

KHÓA VÀ CÁC DẠNG CHUẨN TRONG CÁC SƠ ĐỒ QUAN HỆ

NGUYỄN BÁ TƯỜNG

Abstract. The key of Relation Scheme is played a very important part for problems and matters in relation database. In this paper, we present several methods to seek keys and their application for determining normal forms in Relation Scheme.

Tóm tắt. Bài báo trình bày một số phương pháp tìm khóa của sơ đồ quan hệ và ứng dụng của chúng vào việc xác định những dạng chuẩn của các sơ đồ quan hệ.

1. MỞ ĐẦU

Như chúng ta đã biết khóa đóng một vai trò hết sức quan trọng trong các bài toán liên quan đến sơ đồ quan hệ (SĐQH) $W = \langle R, \Omega \rangle$, với R là tập thuộc tính, Ω là tập các ràng buộc dạng phụ thuộc hàm, phụ thuộc đa trị hoặc phi phụ thuộc hàm [3-7]. Ví dụ khi cần xét $W = \langle R, \Omega \rangle$ thuộc dạng nào: 2NF, 3NF, BCNF,...? nói chung chúng ta phải biết khóa của W và suy ra tập F_n các thuộc tính thứ cấp (các thuộc tính không khóa). Trong [5], [6], [7],... các tác giả đã xét thuật toán tìm khóa (xem Thuật toán 1 trong phần thuật toán tìm khóa sau đây), tuy nhiên thuật toán trên chỉ cho phép chúng ta tìm một khóa, việc tìm một khóa khác bằng thuật toán này và thay đổi thứ tự loại bỏ các thuộc tính nói chung là không đơn giản. Một vấn đề hết sức quan trọng ở đây là cách tìm khóa theo thuật toán trên không cho ta khẳng định: “đã tìm hết khóa của W ”.

Cụ thể hơn, chúng ta xét ví dụ sau:

Cho SĐQH $W = \langle R, F \rangle$, với $R = \{A, B, C, D, E, G\}$ và tập phụ thuộc hàm $F = \{AB \rightarrow C, D \rightarrow EG, C \rightarrow A, BE \rightarrow C, BC \rightarrow D, CG \rightarrow BD, ACD \rightarrow B, CE \rightarrow AG\}$. Hãy xét xem W thuộc dạng nào? Bài toán đòi hỏi chúng ta phải tìm hết khóa của W , tức chúng ta phải xác định hết các thuộc tính khóa và không khóa. Bằng Thuật toán 1 để tìm các khóa ta sẽ gặp khó khăn và lúng túng vì không biết ta đã tìm hết khóa chưa.

Trong bài này tác giả muốn trình bày một vài phương pháp để giải quyết vấn đề nêu trên một cách đơn giản trong một số lớp các sơ đồ quan hệ.

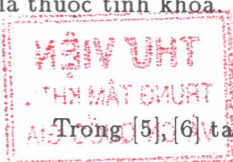
2. ĐỊNH NGHĨA KHÓA

Cho SĐQH $W = \langle R, F \rangle$ với R là tập các thuộc tính. F là tập các phụ thuộc hàm trên R . Tập con K các thuộc tính của R được gọi là khóa (tối thiểu) của W nếu K có bao đóng bằng R và nếu bớt khỏi K dù chỉ một phần tử thì tập còn lại có bao đóng khác R . Nói cách khác K là khóa của W nếu K là tập bé nhất có bao đóng bằng R .

Ví dụ quay lại SĐQH đã nêu ở trên $W = \langle R, F \rangle$ và $R = \{A, B, C, D, E, G\}$, $F = \{AB \rightarrow C, D \rightarrow EG, C \rightarrow A, BE \rightarrow C, BC \rightarrow D, CG \rightarrow BD, ACD \rightarrow B, CE \rightarrow AG\}$, ta dễ dàng thấy rằng các tập $K_1 = \{A, B\}$, $K_2 = \{B, E\}$, $K_3 = \{C, G\}$, $K_4 = \{C, E\}$, $K_5 = \{C, D\}$,... là các khóa của W và như vậy W không có thuộc tính thứ cấp (thuộc tính không khóa), tất cả các thuộc tính của W đều là thuộc tính khóa.

3. THUẬT TOÁN TÌM KHÓA

Trong [5], [6] ta đã biết thuật toán tìm khóa như sau:



Input $W = \langle R, F \rangle$

Output K là một khóa của W .

Algorithm

b1 $K := R$

b2 Lặp lại quá trình loại khỏi K phần tử $A \in K$ mà $(K - A)^+ = R$

Viết cách khác ta có

Thuật toán 1

Begin

$K := R$;

For each $A \in K$ if $(K - A)^+ = R$ then $K := K - A$ else $K := K$

End;

Dùng Thuật toán 1 để tìm một khóa của SDQH $W = \langle R, F \rangle$ nêu trên ta có

b1 (bước 1)

$K := R = \{A, B, C, D, E, G\}$

b2 (bước 2) trước tiên ta chọn thuộc tính A

Vì $(K - A)^+ = R$ nên $K := K - A = \{B, C, D, E, G\}$, tiếp đến ta chọn B

Vì $(K - B)^+ = R$ nên $K := K - A = \{C, D, E, G\}$, sau đây ta chọn C

Vì $(K - C)^+ \neq R$ nên $K := K = \{C, D, E, G\}$, chọn tiếp D

Vì $(K - D)^+ = R$ nên $K := K - D = \{C, E, G\}$, cuối cùng ta chọn E

Vì $(K - E)^+ = R$ nên $K := K - E = \{C, G\}$, và dễ dàng suy ra đây là một khóa vì không thể chọn để bớt tiếp được nữa. Việc thay đổi thứ tự loại bỏ các thuộc tính khỏi K có thể cho ta các khóa khác nhau nhưng không chota kết luận *đã hết khóa chưa?* và chúng ta chưa thể biết W có bao nhiêu thuộc tính khóa và thuộc tính thứ cấp.

4. CÁC DẠNG CHUẨN CỦA CÁC SƠ ĐỒ QUAN HỆ

a. Dạng chuẩn 2 (2NF)

Ta nói sơ đồ quan hệ $W = \langle R, F \rangle$ là 2NF nếu mọi thuộc tính thứ cấp phụ thuộc hoàn toàn vào khóa. Nói cách khác W là 2NF nếu không có tập con thực sự của khóa kéo theo thuộc tính thứ cấp.

b. Dạng chuẩn 3 (3NF)

Sơ đồ quan hệ $W = \langle R, F \rangle$ là 3NF nếu trong W không có hiện tượng một tập con X của R , mà X có bao đóng khác R , kéo theo thuộc tính thứ cấp, tức là trong W (hoặc chính xác hơn trong F) không tồn tại phụ thuộc hàm dạng: $X \rightarrow x \notin X$ với $X^+ \neq R$ và x là thuộc tính thứ cấp.

c. Dạng chuẩn Boyce-Codd (BCNF)

Sơ đồ quan hệ $W = \langle R, F \rangle$ là BCNF nếu trong W không tồn tại phụ thuộc hàm dạng: $X \rightarrow x$ với $x \notin X$ và $X^+ \neq R$.

5. MỘT SỐ TÍNH CHẤT CỦA CÁC THUỘC TÍNH KHÓA

Sau đây chúng ta sẽ trình bày một số các tính chất của các thuộc tính khóa, trên cơ sở đó cho phép chúng ta trong một số bài toán cụ thể có thể tìm khóa một cách đơn giản và khẳng định được mệnh đề *đã tìm hết khóa của W* .

Bổ đề 1. Cho SDQH $W = \langle R, F \rangle$ với R là tập các thuộc tính, F là tập phụ thuộc hàm, nếu K là một khóa bất kỳ của W thì K phải chứa tất cả các thuộc tính của R mà chúng không xuất hiện trong vé trái cũng như vé phải của tập F .

Điều kết luận của bổ đề là hiển nhiên theo định nghĩa của khóa, nếu K không chứa các thuộc tính không xuất hiện trong tập F thì K^+ không thể bằng R .

Bài toán 1. Cho $W = \langle R, F \rangle$, với $R = \{A, B, C, D, E, G, H\}$,
 $F = \{PA \rightarrow B, C \rightarrow DE, BD \rightarrow E, AC \rightarrow EG, B \rightarrow G\}$.
 Chứng minh rằng mọi khóa của W đều chứa H .

Theo Bổ đề 1 khi đó ta có ngay mọi khóa của W phải chứa thuộc tính H hay nói cách khác $K = \{H, \dots\}$, K là khóa bất kỳ.

Bổ đề 2. Cho $SDQH W = \langle R, F \rangle$ với R là tập các thuộc tính, F là tập phụ thuộc hàm, nếu K là một khóa của W thì K phải chứa tất cả các thuộc tính của R mà các thuộc tính đó chỉ xuất hiện ở vế trái của tập phụ thuộc hàm F .

Kết luận của bổ đề này cũng hoàn toàn suy từ định nghĩa của khóa.

Bài toán 2. Cho $W = \langle R, F \rangle$ với R và F như trong Bài toán 1. Chứng minh rằng mọi khóa của W đều chứa A, C .

Theo Bổ đề 2 ta có ngay kết luận của bài toán.

Kết hợp cả hai Bổ đề 1 và 2 ta có mọi khóa của W đều chứa A, C, H , hay nói cách khác nếu K là khóa của W thì $K = \{A, C, H, \dots\}$.

Bài toán 3. Chứng minh rằng $SDQH$ trong Bài toán 1 (Bài toán 2) có duy nhất một khóa.

Thật vậy theo Bổ đề 1 và Bổ đề 2 thì mọi khóa của W đều chứa các thuộc tính $\{A, C, H\}$, thế nhưng ta lại thấy $K = \{A, C, H\}$ có bao đóng bằng R . Vậy K là khóa của W , do tính chất tất cả các khóa đều chứa K nên W chỉ có một khóa duy nhất.

Bổ đề 3. Cho $W = \langle R, F \rangle$ là một $SDQH$, với R là tập thuộc tính, F là tập phụ thuộc hàm, nếu thuộc tính A của R chỉ xuất hiện trong vế phải của tập F thì A phải thuộc tập thứ cấp F_n của W .

Kết luận này cũng hiển nhiên suy ra từ định nghĩa của khóa.

Cũng xét $SDQH$ trong Bài toán 1 ta thấy ngay rằng E là thuộc tính thứ cấp hay nói cách khác $F_n = \{E, \dots\}$.

Bài toán 4. Cho $SDQH W = \langle R, F \rangle$ với $R = \{A, B, C, D, E, G, H, I, L\}$ và $F = \{A \rightarrow BC, DE \rightarrow GH, G \rightarrow H, H \rightarrow I\}$. Xét xem W thuộc dạng chuẩn nào? (định nghĩa các dạng chuẩn 2NF, 3NF, 4NF, BCNF, ... xem trong [3], [6], [7], ...).

Theo các Bổ đề 1, 2 thì mọi khóa K của W phải chứa A, L . Hơn nữa mọi khóa của W đều chứa A thì các thuộc tính B và C phải thuộc tập thứ cấp F_n (vì $A \rightarrow BC$). Từ đây ta dễ dàng thấy rằng W không là 2NF vì có tập con thực sự của khóa kéo theo các phần tử thứ cấp ($A \rightarrow BC$, A là tập con thực sự của khóa và B, C là các thuộc tính thứ cấp). Vậy W cũng không là 3NF, BCNF, 4NF, 5NF, DK/NF và W chỉ là 1NF.

Bài toán 5. Cho $SDQH W = \langle R, F \rangle$ với $R = \{A, B, C, D, E, G, H, I, K, L\}$ và tập phụ thuộc hàm, $F = \{A \rightarrow H, DE \rightarrow B, AC \rightarrow DE, D \rightarrow I, IA \rightarrow BC, AH \rightarrow BC, GA \rightarrow BH, GEH \rightarrow BD\}$. Hãy xét xem W có là 3NF không?

Theo các Bổ đề 1, 2 thì mọi khóa của W phải chứa các thuộc tính A, G, H, L . Theo Bổ đề 3 ta có B là thuộc tính thứ cấp vì có $GA \rightarrow BH$. Vậy ta dễ dàng suy ra rằng W không là 3NF vì trong W có tập thuộc tính $\{D, E\}$ với bao đóng của nó khác R nhưng nó lại kéo theo thuộc tính thứ cấp B .

Như vậy các Bổ đề 1, 2, 3 đã giúp chúng ta xác định nhanh các sơ đồ quan hệ thuộc dạng chuẩn nào. Khi xét một $SDQH W = \langle R, F \rangle$ thuộc dạng chuẩn nào trước tiên chúng ta phải xem các sơ đồ

quan hệ có thỏa mãn các bổ đề trên không? Nói chung trong rất nhiều trường hợp các sơ đồ quan hệ sẽ thỏa mãn một trong các bổ đề trên.

6. KẾT LUẬN

Trên đây chúng ta đã xét một số tính chất nội tại của các thuộc tính khóa và thuộc tính không khóa của một sơ đồ quan hệ. Thực chất nội dung của bài viết này chúng tôi muốn nêu một số ý sau đây:

1. Làm sáng tỏ thêm các khái niệm nội tại của các khóa.
2. Xét một số tính chất quan trọng của các thuộc tính khóa, không khóa.
3. Thông qua các bổ đề chúng ta biết được các tính chất của khóa và đồng thời cho ta một số phương pháp tìm khóa nhanh.
4. Qua các bổ đề chúng ta có được trong rất nhiều trường hợp khẳng định *đã tìm hết khóa*.
5. Qua các bổ đề và các bài toán minh họa, chúng ta thấy bài viết thực sự có giá trị trong các bài toán về xác định dạng chuẩn của một sơ đồ quan hệ.

TÀI LIỆU THAM KHẢO

- [1] Codd E. F., A relational model of data for large relational database, *Communication of the ACM* **13** (6) (1970).
- [2] Date C. J., *Introduction to Database System*, 3rd ed., Reading, Mass. Addison-Wesley, 1981.
- [3] Fred R. Mcfadden and Jeffrey A. Hoffer, *Modern Database Management*, The Benjamin/Cummings Publishing Company, INC, 1994.
- [4] Lê Văn Bào, Hồ Thuần, Hồ Cẩm Hà, Structure of multivalued dependencies in relation scheme, *Journal of Computer Science and Cybernetics* **1** (1998).
- [5] Nguyễn Bá Tường, *Lý thuyết Cơ sở Dữ liệu*, Giáo trình in tại HVKTQS, 1999.
- [6] Ullman J. D., *Principles of Database and Knowledges*, Vol. 1, 2, Computer Science Press, Rocklives, MD, 1988.
- [7] Vũ Đức Thi, *Cơ sở dữ liệu - Kiến thức và Thực hành*, NXB Thống kê, 1997.

Nhận bài ngày 18 - 2 - 2000

Nhận lại sau khi sửa ngày 25 - 8 - 2000

Khoa Công nghệ thông tin
Học viện Kỹ thuật quân sự