



CHƯƠNG 3

LÝ THUYẾT THIẾT KẾ CƠ SỞ DỮ LIỆU

chuc.nv@due.edu.vn

3.1. Phụ thuộc hàm

Định nghĩa phụ thuộc hàm (Functional Dependence - FD)

Cho LĐQH $R(U)$, $X, Y \subseteq U$. Ta nói rằng X xác định Y hay Y phụ thuộc hàm vào X , ký hiệu là $X \rightarrow Y$ nếu với mọi quan hệ $r \subseteq R(U)$, lấy 2 bộ bất kỳ $t, u \subseteq r$ nếu $t(X) = u(X)$ thì $t(Y) = u(Y)$.

Ví dụ: Nếu ta biết rằng phụ cấp chức vụ (PCCV) phụ thuộc hàm vào chức vụ (CV), có nghĩa là nếu có 2 người giống nhau trên CV thì phải bằng nhau trên PCCV.

Ta có phụ thuộc hàm $CV \rightarrow PCCV$

3.1. Phụ thuộc hàm

Định nghĩa phụ thuộc hàm (Cont)

- $X \rightarrow Y$ gọi là phụ thuộc hàm tầm thường (*trivial*) nếu $Y \subseteq X$
- $X \rightarrow Y$ gọi là phụ thuộc hàm nguyên tố nếu X là tập nhỏ nhất có thể xác định được Y . Hay nói cách khác không tồn tại tập con thực sự Z của X mà $Z \rightarrow Y$
- Tập thuộc tính $K \subseteq U$ được gọi là khóa của $R(U)$ nếu $K \rightarrow U$ là phụ thuộc hàm nguyên tố
- Tính chất của phụ thuộc hàm cần thỏa mọi quan hệ của LĐQH
- Phụ thuộc hàm là qui luật của thế giới thực, ta không thể chứng minh chúng.
- Phụ thuộc hàm mô tả sự phụ thuộc dữ liệu giữa các tập thuộc tính.

3.1. Phụ thuộc hàm

Suy diễn phụ thuộc hàm (phụ thuộc hàm logic)

Cho LĐQH R với tập thuộc tính U và tập phụ thuộc hàm F (k/h: $R(U, F)$, $X, Y \subseteq U$).

Ta nói rằng F suy diễn logic $X \rightarrow Y$ hay $X \rightarrow Y$ được suy diễn logic từ F , ký hiệu $F \models X \rightarrow Y$ nếu mọi quan hệ r thuộc $R(U)$ thỏa mãn F thì r cũng thỏa $X \rightarrow Y$.

Cho lược đồ qua hệ $R(U, F)$, với $U=ABC$ và tập phụ thuộc hàm $F=\{A \rightarrow B, B \rightarrow C\}$

Khi đó phụ thuộc hàm $A \rightarrow C$ được suy diễn logic ra từ F , ký hiệu là $F \models A \rightarrow C$

3.1. Phụ thuộc hàm

Bao đóng của các tập phụ thuộc hàm

Cho LĐQH $R(U, F)$, bao đóng của F là tập tất cả các phụ thuộc hàm được suy diễn logic ra từ F . Ký hiệu là F^+ .

$$F^+ = \{X \rightarrow Y \mid X, Y \subseteq U : F \models X \rightarrow Y\}.$$

- Bao đóng của tập phụ thuộc hàm cho biết năng lực suy diễn thông tin của tập phụ thuộc hàm đó.
- Trong thực tế thì việc tìm tất cả các PTH được suy ra từ F là rất phức tạp khi số thuộc tính và số PTH trong F lớn. Một cách hiệu quả hơn là người ta đi kiểm tra một PTH $X \rightarrow Y$ có được suy ra từ F không (có thuộc F^+ không). Bài toán này gọi là bài toán thành viên.

3.1. Phụ thuộc hàm

Tiên đề Armstrong (Armstrong's axioms)

Cho LĐQH $R(U, F)$

Tiên đề phản xạ (reflexivity): Nếu $Y \subseteq X \subseteq U$ thì $X \rightarrow Y$.

Tiên đề tăng trưởng (augmentation): Nếu $X \rightarrow Y$ và $Z \subseteq U$ thì $XZ \rightarrow YZ$.

Tiên đề bắc cầu (transitivity): Nếu $X \rightarrow Y$ và $Y \rightarrow Z$ thì $X \rightarrow Z$.

3.1. Phụ thuộc hàm

Tiên đề Armstrong (Armstrong's axioms)

Từ hệ tiên đề Armstrong, người ta chứng minh được các qui tắc sau đây gọi là các qui tắc suy dẫn mở rộng.

Cho LĐQH $R(U, F)$

Luật hợp (Union rule): Nếu $X \rightarrow Y$ và $X \rightarrow Z$ thì $X \rightarrow YZ$.

Luật tách (Decomposition rule): Nếu $X \rightarrow Y$ và $Z \subseteq Y$ thì $X \rightarrow Z$.

Luật tựa bắc cầu (Pseudo transitivity rule): Nếu $X \rightarrow Y$ và $WY \rightarrow Z$ thì $WX \rightarrow Z$.

3.2. Bao đóng của tập thuộc tính (Closure of Attribute Set)

Định nghĩa bao đóng của tập thuộc tính

Cho LĐQH $R(U, F)$, Tập thuộc tính $X \subseteq U$. Bao đóng của tập thuộc tính X là tập tất cả các thuộc tính $A \in U$ sao cho $X \rightarrow A$ được suy dẫn logic từ F . Bao đóng của X ký hiệu là X^+ .

$$X^+ = \{ A \in U : F \models X \rightarrow A \}$$

Nói cách khác bao đóng của tập thuộc tính X là bao gồm tất cả các thuộc tính phụ thuộc hàm vào X .

Bao đóng của tập thuộc tính cho biết năng lực suy diễn thông tin của tập thuộc tính đó.

3.2. Bao đóng của tập thuộc tính (Closure of Attribute Set)

Tính bao đóng của tập các thuộc tính.

Input: Lược đồ quan hệ $R(U, F)$, $X \subseteq U$.

Output: X^+ .

Method:

Tính dãy các tập X^0, X^1, \dots, X^n như sau:

$$X^0 = X$$

$$X^i = X^{i-1} \cup \{A \mid A \in Z, Y \rightarrow Z \in F, Y \subseteq X^{i-1}\}$$

$$X^n = X^{n-1} \text{ dừng}$$

$$X^+ = X^n$$

3.3. Tìm khóa của lược đồ quan hệ

Algorithm FindOneKey

Input: - Tập thuộc tính U
- Tập PTH F

Output: - Khóa $K \subseteq U$ thỏa

$$K^+ = U$$

$$\forall A \in K: (K - \{A\})^+ \neq U$$

Method

Begin

$K := U;$

for each attribute A in K do

if $(K - \{A\})^+ = U$ then

$K := K - \{A\}$

endif;

endfor;

return K;

End FindOneKey;

Thuật toán tìm một khóa

Tư tưởng của thuật toán:

Xuất phát từ một siêu khóa K tùy ý của LĐQH $R(U)$, duyệt lần lượt các thuộc tính A của K , nếu $(K - \{A\})^+ = U$ thì loại thuộc tính A khỏi K .

3.3. Tìm khóa của lược đồ quan hệ

Ví dụ về thuật toán tìm một khóa của LĐQH

Cho LĐQH $R(ABCDE)$ và tập PTF $F = \{A \rightarrow B, B \rightarrow CD, C \rightarrow A, AB \rightarrow DE\}$. Tìm một khóa của lược đồ quan hệ trên.

Đặt $K = ABCDE$

Xét thuộc tính A

$\{K - A\}^+ = ABCDE = U$ nên $K = BCDE$

Xét thuộc tính B

$\{K - B\}^+ = ABCDE$ nên $K = CDE$

Xét thuộc tính C

$\{K - C\}^+ = DE$ nên $K = CDE$

Xét thuộc tính D

$\{K - D\}^+ = ABCDE$ nên $K = CE$

Xét thuộc tính E

$\{K - E\}^+ = ABCDE$ nên $K = C$

Vậy C là một khóa của lược đồ quan hệ đã cho

3.3. Tìm khóa của lược đồ quan hệ

Thuật toán tìm tất cả các khóa của lược đồ quan hệ

Input: $R(U, F)$

Output: Tập tất cả các khóa

Tư tưởng: Tìm tất cả các siêu khóa của $R(U, F)$ và loại các siêu khóa không tối thiểu

Method:

Bước 0: Xác định tập nguồn (TN) và tập trung gian (TG) như sau:

TN: Gồm tất cả các thuộc tính chỉ xuất hiện ở vế trái và không xuất hiện ở vế phải của tập phụ thuộc hàm, và những thuộc tính không tham gia vào bất kỳ một phụ thuộc hàm nào.

TG: Gồm tất cả các thuộc tính vừa tham gia vào vế trái vừa tham gia vào vế phải của tất cả các PTH trong F .

$S = \{ \}$: Tập chứa các khóa

3.3. Tìm khóa của lược đồ quan hệ

Thuật toán tìm tất cả các khóa của lược đồ quan hệ (Cont)

Bước 1: Nếu tập $TG = \emptyset$ thì khóa là tập TN. Kết thúc, ngược lại qua bước 2

Bước 2:

- Tìm tất cả các tập con X_i của tập TG
- Tìm siêu khóa S_i như sau: Nếu $(X_i \cup TN)^+ = U$ thì $S_i = X_i \cup TN$
- Thêm khóa S_i vào tập S: $S = S \cup S_i$

Bước 3: Loại bỏ các siêu khóa không tối thiểu trong S, tập còn lại là khóa

3.3. Tìm khóa của lược đồ quan hệ

Tìm tất cả các khóa của lược đồ quan hệ sau:

$R(CSZ); F=\{CS \rightarrow Z; Z \rightarrow C\}$

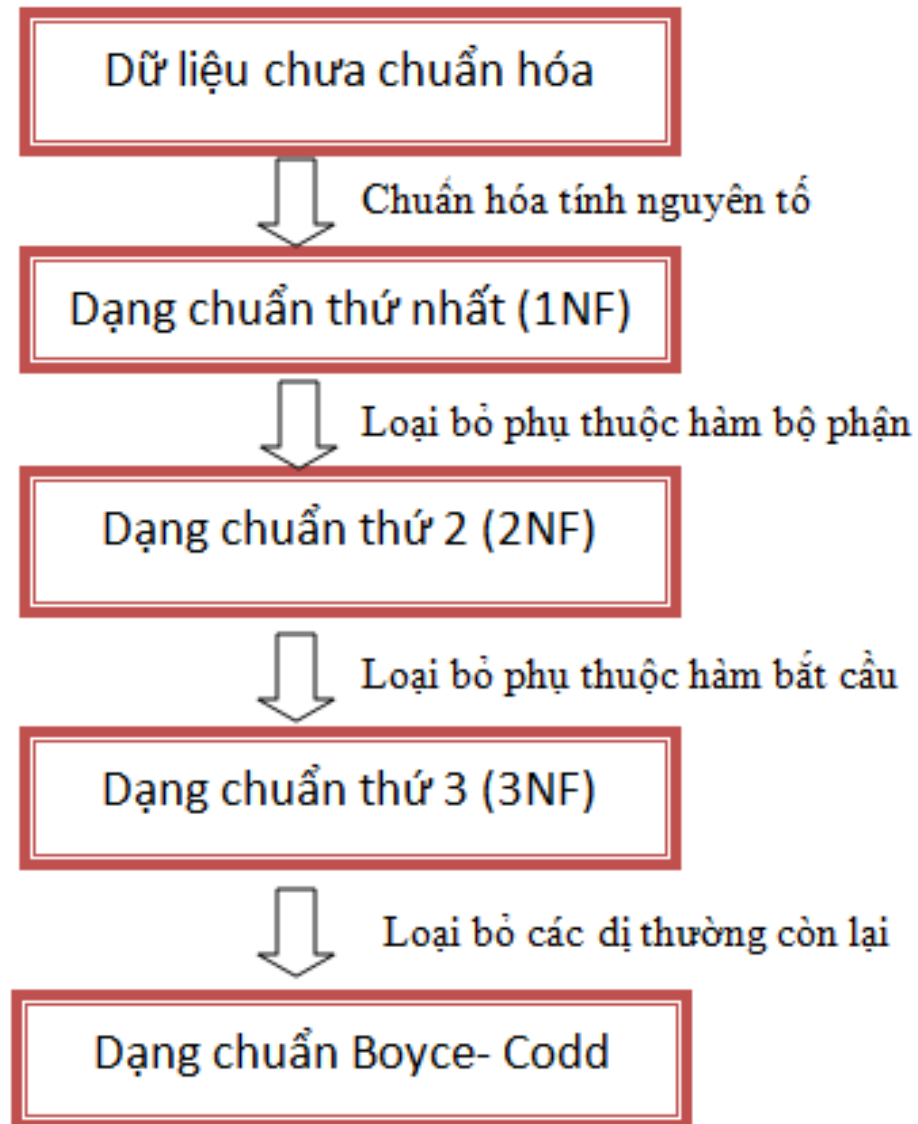
Xác định tập nguồn và tập trung gian

$TN= \{S\}; TG=\{C,Z\}$

TN	X_i	$TN \cup X_i$	$(TN \cup X_i)^+$	Siêu khóa	Khóa
S	\emptyset	S	S		
S	C	CS	CSZ	CS	CS
S	Z	SZ	CSZ	SZ	SZ
S	CZ	CSZ	CSZ	CSZ	

Lược đồ có 2 khóa là CS và SZ

3.4. Chuẩn hóa dữ liệu



3.4. Chuẩn hóa dữ liệu

Thuộc tính khóa (key attribute) và thuộc tính không khóa (non-key attribute): Cho LĐQH $R(U, F)$, $A \in U$ được gọi là thuộc tính khóa nếu A là thành phần của một khóa nào đó. Ngược lại A gọi là thuộc tính không khóa.

Ví dụ: $R(U, F)$, $U=ABCD$, $F=\{AB \rightarrow C, C \rightarrow D\}$

Khóa là AB . Thuộc tính A và B gọi là thuộc tính khóa còn C và D gọi là thuộc tính không khóa.

Phụ thuộc hàm đầy đủ (Fully Functional Dependence): Cho LĐQH $R(U, F)$. $X, Y \subseteq U$. Ta nói Y phụ thuộc hàm đầy đủ vào X nếu PTH $X \rightarrow Y$ là phụ thuộc hàm nguyên tố.

3.4. Chuẩn hóa dữ liệu

Dạng chuẩn thứ nhất (First Normal Form - 1NF)

Một lược đồ quan hệ R được gọi là ở dạng chuẩn 1NF khi và chỉ khi mọi thuộc tính của R là nguyên tố.

Một thuộc tính được gọi là nguyên tố khi:

- Đơn trị
- Không thể phân rã thành các thuộc tính nhỏ hơn
- Không thể suy dẫn từ các thuộc tính khác

Chú ý: Tính nguyên tố của thuộc tính phụ thuộc vào việc tổ chức dữ liệu và yêu cầu xử lý.

3.4. Chuẩn hóa dữ liệu

Dạng chuẩn thứ nhất (First Normal Form - 1NF)

Employees						
Eno	Full name	Skill	DateOfBirth	Age	Gender	Salary
E001	Nguyen Van Chau	SQL, C#	20/11/1980	38	Male	1200
E002	Ngo Van Chung	Java	15/10/1988	30	Male	1000
E003	Le Thi Nga	SQL	22/07/1990	28	Female	900
E004	Nguyen Dai Khoa	C#, R	18/09/1989	29	Male	1100
E005	Le Thi Nguyet	Java	24/01/1992	26	Female	950

Employees đã đạt 1NF chưa? Vì sao?

Chuyển Employees về 1NF

3.4. Chuẩn hóa dữ liệu

Dạng chuẩn thứ hai (Second Normal Form – 2NF)

Định nghĩa dạng chuẩn 2NF

Lược đồ quan hệ R được gọi là ở dạng chuẩn 2NF nếu R ở dạng chuẩn 1NF và mọi thuộc tính không khóa của R phụ thuộc đầy đủ vào khóa chính.

Hay nói cách khác: LĐQH R được gọi là ở dạng chuẩn 2NF nếu nó ở dạng chuẩn 1NF và không tồn tại phụ thuộc hàm bộ phận (partial dependency).

3.4. Chuẩn hóa dữ liệu

Kiểm tra một lược đồ đạt 2NF:

Tìm tất cả các khóa K của $R(U, F)$: $K = \{K_1, K_2, \dots, K_n\}$

Xác định tập thuộc tính khóa (KA: Key Attributes): $KA = \{K_1 \cup K_2 \cup \dots \cup K_n\}$

Xác định tập thuộc tính không khóa (NKA: Non Key Attributes)
 $NKA = U - KA$

Mọi thuộc tính A thuộc NKA, kiểm tra A có phụ thuộc vào bộ phận của khóa không.

Orders(Ono, Odate, Cno, Cname, Ino, Iname, Price, Amount)

$F = \{ Ono \rightarrow Odate, Ono \rightarrow Cno, Cno \rightarrow Cname, Ino \rightarrow (Iname, Price), (Ono, Ino) \rightarrow Amount \}$

Orders đạt 2NF không? Vì sao?

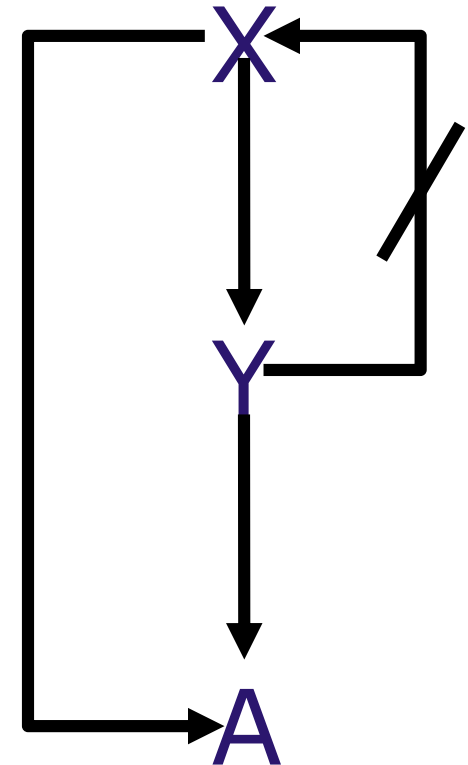
3.4. Chuẩn hóa dữ liệu

Dạng chuẩn thứ 3 (Third Normal Form – 3NF)

Phụ thuộc hàm bất cầu:

Cho $R(U, F)$, X là tập con thuộc tính của U , A là một thuộc tính của U . A được gọi là phụ thuộc hàm bất cầu vào X trên $R(U, F)$ nếu tồn tại một tập con Y của U sao cho:

$X \rightarrow Y$, $Y \rightarrow A$ nhưng Y không xác định X và A không thuộc XY



3.4. Chuẩn hóa dữ liệu

Định nghĩa dạng chuẩn 3NF

Lược đồ quan hệ R được gọi là ở dạng chuẩn 3NF nếu nó ở dạng chuẩn 2NF và mỗi thuộc tính không khóa của R không phụ thuộc hàm bất cầu vào khóa chính

Nhân Viên				
MãNV	Họ lót	Tên	Hệ Số Lương	Phụ Cấp
NV001	Nguyễn Văn	Chung	6	1000
NV002	Nguyễn Đại	Khoa	7	1500
NV003	Tran Văn	Tiến	6	1000
NV004	Lê Thị	Hà	7	1500


```
graph LR;
  A[MãNV] --> B[Họ lót];
  A --> C[Tên];
  A --> D[Hệ Số Lương];
  A --> E[Phụ Cấp];
  D --> E;
```

Lược đồ quan hệ Nhân Viên đạt dạng chuẩn 3NF không? Vì sao

3.4. Chuẩn hóa dữ liệu

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

Lược đồ quan hệ $R(U, F)$ gọi là ở dạng chuẩn BCNF nếu mọi phụ thuộc hàm không tầm thường đều có vế trái là khóa.

Nói cách khác: LĐQH $R(U, F)$ được gọi là ở dạng chuẩn BCNF nếu mọi PHT $X \rightarrow A \in F$ mà $A \notin X$ thì X phải là một khóa của lược đồ quan hệ $R(U, F)$.

Ví dụ: $MonHoc(MaMH, TenMH, SoTC)$

$F = \{MaMH \rightarrow TenMH, MaMH \rightarrow SoTC\}$

Lược đồ $MonHoc$ với tập PTH như trên đạt dạng chuẩn BCNF