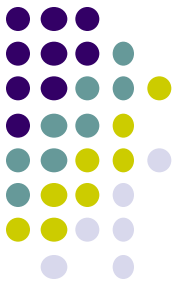




Chương 7

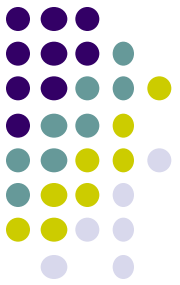
Tối ưu hoá câu hỏi truy vấn

Tối ưu hoá câu hỏi



- 7.1. Các nguyên tắc tổng quát để tối ưu hóa câu hỏi
- 7.2. Một số thuật toán tối ưu



Các nguyên tắc tổng quan



- **Xét một ví dụ đơn giản sau đây :**
- Cho hai quan hệ $R(A,B)$ với n bản ghi và $S(C,D)$ với m bản ghi.
Tích Đề-các của R và S là một quan hệ $Q(A,B,C,D)$ có $n*m$ bản ghi.
- Câu hỏi "*Lấy giá trị của thuộc tính A sao cho $B=C$ và $D=50$* ".
- $((R(A,B) \times S(C,D)) : (B=C \wedge D=50)) [A]$
- Nếu đưa phép chọn $D=50$ vào bên trong phép tích Đề-các sẽ được:
- $((R(A,B) \times (S(C,D) : (D=50))) : (B=C)) [A]$
- và sau đó chuyển phép chọn $B=C$ của tích Đề-các thành phép "kết nối bằng" chúng ta thu được:
- $(R(A,B) \bowtie_{B=C} S(C,D) : (D=50)) [A]$

Các nguyên tắc tổng quan



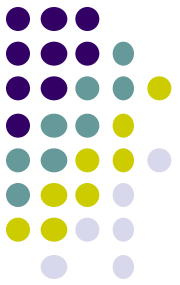
- Rõ ràng, phép tính cuối cùng sẽ đỡ tốn kém thời gian hơn rất nhiều.
-  Việc biến đổi câu hỏi thành câu hỏi tương đương để giảm bớt thời gian trả lời câu hỏi dựa trên nguyên tắc thực hiện phép chọn càng sớm càng tốt.
-  Trình tự thực hiện các phép tính sẽ đóng một vai trò quan trọng quá trình tổ chức câu hỏi.

Các nguyên tắc tổng quan...



- **Sáu chiến lược tổng quan của J. D. Ullman [4] :**
- **1. Thực hiện phép chọn càng sớm càng tốt.**
- Biến đổi câu hỏi để đưa phép chọn vào thực hiện trước nhằm làm giảm bớt kích cỡ của kết quả trung gian và do vậy chi phí phải trả cho việc truy nhập bộ nhớ thứ cấp cũng như lưu trữ của bộ nhớ chính sẽ nhỏ đi.
- **2. Tổ hợp những phép chọn xác định với phép tích Đề-các thành phép kết nối.**
- Phép kết nối, đặc biệt là phép kết nối bằng (*Equi Join*) có thể được thực hiện ít tốn kém hơn nhiều so với phép tích Đề-các trên cùng các quan hệ.

Các nguyên tắc tổng quan...



- **3. Tổ hợp dãy các phép toán quan hệ một ngôi như các phép chọn và phép chiếu.**
- Dãy các phép một ngôi như phép chọn, phép chiếu mà kết quả của chúng phụ thuộc vào các bộ của một quan hệ độc lập thì có thể nhóm các phép đó lại.
- **4. Tìm các biểu thức con chung trong một biểu thức.**
- Nếu kết quả của một biểu thức con chung là một quan hệ không lớn và nó có thể được đọc từ bộ nhớ thứ cấp với ít thời gian thì nên tính toán trước biểu thức đó chỉ một lần. Nếu biểu thức con chung có liên quan tới một phép kết nối thì trong trường hợp tổng quát không thể thay đổi được nó bằng cách "đẩy" phép chọn vào trong.



Các nguyên tắc tổng quan...

- **5. Tiền xử lý các quan hệ / bảng (*Table Preprocessing*).**
- Có hai vấn đề quan trọng cần xử lý trước cho các quan hệ là sắp xếp trước các bộ giá trị theo thứ tự vật lý và sắp xếp logic - tức là thiết lập các bảng chỉ mục (*Index*) cho các bản ghi.
- **6. Đánh giá trước khi thực hiện tính toán.**
- Mỗi khi thực hiện phép toán, thì cần tính toán xem chi phí (*Cost*) các phép tính đó (thường tính theo số phép toán, thời gian, hoặc/và dung lượng bộ nhớ cần thiết so với kích thước của các quan hệ, từ đó xác định được chi phí tổng thể phải trả cho các cách khác nhau khi thực hiện các câu hỏi).

Biểu thức tương đương

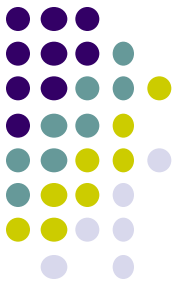


- Biểu thức trong ngôn ngữ ĐSQH có các hạng thức là *biến quan hệ* R_1, \dots, R_n ; các *quan hệ hằng*, được xác định như là một ánh xạ từ các k -bộ của các quan hệ (r_1, \dots, r_k) trong đó r_i là quan hệ trên lược đồ r_i và thay thế r_i vào R_i khi đánh giá biểu thức.
- Hai biểu thức E_1 và E_2 được gọi là **tương đương (*Equivalent*)**, viết tắt là $E_1 \equiv E_2$, nếu chúng biểu diễn cùng một ánh xạ, nghĩa là, nếu chúng ta thay thế cùng các quan hệ cho tên các lược đồ tương ứng ở hai biểu thức E_1 và E_2 , thì chúng sẽ cho ra cùng một kết quả.

Các quy tắc



- **1. Quy tắc giao hoán của phép kết nối và tích Đề-các**
- Nếu E1 và E2 là hai biểu thức quan hệ và F là điều kiện trên các thuộc tính của E1 và E2 thì:
 - $E1 \bowtie E2 \equiv E2 \bowtie E1$ // Tính giao hoán của kết nối
 - $E1 * E2 \equiv E2 * E1$ // Tính giao hoán của kết bằng
 - $E1 \times E2 \equiv E2 \times E1$ // Tính giao hoán của tích Đề-các.
- Chú ý: Nếu quan niệm quan hệ là tập các bộ (có thứ tự thuộc tính cố định) thì phép q -kết, kết tự nhiên và tích Đề-các không thể giao hoán được vì thứ tự các thuộc tính trong quan hệ kết quả bị thay đổi.
- **2. Quy tắc kết hợp của phép kết nối và tích Đề-các.**
- Nếu E1, E2 và E3 là các biểu thức quan hệ: F1, F2 là điều kiện thì:
 - $(E1 \bowtie E2) \bowtie E3 \equiv E1 \bowtie (E2 \bowtie E3)$
 - $(E1 * E2) * E3 \equiv E1 * (E2 * E3)$
 - $(E1 \times E2) \times E3 \equiv E1 \times (E2 \times E3)$



Các quy tắc...

- Dãy các phép chiếu có thể tổ hợp lại thành một phép chiếu:
- **3. Dãy các phép chiếu**
- $(E [B1 \dots Bm]) [A1 \dots An] \equiv E [A1 \dots An]$
- Ở đây, các thuộc tính $A1, \dots, An$ phải nằm trong tập các thuộc tính $B1, \dots, Bm$.
- Dãy các phép chọn có thể tổ hợp thành một phép chọn để kiểm tra tất cả các điều kiện cùng một lúc và được biểu diễn như sau:
- **4. Dãy các phép chọn:**
- $(((E : (f1)) : f2) : \dots) : fn \equiv E : (f1 \wedge f2 \dots \wedge fn)$
- **5. Giao hoán phép chọn và phép chiếu:**
- $(E [A1 \dots An] : (f)) \equiv (E : (f)) [A1 \dots An]$
- Nếu điều kiện chọn f liên quan tới các thuộc tính $B1, \dots, Bm$ mà không nằm trong tập thuộc tính $A1, \dots, An$ thì:
- $(E [A1 \dots An]) : (f) \equiv ((E [A1 \dots An B1 \dots Bm]) : (f)) [A1 \dots An]$

Các quy tắc...



- **6. Giao hoán phép chọn và tích Đề-các:**
- Nếu tất cả các thuộc tính của F là thuộc tính của $E1$ thì:
$$(E1 \times E2) : (f) \equiv (E1 : (f)) \times E2$$
- Nếu F có dạng $f = f1 \wedge f2$ trong đó $f1$ chỉ liên quan tới các thuộc tính của $E1$; $f2$ chỉ liên quan tới các thuộc tính của $E2$, thì có thể sử dụng các luật 1,4, 6 để có: $(E1 \times E2) : (f) \equiv (E1 : (f1)) \times (E2 : (f2))$
- Nếu $f1$ chỉ liên quan tới các thuộc tính của $E1$, nhưng $f2$ liên quan tới các thuộc tính của cả $E1$ và $E2$ thì: $(E1 \times E2) : (f) \equiv ((E1 : (f1)) \times E2) : (f2)$

Các quy tắc...



- **7. Giao hoán phép chọn và một phép hợp:**
- Nếu $E = E1 \cup E2$; Khi đó: $(E1 \cup E2) : (f) \equiv (E1 : (f)) \cup (E2 : (f))$
- **8. Giao hoán phép chọn và một phép hiệu tập hợp**
- $(E1 - E2) : (f) \equiv (E1 : (f)) - (E2 : (f))$
- Như đã nêu trong luật 7, nếu tên các thuộc tính của $E1$ và $E2$ là khác nhau thì cần thay thế các thuộc tính trong f ở vế phải biểu thức tương đương tương ứng với $E1$.
- ➡ Quy tắc đẩy phép chọn xuống trước phép kết nối suy ra từ quy tắc 4, 5, 6.
- ➡ Quy tắc đẩy phép chiếu xuống trước phép tích Đề-các hoặc phép hợp cũng tương tự như quy tắc 6, 7.

Các quy tắc...



- **9. Giao hoán một phép chiếu với một phép tích Đề-các:**
-
- Gọi E_1, E_2 là hai biểu thức quan hệ, $A_1 \dots A_n$ là tập các thuộc tính trong đó $B_1, \dots B_m$ là các thuộc tính của E_1 , các thuộc tính còn lại C_1, \dots, C_k thuộc E_2 . Khi đó: $(E_1 \times E_2) [A_1 \dots A_n] \equiv E_1 [B_1 \dots B_m] \times E_2 [C_1 \dots C_k]$
- **10. Giao hoán một phép chiếu với một phép hợp:**
- $(E_1 \cup E_2) [A_1 \dots A_n] \equiv E_1[A_1 \dots A_n] \cup E_2[A_1 \dots A_n]$

Ví dụ về thuật toán tối ưu hoá biểu thức quan hệ



- Áp dụng các quy tắc nêu trên để tối ưu hóa.
- Luôn **đẩy phép chọn và phép chiếu xuống mức càng sâu càng tốt**, để từ đó có thể tổ chức thành **một phép chọn theo sau một phép chiếu**. **Nhóm các phép chọn và phép chiếu** lại trong một nhóm để thực hiện trước như phép hợp, tích Đề-các, hiệu tập hợp v.v...
- Trường hợp đặc biệt khi một phép tính hai ngôi có các hạng thức chứa phép chọn/chiếu được áp dụng đối với lá của cây. Cần xem xét cẩn thận tác động của phép tính hai ngôi vì một số trường hợp phải liên kết phép chọn/chiếu với phép hai ngôi đó.

Ví dụ về thuật toán tối ưu hoá biểu thức quan hệ...



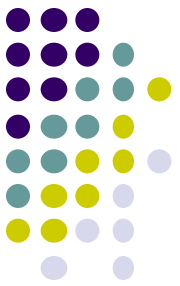
- **Kết quả đầu ra (*Output*) của thuật toán** là một chương trình bao gồm các bước như sau:
 - a. Áp dụng của một phép chọn hoặc một phép chiếu đơn giản.
 - b. Áp dụng của một phép chọn và một phép chiếu hoặc
 - c. Áp dụng của một tích Đề-các, phép hợp hoặc phép hiệu tập hợp cho hai hạng thức mà trước đó các phép chọn hoặc các phép chiếu đã được áp dụng cho một hoặc cả hai hạng thức.

Ví dụ về thuật toán tối ưu hoá biểu thức quan hệ...

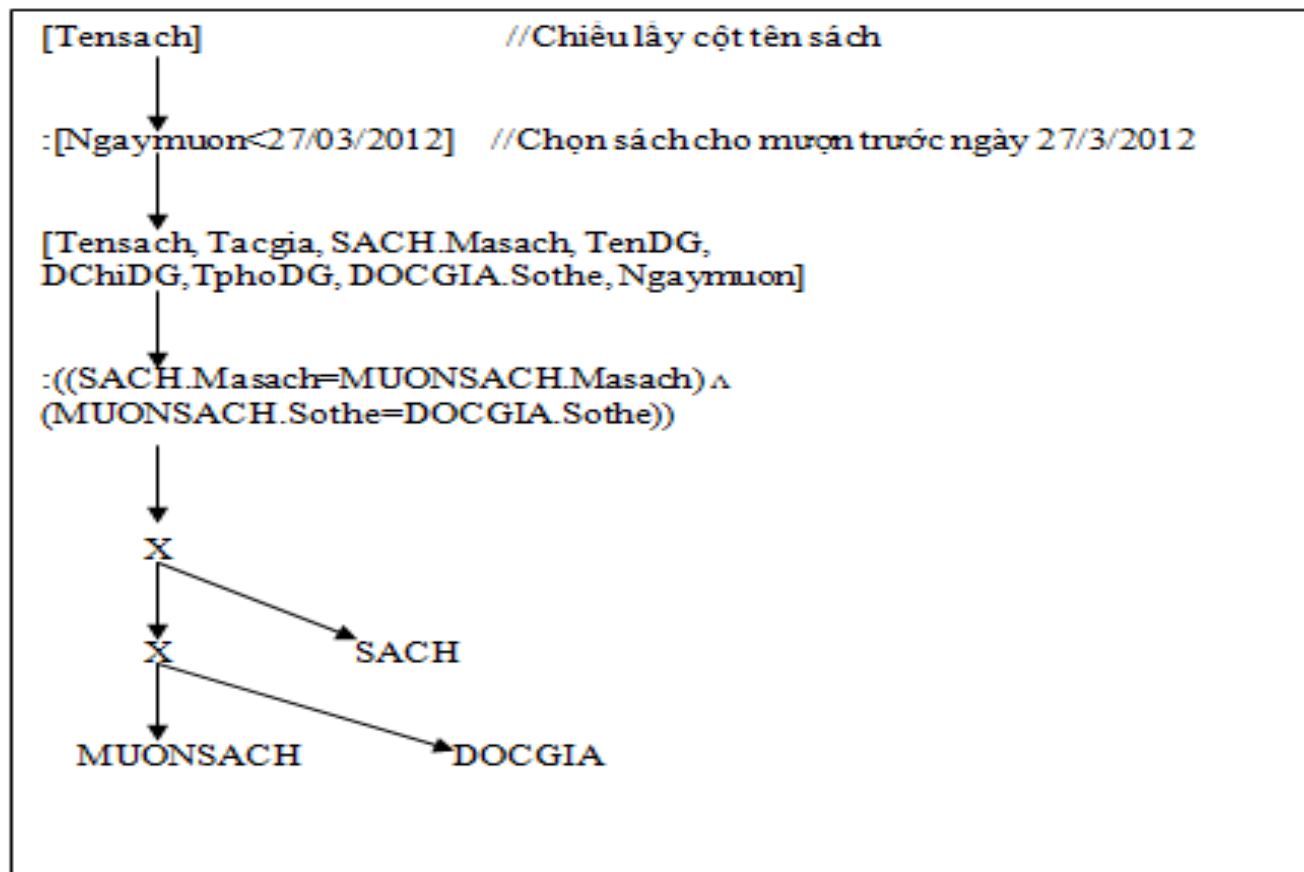


- Xét CSDL quản lý thư viện gồm các quan hệ sau:
- 1. **SACH** (Tensach, Tacgia, NhaXB, Masach)
- 2. **NHAXUATBAN** (NhaXB, Diachi, Thanhpho).
- 3. **DOCGIA** (TenDG, DchiDG, TphoDG, Sothe).
- 4. **MUONSACH** (Sothe, Masach, Ngaymuon).
- Giả thiết có một khung nhìn (VIEW) theo dõi các sách được mượn, TDMUON, bao gồm một số thông tin bổ sung về sách được mượn, là kết quả của kết nối tự nhiên của quan hệ SACH, DOCGIA, MUONSACH, chẳng hạn được xác định qua biểu thức quan hệ:
- $((SACH \times DOCGIA \times MUONSACH) : (f))[\{S\}]$
- Ở đây: $f = (DOCGIA.Sothe = MUONSACH.Sothe) \wedge (SACH.Masach = MUONSACH.Masach)$.
- và S là tập các thuộc tính: $S = \{Tensach, Tacgia, NhaXB, SACH.Masach, Tên_ĐG, DchiDG, TphoDG, DOCGIA.Sothe, Ngaymuon\}$

Ví dụ về thuật toán tối ưu hoá biểu thức quan hệ...



- Câu hỏi: Cho danh sách những cuốn sách đã cho mượn trước 27/03/2012. Biểu thức quan hệ được viết như sau: (TDMUON : (Ngaymuon < 27/03/2012))[Tensach]



Hình 6.1. Biểu diễn cây của biểu thức hỏi

Ví dụ về thuật toán tối ưu hoá biểu thức quan hệ...



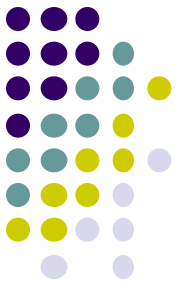
- *Bước thứ nhất*: tách phép chọn f thành hai phép chọn với điều kiện:
- $SACH.Masach = MUONSACH.Masach$ và
- $MUONSACH.Sothe = DOCGIA. Sothe$
- Có 3 phép chọn cần "đẩy" xuống mức thấp hơn.
- Phép chọn với ĐK **Ngaymuon** < 27/03/2012 được đẩy xuống dưới phép chiếu và hai phép chọn kia bằng cách áp dụng các quy tắc 4,5. Phép chọn đầu được áp dụng cho tích Đề-các $((MUONSACH \times DOCGIA) \times SACH)$. Vì thuộc tính **Ngaymuon** trong phép chọn chỉ có ở quan hệ MUONSACH nên có thể thay thế:
- $((MUONSACH \times DOCGIA) \times SACH) : (Ngaymuon < 27/03/2012)$
- bằng biểu thức:
- $((MUONSACH \times DOCGIA) : (Ngaymuon < 27/03/2012)) \times SACH)$

Ví dụ về thuật toán tối ưu hoá biểu thức quan hệ...



- và tiếp tục đẩy xuống ta được biểu thức:
- $((MUONSACH : (\text{Ngaymuon} < 27/03/2012)) \times DOCGIA) \times SACH)$
- Như vậy đã đẩy được phép chọn theo ngày mượn xuống sâu có thể.
- Tiếp tục đẩy phép chọn với điều kiện **SACH.Masach** = **MUONSACH.Masach** xuống mức thấp nhất có thể. Không thể đẩy phép chọn này xuống dưới tích Đề-các vì nó liên quan tới một thuộc tính của quan hệ SÁCH và một thuộc tính thuộc quan hệ MUONSACH.

Ví dụ về thuật toán tối ưu hoá biểu thức quan hệ...



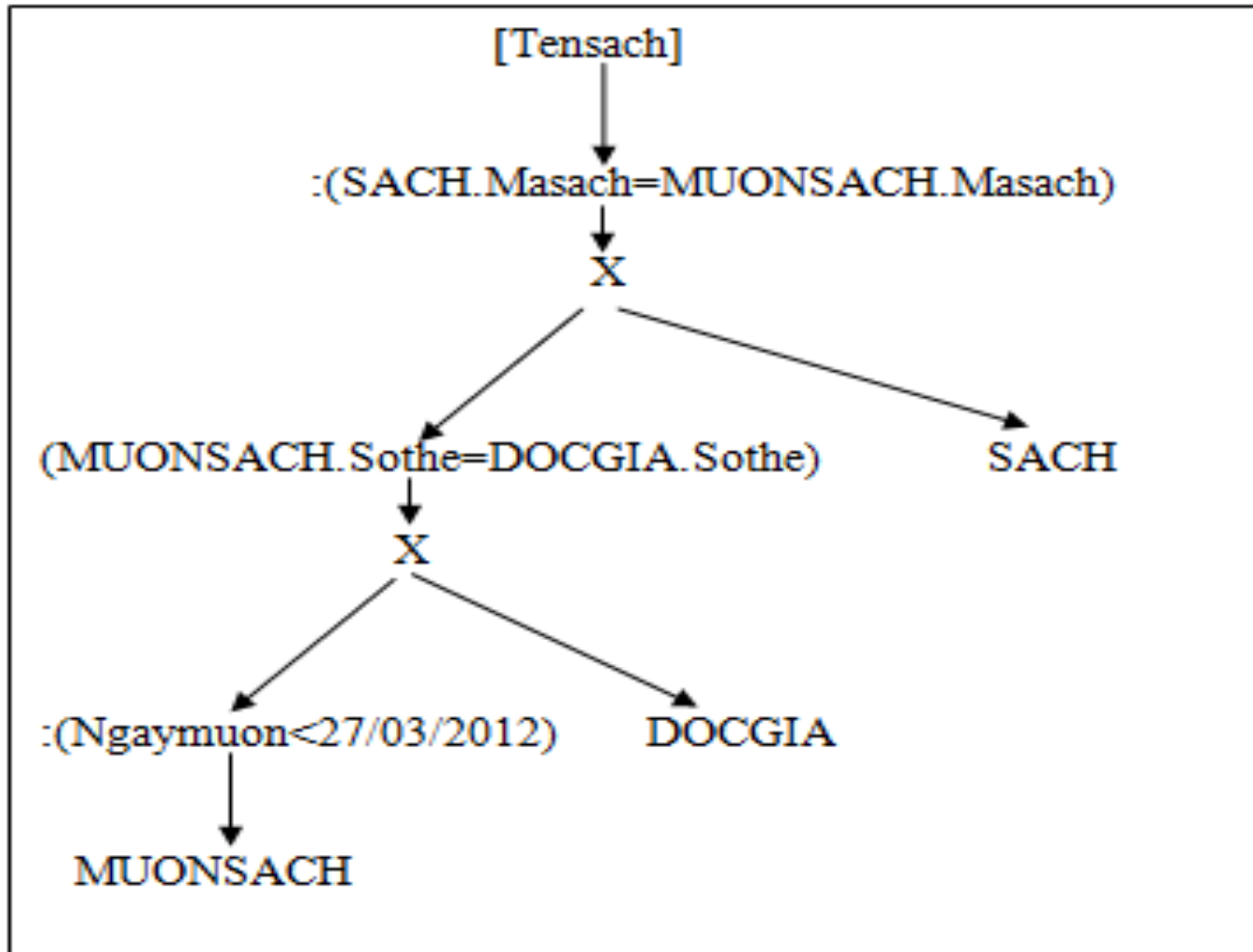
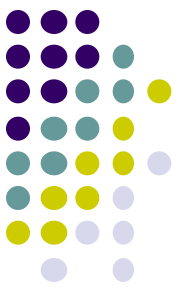
- Do vậy phép chọn: **MUONSACH.Sothe = DOCGIA.Sothe**
- có thể đẩy xuống để áp dụng cho tích Đề-các:
- **(MUONSACH x DOCGIA) : (: (Ngaymuon < 27/03/2012)**
- *Bước tiếp theo*: Tổ hợp hai phép chiếu thành một phép chiếu là **[Tensach]** nhờ luật 3 . Kết quả được cho như trong hình 6.2. Sau đó áp dụng quy tắc mở rộng 5 thay thế:
- **:(MUONSACH.Sothe = DOCGIA.Sothe)** và chiếu **[Tensach]**
- nhờ dãy phép toán:
- **[Tensach, SACH.Masach, MUONSACH.Masach] (1)**
- **:(SÁCH.Masach = MUONSACH.Masach)(2)**
- rồi chiếu để lấy tên sách:
- **[Tensach] (3)**

Ví dụ về thuật toán tối ưu hoá biểu thức quan hệ...



- Áp dụng quy tắc 9 để thay thế biểu thức đầu tiên, biểu thức số (1), của phép chiếu nhờ phép chiếu
- : **[Tensach, SACH.Masach]**
- áp dụng cho quan hệ SACH và **[MUONSACH.Masach]** áp dụng cho hạng thức phía trái của tích Đề-các trong hình 6.2.
- Phép chiếu cuối và phép chọn có thể áp dụng quy tắc mở rộng 5 để có dãy: **[MUONSACH.Masach, MUONSACH.Sothe, DOCGIA.Sothe]** (5)
- : **(MUONSACH. Sothe = DOCGIA.Sothe)** (6)
- **[MUONSACH.Masach] Masach]** (7)
- Phép chiếu đầu, số (5), được phân tách chuyển xuống tích Đề-các nhờ quy tắc 9 .
- Một phần phép chiếu **DOCGIA.Sothe** xuống hạng thức **DOCGIA** vì là thuộc tính của quan hệ này.

Ví dụ về thuật toán tối ưu hoá biểu thức quan hệ...



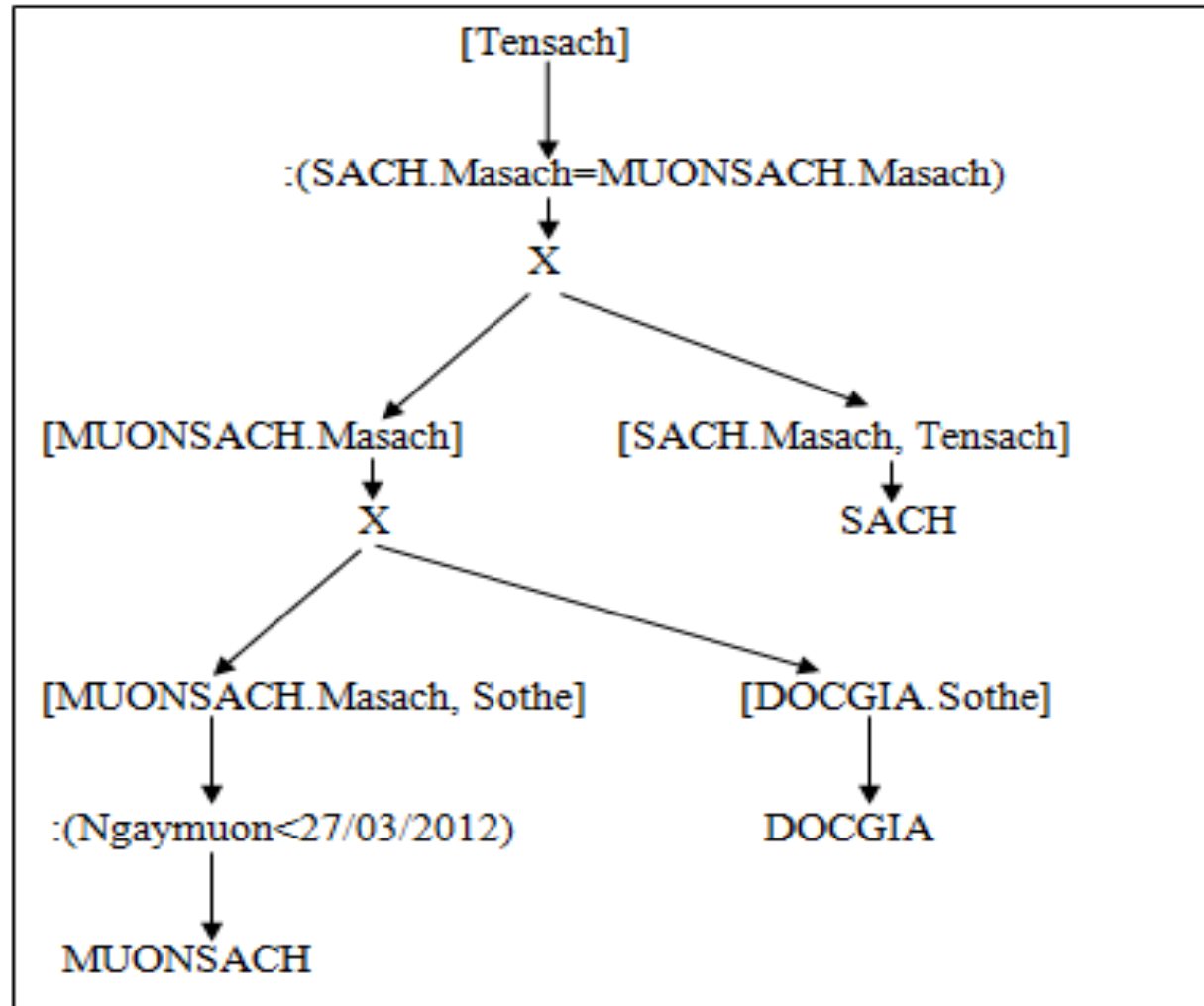
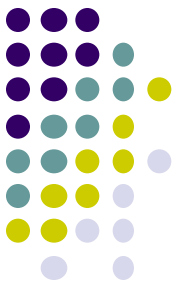
Hình 6.2. Cây với tổ hợp phép chọn và phép chiếu

Ví dụ về thuật toán tối ưu hoá biểu thức quan hệ...



- Phần còn lại là phép chiếu để lấy 3 thuộc tính:
- [MUONSACH.Masach, MUONSACH.Sothe, Ngaymuon]
- được đẩy xuống hạng thức MUONSACH. Các thuộc tính không phù hợp sẽ bị loại bỏ. Biểu diễn cây cuối cùng của biểu thức như trong hình 6.3.
- Nhóm các phép tính bằng đường mũi tên gián đoạn. Mỗi tích Đề-các được tổ hợp với phép chọn để tạo thành một phép kết nối bằng nhau (*Equi Join*) rất có hiệu quả. Đặc biệt phép chọn trên quan hệ MUONSACH và phép chiếu của DOCGIA để lấy thuộc tính **Sothe** ở phía dưới là đủ để tổ hợp với tích Đề-các. Thứ tự thực hiện của cây biểu thức trong các hình 6.1, 6.2 và 6.3 là từ dưới lên: Nhóm các phép toán nằm ở phía dưới được thực hiện trước các phép toán ở phía trên.

Ví dụ về thuật toán tối ưu hoá biểu thức quan hệ...



Hình 6.3. Cây kết quả biểu diễn việc phân nhóm các biểu thức

Thuật toán tối ưu hoá câu hỏi trong ngôn ngữ ĐSQH



- Bảng sau là phép biến đổi tương đương đối với các phép Hội (*Union*), Trừ (*Minus*), Giao (*Intersect*), Tích Đề-các (*Cartesian*), Chia (*Division*), Chiếu (*Projection*), Chọn (*Selection*).
- (B1). Kết tự nhiên tương đương với dãy phép tích Đề-các, phép chọn, phép chiếu: $Q1(A,B) * Q2(B,C) \equiv (Q1 \times Q2 : (Q1[B]=Q2[B])) [A,B,C]$
- (B2). Phép *theta* kết tương đương với dãy phép tích Đề-các và phép chọn với điều kiện *theta*: $Q1(A,B) Q2(C,D) \equiv (Q1 \times Q2) : (B \theta D)$
- (B3). Phép giao tương đương với phần bù (*Complement*) của hội hai phần bù của 2 QH:
$$Q1 \cap Q2 \equiv \neg ((\neg Q1) \cup (\neg Q2))$$
- (B4). Phép bù của một QH tương đương với tích Đề-các của các phép chiếu trên từng thuộc tính của QH trừ đi các bộ giá trị đã có trong thể hiện của QH:
$$\neg Q(X1, \dots, Xn) \equiv (Q[X1] \times \dots \times Q[Xn]) - Q(X1, \dots, Xn)$$
- (B5). Thương của 2 QH tương đương với hiệu của các QH trung gian sau:
$$Q1(A,B) \div Q2(A) = Q1[B] - ((Q1[B] \times Q2[A] - Q1(A,B)) [B])$$

Thuật toán tối ưu hoá câu hỏi trong ngôn ngữ ĐSQH...

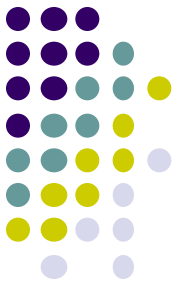


- *Input*: Sơ đồ cú pháp câu hỏi bằng ngôn ngữ ĐSQH.
- *Output*: Sơ đồ cú pháp tối ưu.
- Bước 1. Áp dụng các phép biến đổi tương đương nêu trong bảng (B1) đến (B5).
- Bước 2. Áp dụng luật 4 biến đổi dãy các phép chọn tương đương: tách phép chọn thành các phép chọn con.
- Bước 3. Đối với mỗi phép chọn, áp dụng các luật 5,6,7,8 để đẩy phép chọn đó xuống càng sâu càng tốt.
- Bước 4. Đối với mỗi phép chiếu, áp dụng các quy tắc 3,9,10 để đẩy phép chiếu đó xuống càng sâu càng tốt.
- Bước 5. Tập trung các phép chọn nhằm áp dụng luật 4
- Áp dụng luật 3 để loại bỏ bớt các phép chiếu vô ích.
- Tập trung các phép chọn với tích Đề-các, nếu được, để chuyển thành phép kết tự nhiên hay *theta* kết bằng cách áp dụng các luật 3,6 .

Thuật toán tối ưu hoá câu hỏi trong ngôn ngữ ĐSQH...



- **Nhận xét:**
- 1. Thuật toán trên chủ yếu nhằm giảm khối lượng dữ liệu trung gian chứ không chỉ ra thứ tự thực hiện các phép kết.
 - Ví dụ:
 - $(Q1(A,B) * Q2(B,C)) * Q3(A,C) \equiv$
 - $(Q1(A,B) * Q3(A,C)) * Q2(B,C)$
- 2. Thuật toán này không cho chúng ta một kết quả tối ưu mà nó chỉ đưa ra một giải pháp tốt.
- 3. Các phép biến đổi chỉ dựa trên các phép toán cơ bản là Hội (*Union*), Trừ (*Minus*), Giao (*Intersect*), Tích Đề-các (*Cartesian*), Chia (*Division*), Chiếu (*Projection*) và Chọn (*Selection*) mà chúng ta còn có thể thực hiện các phép biến đổi dựa trên các phép toán khác nữa.



Tổng kết và ôn tập