

CSDL – BÀI TẬP LUYỆN TẬP

Phụ thuộc hàm và Dạng chuẩn

Phụ thuộc hàm: $X \rightarrow Y$

(X, Y: Hai tập thuộc tính trên cùng lược đồ quan hệ R

Tập thuộc tính X xác định tập thuộc tính Y trên R

Tập Y phụ thuộc vào X trên R

Một giá trị tập X cho ra một giá trị duy nhất trên tập Y)

Ví dụ trên lược đồ QLDT: GiaoVien có các PTH:

mãgv \rightarrow họten, gvqlcm

gvqlcm \rightarrow mãgv: có phải PTH, nếu có quy tắc nghiệp vụ là mỗi người chỉ quản lý 1 người khác

Khoá: Siêu khoá (superkey), Khoá (ứng viên, key/candidate key), Khoá chính (primary key)

- Siêu khoá: tập thuộc tính bất kỳ xác định duy nhất 1 dòng dữ liệu (xác định tất cả thuộc tính của quan hệ),

$R(ABCDE)$ cho f: $AB \rightarrow ABCDE = R$ thì AB là một siêu khoá

- Khoá ứng viên (khoá): là siêu khoá nhỏ nhất

$R(ABCDE), F = \{f_1: AB \rightarrow CD, f_2: B \rightarrow E, f_3: C \rightarrow D\}$

SK: ABCDE, ABC ($\rightarrow ABCED = R$), ABE, ...

K (sẽ chứa AB): $AB \rightarrow ABCDE = R$: AB là tập thuộc tính nhỏ nhất suy ra toàn bộ

R (AB là siêu khoá nhỏ nhất \rightarrow AB là khoá ứng viên của quan hệ)

- Cách tìm khoá ứng viên:

- B1: Tìm tập thuộc tính chỉ nằm bên vế trái ($LHS = AB$) hoặc không xuất hiện trong tập PTH (\rightarrow Bắt buộc phải có trong khoá vì không có thuộc tính nào xác định được LHS ngoại trừ chính nó)
- B2: Tìm tập thuộc tính chỉ nằm bên vế phải ($RHS = DE$) (\rightarrow Hiển nhiên không phải là thuộc tính khoá vì không xác định được thuộc tính nào cả)
- B3: Kiểm tra thử LHS đủ là khoá chưa? OK: LHS là khoá duy nhất.
- B4: Tìm tập thuộc tính trung gian ($I = C$) là tập vừa nằm bên trái vừa nằm bên phải của bất kỳ PTH nào trong tập F.

Thuộc tính Khoá và không khoá:

- Thuộc tính khoá: thuộc ít nhất 1 khoá ứng viên
- Thuộc tính không khoá: không thuộc bất kỳ khoá ứng viên nào cả
- VD1: R(ABCDE), $F = \{f1: AB \rightarrow CD, f2: B \rightarrow E, f3: C \rightarrow D\}$
 - TT khoá: A, B
 - TT không khoá: CDE
- VD2: R(ABCDEF), $K = \{AB, BC, BD\}$
 - TT khoá: A, B, C, D
 - TT không khoá: E, F

Dạng chuẩn: các tiêu chí để xác định mức độ chuẩn hoá (mức độ tổ chức dữ liệu) của một lược đồ CSDL, mức cao: ít trùng lặp thông tin; mức thấp: trùng lặp nhiều (→ các bất thường trong việc thêm/xoá/sửa).

Mức độ dạng chuẩn	Tiêu chí	Vi phạm <i>Các tập thuộc tính gây vi phạm chính là tập làm trùng lặp thông tin.</i>	Nguyên tắc chuẩn hoá
DC1 (1NF)	Không đa trị, không kết hợp.	Có ít nhất một hoặc một tập thuộc tính bị đa trị hoặc kết hợp.	Đa trị: <ul style="list-style-type: none">- C1: Đem tập trùng lặp đa trị tách ra, kéo theo khoá của quan hệ cũ.- C2: Đem các dòng dữ liệu vi phạm đa trị tách ra thành nhiều dòng. Kết hợp: <ul style="list-style-type: none">- Tách thành phần con thành cột khác nhau.

Mức độ dạng chuẩn	Tiêu chí	Vi phạm <i>Các tập thuộc tính gây vi phạm chính là tập làm trùng lặp thông tin.</i>	Nguyên tắc chuẩn hoá
DC2 (2NF)	Không tồn tại phụ thuộc riêng phần giữa thuộc tính không khoá và bất kỳ khoá ứng viên nào.	Có ít nhất một thuộc tính không khoá phụ thuộc một phần vào khoá hay: Tìm thấy ít nhất một $X \rightarrow Y$ trong đó: X là phần con của khoá và Y là không khoá.	- Tách tất cả các trường hợp $X \rightarrow Y$ vi phạm (tức là $X \rightarrow Y$ có X là con của khoá và Y là không khoá) ra (giữ lại X ở bảng cũ để kết bảng).
DC3 (3NF)	Không tồn tại phụ thuộc bắc cầu giữa không khoá và bất kỳ khoá ứng viên nào. ➔ Các tập thuộc tính không khoá suy ra lẫn nhau.	Có ít nhất một thuộc tính không khoá phụ thuộc vào một hay một tập thuộc tính không khoá khác: Tìm thấy ít nhất một $X \rightarrow Y$ trong đó: X không chứa khoá và Y là không phải thuộc tính khoá. $R(\underline{AB}/\underline{CDE})$ $F = \{$ $f1: AB \rightarrow C;$ $f2: AB \rightarrow D;$ $f3: AB \rightarrow E;$ $f4: C \rightarrow A;$ $f5: C \rightarrow B;$ $f6: E \rightarrow A \}$ Đạt DC2 vì không có PTH có vế trái là con khoá và vế phải là thuộc tính không khoá. Đạt DC3 vì không có PTH vế trái không chứa khoá và vế phải là thuộc tính không khoá.	- Tách tất cả các trường hợp $X \rightarrow Y$ vi phạm ra (giữ lại X ở bảng cũ để kết bảng).

Mức độ dạng chuẩn	Tiêu chí	Vi phạm <i>Các tập thuộc tính gây vi phạm chính là tập làm trùng lặp thông tin.</i>	Nguyên tắc chuẩn hoá
DCBC (BCNF)	Mọi phụ thuộc hàm về trái phải là siêu khoá (chứa bất kỳ khoá ứng viên nào).	Tìm thấy ít nhất một $X \rightarrow Y$ trong đó: X không chứa khoá. Ví dụ trên vi phạm BC vì f6: $E \rightarrow A$ có E không phải là siêu khoá.	Đem $E \rightarrow A$ tách ra bảng mới dữ lại E để kết bảng và xác định lại tập PTH cho từng quan hệ mới.

1. Cho bảng dữ liệu SINHVIÊN(MãSV, ThôngTin, KếtQuả). Trong đó thuộc tính ThôngTin bao gồm {TênSV, Ngày Sinh, Email, (Điện thoại1, Điện thoại 2, ...)}, thuộc tính KếtQuả bao gồm {Môn học, Lần học, Điểm} lưu trữ môn học sinh viên đã tham gia học và điểm của môn học đó cho từng lần khác nhau.

- Giải thích tại sao danh sách SINHVIÊN không đạt dạng chuẩn 1. Hãy thiết kế lại để được quan hệ SINHVIÊN đạt dạng chuẩn 1 và vẽ lược đồ quan hệ.
- Từ kết quả câu a), chỉ ra các điểm trùng lặp thông tin? Quan hệ có đạt dạng chuẩn 2 hay 3 không? Giải thích. Hãy thiết kế lại để được bảng SINHVIÊN đạt dạng chuẩn 3 và vẽ lược đồ quan hệ.

2. Xét bảng NHÂNVIÊN sau, với TÊNNV là khóa chính:

TÊNNV	DỰÁN	PHÒNGBAN
Anh	Phân công việc	Phòng thí nghiệm mạng
Bình	Định tuyến mạng	Phòng thí nghiệm mạng
Duy	Thiết kế khảo sát	Phòng quản lý nhân sự
Hạ	Phân công việc	Phòng thí nghiệm mạng
Tâm	Thiết kế khảo sát	Phòng quản lý nhân sự
Minh	Định tuyến mạng	Phòng thí nghiệm mạng
Hùng	Dịch vụ phân tích tin tức	Phòng quản lý nhân sự

- Chỉ ra các điểm trùng lặp thông tin và giải thích vì sao NHÂNVIÊN không đạt DC3?
- Hãy chuẩn hóa NHÂNVIÊN thành dạng chuẩn 3.

3. Một hệ thống thông tin quản lý sinh viên bao gồm các quan hệ sau đây:

SINHVIÊN (MSSV, HỌTÊN, SỐCMND, ĐỊACHỈ, NGÀY SINH, TÊNNGÀNH)

F1 = {

f1: MSSV \rightarrow HỌTÊN, SỐCMND, ĐỊACHỈ, NGÀY SINH, TÊNNGÀNH;

f2: SỐCMND \rightarrow MSSV }

NGÀNH (TÊNNGÀNH, NĂM, SLTĐ, SLSV, MÃPHG, TÊNPHG, TÒANHÀ)

F2 = {

f3: TÊNNGÀNH, NĂM \rightarrow SLSVTỔĐA;

f4: TÊNNGÀNH \rightarrow SLSV, MÃPHG;

f5: MÃPHG \rightarrow TÊNPHG, TÒANHÀ }

KQ (MÃMH, TÊNMH, MÔTẢ, SỐTC, HK, NĂM, TÊN GV, ĐIỀMMÔN, MSSV)

F3 = {

f6: MÃMH \rightarrow TÊNMH, MÔTẢ, SỐTC;

f7: MÃMH, HK, NĂM \rightarrow TÊN GV;

f8: MÃMH, HK, NĂM, MSSV \rightarrow ĐIỀMMÔN }

a) Xác định khoá chính cho từng quan hệ.

c) Chỉ ra các trùng lặp thông tin và cho ví dụ về các dị thường dữ liệu có thể gặp phải.

d) Xác định dạng chuẩn của lược đồ CSDL bên trên. Giải thích.

e) Đề xuất lược đồ mới đạt tối thiểu dạng chuẩn 3. Giải thích.

4. Hãy xác định các khoá và dạng chuẩn của từng Q_i, F_i , được cho như sau:

a) $Q_1(ABCDEFGH), F_1 = \{A \rightarrow H; AB \rightarrow C; BC \rightarrow D; G \rightarrow B\}$

b) $Q_2(ABCSXYZ), F_2 = \{S \rightarrow A; AX \rightarrow B; S \rightarrow B; BY \rightarrow C; CZ \rightarrow X\}$

c) $Q_3(ABCD), F_3 = \{A \rightarrow B; BC \rightarrow D; D \rightarrow A\}$

Chuẩn hoá các lược đồ bên trên về DC3 (3NF) hoặc DCBC (BCNF).