

5. BÀI TẬP VỀ chuẩn HOÁ

MỤC TIÊU CỦA BÀI NÀY GIÚP NGƯỜI HỌC

- Phân biệt các dạng chuẩn của quan hệ.
- Xác định một lược đồ ở dạng chuẩn nào.
- Vận dụng giải các bài tập về chuẩn hóa quan hệ (Đưa các lược đồ quan hệ (quan hệ) từ dạng chuẩn thấp lên dạng chuẩn cao hơn).
- Kiểm tra được một phép tách lược đồ aqua nhẹ có mất thông tin không.

A/ NHẮC LẠI LÝ THUYẾT

I. CÁC ĐỊNH NGHĨA, TÍNH CHẤT

1. Dạng chuẩn 1 (1NF - first normal form)

Một lược đồ quan hệ $\alpha = (U, F)$ được gọi là ở dạng chuẩn một (1NF) nếu và chỉ nếu tất cả miền giá trị của các thuộc tính của R đều nguyên tố (không thể phân chia được).

Chú ý:

Tính không thể phân chia được chỉ có tính chất tương đối.

Định nghĩa này cho thấy ngay rằng bất kỳ quan hệ chuẩn hóa nào cũng ở 1NF.

2. Dạng chuẩn 2 (2NF- Second normal form)

Trước khi nghiên cứu dạng chuẩn thứ 2, ta xét Ví dụ sau đây:

Xét CSDL gồm 2 lược đồ quan hệ THI(MONTHI,GIAOVIEN) và SINHVIEN(MONTHI, MSSV, TEN, TUOI, DCHI, DIEM) phản ánh thông tin về kết quả thi của một đơn vị nào đó.

Trong quan hệ THI thì MONTHI là khóa và trong quan hệ SINHVIEN thì MONTHI và MSSV là khóa.

ở quan hệ thứ hai dễ nhận thấy rằng MONTHI, MSSV,DIEM xác định kết qu thi của sinh viên còn MSSV,TEN, TUOI, DCHI xác định đối tượng dự thi

Xét các hiện hành của 2 lược đồ quan hệ THI và SINHVIEN như sau:

THI	
MONTHI	GIAOVIEN
Toán	Thầy Công
Lý	Thầy Hứa
Hóa	Thầy Giao

SINHVIEN					
MONTHI	MSSV	TEN	TUOI	DCHI	DIEM
Toán	11	Lan	20	HN	8.0
Toán	12	Hue	21	HY	7.5
Hóa	11	Lan	20	HN	7.0
Hóa	12	Hue	21	HY	6.0
Lý	11	Lan	20	HN	5.0
Lý	13	An	22	BN	4.0

3. Dạng chuẩn 3 (3NF- Second normal form)

Định nghĩa: Cho lược đồ quan hệ $\alpha = (U, F)$, lược đồ α được gọi là ở dạng chuẩn 3, kí hiệu là 3NF, nếu như lược đồ ở dạng chuẩn 1NF và các thuộc tính không khoá của α là không phụ thuộc hàm bậc cầu vào khoá chính.

4. Dạng chuẩn Boyce Codd (BCNF- Boyce Codd normal form)

Định nghĩa: Cho lược đồ quan hệ $\alpha = (U, F)$, lược đồ α được gọi là ở dạng chuẩn Boyce Codd, kí hiệu là BCNF, nếu như lược đồ ở dạng chuẩn 1NF và nếu $X \rightarrow Y \in F^+$ ($Y \not\subset X$) thì X phải là siêu khoá của lược đồ.

5. Tách lược đồ quan hệ

Định nghĩa: Phép tách lược đồ quan hệ $\alpha = (U, F)$ là phép thay thế nó bằng một tập các lược đồ con $\alpha_i = (U_i, F_i)$, $i=1, \dots, k$ với điều kiện

$U_i \neq \emptyset \forall i=1, \dots, k, \cup U_i = U, F_i = F/U_i$, F_i là hình chiếu của F lên tập thuộc tính U_i

Phép tách đó được ký hiệu là $\sigma = \{U_1, U_2, \dots, U_k\}$

Kí hiệu $\alpha = (U, F)$, $\sigma = \{U_1, U_2, \dots, U_k\}$ là một phép tách khi đó R là một quan hệ trên U, kí hiệu $m_\sigma(R) = R[U_1] * R[U_2] * \dots * R[U_k]$

Định nghĩa: phép tách kết nối không mất thông tin

Cho lược đồ quan hệ $\alpha = (U, F)$ và phép tách $\delta = \{U_1, U_2, \dots, U_k\}$ đối với lược đồ đó. phép tách δ được gọi là phép tách kết nối không mất thông tin nếu mọi quan hệ R trên U thì ta có $m_\delta(R) = R$, ngược lại nếu $m_\delta(R) \neq R$ thì ta nói phép tách δ là phép tách mất thông tin.

6. Thuật toán kiểm tra phép tách kết nối có mất thông tin hay không?

Dữ liệu vào:

- Tập thuộc tính U
- Tập phụ thuộc hàm F
- Phép tách $\delta = \{U_1, U_2, \dots, U_k\}$

Ra:

Xác định liệu phép tách δ có mất thông tin hay không?

Phương pháp:

Giả sử $U = \{A_1, A_2, \dots, A_n\}$, ta xây dựng một bảng gồm k dòng n cột ($n = |U|$, $k = |\delta|$), cột thứ i của bảng ứng với thuộc tính A_i , hàng thứ j của bảng ứng với lược đồ con $\alpha_i = (U_i, F_i)$, tại hàng i và cột j ta điền kí hiệu a_j (ta gọi kí hiệu a_j là tín hiệu chính) nếu thuộc tính $a_j \in U_i$, nếu không ta điền b_{ij} (ta gọi b_{ij} là tín hiệu phụ).

Bây giờ ta biến đổi bảng như sau:

Với mỗi phụ thuộc hàm $X \rightarrow Y \in F$, nếu trong bng có hai hàng giống nhau trên tập thuộc tính X thì ta cần làm chúng giống nhau trên tập thuộc tính Y theo quy tắc sau:

- Nếu một trong hai giá trị là tín hiệu phụ thì ta sửa lại tín hiệu phụ thành tín hiệu chính tức là sửa b_{ij} thành a_j
- Nếu cả hai là tín hiệu phụ thì ta sửa lại hai tín hiệu đó bằng một trong các kí hiệu b_{ij} , tức là sửa lại chỉ số cho giống nhau.

Tiếp tục áp dụng các phụ thuộc hàm trong bảng (kể c các phụ thuộc hàm đã được sử dụng) cho tới khi không còn áp dụng được nữa.

Quan sát trong bảng cuối cùng: nếu xuất hiện ít nhất một hàng gồm toàn tín hiệu chính (hàng gồm toàn kí hiệu a) thì phép tách kết nối là không mất thông tin, trường hợp ngược lại là kết nối mất thông tin.

7. Phương pháp chuẩn hóa dữ liệu

7.1. Thuật toán tách lược đồ thành 3NF

Input: Lược đồ quan hệ $\alpha = (U, F)$

Output: Các lược đồ ở dạng 3NF

$(U_1, K_1), (U_2, K_2), \dots, (U_n, K_n)$ thỏa mãn:

a) Quan hệ R trên U thì $R[U_1] * R[U_2] * \dots * R[U_n] = R$

b) K_1, K_2, \dots, K_n là các khoá của lược đồ con tương ứng

Phương pháp:

1. Tìm một khóa K của lược đồ α

2. Tìm một phủ G tối thiểu của F

$G = \{K_1 \rightarrow A_1, K_2 \rightarrow A_2, \dots, K_p \rightarrow A_p\}$

3. Ghép các phụ thuộc hàm có cùng vế trái trong G để thu được phủ

$G = \{K_1 \rightarrow X_1, K_2 \rightarrow X_2, \dots, K_n \rightarrow X_n\}$

4. Phép tách $\delta = \{K_1 X_1, K_2 X_2, \dots, K_n X_n\}$ nếu khóa K không có mặt trong thành phần nào của δ thì thêm thành phần K vào δ .

5. Return δ

7.2. Tách không mất thông tin thành các lược đồ ở dạng BCNF

Cho lược đồ $\alpha = (U, F)$, và phép tách $\delta = \{U_1, U_2, \dots, U_k\}$, phép tách một lược đồ thành một tập các lược đồ ở dạng BCNF là phép tách thỏa mãn:

- Phép tách δ là phép tách kết nối không mất thông tin
- Tất cả các lược đồ con $\alpha_i = (U_i, F_i)$ đều ở dạng BCNF

Phương pháp :

Xuất phát từ một phụ thuộc hàm $X \rightarrow A$ nào đó của F, phụ thuộc hàm $X \rightarrow A$ này vi phạm điều kiện BCNF, ta xây dựng phép tách $\delta = \{U_1, U_2\}$, tương ứng với lược đồ α_1 và α_2 sao cho:

- Phép tách đó là phép tách kết nối không mất thông tin
- Phụ thuộc hàm $X \rightarrow A$ là phụ thuộc hàm của lược đồ α_1 và nó thỏa mãn điều kiện của BCNF trong lược đồ này
- Nếu như các lược đồ α_1 và α_2 vẫn chưa ở dạng BCNF thì tiếp tục quá trình đó, thì các điều vi phạm BCNF đều bị loại bỏ, cuối cùng ta thu được một tập các lược đồ con đều ở dạng BCNF và quá trình tách luôn luôn đảm bảo phép tách kết nối không mất thông tin.

Cơ sở của thuật toán trên là do giả thiết lược đồ $\alpha = (U, F)$ chưa ở dạng BCNF do đó tồn tại phụ thuộc hàm $X \rightarrow A, A \rightarrow X, X$ không phải là siêu khoá
 $U_1 = XA, U_2 = U \setminus A$

Nhận xét

$X = U_1 \cap U_2, U_1 \setminus U_2 = A$, đã có $X \rightarrow A$ do đó $U_1 \cap U_2 \rightarrow U_1 \setminus U_2$ theo định lý ở phần trên thì phép tách $\delta = \{U_1, U_2\}$ là phép tách có kết nối không mất thông tin. Vì $U_1 = XA$ và phụ thuộc hàm $X \rightarrow A$ là duy nhất trên lược đồ $\alpha_1 = (U_1, F_1)$ nên X là siêu khoá.

Nếu α_1, α_2 chưa ở dạng BCNF thì ta áp dụng quá trình tách tương tự. Cuối cùng ta thu được một tập các lược đồ ở dạng BCNF và quá trình tách là không mất thông tin.

Ví dụ:

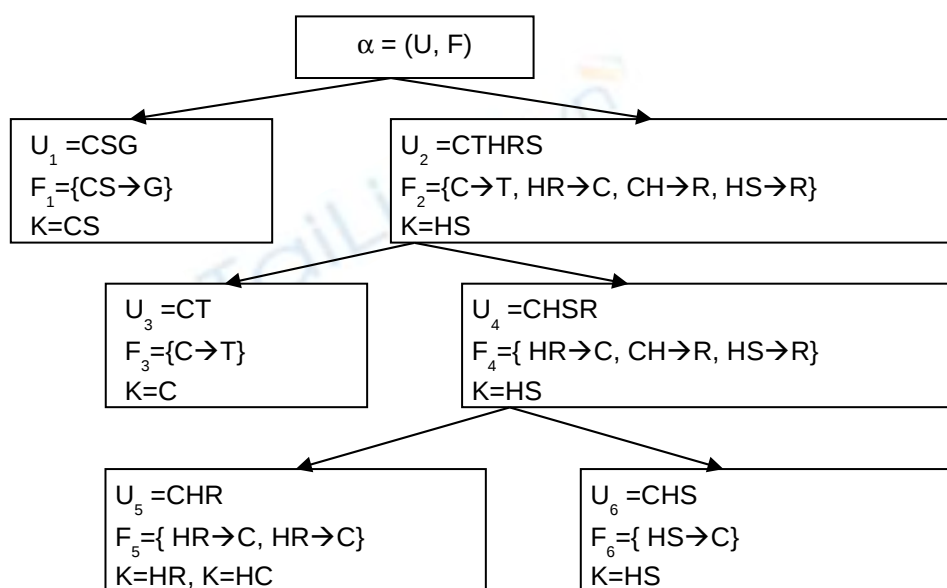
Cho lược đồ $\alpha = (U, F)$ với

$U = CRHTSG$ (C : Course, T : Teacher, H : Hour, R : Room, S : Student, G : Group)

$F = \{C \rightarrow T, HR \rightarrow C, CH \rightarrow R, CS \rightarrow G, HS \rightarrow R\}$

Nhận xét

- Lược đồ này có duy nhất một khoá là HS
- Lược đồ này chưa ở dạng BCNF
- Ta thấy trong lược đồ $\alpha = (U, F)$ có phụ thuộc hàm $CS \rightarrow G$ vi phạm điều kiện BCNF nên ta tách lược đồ thành các lược đồ $U_1 = CSG$, $U_2 = CTHRS$
- Ta thấy trong lược đồ $\alpha_2 = (U_2, F_2)$ có phụ thuộc hàm $C \rightarrow T$ vi phạm điều kiện BCNF nên ta tách lược đồ thành các lược đồ $U_3 = CT$, $U_4 = CHRS$
- Ta thấy trong lược đồ $\alpha_4 = (U_4, F_4)$ có phụ thuộc hàm $CH \rightarrow R$ vi phạm điều kiện BCNF nên ta tách lược đồ thành các lược đồ $U_5 = CHR$, $U_6 = CHS$



Như vậy phép tách cuối cùng là $\delta = \{ CSG, CT, CHR, CHS \}$

III. MỘT SỐ LƯU Ý

- Tầm quan trọng của việc chuẩn hóa dữ liệu.
- Phân biệt các dạng chuẩn, phương pháp tách quan hệ ở dạng chuẩn thấp lên dạng chuẩn cao hơn.
- Thuật toán kiểm tra phép tách có mất thông tin không?

B/ BÀI TẬP MẪU

Bài số 1: Kiểm tra phép tách có mất thông tin hay không?

Cho lược đồ quan hệ $\alpha = (U, F)$ với

$U = \{A_1, A_2, A_3, A_4, A_5\}$

$F = \{A_1 \rightarrow A_2 A_3, A_2 A_4 \rightarrow A_5, A_2 \rightarrow A_3\}$

$\delta = \{A_1 A_2 A_4, A_2 A_3, A_1 A_4 A_5\}$

Hỏi rằng phép tách δ trên có kết nối không mất thông tin không?

Hướng dẫn:

Áp dụng thuật toán kiểm tra phép tách có mất thông tin không, ta tiến hành từng bước.

Giải:

Xây dựng bảng gồm 3 dòng 5 cột

- Điền các tín hiệu vào bảng

	A₁	A₂	A₃	A₄	A₅
U₁	a ₁	a ₂	b ₁₃	a ₄	b ₁₅
U₂	b ₁₂	a ₂	a ₃	b ₂₄	b ₂₅
U₃	a ₁	b ₃₂	b ₃₃	a ₄	a ₅

- Biến đổi bảng trên dựa vào tập phụ thuộc hàm F

+ Sử dụng phụ thuộc hàm $A_1 \rightarrow A_2 A_3$ ta biến đổi bảng

	A₁	A₂	A₃	A₄	A₅
U₁	a ₁	a ₂	b ₁₃	a ₄	b ₁₅
U₂	b ₁₂	a ₂	a ₃	b ₂₄	b ₂₅
U₃	a ₁	a ₂	b ₁₃	a ₄	a ₅

+ Sử dụng phụ thuộc hàm $A_2 A_4 \rightarrow A_5$

	A₁	A₂	A₃	A₄	A₅
U₁	a ₁	a ₂	b ₁₃	a ₄	a ₅
U₂	b ₁₂	a ₂	a ₃	b ₂₄	b ₂₅
U₃	a ₁	a ₂	b ₁₃	b ₄	a ₅

+ Sử dụng phụ thuộc hàm $A_2 \rightarrow A_3$

	A₁	A₂	A₃	A₄	A₅
U₁	a ₁	a ₂	a ₃	a ₄	a ₅
U₂	B ₁₂	a ₂	a ₃	b ₂₄	b ₂₅
U₃	a ₁	a ₂	a ₃	a ₄	a ₅

Trong bảng này có hàng cuối cùng gồm toàn các tín hiệu chính, do vậy phép tách δ là phép tách kết nối không mất thông tin.

C/ BÀI TẬP TỰ GIẢI

Bài tập 1:

Dùng kỹ thuật bảng kiểm tra phép tách sau có mất thông tin không

- a) $\alpha=(U, F)$ với $U=ABCD$, $F=\{A \rightarrow B, AC \rightarrow D\}$, $\delta=\{AB, ACD\}$
- b) $\alpha=(U, F)$ với $U=ABCDE$, $F=\{A \rightarrow C, B \rightarrow C, C \rightarrow D, DE \rightarrow C, CE \rightarrow A\}$, $\delta=\{AD, AB, BE, CDE\}$
- c) Xác định và giải thích dạng chuẩn cao nhất của lược đồ quan hệ $\alpha=(U, F)$ với $U=ABCD$, $F=\{A \rightarrow C, D \rightarrow B, C \rightarrow ABD\}$

Bài tập 2:

Cho lược đồ quan hệ $\alpha=(U, F)$ với

$U=ABCDEFGH$

$F=\{CD \rightarrow H, E \rightarrow B, D \rightarrow G, BH \rightarrow E, CH \rightarrow DG, C \rightarrow A\}$

Hỏi rằng phép tách $\delta=(ABCDE, BCH, CDEGH)$ có kết nối mất thông tin không.

Bài tập 3:

Cho lược đồ quan hệ $\alpha=(U, F)$ với

$U=ABCD$, $F=\{D \rightarrow B, C \rightarrow A, B \rightarrow ACD\}$

Xác định dạng chuẩn cao nhất của lược đồ quan hệ trên

Bài tập 4:

Cho lược đồ quan hệ $\alpha=(U, F)$ với

$U=ABCD$, $F=\{CD \rightarrow B, A \rightarrow C, B \rightarrow ACD\}$

Xác định dạng chuẩn cao nhất của lược đồ quan hệ trên

Bài tập 5:

Cho $\alpha=(u, F)$ với

$U=ABCDE$ và

$F=\{A \rightarrow C, B \rightarrow C, A \rightarrow D, DE \rightarrow C, CE \rightarrow A\}$

kiểm tra tính kết nối không mất thông tin đối với phép tách

$\delta=\{AD, AB, BE, CDE, AE\}$

Bài tập 6:

Cho $\alpha=(u, F)$ với

$U=ABCDEF$ và

$F=\{AB \rightarrow C, C \rightarrow B, ABD \rightarrow E, F \rightarrow A\}$

kiểm tra tính kết nối không mất thông tin đối với phép tách

$\delta=\{BC, AC, ABDE, ABDF\}$

Bài tập 7:

Cho $\alpha=(u, F)$ với

$U=ABCDEG$

$F=\{D \rightarrow G, C \rightarrow A, CD \rightarrow E, A \rightarrow B\}$

kiểm tra tính kết nối không mất thông tin đối với phép tách

$\delta=\{DG, AC, SCE, AB\}$

Bài tập 8:

Cho $\alpha=(u, F)$ với

$U=ABCDE$ và

$F=\{A \rightarrow C, B \rightarrow C, C \rightarrow D, DE \rightarrow C, CE \rightarrow A\}$

kiểm tra tính kết nối không mất thông tin đối với phép tách

$\delta = \{AC, CD, BE, BC, AE\}$

Bài tập 9:

Cho (U, F) với

$U = XYZW$ và tập

$F = \{Y \rightarrow W, W \rightarrow Y, XY \rightarrow Z\}$

Dạng chuẩn cao nhất của lược đồ là gì?

Bài tập 10:

Cho (U, F) với

$U = ABCDEG$ và tập phụ thuộc hàm

$F = \{AB \rightarrow C, AC \rightarrow E, EG \rightarrow D, AB \rightarrow G\}$

$\delta = \{DEG, ABDEG\}$

Phép tách trên có mất thông tin không?

Hãy chứng minh mọi quan hệ chỉ có 2 thuộc tính để ở dạng chuẩn BCNF?

Bài tập 11:

Xét quan hệ $R(ABCDE)$ và tập phụ thuộc hàm

$F = \{AB \rightarrow CE, E \rightarrow AB, C \rightarrow D\}$

Hãy tìm dạng chuẩn cao nhất của lược đồ?

Bài tập 12:

Xét quan hệ $R(ABCDEFG)$ và tập phụ thuộc hàm

$F = \{A \rightarrow B, C \rightarrow DG, AC \rightarrow E, D \rightarrow G\}$

- Hãy tìm khoá của lược đồ
- Hãy tìm dạng chuẩn cao nhất của lược đồ

Bài tập 13:

Xét quan hệ $R(ABCD)$ và tập phụ thuộc hàm

$F = \{AB \rightarrow D, AC \rightarrow BD, B \rightarrow C\}$

Hãy tìm dạng chuẩn cao nhất của lược đồ

Bài tập 14:

Cho $\alpha = (u, F)$ với

$U = ABCDEF$

$F = \{AB \rightarrow C, C \rightarrow B, ABD \rightarrow E, F \rightarrow A\}$

Lược đồ có ở dạng BCNF không