

BÀI GIẢNG CƠ SỞ DỮ LIỆU

9. Các thuật toán thiết kế lược đồ CSDL quan hệ

Nguyễn Hải Châu

Khoa Công nghệ Thông tin
Trường Đại học Công nghệ, ĐHQGHN

Tách lược đồ quan hệ & tính không đầy đủ của dạng chuẩn

- Chuẩn hóa: Tách lược đồ quan hệ $R = \{A_1, A_2, \dots, A_n\}$ thành tập các lược đồ quan hệ $D = \{R_1, R_2, \dots, R_m\}$. D được gọi là một phép tách của R
- Chúng ta phải đảm bảo rằng mỗi thuộc tính A_i của R phải xuất hiện ít nhất trong một lược đồ quan hệ R_j
- Điều kiện bảo toàn thuộc tính: $\forall i = \overline{1, n}, \exists j$ sao cho A_i là một thuộc tính của R_j

Tách lược đồ quan hệ & tính không đầy đủ của dạng chuẩn

- Xét hai lược đồ quan hệ:
 - $NV_DD(\underline{Ten}, \underline{DDDA})$
 - $NV_DA1(\underline{MaNV}, \underline{MaDA}, Sogio, TenDA, DDDA)$
- Mặc dù NV_DD là lược đồ quan hệ ở dạng chuẩn BCNF, nhưng khi nối tự nhiên NV_DD với NV_DA1 , chúng ta vẫn thu được các bộ giả: Ngữ nghĩa của NV_DD không rõ ràng

Phép tách và sự bảo toàn phụ thuộc

- Cho trước tập phụ thuộc hàm $\mathcal{F}$ trên R và phép tách $D = \{R_1, R_2, \dots, R_m\}$
- Phép chiếu của $\mathcal{F}$ trên $R_i, i = \overline{1, m}$, ký hiệu là $\pi_{R_i}(\mathcal{F})$, là một tập các phụ thuộc hàm $X \rightarrow Y$ trong $\mathcal{F}^+$ sao cho các thuộc tính trong $X \cup Y$ đều có trong R_i
- Ta nói rằng phép tách D bảo toàn phụ thuộc đối với $\mathcal{F}$ nếu:

$$\{\pi_{R_1}(\mathcal{F}) \cup \pi_{R_2}(\mathcal{F}) \cup \dots \cup \pi_{R_m}(\mathcal{F})\}^+ = \mathcal{F}^+ \quad (1)$$

Định lý

Luôn luôn tìm được một phép tách bảo toàn phụ thuộc D đối với tập phụ thuộc hàm $\mathcal{F}$ sao cho mỗi quan hệ $R_i \in D$ là 3NF.

Thuật toán tìm một phép tách bảo toàn phụ thuộc

Thuật toán 1: Tìm một phép tách bảo toàn phụ thuộc

Vào: Lược đồ quan hệ R , tập phụ thuộc hàm $\mathcal{F}$

Ra: Phép tách D bảo toàn phụ thuộc đối với $\mathcal{F}$

- 1 Tìm phủ tối thiểu $\mathcal{G}$ của $\mathcal{F}$ (thuật toán 3);
 - 2 **for** mỗi vế trái X của các phụ thuộc hàm có dạng $X \rightarrow A_i, i = \overline{1, k}$ **do**
 - 3 | Tạo một lược đồ trong D với các thuộc tính $\{X, A_1, A_2, \dots, A_k\}$;
 - 4 **end**
 - 5 Đặt các thuộc tính còn lại (những thuộc tính chưa được đặt vào quan hệ nào) vào một quan hệ đơn để đảm bảo tính chất bảo toàn thuộc tính;
-

Ví dụ tìm phép tách bảo toàn phụ thuộc

- Lược đồ quan hệ $R(\underline{A}, B, C, D)$ có tập phụ thuộc hàm $\mathcal{F} = \{A \rightarrow BCD, BC \rightarrow DA, D \rightarrow B\}$. Lược đồ này có hai khóa dự tuyển là A và BC .
- Áp dụng thuật toán 3 (tìm phủ tối thiểu của tập phụ thuộc hàm), chúng ta tìm được phủ tối thiểu của $\mathcal{F}$ là:
 $\mathcal{G} = \{A \rightarrow C, A \rightarrow D, BC \rightarrow D, BC \rightarrow A, D \rightarrow B\}$
- Áp dụng thuật toán 4, chúng ta có phép tách bảo toàn phụ thuộc:
 $D = \{R_1(\underline{A}, C, D), R_2(\underline{B}, \underline{C}, D, A), R_3(\underline{D}, B)\}$
- R_1, R_2, R_3 đều là 3NF

Phép tách và tính chất kết nối không mất mát

- Cho trước lược đồ quan hệ $R(A_1, A_2, \dots, A_n)$, tập phụ thuộc hàm $\mathcal{F}$ trên R và phép tách $D = \{R_1, R_2, \dots, R_m\}$
- D có tính chất nối không mất mát (không phụ thêm) đối với $\mathcal{F}$ nếu với mỗi trạng thái quan hệ $r(R)$ thỏa mãn $\mathcal{F}$ thì

$$\pi_{R_1}(r) * \pi_{R_2}(r) * \dots * \pi_{R_m}(r) = r \quad (2)$$

trong đó $*$ là phép nối tự nhiên

- Nghĩa của từ mất mát ở đây là mất mát thông tin chứ không phải mất các bộ giá trị. Vì vậy, với tính chất này ta nên gọi chính xác hơn là tính chất nối không phụ thêm

Thuật toán kiểm tra tính chất nổi không mất mát

Thuật toán 2: Thuật toán kiểm tra tính chất nổi không mất mát

Vào: Lược đồ $R(A_1, A_2, \dots, A_n)$, tập phụ thuộc hàm $\mathcal{F}$, phép tách $D = \{R_1, R_2, \dots, R_m\}$

Ra: *True* nếu D có tính chất nổi không mất mát, *False* nếu ngược lại

- 1 Tạo một ma trận S có m hàng, n cột;
- 2 Đặt $S(i, j) = a_j$ nếu A_j thuộc về R_i và $S(i, j) = b_{ij}$ nếu ngược lại, trong đó a_j, b_{ij} là các ký hiệu khác nhau;
- 3 Lặp lại từ dòng 4 đến 14 cho đến khi S không thay đổi:
- 4 **for** mọi phụ thuộc hàm $X \rightarrow Y$ trong $\mathcal{F}$ **do**
 - 5 **for** mọi hàng h trong S có cùng ký hiệu ở các cột c ứng với các thuộc tính trong X **do**
 - 6 Gọi d là các cột của S ứng với các thuộc tính trong Y ;
 - 7 **if** tồn tại hàng h chứa ký hiệu a_k ở các cột d **then**
 - 8 | Gán giá trị a_k cho các cột d của các hàng h khác;
 - 9 **end**
 - 10 **else**
 - 11 | Chọn b_{ij} ở các cột d của một hàng h và gán cho cột d ở các hàng h khác;
 - 12 **end**
 - 13 **end**
 - 14 **end**
 - 15 Nếu có một hàng của S chỉ chứa các giá trị a return *True*, ngược lại return *False*;

Ví dụ kiểm tra nổi không mất mát 1

Ta có lược đồ quan hệ $R(\text{MaNV}, \text{TenNV}, \text{MaDA}, \text{TenDA}, \text{DDDA}, \text{Sogio})$,
phép tách $D = \{R_1(\text{TenNV}, \text{DDDA}),$
 $R_2(\text{MaNV}, \text{MaDA}, \text{Sogio}, \text{TenDA}, \text{DDDA})\}$ và tập phụ thuộc hàm
 $\mathcal{F} = \{\text{MaNV} \rightarrow \text{TenNV}, \text{MaDA} \rightarrow \{\text{TenDA}, \text{DDDA}\},$
 $\{\text{MaNV}, \text{MaDA}\} \rightarrow \text{Sogio}\}$

- Ma trận khởi tạo:

	MaNV	TenNV	MaDA	TenDA	DDDA	Sogio
R_1	b_{11}	a_2	b_{13}	b_{14}	a_5	b_{16}
R_2	a_1	b_{22}	a_3	a_4	a_5	a_6

- Thuật toán không làm thay đổi ma trận nên dừng ngay từ lần lặp đầu tiên
- Không có hàng nào của ma trận chứa toàn ký tự a , do đó phép tách là mất mát

Ví dụ kiểm tra tính chất nổi không mất mát 2

Ta có lược đồ quan hệ $R(MaNV, TenNV, MaDA, TenDA, DDDA, Sogio)$, phép tách $D = \{R_1(MaNV, TenNV), R_2(MaDA, TenDA, DDDA), R_3(MaNV, MaDA, Sogio)\}$ và tập phụ thuộc hàm $\mathcal{F} = \{MaNV \rightarrow TenNV, MaDA \rightarrow \{TenDA, DDDA\}, \{MaNV, MaDA\} \rightarrow Sogio\}$

- Ma trộn khởi tạo:

	MaNV	TenNV	MaDA	TenDA	DDDA	Sogio
R_1	a_1	a_2	b_{13}	b_{14}	b_{15}	b_{16}
R_2	b_{21}	b_{22}	a_3	a_4	a_5	b_{26}
R_3	a_1	b_{32}	a_3	b_{34}	b_{35}	a_6

Ví dụ kiểm tra tính chất nổi không mất mát 2

- Ma trận sau khi áp dụng lần lặp 1 (với phụ thuộc hàm $MaNV \rightarrow TenNV$)

	MaNV	TenNV	MaDA	TenDA	DDDA	Sogio
R_1	a_1	a_2	b_{13}	b_{14}	b_{15}	b_{16}
R_2	b_{21}	b_{22}	a_3	a_4	a_5	b_{26}
R_3	a_1	b_{32} a_2	a_3	b_{34}	b_{35}	a_6

- Ma trận sau khi áp dụng lần lặp 2 (với phụ thuộc hàm $MaDA \rightarrow \{TenDA, DDDA\}$)

	MaNV	TenNV	MaDA	TenDA	DDDA	Sogio
R_1	a_1	a_2	b_{13}	b_{14}	b_{15}	b_{16}
R_2	b_{21}	b_{22}	a_3	a_4	a_5	b_{26}
R_3	a_1	a_2	a_3	b_{34} a_4	b_{35} a_5	a_6

- Sau lần lặp 2, hàng 3 của ma trận chứa toàn các giá trị a , do đó phép tách là không mất mát

Tính chất của các phép tách thỏa mãn nối không mất mát

Tính chất

Một phép tách $D = \{R_1, R_2\}$ của lược đồ quan hệ R có tính chất nối không mất mát đối với một tập phụ thuộc hàm $\mathcal{F}$ trên R khi và chỉ khi

- Hoặc $(R_1 \cap R_2) \rightarrow (R_1 - R_2) \in \mathcal{F}^+$
- Hoặc $(R_1 \cap R_2) \rightarrow (R_2 - R_1) \in \mathcal{F}^+$

Tính chất

Nếu một phép tách $D = \{R_1, R_2, \dots, R_m\}$ của lược đồ quan hệ R có tính chất nối không mất mát đối với một tập phụ thuộc hàm $\mathcal{F}$ trên R và nếu một phép tách $D_1 = \{Q_1, Q_2, \dots, Q_k\}$ của R_i có tính chất nối không mất mát đối với $\pi_{R_i}(\mathcal{F})$ thì phép tách

$D_2 = \{R_1, R_2, \dots, R_{i-1}, Q_1, Q_2, \dots, Q_k, R_{i+1}, \dots, R_m\}$ của R có tính chất nối không mất mát đối với $\mathcal{F}$

Tách lược đồ thành BCNF với tính chất nổi không mất mát

Thuật toán 3: Tách quan hệ thành các quan hệ BCNF với tính chất nổi không mất mát

Vào: Lược đồ quan hệ R và tập phụ thuộc hàm $\mathcal{F}$

Ra: Phép tách D chứa các lược đồ BCNF với tính chất nổi không mất mát

- 1 Đặt $D = R$;
- 2 **for** mỗi lược đồ quan hệ Q không là BCNF trong D **do**
- 3 Tìm phụ thuộc hàm $X \rightarrow Y$ trong Q vi phạm BCNF;
- 4 Thay thế Q bởi hai lược đồ quan hệ mới là $Q - Y$ và $X \cup Y$;
- 5 **end**

Để thực hiện bước 3 chúng ta sử dụng định nghĩa BCNF tổng quát: Với mỗi phụ thuộc hàm $X \rightarrow Y$, nếu X không là siêu khóa và X^+ không chứa tất cả các thuộc tính trong Q thì $X \rightarrow Y$ vi phạm BCNF.

Ví dụ

Xét lược đồ quan hệ $R(A, B, C, D, E, F)$ có tập phụ thuộc hàm $\mathcal{F} = \{A \rightarrow BCDEF, BC \rightarrow ADEF, B \rightarrow F, D \rightarrow E, D \rightarrow B\}$: R có hai khóa dự tuyển là A và BC

- ❶ Ta có $B \rightarrow F$ vi phạm BCNF do B không là siêu khóa, do đó R được tách thành:
 - $R_1(B, F)$ với phụ thuộc hàm $B \rightarrow F$
 - $R_2(A, B, C, D, E)$ với các phụ thuộc hàm $A \rightarrow BCDE, BC \rightarrow ADE, D \rightarrow E, D \rightarrow B$
- ❷ Do $D \rightarrow E$ vi phạm BCNF (D không là siêu khóa), R_2 được tách thành:
 - $R_{21}(D, E)$ với phụ thuộc hàm $D \rightarrow E$
 - $R_{22}(A, B, C, D)$ với các phụ thuộc hàm $A \rightarrow BCD, BC \rightarrow AD, D \rightarrow B$
- ❸ Do $D \rightarrow B$ vi phạm BCNF (D không phải là siêu khóa) nên R_{22} được tách thành:
 - $R_{221}(D, B)$ với phụ thuộc hàm $D \rightarrow B$
 - $R_{222}(A, B, D)$ với phụ thuộc hàm $A \rightarrow BD$ ($BC \rightarrow AD$ bị mất)

Vậy ta có phép tách $D = \{R_1, R_{21}, R_{221}, R_{222}\}$: có tính chất **nổi không mất thông tin** nhưng **không bảo toàn phụ thuộc**

Thuật toán tổng hợp quan hệ

Thuật toán 4: Tổng hợp quan hệ với tính chất bảo toàn phụ thuộc và nối không mất mát

Vào: Lược đồ quan hệ R và tập phụ thuộc hàm $\mathcal{F}$

Ra: Phép tách D chứa các lược đồ 3NF với tính chất nối không mất mát và bảo toàn phụ thuộc

- 1 Tìm phủ tối thiểu $\mathcal{G}$ của $\mathcal{F}$;
- 2 **for** mỗi vế trái của một phụ thuộc hàm trong $\mathcal{G}$ **do**
- 3 Tạo một lược đồ quan hệ $S(X, A_1, \dots, A_k)$ trong D với
 $X \rightarrow A_i, i = \overline{1, k}$ là các phụ thuộc hàm trong $\mathcal{G}$;
- 4 **end**
- 5 **if** không có lược đồ quan hệ nào trong D chứa một khóa của R **then**
- 6 Tìm một khóa K của R ;
- 7 Tạo thêm một lược đồ quan hệ T trong D chứa tất cả các thuộc tính của K ;
- 8 **end**

Ví dụ tổng hợp quan hệ: thiết kế 1

Ta có lược đồ quan hệ $R(P, C, L, A)$ và tập phụ thuộc hàm $\mathcal{F} = \{P \rightarrow LCA, LC \rightarrow AP, A \rightarrow C\}$

- Phủ tối thiểu của $\mathcal{F}$: $\mathcal{G} = \{P \rightarrow LC, LC \rightarrow AP, A \rightarrow C\}$
- Cách tổng hợp 1 (trước khi xóa đi các lược đồ quan hệ dư thừa):
 $R_1(\underline{P}, L, C), R_2(\underline{L}, \underline{C}, A, P), R_3(\underline{A}, C)$
- R_1 và R_3 là dư thừa
- Do đó kết quả tổng hợp là $R_2(\underline{L}, \underline{C}, A, P)$

Ví dụ tổng hợp quan hệ: thiết kế 2

Ta vẫn xét lược đồ quan hệ $R(P, C, L, A)$ và tập phụ thuộc hàm $\mathcal{F} = \{P \rightarrow LCA, LC \rightarrow AP, A \rightarrow C\}$

- Phủ tối thiểu của $\mathcal{F}$: $\mathcal{G} = \{P \rightarrow LA, LC \rightarrow P, A \rightarrow C\}$
- Cách tổng hợp 2 (trước khi xóa đi các lược đồ quan hệ dư thừa):
 $S_1(\underline{P}, A, L), S_2(\underline{L}, \underline{C}, P), S_3(\underline{A}, C)$
- Kết quả tổng hợp có 3 lược đồ quan hệ ở 3NF

- Không phải lúc nào cũng có thể tìm được một phép tách thành các lược đồ quan hệ bảo toàn phụ thuộc và mỗi lược đồ trong phép tách là ở BCNF
- Thuật toán tổng hợp quan hệ tìm được phép tách thành các lược đồ quan hệ 3NF có tính chất bảo toàn phụ thuộc và nối không mất mát
- Do đó nên thiết kế các lược đồ quan hệ ở 3NF nhằm đạt được cả hai tính chất quan trọng:
 - Bảo toàn phụ thuộc
 - Nối không mất mát