Tạp chí Tin học và Điều khiển học, T.24, S.1 (2008), 42–49

MỘT MÔ HÌNH DỮ LIỆU ĐỐI VỚI CÁC TRUY VẤN ĐỊNH LƯỢNG

NGUYỄN KIM ANH

Khoa Công nghệ thông tin - Trường Đại học Bách khoa Hà nội

Abstract. We extend the relational model of data to consider classes as attribute values, thereby to allow eary and convernient the representation of quantified queries. Facts regarding object classes can be stored and manipulated in the same way as facts regarding object instances. Our model is upwards compatible with the standard relational model.

Tóm tắt. Chúng tôi mở rộng mô hình dữ liệu quan hệ để cho phép các lớp như các giá trị thuộc tính và bởi vậy, cho phép biểu diễn các truy vấn định lượng một cách dễ dàng và thuận tiện hơn. Các sự kiện xem xét các lớp đối tượng có thể được lưu trữ và thao tác giống như các sự kiện xem xét các thể hiện đối tượng. Mô hình của chúng tôi là tương thích với mô hình quan hệ chuẩn.

1. GIỚI THIỆU

Phép chia là một phép toán của đại số quan hệ thực hiện một thao tác kiểm tra chứa tập nhiều-một, ở đây, một tập biểu diễn một nhóm các bộ. Phép chia tổng quát [5] mở rộng phép chia thường với việc thực hiện các thao tác kiểm tra chứa tập nhiều-nhiều. Phép toán này cũng tương tự như phép kết nối chứa tập đối với các cơ sở dữ liệu quan hệ (CSDL) không ở dạng chuẩn 1. Hiện tại, không một phép chia nào được cài đặt tường minh trong các hệ quản trị CSDL (DBMS), mặc dù các phép toán này giải quyết nhiều vấn đề ứng dụng quan trọng có ý nghĩa trong thực tiễn. Như chúng ta đã biết, các truy vấn định lượng, tức là các truy vấn với lượng từ hay hạn chế số đều đòi hỏi phải thực hiện một phép chia cùng với một phép kết tập nào đó. Những người sử dụng bình thường hay không chuyên nghiệp thường rất khó khăn và vất vả khi phải biểu diễn các truy vấn định lượng. Chúng tôi cho rằng, các ứng dụng quan tâm đến lượng từ sẽ ngày càng tăng lên trong một tương lai không xa. Bài báo này đề cập đến việc mở rộng mô hình dữ liệu quan hệ cho phép các lớp như các giá trị thuộc tính và bởi vậy, cho phép biểu diễn các truy vấn định lượng một cách dễ dàng và thuận tiện hơn.

Sự phân cấp được xem như một kỹ thuật cấu trúc dữ liệu rất phổ biến. Các cấu trúc cây cung cấp một kỹ thuật biểu diễn giàu ngữ nghĩa đối với việc quản lý các thông tin phức tạp. Sự phân cấp cũng đã được đề nghị trong ngữ cảnh thiết kế CSDL như một kỹ thuật biểu diễn tri thức thế giới thực. Trong bài báo này, chúng tôi cố gắng đưa khái niệm phân cấp lớp vào mô hình quan hệ thông qua việc cho phép các lớp được sử dụng như các giá trị thuộc tính trong một quan hệ. Đặc biệt, chúng tôi cũng đưa vào khái niệm khẳng định âm để cho phép các ngoại lệ được xảy ra. Hai phép toán mới được yêu cầu là: phép nhóm để gộp nhiều bộ của quan hệ vào một vài bộ đơn và phép mở nhóm để thu được một thể hiện đầy đủ của quan hệ. Mô hình dữ liệu thu được là tương thích với mô hình quan hệ chuẩn. Các CSDL đang tồn tại có thể tiếp tục được sử dụng theo mô hình của chúng tôi và ý nghĩa của các phép toán quan hệ thông thường không hề thay đổi.

Nhiều ưu thế có thể đạt được nhờ việc khai thác sự phân cấp trong biểu diễn dữ liệu [1]. Chúng ta có thể lưu trữ chỉ một bộ đơn với tên của lớp để thay thế cho nhiều bộ với các thành viên cấu thành nên lớp đó. Mô hình dữ liệu được trình bày trong bài báo này có thể xem như một giao diện cung cấp các thao tác bậc cao so với mô hình quan hệ chuẩn với sự hỗ trợ của phân cấp các lớp đối tượng. Việc cung cấp các thao tác bậc cao cho phép người dùng có thể biểu diễn các truy vấn một cách đơn giản hơn với ít thao tác hơn và hơn nữa, các truy vấn được đánh giá hiệu quả hơn.

Nội dung bài báo được trình bày như sau: phần 2 mở đầu với một số khái niệm cơ bản và định nghĩa một mô hình dữ liệu mở rộng đối với CSDL quan hệ. Phần 3 trình bày các phép toán đại số đối với mô hình dữ liệu mở rộng và phần 4 biểu diễn các truy vấn định lượng sử dụng các phép toán đại số của mô hình dữ liệu mở rộng. Cuối cùng, phần 5 đưa ra một vài đánh giá và kết luận.

2. MÔ HÌNH DỮ LIỆU MỞ RỘNG

Trước tiên, chúng tôi nhắc lại ngắn gọn một số khái niệm cơ bản. Cho một tập các thuộc tính $N = \{A_1, A_2, ..., A_n\}$, mỗi thuộc tính A_i được kết hợp với một miền D_i với $i = 1...n$. Kí hiệu D^* là tích Đề-các của các miền. Một quan hệ xác định trên tập thuộc tính $\{A_1, A_2, ..., A_n\}$ là một tập con nào đó của D^* và một bộ của quan hệ là một ánh xạ từ N đến D^*.

Đối với một số miền của thuộc tính, đặc biệt là miền của các thuộc tính khóa hay các thuộc tính định danh đối tượng, chúng tôi xác định một tập các miền con của nó ứng với một sự phân nhóm các đối tượng được xác định bởi thuộc tính này và xây dựng một cấu trúc phân cấp đối với các miền con hay các lớp đối tượng này. Một lớp, theo chúng tôi, là một tập C sao cho $C \subseteq D(D$ là miền của thuộc tính). Chúng tôi sẽ viết $\forall C$ hay đơn giản là C như giá trị của thuộc tính này trong một bộ, điều này có nghĩa là quan hệ phải chứa mọi bộ ứng với mỗi giá trị x mà $x \in C$ với các giá trị trên các thuộc tính khác vẫn giữ nguyên. Cấu trúc phân cấp đối với một miền D là một cấu trúc cây với gốc của cây chính là miền D và các cạnh đi từ một lớp tổng quát hơn đến các lớp đặc biệt hơn của nó. Các giá trị thuộc D là các lớp chỉ có duy nhất một đối tượng và được đặt ở các nút lá của cây.

Bất kể một cố gắng nào đối với việc tổng quát hóa tri thức thế giới thực tất sẽ thấy cần đưa vào các ngoại lệ. Mô hình dữ liệu của chúng tôi cũng cho phép một dạng ngoại lệ thông qua việc kết hợp mỗi bộ với một giá trị chân lý. Một bộ được gọi là bộ dương hay khẳng định dương nếu giá trị chân lý của bộ là 'true' (đúng), ngược lại, bộ với giá trị chân lý 'false' (sai) sẽ được gọi là bộ âm hay khẳng định âm. Tất nhiên, các khẳng định âm ứng với các sự kiện không tồn tại trong CSDL. Với sự có mặt của các khẳng định âm, một bộ x với các giá trị nguyên tố được chứa trong một bộ dương của quan hệ chỉ là đúng khi không tồn tại một khẳng định âm chứa x. Chúng ta có thể đưa ra một định nghĩa tổng quát cho khái niệm quan hệ nhóm trong mô hình dữ liệu mở rộng như sau.

Định nghĩa 1. Cho một tập các thuộc tính $N = \{A_1, A_2, ..., A_n\}$, mỗi thuộc tính A_i được kết hợp với một miền D_i và có thể có một cấu trúc phân cấp H_{Di} với $i = 1..n$. Kí hiệu D_i^+ là tập tất cả các nút của H_{Di} đối với các thuộc tính A_i có cấu trúc phân cấp, $D_i^+ = D_i$ đối với các thuộc tính A_i không có cấu trúc phân cấp, $TF = \{'true', 'false'\}$ và kí hiệu D^* là tích Đề-các của các miền D_i^+ và TF. Một quan hệ nhóm R xác định trên tập thuộc tính $\{A_1, A_2, ..., A_n, T\}$ là một tập con nào đó của D^* và một bộ của quan hệ là một ánh xạ từ $N \cup T$ đến D^* với $T \notin N$ là tên của thuộc tính bổ sung để lưu trữ giá trị chân lý của bộ.

3. CÁC PHÉP TOÁN ĐẠI SỐ ĐỐI VỚI MÔ HÌNH DỮ LIỆU MỞ RỘNG

Trong phần trước, chúng ta đã đưa ra khái niệm quan hệ nhóm, một sự mở rộng của khái niệm quan hệ kinh điển. Trong trường hợp tất cả các thuộc tính của một quan hệ nhóm đều không có cấu trúc phân cấp và giá trị chân lý của mọi bộ trong quan hệ đều là 'true' thì quan hệ nhóm này chính là quan hệ kinh điển với việc bổ sung thêm thuộc tính T nhận giá trị 'true' đối với mọi bộ của quan hệ. Chúng tôi gọi quan hệ nhóm này là quan hệ nhóm nguyên tố và quan hệ nhóm nguyên tố sau khi bỏ đi thuộc tính T là quan hệ nguyên tố. Như vậy, có thể xem quan hệ kinh điển là một trường hợp đặc biệt của quan hệ nhóm tổng quát. Trong phần này, chúng tôi sẽ đưa ra các phép toán đại số trên mô hình mở rộng này. Hai phép toán mới: phép nhóm và phép mở nhóm là cần thiết để định nghĩa các phép toán đại số trên các quan hệ nhóm. Chú ý rằng, mỗi quan hệ nhóm phải tương đương với một quan hệ nguyên tố duy nhất đối với các cấu trúc phân cấp đã cho. Có nghĩa là, có một mô hình duy nhất với các giá trị nguyên tố (các giá trị thuộc miền gốc của thuộc tính) thỏa mãn quan hệ đã cho. Một phép toán nào đó trên các quan hệ nhóm sẽ có cùng tác động bất kể nó được thực hiện trên các quan hệ nhóm hay trên các quan hệ nguyên tố tương đương. Theo nghĩa này, ngữ nghĩa của các phép toán quan hệ là không bị thay đổi ngay cả trong trường hợp với các quan hệ nhóm.

Để trình bày các phép toán trên các quan hệ nhóm, trước tiên, chúng tôi trình bày một thuật toán khá đơn giản cho phép dịch một tập các giá trị trong một D^+ với H_D tương ứng của một thuộc tính A nào đó thành các lớp tổng quát hơn trong H_D và cho phép có ngoại lệ. Với các loại câu truy vấn định lượng khác nhau, chúng ta có thể sử dụng các thuật toán dịch khác nhau với các tiêu chí nhóm khác nhau. Ở đây, chúng tôi chỉ quan tâm đến các truy vấn với lượng từ $\forall$ và bởi vậy, tiêu chí nhóm của chúng tôi là nhóm tất cả các nút dương có cùng cha lại khi số các ngoại lệ hay các nút âm của chúng là nhỏ hơn. Trong thực tế, với một sự phân nhóm các đối tượng hợp lý, số các ngoại lệ thường là khá nhỏ hay thực chất là không đáng kể.

Cho một thuộc tính A được kết hợp với miền D có một cấu trúc phân cấp H_D và $S \subset D^+$ với D^+ - kí hiệu tập tất cả các nút của H_D.

Thuật toán 1. Tính các lớp tổng quát của $S \subset D^+$ và các ngoại lệ.

Vào: Thuộc tính A, miền D, cấu trúc phân cấp H_D và $S \subset D^+$.
Ra: Tính Lớp (S) và $NgL(S)$.
Cách tính:
• Tìm nút cha chung nhỏ nhất của S đối với H_D và kí hiệu nút này là Rt.
• Kí hiệu cây con của H_D với nút gốc là Rt và các nút lá là các nút hoặc thuộc S hoặc là các nút anh em của các nút trong S là T.
• Đánh dấu các nút lá thuộc S là $+$
• Đánh dấu các nút lá không thuộc S là $-$
• Đối với mỗi nút trong N của T, kí hiệu cây con của T gốc tại N là T_N và kí hiệu S_N^+ là tập các nút lá được đánh dấu $+$, S_N^- là tập các nút lá được đánh dấu - đối với T_N. Tính $x = |S_N^+|$ và $y = |S_N^-|$. Một nút N có $x > y$ được gọi là nút tốt, ngược lại được gọi là nút tồi.
• Đặt $N := Rt$. Gọi thủ tục Chọn (N, T_N, S_N^+, S_N^-)
• Thủ tục chọn nút Chọn $(N, T_N,\ S_N^+,\ S_N^-)$:

1) Nếu N là nút lá và được đánh dấu $+$ thì Lớp$(S_N^+) = N,\ NgL(S_N^+) = \emptyset$

2) Nếu N là nút tốt có m con trực tiếp là $N_1, N_2, ..., N_m$ và không mất tính tổng quát, giả sử $N_1, N_2, ..., N_k$ là các nút tồi và $N_{k+1}, N_2, ..., N_m$ là các nút tốt.

Nếu $1+\sum_1^k y_i < m-k+\sum_1^k x_i$ với $x_i = |S_{Ni}^+|$ và $y_i = |S_{Ni}^-|$ thì Lớp$(S_N^+) = N,\ NgL(S_N^+) = S_N^-$ và kết thúc.

Ngược lại, thực hiện Chọn $(N_i, T_{Ni}, S_{Ni}^+, S_{Ni}^-)$ với $i = 1..m$ và đặt Lớp$(S_N^+) = \cup_1^m$ Lớp$(S_{Ni}^+),\ NgL(S_N^+) = \cup_1^m NgL(S_{Ni}^+)$.

3) Nếu N là nút tồi có m con trực tiếp là $N_1, N_2, ..., N_m$, thực hiện Chọn$(N_i, T_{Ni}, S_{Ni}^+, S_{Ni}^-)$ với $i = 1..m$ và đặt Lớp$(S_N^+) = \cup_1^m$ Lớp$(S_{Ni}^+),\ NgL(S_N^+) = \cup_1^m NgL(S_{Ni}^+)$.

• Đặt Lớp$(S) =$ Lớp(S_{Rt}^+) và $NgL(S) = NgL(S_{Rt}^+)$

Mệnh đề 1. *Cho một thuộc tính A được kết hợp với miền D có một cấu trúc phân cấp H_D và $S \subset D^+$. Kí hiệu ĐTượng (S) là tập tất cả các đối tượng thuộc các lớp trong S thì ĐTượng (S) = ĐTượng (Lớp(S)) $\setminus$ ĐTượng ($NgL(S)$).*

Chứng minh. Dễ thấy, do cấu trúc phân cấp H_D, tập tất cả các đối tượng thuộc một lớp chính là hợp của tập tất cả các đối tượng thuộc các lớp con của nó. Bởi vậy, chúng ta có ĐTượng (Lớp(S)) = ĐTượng (S) $\cup$ĐTượng ($NgL(S)$). Hơn nữa, ĐTượng (S) và ĐTượng ($NgL(S)$) là rời nhau. Do vậy, ta có điều phải chứng minh. ■

Phép mở nhóm: Cho quan hệ nhóm $R(X, Y, T)$ với các thuộc tính $A \in Y$ được kết hợp với một miền D và có một cấu trúc phân cấp H_D và các thuộc tính thuộc X không được kết hợp với một cấu trúc phân cấp nào. X hoặc Y có thể bằng $\emptyset$ nhưng $XY \neq \emptyset$.

Giả sử $Y = \{A_1, A_2, ..., A_k\}$, $v[Y] = (C_1, C_2, ..., C_k)$ và một bộ nguyên tố $t[Y] = (c_1, c_2, ..., c_k)$ với $c_i \in D_i$ - miền của A_i. Chúng ta viết $t[Y] \in v[Y]$ nếu với $i = 1..k$ thì $c_i \in C_i$. Phép mở nhóm đối với quan hệ nhóm R, kí hiệu là MởNhóm(R) được định nghĩa như sau:

MởNhóm$(R) = \{t[XY]/\exists v \in R(t[X] = v[X] \wedge t[Y] = y \wedge y \in v[Y] \wedge v[T] = \text{‘}true\text{'} \wedge\ \nexists v' \in$

$R\ (v'[X] = v[X] \wedge v'[T] = \text{'}false' \wedge y \in v'[Y]))\}$.

Chú ý: Phép mở nhóm trả lại kết quả là một quan hệ nguyên tố thông thường, quan hệ chỉ chứa các giá trị nguyên tố và không chứa thuộc tính T lưu giá trị chân lý của bộ. Hơn nữa, phép toán MởNhóm đảm bảo mỗi quan hệ nhóm là tương đương với một quan hệ nguyên tố duy nhất đối với các cấu trúc phân cấp đã cho.

Phép nhóm: Cho một quan hệ nhóm R xác định trên tập thuộc tính U với một thuộc tính $A \in U$ nào đó được kết hợp với một miền D và có một cấu trúc phân cấp H_D. Kí hiệu R^+ là tập các bộ dương của R và R là tập các bộ âm của R. Phép nhóm quan hệ nhóm R theo thuộc tính A, kí hiệu là $\text{Nhóm}_A(R)$ được định nghĩa như sau:

$\text{Nhóm}_A(R^+) = \{t/(v \in R^+(t[U\backslash TA] = v[U\backslash TA] \wedge ((t[A] = \{C/\exists C \in \text{Lớp }(S)\} \wedge t[T] = \text{'}true') \vee (t[A] = \{C/\exists C \in NgL(S)\} \wedge t[T] = \text{'}false')) \wedge S = \cup v[A]$ với các $v \in R^+$ thỏa $t[U\backslash A] = v[U\backslash A])\}$.

$\text{Nhóm}_A(R) = \text{Nhóm}_A(R^+) \cup R^-$.

Chú ý: Trong quá trình nhóm, chúng ta luôn cố gắng nhóm nhiều nhất có thể các lớp con xuất hiện trong thuộc tính A của các bộ dương thành một lớp tổng quát hơn với việc chấp nhận một số ngoại lệ không đáng kể nên tập các bộ thuộc $\text{Nhóm}_A(R^+)$ và R^- là rời nhau.

Mệnh đề 2. *Phép toán Nhóm bảo toàn nội dung thông tin của quan hệ nhóm ban đầu. Có nghĩa là với một quan hệ nhóm R và một thuộc tính A có một cấu trúc phân cấp thì MởNhóm(Nhóm$_A$(R)) = MởNhóm(R).*

Chứng minh. Theo định nghĩa của các phép toán MởNhóm, Nhóm và tính đúng đắn của Thuật toán 1 được khẳng định trong mệnh đề 1, mỗi bộ x với các giá trị nguyên tố thuộc MởNhóm(Nhóm$_A(R)$) cũng thuộc MởNhóm(R) và ngược lại. ■

Hệ quả 1. *Cho R là quan hệ nhóm nguyên tố và một thuộc tính A có một cấu trúc phân cấp. Gọi R' là quan hệ nguyên tố đối với R thì MởNhóm(NhómA(R)) = R'.*

Hệ quả 1 được suy ra trực tiếp từ Mệnh đề 2.

Phép kết nối tự nhiên: Cho hai quan hệ nhóm $R(U)$ và $S(V)$ với $U \cap V = YZ$ và $Y = \{A_1, A_2, ..., A_k\}$. Các thuộc tính $A_i \in Y$ được kết hợp với một miền Di và có một cấu trúc phân cấp H_{Di}, $i = 1..k$. Các thuộc tính thuộc Z không được kết hợp với một cấu trúc phân cấp nào. Z có thể bằng $\emptyset$ và k có thể bằng 0. Phép kết nối tự nhiên của 2 quan hệ nhóm R và S, kí hiệu là $R * S$ được định nghĩa như sau:

$R * S = \{t[UV]/\exists u \in R \exists v \in S(t[U\backslash YT] = u[U\backslash YT] \wedge t[V\backslash YT] = v[V\backslash YT] \wedge (t[A_i] = u[A_i] \cap v[A_i],\ i = 1..k) \wedge t[T] = u[T] * v[T]\}$, ở đây $\text{'}true' * \text{'}true' = \text{'}true'$ và $\text{'}true' * \text{'}false' = \text{'}false' * \text{'}true' = \text{'}false'$.

Phép chiếu: Cho quan hệ nhóm $R(X, Y, T)$ với các thuộc tính $A \in Y$ được kết hợp với một miền D và có một cấu trúc phân cấp H_D và các thuộc tính thuộc X không được kết hợp với một cấu trúc phân cấp nào. X và Y có thể bằng $\emptyset$. Chúng tôi định nghĩa hai phép chiếu trên quan hệ nhóm R phụ thuộc vào tập thuộc tính chiếu: Π^0 là phép chiếu với tập thuộc tính chiếu chỉ liên quan đến các thuộc tính không có cấu trúc phân cấp và Π^1 là phép chiếu với tập thuộc tính chiếu có liên quan đến các thuộc tính có cấu trúc phân cấp.

$\Pi^0_Z(R) = \{t[ZT]/t \in R^+$ với $Z \subseteq X$ và $Z \neq \emptyset\}$, ở đây R^+ là tập các bộ dương của R.

$\Pi^1_Z(R) = \{t[ZT]/t \in R$ với $Z \subseteq XY$ và $Z \neq \emptyset\}$.

Phép chọn: Cho quan hệ nhóm $R(X, Y, T)$ với các thuộc tính $A \in Y$ được kết hợp với một miền D và có một cấu trúc phân cấp H_D và các thuộc tính thuộc X không được kết hợp với một cấu trúc phân cấp nào. Chúng tôi định nghĩa hai phép chọn trên quan hệ nhóm R phụ thuộc vào dạng của điều kiện chọn: σ^0 là phép chọn với điều kiện chọn chỉ liên quan đến các thuộc tính không có cấu trúc phân cấp và σ^1, σ^2 là phép chọn với điều kiện chọn liên quan đến các thuộc tính có cấu trúc phân cấp. Kí hiệu R^+ là tập các bộ dương của R, R^- là tập các bộ âm của R. Để đơn giản, giả thiết điều kiện chọn có liên quan đến các thuộc tính có cấu trúc phân cấp có dạng $A\theta C$ với $A \in Y$, $C \in H_D$ và $\theta \in \{=, \subseteq, \supseteq\}$.

$\sigma^0_{F(X)}(R) = \{t/t \in R \wedge F(t) = \text{đúng}\}$

$\sigma^1_{A=C}(R) = \{t/\exists u \in R(u[A] \cap C \neq \emptyset \wedge t[U\backslash A] = u[U\backslash A]) \wedge t[A] = u[A] \cap C\}$

$\sigma^2_{A=C}(R) = \{t/\exists u \in R^+(u[A] = C \wedge t[U] = u[U]) \vee \exists v \in R^-(v[A] \subseteq C \wedge t[U] = u[U])\}$.

$\sigma^2_{A\supseteq C}(R) = \{t/\exists u \in R^+(u[A] \supseteq C \wedge t[U] = u[U]) \vee \exists v \in R^-(v[A] \subseteq C \wedge t[U] = u[U])\}$.

$\sigma^2_{A\subseteq C}(R) = \{t/\exists u \in R^+(u[A] \supseteq C \wedge t[U] = u[U]) \vee \exists v \in R^-(v[A] \subseteq C \wedge t[U] = u[U]) \wedge v \notin R'^-\}$. với R'^- là các bộ âm của phép nhóm đầu tiên theo A)}.

Chú ý: Phép chọn σ^0 và σ^1 được sử dụng trong trường hợp $X \neq \emptyset$ còn σ^2 được sử dụng trong trường hợp $X = \emptyset$ và quan hệ R đã được nhóm theo một thuộc tính $A \in Y$ đầu tiên, sau đó có thể tiếp tục nhóm theo các thuộc tính khác trong Y.

Các phép toán tập hợp: Để thực hiện các phép toán tập hợp: hợp, trừ và giao, trước tiên chúng ta thực hiện phép toán MởNhóm đối với hai quan hệ toán hạng, sau đó thực hiện các phép toán hợp, trừ, giao tương ứng của đại số quan hệ kinh điển trên hai quan hệ thu được. Cuối cùng, thực hiện phép Nhóm trên quan hệ kết quả theo các thuộc tính mong muốn, chúng ta sẽ thu được quan hệ nhóm kết quả của các phép toán tập hợp này.

Định lý 1. *Các phép toán đại số: chọn (σ^0), chiếu $\Pi^{(0)}$ và kết nối tự nhiên trên các quan hệ nhóm là bảo toàn ngữ nghĩa bất kể nó được thực hiện trên các quan hệ nhóm hay trên các quan hệ nguyên tố tương đương.*

Chứng minh. Giả sử P - kí hiệu phép chọn (σ^0) hoặc phép chiếu $\Pi^{(0)}$. Dựa trên định nghĩa của phép chọn và phép chiếu đối với quan hệ nhóm, chúng ta có: MởNhóm($P(R)$)= P(MởNhóm(R)). Tương tự với phép kết nối tự nhiên, mỗi bộ x ∈MởNhóm($R * S$) khi và chỉ khi x phải được chứa trong một bộ dương của ($R * S$) và không tồn tại một bộ âm của ($R * S$) chứa x, tức là x không thể được ghép bởi một bộ âm của R hoặc một bộ âm của S. Điều này có nghĩa là $x \in$ MởNhóm(R)∗MởNhóm(R). Do vậy, MởNhóm($R * S$) = MởNhóm(R)∗MởNhóm(R). ■

Chú ý: Phép toán Π^1 sẽ không còn bảo toàn ngữ nghĩa khi nó được áp dụng sau một phép chọn σ^2. Định lý 2 sau đây cho thấy, một phép toán Π^1 sẽ được áp dụng ngay sau một phép chọn σ^2 có tác động tương đương với một phép chia của đại số quan hệ kinh điển.

Mệnh đề 3. *Cho quan hệ nhóm $R(U)$ với một thuộc tính $A \in U$ được kết hợp với một miền D và có một cấu trúc phân cấp H_D. Khi đó, ta có MởNhóm($\sigma^1_{A=C}(R)$) = $\sigma_{A\in C}$*

(MởNhóm(R)), ở đây $A \in C$ *kí hiệu biểu thức logic* $A = c_1 \vee A = c_2 \vee ... \vee A = c_k$ *với* $C = \{c_1, c_2, ..., c_k\}$.

Có thể dễ dàng chứng minh mệnh đề này dựa trên định nghĩa của phép toán σ^1.

Định lý 2. *Cho một quan hệ nhóm R xác định trên tập thuộc tính* $U = \{X, A, T\}$ *với thuộc tính A có một cấu trúc phân cấp. Quan hệ R đã được nhóm theo A đầu tiên. Khi đó MởNhóm(*$\Pi^1_X \sigma^2_{A \supseteq C}(R)$*) = MởNhóm(R)* $\div S$ *với S là một quan hệ thông thường xác định trên một thuộc tính A và chứa các bộ tương ứng với các đối tượng thuộc lớp C.*

Chứng minh. Ta có mỗi bộ $x \in$MởNhóm$(\Pi_X \sigma^2_{A \supseteq C}(R))$ khi và chỉ khi x phải được chứa trong một bộ dương của $\sigma^2_{A \supseteq C}(R)$ và không tồn tại một bộ âm của $\sigma^2_{A \supseteq C}(R)$ chứa x, tức là $\exists u \in R^+(u[A_k] = C) - u$ chứa x, và không $\exists v \in R^-(v[A_k] \subseteq C) - v$ chứa x và x chỉ xác định trên tập X. Điều này có nghĩa là $x \in$MởNhóm$(R) \div S$. Ta có điều phải chứng minh. ■

4. BIỂU DIỄN CÁC TRUY VẤN ĐỊNH LƯỢNG CỦA MÔ HÌNH DỮ LIỆU MỞ RỘNG

Cho một CSDL với các quan hệ nguyên tố: *S1(S#, SNAME), S2(S#, CITY), P1(P#, PNAME), P2(P#, COLOR), P3(P#, WEIGHT), SP(S#,P#)*, ở đây thuộc tính $S\#$ được kết hợp với miền D_S và và có một cấu trúc phân cấp H_Ds và $P\#$ được kết hợp với miền D_P và và có một cấu trúc phân cấp H_{Dp}.

Các truy vấn định lượng

- Cho biết các lớp mặt hàng có màu đỏ: $(\Pi^1_{P\#}(\sigma^0_{COLOR} = \text{Đỏ}(\text{Nhóm}_{P\#}P2)))$
- Cho biết thông tin về địa chỉ của tất cả các các hãng thuộc lớp D :

  $\sigma^1_{S\#=D}(\text{Nhóm}_{S\#}S2)$ - kết quả này chứa đựng đầy đủ thông tin mang tính giải thích về địa chỉ của tất cả các các hãng thuộc lớp D.
- Nếu chúng ta không cần quan tâm đến địa chỉ đích xác của từng hãng, chúng ta có thể biểu diễn như sau: $\Pi^0_{CITY}(\sigma^1_{S\#=D}(\text{Nhóm}S\#S2))$.
- Đưa ra tên các hãng cung ứng ít nhất tất cả các mặt hàng thuộc lớp C :

  $\Pi_{SNAME}(S1*$ MởNhóm$(\Pi^1_{S\#}\sigma^2_{P\# \supseteq C}$ $(\text{Nhóm}_{P\#}SP)))$
- Đưa ra tên các hãng chỉ cung ứng tất cả các mặt hàng thuộc lớp C :

  $\Pi_{SNAME}(S1*$ MởNhóm$(\Pi^1_{S\#}\sigma^2_{P\#=C}$ $(\text{Nhóm}_{P\#}SP)))$
- Đưa ra thông tin về địa chỉ của các hãng cung ứng nhiều nhất tất cả các mặt hàng thuộc lớp C: $\text{Nhóm}_{S\#}S2 * (\Pi^1_{S\#}\sigma^2_{P\#} \supseteq C$ $(\text{Nhóm}_{S\#}$ $(\text{Nhóm}_{P\#}SP)))$.

5. ĐÁNH GIÁ VÀ KẾT LUẬN

Với việc đưa khái niệm phân cấp lớp vào mô hình quan hệ thông qua việc cho phép các tên lớp được sử dụng như các giá trị thuộc tính trong một quan hệ và chấp nhận các ngoại lệ, một số dạng truy vấn định lượng trên các lớp đối tượng đã có thể được biểu diễn một cách dễ dàng thông qua các phép toán đại số trên các quan hệ nhóm. Như vậy, mô hình dữ liệu mở rộng này có thể xem như một giao diện cung cấp các thao tác bậc cao so với mô hình quan hệ chuẩn với sự hỗ trợ của phân cấp các lớp đối tượng. Việc cung cấp các thao tác bậc cao cho phép người dùng có thể biểu diễn các truy vấn một cách đơn giản hơn với ít thao tác hơn và hơn nữa, các truy vấn được đánh giá hiệu quả hơn. Đặc biệt, so với các mô hình dữ liệu quan hệ lồng, các quan hệ lồng không ở dạng chuẩn 1, việc cài đặt và đánh giá các biểu thức đại số lồng khá phức tạp và thường phải biến đổi thành các biểu thức đại số phẳng [3, 4]. Trong mô hình dữ liệu mở rộng của chúng tôi, các sự kiện xem xét các lớp đối tượng có thể được lưu trữ và thao tác giống như các sự kiện xem xét các thể hiện đối tượng. Mô hình của chúng tôi là tương thích với mô hình quan hệ chuẩn. Tuy nhiên, một trong nhưng khó khăn và hạn chế khi áp dụng mô hình dữ liệu mở rộng này là chúng ta cần xác định được các cấu trúc phân cấp H_D đối với miền D của một thuộc tính định danh A nào đó. Trong thực tế, nếu một sự phân nhóm các đối tượng hợp lý được cung cấp từ chính những người sử dụng của hệ và số các ngoại lệ là khá nhỏ hay thực chất là không đáng kể, các truy vấn định lượng trên các lớp đối tượng sẽ được biểu diễn một cách dễ dàng thông qua các phép toán đại số trên các quan hệ nhóm. Ngoài ra, chúng tôi cũng cho rằng, chúng ta có thể áp dụng một số thuật toán tự động sản sinh ra các cấu trúc phân cấp dựa trên chính CSDL của hệ với tiêu chí tối thiểu hoá lực lượng các quan hệ nhóm thu được. Bài báo [2] đề cập đến một giải thuật cho phép nhóm các giá trị thuộc tính dựa trên các mẫu có thể là một tham kháo tốt cho ý đồ này. Cuối cùng, chúng tôi hy vọng có thể tiếp tục phát triển và thử nghiệm mô hình này với nhiều dạng truy vấn định lượng khác nhau để đáp ứng được các nhu cầu khai thác thông tin đa dạng và phong phú của người dùng.

TÀI LIỆU THAM KHẢO

[1] H. V. Jagadish, Incorporating hierarchy in a relational model of data, *Proceedings ACM-SIGMOD International Conference on Management of Data*, 1989 (78–87). NOI NAO???

[2] M. Merzbacher, W. W. Chu, Pattern-Based Clustering for Database Attribute Values, *Proceedings AAAI Workshop on Knowledge Discovery in Database*, 1993. NOI NAO???

[3] W. Y. Mok, D. W. Embley, A normal formal for precisely characterizing redundancy in nested relations, *ACM TODS* **21** (1996) 77–106.

[4] J. Paredaens, D. V. Gucht, Converting nested algebra expressions to flat algebra expressions, *ACM TODS* **17** (1992) 66–93.

[5] R. Rantzau, C. Mangold, Laws for rewriting queries containing division operators, *Proceedings ICDE*, USA, 2006.

Nhận bài ngày 15 - 8 - 2007