $q_1, q_2 \in \{\forall, \exists\}$ *và* $\theta \in \{=, \neq, >, <, \geq, \leq\}$.

Ví dụ 2. Cho phép chọn biểu thức đường dẫn

$$sv.thuocLop.chuNhiem.hoTen_{\exists} = {}_{\exists}'MaiAnh'$$

là phép chọn biểu thức đường dẫn có kiểu $= (\exists)$, phép chọn này cho kết quả gồm các sinh viên sv có thầy giáo chủ nhiệm có họ tên là Mai Anh.

$\{sv|sv.thuocLop.chuNhiem.hoTen_{\exists} =_{\exists}' MaiAnh'\}$

$= \pi_{(SINHVIEN)}(SINHVIEN \bowtie LOP \bowtie \sigma_{(ns.hoTen='MaiAnh')}(NHANSU))$, trong đó, $ns \in NHANSU$.

Định nghĩa 6. Phép so sánh biểu thức đường dẫn có dạng $s_{q_1}\theta_{q_2}t$, trong đó, $q_1, q_2 \in \{\forall, \exists\}$, s và t là các biểu thức đường dẫn, $\theta \in \{=, \neq, >, <, \geq, \leq\}$. Nếu s là biểu thức đường dẫn đơn trị thì q_1 có thể là $\exists$ hay $\forall$. Trong cả hai trường hợp, q_1 và q_2 luôn được chọn mặc định là $\exists$. Phép so sánh biểu thức đường dẫn được định nghĩa dưới dạng $\theta(q_1, q_2)$.

Ước lượng phép so sánh hai biểu thức đường dẫn: $o.A_1...A_n$ và $o'.B_1...B_m$ được biến đổi thành phép kết nối hiển giữa $O_0 \bowtie ... \bowtie O_{n-1}$ và $O_0' \bowtie ... \bowtie O_{m-1}'$, trong đó phép toán so sánh trở thành điều kiện kết nối.

Do đó, phép so sánh hai biểu thức đường dẫn có thể được xem như là một phép kết nối hiển giữa hai biểu thức đường dẫn, trong đó, phép toán so sánh θ chính là phép toán kết nối.

Vấn đề đặt ra ở đây là: *khi nào và làm cách nào để đơn giản hóa điều kiện kết nối hiển và đẩy chúng vào trước các kết nối ẩn giữa các lớp được kết hợp với các thuộc tính trong các biểu thức đường dẫn tương ứng.* Định lý 2 sau chỉ ra các chiến lược tối ưu khác nhau, được áp dụng dựa vào các kiểu khác nhau của phép kết nối hiển.

Định lý 2. ([9])

Nếu phép so sánh biểu thức đường dẫn $o.A_1...A_nq_1\theta q_2o'.B_1...B_m$ có kiểu $\theta(\exists, \exists)$ thì

$\{(o, o')|o.A_1...A_nq_1\theta q_2o'.B_1...B_m, o \in O_0, o' \in O_0'\}$

$= \pi_{(O,O')}(\sigma_{(o.A_1...A_nq_1\theta q_2o'.B_1...B_m)}(O_0 \bowtie ... \bowtie O_n \bowtie O_m' \bowtie ... \bowtie O_0'))$

$= \pi_{(O,O')}(O_0 \bowtie ... \bowtie \sigma_{(o_nq_1\theta q_2o_m')}(O_n \bowtie O_m') \bowtie ... \bowtie O_0').$

Nếu phép so sánh biểu thức đường dẫn $o.A_1...A_nq_1\theta q_2o'.B_1...B_m$ có kiểu $\theta(\exists, \forall)$ thì

$\{(o, o')|o.A_1...A_nq_1\theta q_2o'.B_1...B_m, o \in O_0, o' \in O_0'\}$

$= \pi_{(O,O')}(\sigma_{(o.A_1...A_nq_1\theta q_2o'.B_1...B_m)}(O_0 \bowtie ... \bowtie O_n \bowtie O_m' \bowtie ... \bowtie O_0'))$

$= \pi_{(O,O')}(O_0 \bowtie ... \bowtie \sigma_{(o_nq_1\theta q_2o'.B_1...B_m)}(O_n \bowtie O_m' \bowtie ... \bowtie O_0').$

Nếu phép so sánh biểu thức đường dẫn $o.A_1...A_nq_1\theta q_2o'.B_1...B_m$ có kiểu $\theta(\forall, \exists)$ thì

$\{(o, o')|o.A_1...A_nq_1\theta q_2o'.B_1...B_m, o \in O_0, o' \in O_0'\}$

$= \pi_{(O,O')}(\sigma_{(o.A_1...A_nq_1\theta q_2o'.B_1...B_m)}(O_0 \bowtie ... \bowtie O_n \bowtie O_m' \bowtie ... \bowtie O_0'))$

$= \pi_{(O,O')}(\sigma_{(o.A_1...A_nq_1\theta q_2o_m')}(O_0 \bowtie ... \bowtie O_n \bowtie O_m') \bowtie ... \bowtie O_0').$

Nếu phép so sánh biểu thức đường dẫn $o.A_1...A_nq_1\theta q_2o'.B_1...B_m$ có kiểu $\theta(\forall, \forall)$ thì

$\{(o, o')|o.A_1...A_nq_1\theta q_2o'.B_1...B_m, o \in O_0, o' \in O_0'\}$

$= \pi_{(O,O')}(\sigma_{(o.A_1...A_nq_1\theta q_2o'.B_1...B_m)}(O_0 \bowtie ... \bowtie O_n \bowtie O_m' \bowtie ... \bowtie O_0')).$

trong đó, $o.A_1...A_n$ và $o'.B_1...B_m$ là các biểu thức đường dẫn được kết hợp với các lớp $O_0, ..., O_n$ và $O_0', ..., O_m'$ tương ứng, $o_n \in O_n, o_m' \in O_m'$; $q_1, q_2 \in \{\forall, \exists\}$ và $\theta \in \{=, \neq, >, <, \geq, \leq\}$.

Ví dụ 3. Cho phép kết nối biểu thức đường dẫn

$$k.truongKhoa.hoTen_{\exists} = {}_{\exists}l.chuNhiem.hoTen$$

là một phép kết nối biểu thức đường dẫn có kiểu $= (\exists, \exists)$, phép kết nối này có nghĩa là trưởng khoa của khoa k có họ tên trùng với họ tên của giáo viên chủ nhiệm lớp l bất kỳ nào đó.

$\{(k,l)|k.truongKhoa.hoTen_{\exists} =_{\exists} l.chuNhiem.hoTen, k \in KHOA, l \in LOP\}$

$= \pi_{(KHOA,LOP)}(KHOA \bowtie \sigma_{(ns.hoTen=ns'.hoTen)}(NHANSU \bowtie NHANSU') \bowtie LOP)$, trong đó, $ns \in NHANSU, ns' \in NHANSU'$.

3. TRUY VẤN VỚI CÁC PHÉP CHỌN ĐƯỜNG DẪN DẠNG HỘI HOẶC TUYỂN

Trong các hệ thống cơ sở dữ liệu lớn thường có rất nhiều truy vấn phức tạp, các truy vấn này thường dài và có các vị từ phức tạp. Vì vậy, vấn đề đặt ra là các trình tối ưu phải tìm ra được các lược đồ thực thi sao cho hiệu quả nhất, mặc dù các biểu thức truy vấn phức tạp. Có nhiều nghiên cứu về vấn đề xử lý một số lượng lớn các kết nối, nhưng vấn đề tối ưu các điều kiện chọn phức tạp liên quan đến các phép toán lôgic AND và OR vẫn chưa được quan tâm đúng mức. Trong phần này chúng tôi tập trung vào vấn đề tối ưu các truy vấn chứa các phép chọn biểu thức đường dẫn phức tạp.

Định nghĩa 7. Cho hai biểu thức đường dẫn $P_1 = o_0.A_1...A_n$ và $P_2 = o_0.A'_1...A'_m$, hai đường dẫn P_1 và P_2 được gọi là chồng lấp nếu tồn tại $j < min(n,m)$ sao cho: $A_i = A'_i (1 \le i \le j)$ và $A_i \neq A'_i$ nếu $i > j$. Biểu thức $P = o_0.A_1...A_i = o_0.A'_1...A'_i$ được gọi là biểu thức đường dẫn con chung lớn nhất của P_1 và P_2.

Ví dụ, hai biểu thức đường dẫn sv.thuocLop.khoaHoc và sv.thuocLop.banCanSu.gioiTinh, có biểu thức đường dẫn con chung lớn nhất là sv.thuocLop.

Cho biểu thức $(P_{1q_1}\theta_{1q_2}C_1)\Delta(P_{2q'_1}\theta_{2q'_2}C_2)$ gồm hai phép chọn biểu thức đường dẫn được nối với nhau bởi toán tử lôgic $\Delta \in \{and, or\}$, trong đó, $P_1 = o_0.A_1...A_n$ và $P_2 = o_0.A'_1...A'_m$ là các biểu thức đường dẫn. Khi đó, tập các đối tượng thỏa mãn biểu thức $(P_{1q_1}\theta_{1q_2}C_1)\Delta(P_{2q'_1}\theta_{2q'_2}C_2)$ là tập các đối tượng của biểu thức sau:

$$\{o_0|o_0.A_1...A_nq_1\theta_1q_2C1, o_0 \in O_0\} \triangledown \{o_0|o_0.A'_1...A'_nq'_1\theta_2q'_2C2, o_0 \in O_0\}$$

trong đó, $\triangledown = \begin{cases} \cap & \text{nếu } \Delta \text{ là toán tử } \textit{and} \\ \cup & \text{nếu } \Delta \text{ là toán tử } \textit{or}. \end{cases}$

Ví dụ 4. Cho phép chọn biểu thức đường dẫn dạng hội và tuyển như sau

(sv.gioiTinh$_{\exists} = {}_{\exists}$'Nữ') **and** (sv.thuocLop.thuocKhoa.tenKhoa$_{\exists} = {}_{\exists}$ 'CNTT' **or**
sv.thuocLop.thuocKhoa.truongKhoa.hoTen$_{\exists} = {}_{\exists}$ 'Mai Anh')

kết quả của phép chọn này là tập các sinh viên sv có giới tính là 'Nữ' và thuộc khoa 'CNTT' hay thuộc khoa có chủ nhiệm khoa là Mai Anh.

{sv — sv.gioiTinh$_{\exists} = {}_{\exists}$ 'Nữ', $sv \in SINHVIEN\} \cap$ ({sv — sv.thuocLop.thuocKhoa.tenKhoa$_{\exists} =_{\exists}$

’CNTT’, $sv \in SINHVIEN\} \cup \{$sv.thuocLop.thuocKhoa.truongKhoa.hoTen$_\exists =_\exists$ ’Mai Anh’, $sv \in SINHVIEN\})$

Trong biểu thức trên, ta thấy biểu thức đường dẫn sv.thuocLop.thuocKhoa được ước lượng hai lần bởi hai phép chọn biểu thức đường dẫn:

sv.thuocLop.thuocKhoa.tenKhoa$_\exists =_\exists$ ’CNTT’ và

sv.thuocLop.thuocKhoa.truongKhoa.hoTen$_\exists =_\exists$ ’Mai Anh’

và hai phép chọn biểu thức đường dẫn này có biểu thức đường dẫn con chung lớn nhất là sv.thuocLop.thuocKhoa.

Vậy vấn đề đặt ra ở đây là *khi nào và làm cách nào để cực tiểu hóa chi phí ước lượng các phép chọn biểu thức đường dẫn dạng hội hoặc tuyển có chứa các biểu thức đường dẫn con chung.* Định lý 3 sau đây chỉ ra phương pháp tối ưu đối với các phép chọn biểu thức đường dẫn dạng hội hoặc tuyển có chứa biểu thức đường dẫn con chung.

Định lý 3. *Cho biểu thức* $(P_{1q_1}\theta_{1q_2}C_1)\Delta(P_{2q'_1}\theta_{2q'_2}C_2)$, *trong đó,* $P_1 = o_0.A_1...A_n$ *và* $P_2 = o_0.A'_1...A'_m$ *là các biểu thức đường dẫn,* $\Delta \in \{and, or\}$. *Nếu* P_1 *và* P_2 *tồn tại biểu thức đường dẫn con chung lớn nhất* $P = o_0.A_1...A_j$ *thì tập các đối tượng thỏa mãn biểu thức* $(P_{1q_1}\theta_{1q_2}C_1)\Delta(P_{2q'_1}\theta_{2q'_2}C_2)$ *là tập các đối tượng của biểu thức sau:*

$$\begin{aligned}&\{o_0|o_0.A_1...A_nq_1\theta_1q_2C_1, o_0 \in O_0\} \bigtriangledown \{o_0|o_0.A'_1...A'_nq'_1\theta_2q'_2C_2, o_0 \in O_0\}\\ &\quad = \pi_{(O_0)}(\sigma_{(o_0.A_1...A_nq_1\theta_1q_2C_1)}(O_0 \bowtie ... \bowtie O_{j-1} \bowtie O_j \bowtie O_{j+1} \bowtie ... \bowtie O_n))\bigtriangledown\\ &\qquad\qquad \pi_{(O_0)}(\sigma_{(o_0.A'_1...A'_mq'_1\theta_2q'_2C_2)}(O_0 \bowtie ... \bowtie O_{j-1} \bowtie O_j \bowtie O'_{j+1} \bowtie ... \bowtie O'_m))\end{aligned}$$

Nếu phép chọn biểu thức đường dẫn $(P_{1q_1}\theta_{1q_2}C_1)$ *có kiểu* $\theta_1(\exists)$ *và* $P_{2q'_1}\theta_{2q'_2}C_2$ *có kiểu* $\theta_2(\exists)$ *thì*

$$\begin{aligned}&\{o_0|o_0.A_1...A_nq_1\theta_1q_2C_1, o_0 \in O_0\} \bigtriangledown \{o_0|o_0.A'_1...A'_nq'_1\theta_2q'_2C_2, o_0 \in O_0\}\\ &\quad = \pi_{(O_0)}(\sigma_{(o_0.A_1...A_nq_1\theta_1q_2C_1)}(O_0 \bowtie ... \bowtie O_{j-1} \bowtie O_j \bowtie O_{j+1} \bowtie ... \bowtie O_n))\bigtriangledown\\ &\qquad\qquad \pi_{(O_0)}(\sigma_{(o_0.A'_1...A'_mq'_1\theta_2q'_2C_2)}(O_0 \bowtie ... \bowtie O_{j-1} \bowtie O_j \bowtie O'_{j+1} \bowtie ... \bowtie O'_m))\\ &\quad = \pi_{(O_0)}(O_0 \bowtie ... \bowtie O_{j-1} \bowtie [\pi_{(O_j)}(O_j \bowtie O_{j+1} \bowtie ... \bowtie \sigma_{(o_nq_1\theta_1q_2C_1)}(O_n))\bigtriangledown\\ &\qquad\qquad \pi_{(O_j)}(O_j \bowtie O'_{j+1} \bowtie ... \bowtie \sigma_{(o'_mq'_1\theta_2q'_2C_2)}(O'_m))])\end{aligned}$$

Chứng minh

Đặt $L = \{o_0|o_0.A_1...A_nq_1\theta_1q_2C_1, o_0 \in O_0\} \bigtriangledown \{o_0|o_0.A'_1...A'_nq'_1\theta_2q'_2C_2, o_0 \in O_0\}$

$$\begin{aligned}R = &\pi_{(O_0)}(O_0 \bowtie ... \bowtie O_{j-1} \bowtie [\pi_{(O_j)}(O_j \bowtie O_{j+1} \bowtie ... \bowtie \sigma_{(o_nq_1\theta_1q_2C_1)}(O_n)) \bigtriangledown\\ &\qquad \pi_{(O_j)}(O_j \bowtie O'_{j+1} \bowtie ... \bowtie \sigma_{(o'_mq'_1\theta_2q'_2C_2)}(O'_m))])\end{aligned}$$

Ta cần chứng minh $L = R$.

Giả sử $o_0 \in L$, khi đó $o_0.A_1...A_nq_1\theta_1q_2C_1$ và (hoặc) $o_0.A'_1...A'_mq'_1\theta_2q'_2C_2$ trả về giá trị true. Khi đó, $\exists$ đường dẫn đối tượng $(o_0, o_1, ..., o_n)$ thỏa $o_0.A_1...A_n$ và (hoặc) $(o_0, o'_1, ..., o'_m)$ thỏa $o_0.A'_1...A'_m$. Hay $o_1 = o_0.A_1, ..., o_k \in o_{k-1}.A_k, ..., o_n = o_{n-1}.A_n$ khi A_n là thuộc tính đơn trị hay $o_n \in o_{n-1}.A_n$ khi A_n là thuộc tính đa trị và $o_{n-1}.A_nq_1\theta_1q_2C_1$.

Tương tự, $o'_1 = o_0.A'_1, ..., o'_h \in o'_{h-1}.A'_h, ..., o'_m = o_{m-1}.A'_m$ khi A'_m là thuộc tính đơn trị hay $o'_m \in o_{m-1}.A'_m$ khi A'_m là thuộc tính đa trị và $o'_{m-1}.A'_mq'_1\theta_2q'_2C_2$.

Mặt khác, vì $A_i = A'_i (1 \leq i \leq j)$ nên ta có $o_i = o'_i (1 \leq i \leq j)$. Do đó, $(o_0, ..., o_j, o_{j+1}, ..., o_n)$ là một thể hiện của $O_0 \bowtie ... \bowtie O_j \bowtie O_{j+1} \bowtie ... \bowtie \sigma_{o_nq_1\theta_1q_2C_1}(O_n)$ và $(o_0, ..., o_j, o'_{j+1}, ..., o'_m)$

là một thể hiện của $O_0 \bowtie ... \bowtie O_j \bowtie O'_{j+1} \bowtie ... \bowtie \sigma_{o'_m q'_1 \theta_2 q'_2 C_2}(O'_m)$.

Suy ra, $(o_j, o_{j+1}, ..., o_n)$ là một thể hiện của $O_j \bowtie O_{j+1} \bowtie ... \bowtie \sigma_{o_n q_1 \theta_1 q_2 C_1}(O_n)$ và (hoặc) $(o_j, o'_{j+1}, ..., o'_m)$ là một thể hiện của $O_j \bowtie O'_{j+1} \bowtie ... \bowtie \sigma_{o'_m q'_1 \theta_2 q'_2 C_2}(O'_m)$.

Từ đó ta có: $o_j \in [\pi_{(O_j)}(O_j \bowtie O_{j+1} \bowtie ... \bowtie \sigma_{(o_n q_1 \theta_1 q_2 C_1)}(O_n)) \bigtriangledown \pi_{(O_j)}(O_j \bowtie O'_{j+1} \bowtie ... \bowtie \sigma_{(o'_m q'_1 \theta_2 q'_2 C_2)}(O'_m))]$. Nên $(o_0, ..., o_j)$ là một thể hiện của $O_0 \bowtie ... \bowtie O_{j-1} \bowtie [\pi_{(O_j)}(O_j \bowtie O_{j+1} \bowtie ... \bowtie \sigma_{(o_n q_1 \theta_1 q_2 C_1)}(O_n)) \bigtriangledown \pi_{(O_j)}(O_j \bowtie O'_{j+1} \bowtie ... \bowtie \sigma_{(o'_m q'_1 \theta_2 q'_2 C_2)}(O'_m))]$ và $o_0 \in R$.

Giả sử $o_0 \in R$, khi đó $\exists$ đường dẫn đối tượng $(o_0, o_1, ..., o_j)$ là một thể hiện của $O_0 \bowtie ... \bowtie O_{j-1} \bowtie [\pi_{(O_j)}(O_j \bowtie O_{j+1} \bowtie ... \bowtie \sigma_{(o_n q_1 \theta_1 q_2 C_1)}(O_n)) \bigtriangledown \pi_{(O_j)}(O_j \bowtie O'_{j+1} \bowtie ... \bowtie \sigma_{(o'_m q'_1 \theta_2 q'_2 C_2)}(O'_m))]$, và vì $o_j \in \pi_{(O_j)}(O_j \bowtie O_{j+1} \bowtie ... \bowtie \sigma_{(o_n q_1 \theta_1 q_2 C_1)}(O_n))$ và (hoặc) $o_j \in \pi_{(O_j)}(O_j \bowtie O'_{j+1} \bowtie ... \bowtie \sigma_{(o'_m q'_1 \theta_2 q'_2 C_2)}(O'_m))$ nên tồn tại đường dẫn đối tượng $(o_0, ..., o_j, o_{j+1}, ..., o_n)$ thỏa $O_0 \bowtie ... \bowtie O_j \bowtie O_{j+1} \bowtie ... \bowtie \sigma_{o_n q_1 \theta_1 q_2 C_1}(O_n)$ và (hoặc) $(o_0, ..., o_j, o'_{j+1}, ..., o'_m)$ thỏa $O_0 \bowtie ... \bowtie O_j \bowtie O'_{j+1} \bowtie ... \bowtie \sigma_{o'_m q'_1 \theta_2 q'_2 C_2}(O'_m)$. Do đó, $(o_0, ..., o_j, o_{j+1}, ..., o_n)$ thỏa $o_0.A_1...A_n$ và $o_0.A_1...A_n q_1 \theta_1 q_2 C_1$ là true và (hoặc) $(o_0, ..., o_j, o'_{j+1}, ..., o'_m)$ thỏa $o_0.A'_1...A'_m$ và $o_0.A'_1...A'_m q'_1 \theta_2 q'_2 C_2$ là true. Từ đó ta có $o_0 \in L$.

Vậy $L = R$. ■

Ví dụ 5. Cho phép chọn biểu thức đường dẫn dạng tuyển như sau

sv.thuocLop.thuocKhoa.tenKhoa$_\exists$ $=_\exists$ 'CNTT' **or**

sv.thuocLop.thuocKhoa.truongKhoa.hoTen$_\exists$ $=_\exists$ 'Mai Anh'

Vì cả hai phép chọn biểu thức đường dẫn đều có kiểu $= (\exists)$ và có biểu thức đường dẫn con chung lớn nhất là sv.thuocLop.thuocKhoa, nên biểu thức ước lượng được viết theo Định lý 3 như sau:

$$\pi_{(SINHVIEN)}(SINHVIEN \bowtie LOP \bowtie (\pi_{(KHOA)}(\sigma_{(k.tenKhoa='CNTT')}(KHOA)) \cup \pi_{(KHOA)}(KHOA \bowtie \sigma_{(ns.hoTen='MaiAnh')}(NHANSU)))$$

Trong biểu thức trên, ta thấy biểu thức đường dẫn con chung sv.thuocLop.thuocKhoa chỉ được ước lượng duy nhất một lần trong quá trình thực thi.

4. ĐỒ THỊ TRUY VẤN VÀ LƯỢC ĐỒ THỰC THI

Bằng cách xác định các kiểu khác nhau của các điều kiện ràng buộc trên các phép chọn biểu thức đường dẫn, sau đó áp dụng các chiến lược tối ưu hóa thích hợp như đã nêu trên, như đẩy các ràng buộc vào trong các điều hướng, để có được biểu thức ước lượng với chi phí thấp hơn. Do đó, các ràng buộc có chọn lọc có thể được ước lượng trước để loại bỏ các đối tượng không liên quan trước khi các thao tác điều hướng tốn kém khác được thực hiện.

Có một số phương pháp heuristic được đề xuất để tối ưu hóa các phép chọn, phép kết nối và các phép chọn dạng hội hoặc tuyển trên các biểu thức đường dẫn, dựa trên các Định lý 1, 2 và 3.

Heuristic 1. Nếu phép chọn biểu thức đường dẫn có kiểu $\theta(\exists)$ thì vị từ chọn có thể được đẩy vào lớp chứa thuộc tính cuối cùng trong biểu thức đường dẫn. Phép chọn có thể được thực hiện trước khi thực hiện các kết nối ẩn giữa các lớp kết hợp với các thuộc tính trong biểu thức đường dẫn và nên được thực hiện sớm nhất có thể.

Heuristic 2. Trong phép kết nối biểu thức đường dẫn. Nếu điều kiện kết nối có dạng $\theta(\exists, \exists)$ thì phép kết nối hiển có thể được ước lượng trước khi thực hiện các kết nối ẩn giữa các lớp kết hợp với các thuộc tính trong biểu thức đường dẫn. Nếu điều kiện kết nối có dạng $\theta(\exists, \forall)$ thì phép kết nối hiển chỉ có thể được ước lượng sau khi thực hiện các kết nối ẩn giữa các lớp kết hợp với các thuộc tính trong biểu thức đường dẫn thứ hai. Nếu điều kiện kết nối có dạng $\theta(\forall, \exists)$ thì phép kết nối hiển chỉ có thể được ước lượng sau khi thực hiện các kết nối ẩn giữa các lớp kết hợp với các thuộc tính trong biểu thức đường dẫn thứ nhất.

Heuristic 3. Nếu có nhiều hơn một phép chọn biểu thức đường dẫn dạng hội hoặc tuyển và nếu hai biểu thức đường dẫn này có một biểu thức đường dẫn con chung lớn nhất và cả hai phép chọn đều có kiểu $\theta(\exists)$ thì mỗi phép chọn được viết lại dựa vào Heuristic 1 và các phép kết nối ẩn trong biểu thức đường dẫn con chung lớn nhất được thực hiện duy nhất một lần.

Để xây dựng các lược đồ thực thi truy vấn, [9] đã đưa ra khái niệm đồ thị truy vấn và từ đồ thị truy vấn này tiến hành xây dựng các lược đồ thực thi một cách tổng quát.

4.1. Đồ thị truy vấn

Đồ thị truy vấn được trình bày sau đây, có thể biểu diễn được các phép chọn dạng đơn trị, đa trị, các kết nối đơn trị, đa trị trên các biểu thức đường dẫn. Mặt khác, nó cho phép nhận biết các biểu thức đường dẫn con chung lớn nhất tồn tại giữa các biểu thức đường dẫn một cách trực quan.

Định nghĩa 8. ([9]) Đồ thị truy vấn là một siêu đồ thị $H = (V, E)$, trong đó, V là tập các nút được xây dựng như sau:

1. Các nút tương ứng với tất cả các biến đối tượng xuất hiện trong truy vấn. Các biến đối tượng khác nhau tương ứng với các nút khác nhau, mặc dù chúng thuộc về cùng một lớp.

2. Nếu $t.A_1...A_n$ là một biểu thức đường dẫn trong truy vấn thì thêm các nút tương ứng với dãy các lớp của $A_1, A_2, ..., A_{n-1}$ vào V.

E là tập các cạnh và siêu cạnh được xây dựng như sau:

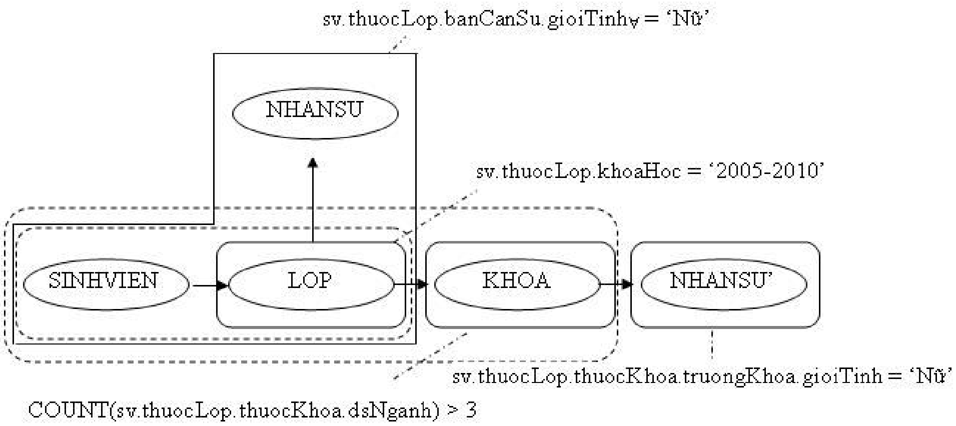
1. Nếu $t.A_1...A_n$ là một biểu thức đường dẫn xuất hiện trong truy vấn thì thêm một cạnh có hướng kết nối nút tương ứng với biến đối tượng t và nút tương ứng của lớp chứa A_1 và các cạnh kết nối hai nút biểu diễn các lớp của A_{i-1} và $A_i (2 \leq i \leq n-1)$.

2. Nếu phép chọn biểu thức đường dẫn xuất hiện trong truy vấn có kiểu $\theta(\exists)$ thì thêm một siêu cạnh chứa nút tương ứng với lớp của thuộc tính cuối cùng trong biểu thức đường dẫn. Cạnh biểu diễn vị từ trên lớp chứa thuộc tính cuối cùng của biểu thức đường dẫn. Nếu phép chọn có kiểu $\theta(\forall)$ thì thêm một siêu cạnh kết nối chứa các nút tương ứng với các lớp kết hợp với biểu thức đường dẫn.

3. Nếu phép kết nối biểu thức đường dẫn xuất hiện trong truy vấn có kiểu $\theta(\exists, \exists)$ thì thêm một siêu cạnh kết nối hai nút tương ứng các lớp chứa các thuộc tính cuối cùng

trong các biểu thức đường dẫn. Cạnh biểu diễn kết nối hiển. Nếu phép kết nối biểu thức đường dẫn có kiểu $\theta(\exists, \forall)$ thì thêm một siêu cạnh kết nối chứa các nút tương ứng lớp của thuộc tính cuối đường dẫn của biểu thức đường dẫn bên trái và các lớp kết hợp với biểu thức đường dẫn bên phải. Nếu phép so sánh biểu thức đường dẫn có kiểu $\theta(\forall, \exists)$ hoặc $\theta(\forall, \forall)$ thì thêm một siêu cạnh kết nối các nút tương ứng các lớp kết hợp với hai biểu thức đường dẫn.

Ví dụ sau minh họa cách một đồ thị truy vấn được xây dựng và từ đó xây dựng các lược đồ thực thi truy vấn.



Hình 1. Đồ thị truy vấn trong Ví dụ 6

Ví dụ 6. Xét cơ sở dữ liệu hướng đối tượng được cho ở Ví dụ 1. Hãy liệt kê danh sách sinh viên của các lớp, học khóa 2005-2010, có ban cán sự lớp là nữ, thuộc khoa có trưởng khoa là nữ và có tổng số ngành lớn hơn 3.

```
SELECT   sv.hoten
FROM     sv SINHVIEN
WHERE    sv.thuocLop.khoaHoc = '2005-2010'
         and sv.thuocLop.banCanSu.gioiTinh∀ = 'Nữ'
         and sv.thuocLop.thuocKhoa.truongKhoa.gioiTinh = 'Nữ'
         and COUNT(sv.thuocLop.thuocKhoa.dsNganh) ¿ 3
```

Đồ thị của truy vấn trên được mô tả trong Hình 1. Mỗi nút trong dồ thị tương ứng với một biến đối tượng hay lớp được biểu diễn bằng hình Ellipse, một cạnh có hướng biểu diễn một kết nối ẩn giữa hai lớp, siêu cạnh được biểu diễn bằng đường khép kính chứa tất cả các nút của siêu cạnh. Ví dụ, siêu cạnh {LOP} biểu diễn phép chọn đơn trị

$$\text{lop.khoaHoc} = \text{'2005-2010'}$$

là kết quả dịch chuyển phép chọn đơn trị

$$sv.thuocLop.khoaHoc_{\exists} = \text{'2005-2010'}$$

{SINHVIEN, LOP, NHANSU} là một siêu cạnh biểu diễn phép chọn đa trị

$$sv.thuocLop.banCanSu.gioiTinh_{\forall} = \text{'Nữ'}$$

Cả bốn đường dẫn (SINHVIEN, LOP), (SINHVIEN, LOP, NHANSU), (SINHVIEN, LOP, KHOA) và (SINHVIEN, LOP, KHOA, NHANSU') có biểu thức đường dẫn con chung lớn nhất là (SINHVIEN, LOP); còn (SINHVIEN, LOP, KHOA) và (SINHVIEN, LOP, KHOA, NHANSU') có đường dẫn con chung lớn nhất là (SINHVIEN, LOP, KHOA). Các biểu thức đường dẫn con chung được bao bởi đường khép kín đứt nét.

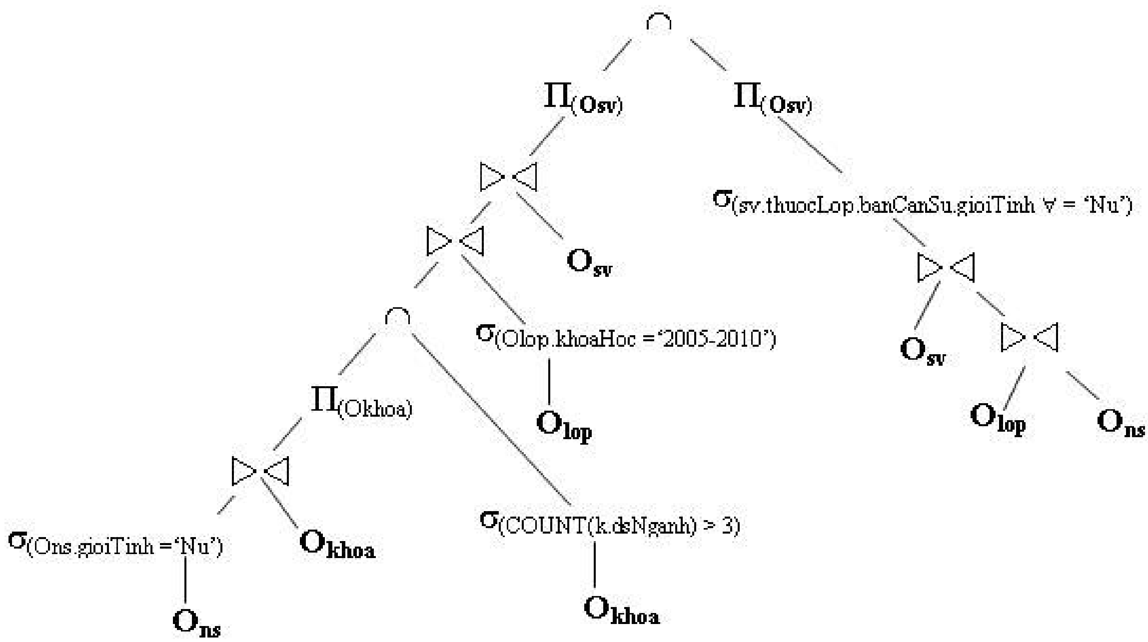
4.2. Lược đồ thực thi truy vấn

Dựa vào đồ thị truy vấn, [9] tiến hành xây dựng các lược đồ thực thi theo các chiến lược đã được đề xuất bởi các heuristic nêu trên, các lược đồ thực thi này được biểu diễn qua các ký hiệu và trải qua nhiều bước rút gon phức tạp, không trực quan. Trong phần này chúng tôi đề xuất phương pháp xây dựng lược đồ thực thi truy vấn dưới dạng cây nhị phân dựa vào một trong hai cách:

- Cách thứ nhất: xây dựng lược đồ thực thi từ đồ thị truy vấn và các heuristic nêu trên.

- Cách thứ hai: xây dựng lược đồ thực thi từ các biểu thức viết lại nhờ áp dụng các Định lý 1, 2 và 3 nêu trên.

Tương tự như cây xử lý tối ưu đại số trong lý thuyết tối ưu hóa truy vấn quan hệ [6], trong đó, các nút ngoài (nút lá) của cây là các nút ứng với các biến đối tượng hay lớp xuất hiện trong đồ thị hay biểu thức sau khi viết lại, các nút trong tương ứng với các phép toán, như: giao $\cap$, hợp $\cup$, chiếu π, chọn σ và kết nối $\bowtie$. Quá trình thực thi truy vấn được thực hiện từ dưới lên và theo thứ tự duyệt trái - gốc - phải hay phải - gốc - trái.



Hình 2. Lược đồ thực thi truy vấn trong Ví dụ 6

Ví dụ 7. Xét đồ thị truy vấn ở Hình 1. Giả sử, kí hiệu O_{sv} là tập các đối tượng trong lớp SINHVIEN, O_{lop} là tập các đối tượng trong lớp LOP, O_{khoa} là tập các đối tượng trong lớp KHOA và O_{ns} là tập các đối tượng trong lớp NHANSU. Ta tiến hành xây dựng lược đồ thực thi theo các bước sau:

- Thực hiện phép chọn đơn trị

(i) $\{sv | sv.thuocLop.thuocKhoa.truongKhoa.gioiTinh =' Nu', sv \in O_{sv}\}$

$= \pi_{(O_{sv})}(\sigma_{(sv.thuocLop.thuocKhoa.truongKhoa.gioiTinh=' Nu')}(O_{sv} \bowtie O_{lop} \bowtie O_{khoa} \bowtie O_{ns}))$

$= \pi_{(O_{sv})}(O_{sv} \bowtie O_{lop} \bowtie O_{khoa} \bowtie \sigma_{(o_{ns}.gioiTinh=' Nu')}(O_{ns})))$

(ii) $\{sv | sv.thuocLop.khoaHoc =' 2005 - 2010', sv \in O_{sv}\}$

$= \pi_{(O_{sv})}(\sigma_{(sv.thuocLop.khoaHoc=' 2005-2010')}(O_{sv} \bowtie O_{lop}))$

$= \pi_{(O_{sv})}(O_{sv} \bowtie \sigma_{(o_{lop}.khoaHoc=' 2005-2010')}(O_{lop}))$ và

(iii) $\{sv | COUNT(sv.thuocLop.thuocKhoa.dsNganh) > 3, sv \in O_{sv}\}$

$= \pi_{(O_{sv})}(\sigma_{(COUNT(sv.thuocLop.thuocKhoa.dsNganh)>3)}(O_{sv} \bowtie O_{lop} \bowtie O_{khoa}))$

$= \pi_{(O_{sv})}(O_{sv} \bowtie O_{lop} \bowtie \sigma_{(COUNT(o_{khoa}.dsNganh)>3)}(O_{khoa}))$

- Thực hiện các kết nối đơn trị nếu có, sau đó áp dụng các Heuristic 1 và Heuristic 3. Từ (i), (ii) và (iii) ta được

(iv) $\pi_{(O_{sv})}(O_{sv} \bowtie \sigma_{(o_{lop}.khoaHoc=' 2005-2010')}(O_{lop}) \bowtie$

$(\pi_{(O_{khoa})}(O_{khoa} \bowtie \sigma_{(o_{ns}.gioiTinh=' Nu')}(O_{ns})) \cap \sigma_{(COUNT(o_{khoa}.dsNganh)>3)}(O_{khoa})))$

- Thực hiện phép chọn đa trị

(v) $\{sv | sv.thuocLop.banCanSu.gioiTinh_{\forall} =' Nu', sv \in O_{sv}\} =$

$\pi_{(O_{sv})}(\sigma_{sv.thuocLop.banCanSu.gioiTinh_{\forall}=' Nu'}(O_{sv} \bowtie O_{lop} \bowtie O_{ns}))$.

Cuối cùng, ta xây dựng được lược đồ thực thi truy vấn từ các biểu thức (iv) và (v) như ở Hình 2.

5. KẾT LUẬN

Trong bài báo này, chúng tôi tập trung nghiên cứu và mở rộng phương pháp tối ưu truy vấn đối tượng chứa các biểu thức đường dẫn, với các phép chọn và kết nối phức tạp. Phương pháp tối ưu phép chọn và kết nối biểu thức đường dẫn dựa vào các quy tắc viết lại, nhằm mục đích đẩy phép chọn và phép kết nối vào trước các phép kết nối ẩn giữa các lớp chứa các thuộc tính trong các biểu thức đường dẫn tương ứng, việc đẩy phép chọn và phép kết nối phụ thuộc vào kiểu của phép chọn và phép kết nối. Một mở rộng phương pháp tối ưu đối với các phép chọn biểu thức đường dẫn dạng hội hoặc tuyển, bằng cách tìm biểu thức đường dẫn con chung lớn nhất giữa các biểu thức đường dẫn và nếu tồn tại thì biểu thức đường dẫn con chung lớn nhất này chỉ được ước lượng duy nhất một lần trong quá trình thực thi. Ưu điểm của phương pháp mở rộng này là cho phép xử lý các phép chọn và kết nối biểu thức đường dẫn bất kỳ dưới dạng hội hoặc tuyển. Mặt khác, bài báo cũng đề xuất phương pháp xây dựng lược đồ thực thi truy vấn từ đồ thị truy vấn hay biểu thức sau khi viết lại dưới dạng cây nhị phân gồm các phép toán đại số một cách tổng quát.

TÀI LIỆU THAM KHẢO

[1] Ahmad Nouri, David J.Wisneski, *Object oriented query path expression to relational outer join translator method*, system, article of manufacture, and computer program product, USA, 2008.

[2] Cetin Ozkan, Asuman Dogac, Cem Evrendilek, *A Heuristic Approach for Optimization of Path Expressions*, Ankara Turkiye, 1999.

[3] Cetin Ozkan, Asuman Dogac, Mehmet Altinel, A cost model for path expressions in object-oriented queries, *ACM SIGMOD*, 1996.

[4] Georges GardarinaYb, Jean-Robert Gruserb, Zhao-Hui Tanga, Cost-based Selection of Path Expression Processing - Algorithms in Object-Oriented Databases, *Proceedings of the 22nd VLDB Conference*, Mumbai(Bombay), India, 1996.

[5] Guoren Wang, Ge Yu, Kunihiko Kaneko, Akifumi Makinouchi, Comparison of parallel algorithms for path expression query in object database systems, *The 7th International Conference on Database Systems for Advanced Applications (DASFAA 2001)*, Hong Kong, China, April 18-21, 2001.

[6] R. Elmasri, S. R. Navathe, *Fundamentals of Database Systems*, Pearson- Addison Wesley, 4th Edition, 2003.

[7] Weimin Chen, Karl Aberer, Erich J. Neuhold, *Efficient Algorithm for Determining the Optimal Execution Strategy for Path Queries in OODBS*, GMD-IPSI, Dolivostr. 15, 64293 Darmstadt, Gemany, 2000.

[8] Wang-Sup Cho, Kyu-Young Whang, Seung-Sun Lee, Yong-Ik Yoon, Query Optimization Techniques Utilizing Path Indexes in Object-Oriented Database Systems, *Proceeding of the Fifth International Conference on Dadatabase Systems for Advanced Applications*, Melbourne, Australia, 1997.

[9] Zhaohui Xie, Optimization of Object Queries Containing Encapsulated Methods, *ACM*, USA, 1993.

Nhận bài ngày 22 - 10 - 2009