PGS. TS. NGUYỄN BÁ TƯỜNG

CƠ SỞ DỮ LIỆU
QUAN HỆ & ỨNG DỤNG

NHÀ XUẤT BẢN THÔNG TIN VÀ TRUYỀN THÔNG

PGS. TS. NGUYỄN BÁ TƯỜNG

CƠ SỞ DỮ LIỆU
QUAN HỆ & ỨNG DỤNG

NHÀ XUẤT BẢN THÔNG TIN VÀ TRUYỀN THÔNG

Mã số: HT 01 HM 11

LỜI NÓI ĐẦU

Ngày nay, khi nhắc đến sự thành công của một doanh nghiệp người ta không thể không nói đến vai trò của hệ thống cơ sở dữ liệu trong doanh nghiệp. Giá trị của dữ liệu là một thứ tài sản quý giá của doanh nghiệp. Doanh nghiệp cần có được những dữ liệu chính xác và kịp thời về các hoạt động của doanh nghiệp, để quản lý và phân tích dữ liệu một cách có hiệu quả nhằm đưa ra được chiến lược cho những hoạt động tiếp theo. Tuy nhiên, nếu không có được sự quản lý một cách hữu hiệu đối với lượng dữ liệu ngày càng lớn này, cũng như khả năng đáp ứng yêu cầu của các câu hỏi tìm kiếm, lượng dữ liệu này sẽ trở nên vô nghĩa. Chính điều này đã thúc đẩy sự cần thiết có được một cơ sở dữ liệu cũng như một hệ quản trị cơ sở dữ liệu hiệu quả và linh hoạt, hỗ trợ doanh nghiệp trong công tác quản lý cũng như tìm kiếm thông tin.

Chúng ta có thể định nghĩa *cơ sở dữ liệu như là một bộ sưu tập có tổ chức của các dữ liệu liên quan logic với nhau và được các hệ ứng dụng của một tổ chức cụ thể nào đó sử dụng*. Cách tiếp cận cơ sở dữ liệu nhấn mạnh đến tính chia sẻ và tích hợp của dữ liệu trong toàn bộ hoạt động của một doanh nghiệp. Theo cách tiếp cận này, dữ liệu được lưu trữ có tổ chức, tập trung dưới dạng một cơ sở dữ liệu và giữa các dữ liệu có sự liên kết logic với nhau.

Nhằm mục đích giới thiệu đến bạn đọc những kiến thức cơ bản nhất về cơ sở dữ liệu quan hệ, NXB Thông tin và Truyền thông trân trọng giới thiệu cuốn sách ***"Cơ sở dữ liệu quan hệ và ứng dụng"*** do PGS.TS Nguyễn Bá Tường, hiện đang công tác và giảng dạy tại Học viện Kỹ thuật Quân sự biên soạn.

Nội dung của Cuốn sách này là hệ thống những vấn đề quan trọng nhất của cơ sở dữ liệu quan hệ, một mô hình dữ liệu xuất hiện trong hầu khắp các bài toán quản lý của doanh nghiệp. Cuốn sách gồm bốn chương.

Chương 1: Các khái niệm cơ bản: nêu một số khái niệm, ký hiệu toán học sẽ sử dụng cho các chương sau. *Chương 2: Cơ sở dữ liệu quan hệ,* trong chương này đề cập đến các khái niệm như quan hệ, phụ thuộc hàm, khóa,

khóa chính, khóa ngoại, dạng chuẩn, ràng buộc toàn vẹn,... Đặc biệt chương này còn đưa thêm các ví dụ ứng dụng cụ thể và các bước thiết kế một cơ sở dữ liệu. *Chương 3: Tập thô và ứng dụng*, giới thiệu những khái niệm cơ bản nhất của lý thuyết tập thô, một công cụ đắc lực của khai thác dữ liệu, khai phá tri thức đang được những người làm tin học cả thế giới quan tâm nghiên cứu, tìm hiểu. Việc nghiên cứu và hiểu biết về tập thô thật sự cần thiết trong nghiên cứu về cơ sở dữ liệu. *Chương 4: Hệ tin và các ứng dụng,* là một trường hợp tổng quát của khái niệm quan hệ. Hệ tin thực chất là một bảng dữ liệu. Việc đưa hệ tin vào lý thuyết cơ sở dữ liệu giúp chúng ta có cách nhìn tổng thể về mô hình dữ liệu quan hệ. Hệ tin giúp chúng ta dễ dàng diễn đạt một số bài toán như thuật toán Quiland, K-Mean, Bayer,...

Hy vọng cuốn sách sẽ là tài liệu bổ ích và lý thú không chỉ dành cho các giảng viên, sinh viên đại học, học viên cao học chuyên ngành công nghệ thông tin mà cả các bạn đọc quan tâm và tìm hiểu về cơ sở dữ liệu.

Nhà xuất bản xin trân trọng giới thiệu cùng bạn đọc và rất mong nhận được ý kiến đóng góp của quý vị. Mọi ý kiến đóng góp xin gửi về ***Nhà xuất bản Thông tin và Truyền thông*** - 18 Nguyễn Du, Hà Nội.

Xin trân trọng cảm ơn./.

NXB THÔNG TIN VÀ TRUYỀN THÔNG

Chương 1

CÁC KHÁI NIỆM CƠ BẢN

1.1. TẬP HỢP VÀ CÁC PHÉP TOÁN TẬP HỢP

Tập hợp tuy không được định nghĩa nhưng khi nói đến tập hợp U nào đó, chúng ta đã hình dung được các phần tử của nó. Trong một số môn học chúng ta đã từng biết và tiếp xúc với những cụm từ tập hợp như: tập các điểm bất động, tập các điểm kỳ dị, tập compact, tập mở, tập đóng, tập sinh viên trong lớp... tất cả những tập như vậy đều được coi là những tập đã "*định nghĩa*" trong lý thuyết tập hợp.

Trong mục 1.1 này ta sẽ nhắc lại một số khái niệm liên quan đến khái niệm tập hợp. Đây cũng là cơ sở để chúng ta phát triển và hiểu các khái niệm về *tập mờ*, *tập thô* được xét trong các phần sau của cuốn sách này. Vậy chúng ta gọi các tập hợp đã được xét cho đến thời điểm này là những tập đã biết theo quan điểm tập hợp trong các tài liệu về toán cao cấp.

Sau đây chúng ta nhắc lại một số phép toán trên các tập hợp.

Để biểu thị một tập X trong không gian U ta thường dùng các phương pháp liệt kê, ví dụ tập $X = \{a, b, c, d\}$ hoặc $X = \{x \in U: V(x)\}$, X gồm những phần tử x của U thỏa mãn điều kiện V(x). Như vậy nói cho tập X ta chắc chắn có cách xác định các phần tử của tập X.

Hợp của hai tập: $X \cup Y$ là tập gồm các phần tử của X hoặc của Y.

Giao của hai tập: $X \cap Y$ là tập gồm các phần tử chung của X và Y.

Hiệu của hai tập: $X \setminus Y$ là tập chỉ chứa những phần tử của X không thuộc Y.

Phần bù của tập X ký hiệu $-X$ là tập gồm những phần tử của không gian U không thuộc X.

$X \subset Y$ hoặc $Y \supset X$ biểu thị X là tập con thực sự của Y.

$X \subseteq Y$ hoặc $Y \supseteq X$ biểu thị X là tập con (có thể bằng Y) của Y.

Ký hiệu $x \in X$ biểu thị phần tử x thuộc tập X.

Ta dùng Card(X) hoặc $|X|$ để chỉ số phần tử của X, 2^U để chỉ họ các tập con của U.

Trong cơ sở dữ liệu quan hệ, nhiều khi thay cho hợp của hai tập thuộc tính $X \cup Y$ ta viết XY.

Cho hai tập B và C. Tích Đề-các của B và C, ký hiệu $B \times C$ là tập tất cả các cặp có thể của B và C, tức $B \times C = \{(a,b): a \in B \text{ và } b \in C\}$.

Ví dụ $B = \{1, 2, 3\}$; $C = \{x, y\}$ khi đó:

$B \times C = \{(1,x), (2,x), (3,x), (1,y), (2,y), (3,y)\}$

$C \times C = \{(x,x), (x,y), (y,y), (y,x)\}$

$B \times B = \{(1,1), (2,2), (3,3), (1,2), (1,3), (2,1), (3,1), (2,3), (3,2)\}$

1.2. PHÂN HOẠCH VÀ PHỦ CỦA KHÔNG GIAN U

Cho U là tập hữu hạn, khác rỗng bất kỳ được coi là một không gian.

Định nghĩa 1.1.a Định nghĩa phân hoạch của U

Họ các tập con $E = \{E_1, E_2, \ldots, E_k\}$, $k \geq 1$ của U được gọi là một phân hoạch của U nếu:

(1) $E_i \neq \phi$; với $i = 1,\ldots,k$

(2) $E_i \cap E_j = \phi$; với $i \neq j$

(3) $\bigcup_{i=1}^{k} E_i = U$

Mỗi E_i ta thường gọi là *một nhóm* của phân hoạch E.

Ví dụ 1.1:

Cho $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

Khi đó $E = \{E_1, E_2, E_3\} = \{\{1, 2\}, \{3, 4, 5, 6\}, \{7, 8, 9\}\}$

hoặc $P = \{P_1, P_2\} = \{\{1, 2\}, \{3, 4, 5, 6, 7, 8, 9\}\}$ là những phân hoạch của U.

Lưu ý:

- U có thể có nhiều phân hoạch khác nhau.
- Số các phân hoạch sẽ rất lớn khi số phần tử của U đủ lớn.
- Khi cho tập U ta chỉ xét 1 phân hoạch cụ thể

Định nghĩa 1.1.b Định nghĩa phủ của U

Họ các tập con $E = \{E_1, E_2, \ldots, E_k\}$, $k \geq 1$ của U được gọi là một phủ của U nếu:

(1) $E_i \neq \phi$; với $i = 1,\ldots,k$

(2) $\bigcup_{i=1}^{k} E_i = U$

Mỗi E_i ta thường gọi là *một nhóm* của phủ E.

Ví dụ 1.2:

Cho U = {1, 2, 3, 4, 5, 6, 7, 8, 9}.

Khi đó E = {E_1, E_2, E_3} = {{1, 2}, {1, 2, 3}, {4, 5, 6}, {7}, {7, 8, 9}} hoặc

P = {P_1, P_2} = {{1, 2}, {1, 2, 3, 4, 5}, {3, 4, 5, 6, 7, 8, 9}} là những phủ của U.

Lưu ý:

- U có thể có nhiều phủ khác nhau.

- Phân hoạch là trường hợp riêng của phủ

- Mỗi phân hoạch của U là một phủ của U nên mọi tính chất đúng cho phủ thì đúng cho phân hoạch

1.3. QUAN HỆ VÀ QUAN HỆ TƯƠNG ĐƯƠNG

Cho $A_1, A_2, \ldots, A_n$; $n \geq 2$ là các tập bất kỳ.

Định nghĩa 1.2 Định nghĩa quan hệ n-ngôi

Tập $R \subseteq A_1 \times A_2 \times \ldots \times A_n$ được gọi là *quan hệ n-ngôi* trên $A_1, A_2, \ldots, A_n$.

Lưu ý:

- Với quan hệ R hai ngôi (n = 2) đôi khi thay cho viết $(a,b) \in R$ ta có thể viết aRb.

- Khi $A_1 = A_2 = B$ quan hệ $R \subseteq B \times B$ được gọi là quan hệ trên B.

Ta nhận xét rằng các quan hệ R là các tập hợp nên có thể thực hiện các phép toán tập hợp trên chúng, ví dụ có thể xét các quan hệ lồng nhau, giao của các quan hệ, hợp các quan hệ, phần bù các quan hệ. Ta nói R là quan hệ con của quan hệ R' trên B, ký hiệu $R \subseteq R'$, nếu mọi $(t, t') \in R$ thì $(t, t') \in R'$. Nói cách khác R là quan hệ con của R' trên B nếu $R, R' \subseteq B \times B$ và R là tập con của R', tức $R \subseteq R'$.

Ví dụ ta lấy B = {1, 2, 3, 4}.

Xét các quan hệ R và R' trên B sau:

R = {(1,2), (2,2), (3,4)}; R' = {(1,2), (4,4), (4,3)}.

Khi đó $R \cap R'$ = {(1,2)};

$R \setminus R'$ = {(2,2), (3,4)}; $R \cup R'$ = {(1,2), (2,2), (3,4), (4,4), (4,3)}.

Phần bù của R, ký hiệu $-R = A \times A \setminus R$.

Định nghĩa 1.3 Quan hệ tương đương

Quan hệ 2-ngôi $R \subseteq B \times B$ được gọi là *quan hệ tương đương* trên B nếu R thỏa mãn 3 điều kiện:

(1) Phản xạ: $\forall a \in B$ aRa

(2) Đối xứng: $\forall a, b \in B$ aRb khi và chỉ khi bRa

(3) Bắc cầu: $\forall a, b, c \in B$ aRb & bRc thì aRc

Khi đó R chia B thành các nhóm tương đương và ta dễ dàng chỉ ra rằng họ các nhóm tương đương đó tạo thành một phân hoạch của B. Ta sẽ ký hiệu phân hoạch đó là B/R và $B/R = \{[a]_R: a \in B\}$ với $[a]_R$ là nhóm các phần tử b mà aRb.

Phân hoạch B/R *là phân hoạch tương đương* (vì nó được sinh ra từ quan hệ tương đương R). Ví dụ gọi B là số sinh viên trong lớp, R là quan hệ cùng quê. Rõ ràng R là quan hệ tương đương và B/R là các nhóm sinh viên cùng quê với nhau.

Ví dụ 1.3:

a. Cho B = {1, 2, 3, 4, 5, 6, 7, 8, 9}. Xét R là quan hệ đồng dư khi chia các phần tử của B cho 3. Khi đó dễ dàng chỉ ra rằng R là quan hệ tương đương trên B và B/R = {{1, 4, 7},{2, 5, 8,},{3, 6, 9}}. Tức là 1, 4, 7 đồng dư 1 khi chia cho 3, còn 2, 5, 8 đồng dư 2 khi chia cho 3 và 3, 6, 9 đồng dư 0 khi chia cho 3.

b. Cho U là tập có 6 sinh viên gồm hai sinh viên nữ là Hoa, Huệ và 4 sinh viên nam là Phú, Quí, Giàu, Sang:

U = {Phú, Quí, Giàu, Sang, Hoa, Huệ}. Gọi R là quan hệ đồng giới, S là quan hệ khác giới.

Khi đó R, S $\subseteq U \times U$ rõ ràng $Card(U \times U) = 36$ và R = {(Phú,Phú), (Quí,Quí), (Giàu,Giàu), (Sang,Sang), (Phú,Quí), (Quí,Phú), (Phú,Giàu), (Giàu,Phú), (Phú, Sang), (Sang,Phú), (Quí,Giàu), (Giàu,Quí), (Quí,Sang), (Sang,Quí), (Giàu,Sang), (Sang,Giàu), (Hoa,Hoa), (Huệ,Huệ), (Hoa,Huệ), (Huệ,Hoa)}.

R là quan hệ tương đương; Card(R) = 20; U/R = {{Phú, Quí, Giàu, Sang}, {Hoa, Huệ}}

S = {(Phú,Hoa), (Hoa,Phú), (Quí,Hoa), (Hoa,Quí), (Giàu,Hoa), (Hoa,Giàu), (Sang,Hoa), (Hoa,Sang), (Phú,Huệ), (Huệ, Phú), (Quí,Huệ), (Huệ,Quí), (Giàu,Huệ), (Huệ,Giàu), (Sang,Huệ), (Huệ,Sang)}.

S là quan hệ không tương đương; Card(S) = 16. Không có phân hoạch U/S

c. Cho B = {1, 2, 3, 4}. Sau đây là các ví dụ về các quan hệ không tương đương trên B:

Xét R_1 là quan hệ chỉ thỏa mãn tính phản xạ:

$R_1 = \{(1,1), (2,2), (3,3), (4,4)\}$

Gọi R_2 là quan hệ chỉ thỏa mãn tính đối xứng:

$R_2 = \{(2,3), (3,2)\}$

Gọi R_3 là quan hệ chỉ thỏa mãn tính bắc cầu:

$R_3 = \{(4,3), (3,2), (2,1), (4,2), (4,1), (3,1)\}$

Dễ dàng thấy rằng:

$R = R_1 \cup R_2$ là quan hệ thỏa mãn hai điều kiện phản xạ và đối xứng.

$S = R_1 \cup R_3$ là quan hệ phản xạ và bắc cầu

Bổ đề 1.1

a. Mọi quan hệ tương đương R trên B thì $B/R = \{[a]_R: a \in B\}$ là một phân hoạch của B.

b. Mọi phân hoạch $E = \{E_1, E_2,..., E_k\}$ của B có một quan hệ tương đương R trên B để $B/R = E$.

Chứng minh:

a. Giả sử $E = B/R = \{[a]_R: a \in B\}$ ta phải chứng minh E là một phân hoạch của B

- Mỗi nhóm $[a]_R$ của E là tập con của B và khác rỗng vì mỗi nhóm $[a]_R$ chứa ít nhất là a.

- Hai nhóm khác nhau $[a]_R$, $[b]_R$ giao với nhau bằng rỗng. Thật vậy giả sử $[a]_R$ khác $[b]_R$ và $[a]_R \cap [b]_R \neq \phi$. Tức chúng có phần tử c chung nhau, khi đó aRc và cRb suy ra aRb và $[a]_R = [b]_R$ điều này vô lý vì đã lấy hai nhóm $[a]_R$, $[b]_R$ khác nhau.

- Hợp của các nhóm trong E bằng B vì $E = B/R = \{[a]_R: a \in B\}$ lấy cho mọi a của B và cũng chỉ của B mà thôi.

b. Giả sử $E = \{E_1, E_2,..., E_k\}$ là một phân hoạch của B.

Ta phải xây dựng quan hệ $R \subset B \times B$ để $E = B/R$

Ta xây dựng $R = E_1 \times E_1 \cup E_2 \times E_2 \cup ... E_k \times E_k$. Khi đó dễ dàng thử lại rằng R thỏa mãn 3 tính chất của quan hệ tương đương.

Ví dụ cho $B = \{1, 2, 3, 4\}$ và cho $E = \{E_1, E_2\} = \{\{1, 2\}, \{3, 4\}\}$.

Khi đó $E_1 \times E_1 = \{(1,1), (2,2), (1,2), (2,1)\}$

và $E_2 \times E_2 = \{(3, 3), (3, 4), (4, 3), (4,4)\}$

và $R = E_1 \times E_1 \cup E_2 \times E_2 = \{(1,1), (2,2), (3,3), (4,4), (1, 2), (2, 1), (4, 3), (3,4)\}$. Rõ ràng R là quan hệ tương đương trên B và $B/R = \{E_1, E_2\}$

Từ bổ đề 1.1 trên ta có nhận xét rằng việc xét một quan hệ tương đương R trên B đôi khi đồng nhất với việc xét một phân hoạch $E = \{E_1, E_2,..., E_k\}$ của B.

Phân hoạch tập B thực chất là một cách phân loại hay gom cụm tập B thành các nhóm theo một tiêu chí nào đó. Ví dụ lấy $B = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Phân loại theo tiêu chí đồng dư 3 ta có $E = \{\{3, 6, 9\}, \{1,4, 7\}, \{2, 5, 8\}\}$. Phân loại theo tiêu chí đồng dư 4 ta có các nhóm: $\{4, 8\}$, $\{1, 5, 9\}$, $\{2, 6\}$, $\{3, 7\}$. Hoặc coi B là tập sinh viên trong lớp và xét R là quan hệ cùng năm sinh thì A được phân hoạch thành các nhóm sinh viên cùng năm sinh, tức cùng tuổi. Nếu phân loại theo giới tính ta có phân hoạch của B là hai nhóm: nhóm nam, nhóm nữ.

Định nghĩa 1.4 Phân hoạch con

Phân hoạch $P = \{P_1, P_2,..., P_l\}$ của U là phân hoạch con của phân hoạch $E = \{E_1,..., E_k\}$ của U, ký hiệu $P << E$ nếu $P \neq E$ và mỗi nhóm P_i của P là một nhóm con của một nhóm E_j nào đó của E.

Ví dụ cho $B = \{1, 2, 3, 4, 5, 6, 7\}$; $P = \{\{1, 2, 3\}, \{4, 5\}, \{6, 7\}\}$ là phân hoạch con của $E = \{\{1, 2, 3\}, \{4, 5, 6, 7\}\}$.

Bổ đề 1.2

Gọi R và S là các quan hệ tương đương trên U. Khi đó ta có $U/R << U/S$ khi và chỉ khi $R \subseteq S$.

Chứng minh:

Giả sử $U/R << U/S$ lấy $(t, s) \in R$ khi đó t và s thuộc một nhóm P_i nào đó của U/R và vì $U/R << U/S$ nên $P_i \subseteq E_j$ và vì vậy $(t,s) \in S$. Vậy $R \subseteq S$.

Ngược lại giả sử $R \subseteq S$ khi đó nếu $(t,s) \in R$ thì $(t,s) \in S$ và vì R, S là các quan hệ tương đương nên t,s thuộc một nhóm P_i nào đó của U/R và t,s thuộc một nhóm E_j nào đó của U/S và $E_i \subseteq P_j$. Vậy $U/R << U/S$.

1.4. ĐẠI SỐ

Định nghĩa 1.5 Đại số

Đại số là một bộ gồm tập U và các phép toán nào đó thỏa mãn trên U.

Ví dụ $D = (N, +, \times)$ là một đại số, với N là tập số tự nhiên, + và $\times$ là các phép toán cộng, nhân tương ứng. $T = (2^U, \cup, \cap)$ với 2^U là họ các tập con của U, $\cup, \cap$ là các phép hợp, giao của các tập trong 2^U, là những đại số.

Định nghĩa 1.6 Tập sắp thứ tự

Cho tập U. Ta nói U là tập được sắp thứ tự nếu trên U xác định quan hệ thứ tự $\leq$ thỏa mãn các điều kiện sau:

(1) Mọi x thì $x \leq x$

(2) Nếu $x \leq y$ và $y \leq z$ thì $x \leq z$

(3) Nếu $x \leq y$ và $y \leq x$ thì $x = y$

(4) Với mọi $x, y \in U$ thì hoặc $x \leq y$ hoặc $y \leq x$.

Ví dụ lấy U là tập các số tự nhiên thì rõ ràng U là tập được sắp thứ tự

Ta dễ dàng thấy rằng nếu U là tập được sắp thứ tự với quan hệ thứ tự $\leq$ thì các tập con B của U cũng được sắp thứ tự với quan hệ thứ tự là $\leq$.

Giả sử U là tập sắp thứ tự với quan hệ thứ tự $\leq$; $B \subset U$.

Phần tử a của U được gọi là *cận trên* của B trong U nếu $x \leq a$ với mọi $x \in B$. Nếu phần tử như vậy tồn tại thì ta nói B được chặn trên trong U. Nếu U có phần tử cận trên a thì a là cận trên duy nhất của tập U; a là phần tử cực đại của U nếu $x \leq a$ với mọi $x \in U$.

Định nghĩa 1.7 Dàn

Một tập được sắp thứ tự được gọi là một dàn nếu mọi cặp phần tử của nó có cận trên và cận dưới.

Định nghĩa 1.8 Topo

Họ các tập con $C = \{C_1, C_2,..., C_k\}$ của U được gọi là topo nếu C thỏa mãn ba điều kiện:

(1) ϕ và O thuộc C

(2) $\cup C_i$ thuộc C

(3) $\cap C_i$ thuộc C

Trong đó: $\cup C_i$ và $\cap C_i$ là hợp và giao của các nhóm C_i nào đó.

1.5. ÁNH XẠ ĐÓNG

Tập đóng hay tập bao đóng giữ vai trò quan trọng trong nghiên cứu các tập thuộc tính của cơ sở dữ liệu quan hệ. Trong phần này ta sẽ nêu vài khái niệm toán học cho các tập đóng.

Cho tập hữu hạn U. Gọi 2^U là họ các tập con của U.

Định nghĩa 1.9 Ánh xạ đóng

Ánh xạ f: $2^U \to 2^U$ được gọi là *ánh xạ đóng (hay hàm đóng)* trên U nếu với mọi tập con X, Y $\subseteq$ U thỏa mãn ba điều kiện sau:

(1) Tính phản xạ: X $\subseteq$ f(X);

(2) Tính đơn điệu: Nếu X $\subseteq$ Y thì f(X) $\subseteq$ f(Y);

(3) Tính lũy đẳng: f(f(X)) = f(X).

Ví dụ 1.4:

Ánh xạ f(X) = X với mọi X $\subseteq$ U là ánh xạ đóng

Ánh xạ f(X) = X $\cup$ X_0 với X_0 là tập cố định của U là ánh xạ đóng

Ánh xạ f(X) = U là ánh xạ đóng.

Định nghĩa 1.10 *Giao, hợp* của các ánh xạ

Cho các ánh xạ f, f': $2^U \to 2^U$ ta xác định các ánh xạ g, h: $2^U \to 2^U$ như sau:

g(X) = f(X) $\cap$ f'(X) gọi là *giao của các ánh xạ f và f'*, ký hiệu g = f $\cap$ f'.

h(X) = f(X) $\cup$ f'(X) gọi là hợp của các ánh xạ f, f', ký hiệu h = f $\cup$ f'.

Định nghĩa 1.11 *Hợp thành* của các ánh xạ

Cho các ánh xạ f, f': $2^U \to 2^U$ ta xác định *ánh xạ hợp thành của f và f'* là ánh xạ T = f(f') như sau: T(X) = f(f'(X)), với X $\subseteq$ U.

Định nghĩa 1.12 Ánh xạ hẹp hơn

Cho các ánh xạ f,g: $2^U \to 2^U$. Ta nói ánh xạ *f hẹp hơn ánh xạ g* và ký hiệu f $\le$ g nếu f(X) $\subseteq$ g(X) với mọi X $\subseteq$ U.

Bổ đề 1.3

Nếu f là một ánh xạ đóng thì với mọi X, Y ta luôn có:

1. f(f(X) $\cup$ Y) = f(X $\cup$ (f(Y)) = f(X $\cup$ Y) = f(f(X) $\cup$ f(Y))
2. f(X $\cup$ Y) $\supseteq$ f(X) $\cup$ f(Y)
3. f(X $\cap$ Y) $\subseteq$ f(X) $\cap$ f(Y).

Chứng minh:

Vì (X $\cap$ Y) $\subseteq$ X nên theo điều kiện (2) của hàm đóng f ta có f(X $\cap$ Y) $\subseteq$ f(X). Tương tự vì (X $\cap$ Y) $\subseteq$ Y nên ta có f(X $\cap$ Y) $\subseteq$ f(Y). Vậy f(X $\cap$ Y) $\subseteq$ f(X) $\cap$ f(Y), điều kiện 3 được chứng minh. Tương tự ta chứng minh điều kiện 2.

Vì $(X \cup Y) \supseteq X$ và $(X \cup Y) \supseteq Y$ nên $f(X \cup Y) \supseteq f(X) \cup f(Y)$

Để chứng minh điều kiện 1. Ta nhận xét rằng: $f(f(X) \cup Y) \supseteq f(X \cup Y)$;

$f(X \cup f(Y)) \supseteq f(X \cup Y)$; $f(f(X) \cup f(Y)) \supseteq f(f(X) \cup Y)$

và $f(f(X) \cup f(Y)) \supseteq f((X) \cup f(Y))$.

Vậy để chứng minh đẳng thức kép 1, ta chỉ cần chứng minh:

$f(X \cup Y) = f(f(X) \cup f(Y))$.

Vì $f(X) \cup f(Y) \supseteq X \cup Y$

nên ta có: $f(f(X) \cup f(Y)) \supseteq f(X \cup Y)$ (*)

Mặt khác $f(X \cup Y) \supseteq f(X)$ và $f(X \cup Y) \supseteq f(Y)$

nên $f(X \cup Y) \supseteq f(X) \cup f(Y) \Rightarrow f(f(X \cup Y)) \supseteq f(f(X) \cup f(Y))$

mà $f(f(X \cup Y)) = f(X \cup Y)$ nên $f(X \cup Y) \supseteq f(f(X) \cup f(Y))$ (**)

Từ (*) và (**) ta có $f(X \cup Y) = f(f(X) \cup f(Y))$. Bổ đề được chứng minh.

Bổ đề 1.4

Giao của hai ánh xạ đóng là một ánh xạ đóng.

Chứng minh:

Giả sử f và f' là các ánh xạ đóng. Ta chứng minh ánh xạ $g(X) = f(X) \cap f'(X)$ là ánh xạ đóng. Thật vậy:

Tính phản xạ: Vì f và f' là các ánh xạ đóng nên $X \subseteq f(X)$ và $X \subseteq f'(X)$ $\Rightarrow X \subseteq f(X) \cap f'(X) = g(X)$

Tính đồng biến: Giả sử $X \supseteq Y$ khi đó vì f và f' là các ánh xạ đóng nên $f(X) \supseteq f(Y)$ và $f'(X) \supseteq f'(Y)$, nên $f(X) \cap f'(X) \supseteq f(Y) \cap f'(Y) \Rightarrow g(X) \supseteq g(Y)$

Tính lũy đẳng: Theo tính phản xạ vừa chứng minh ta có:

$g(g(X)) \supseteq g(X)$ (*)

Mặt khác theo định nghĩa của g(X) ta lại có:

$g(g(X)) = f(g(X)) \cap f'(g(X)) = f(f(X) \cap f'(X)) \cap f'(f(X) \cap f'(X))$,

vì $f(X) \cap f'(X) \subseteq f(X)$ nên $f(f(X) \cap f'(X)) \subseteq f(f(X)) = f(X)$

tương tự có: $f'(f(X) \cap f'(X)) \subseteq f'(f'(X)) = f'(X)$.

Vậy $g(g(X)) \subseteq f(X) \cap f'(X) = g(X) \Rightarrow g(g(X)) \subseteq g(X)$ (**)

Từ (*) và (**) ta có $g(X) = g(g(X))$. Bổ đề được chứng minh.

Bổ đề 1.5

Hợp thành của hai ánh xạ đóng thỏa mãn tính phản xạ và tính đồng biến.

Chứng minh:

Giả sử f, g là các ánh xạ đóng xét ánh xạ hợp thành h = f(g).

- *Kiểm tra tính phản xạ của h*: h(X) = f(g(X)) vì f và g là các ánh xạ đóng nên chúng thỏa mãn tính phản xạ, vì vậy $h(X) = f(g(X)) \supseteq g(X) \supseteq X$.

- *Kiểm tra tính đồng biến của h*: lấy $X \subseteq Y$ ta phải chứng minh $h(X) \subseteq h(Y)$, tức $f(g(X)) \subseteq f(g(Y))$. Vì g là ánh xạ đóng nên $g(X) \subseteq g(Y)$ và vì f là ánh xạ đóng nên $f(g(X)) \subseteq f(g(Y))$.

Bổ đề 1.6

Hợp thành h = f(g) của hai ánh xạ đóng f, g không hẹp hơn ánh xạ thành phần g.

Chứng minh:

Cho f, g: $2^U \rightarrow 2^U$ là các ánh xạ đóng, xét h = f(g). Khi đó h(X) = f(g(X)), vì f là ánh xạ đóng nên $h(X) = f(g(X)) \supseteq g(X)$. Vậy h không hẹp hơn g.

Bổ đề 1.7

Hợp thành của hai ánh xạ đóng f, g là ánh xạ đóng khi phép hợp thành có tính giao hoán, tức là f(g) = g(f)

Chứng minh:

Giả sử f, g là hai ánh xạ đóng.

Giả sử hợp thành của f, g là giao hoán ta chứng minh h = f(g) là ánh xạ đóng. Tính phản xạ và tính đồng biến của h đã được chứng minh trong bổ đề 1.5. Bây giờ ta xét tính lũy đẳng của h. Ta phải chỉ ra h(h(X)) = h(X).

Thật vậy sử dụng tính giao hoán f(g) = g(f) và tính lũy đẳng của các hàm f, g ta có:

$$h(h(X)) = f(g(h(X))) = f(g(f(g(X)))) = g(f(f(g(X))))$$
$$= g(f(g(X))) = f(g(g(X))) = f(g(X)) = h(X).$$ Bổ đề được chứng minh.

Định nghĩa 1.13 Tập đóng

Cho f là ánh xạ đóng trên U, tập con X của U được gọi là *tập đóng* của f nếu X = f(X).

Ví dụ với ánh xạ đồng nhất f(X) = X thì mọi X là tập đóng.

Cho U là tập bất kỳ. Xét G là họ các tập con của U tức là $G \subseteq 2^U$.

Định nghĩa 1.14 Giàn giao

Ta nói G là *một giàn giao* trên U nếu: $\forall X, Y \in G$ thì $X \cap Y \in G$. Nghĩa là G đóng với phép giao (hoặc G bảo toàn phép giao).

Ví dụ ta lấy $G = 2^U$ thì G là một giàn giao trên U.

Định lý 1.1 Định lý về tập sinh của giàn giao

Mọi giàn giao G luôn tồn tại họ bé nhất $Gen(G) = \{X_1, X_2,..., X_k\}$ sao cho mọi tập của G đều được biểu thị qua giao của các tập trong Gen(G).

Chứng minh:

Ta dễ dàng chứng minh định lý 1.1 bằng quy nạp.

Giả sử G chỉ chứa 1 tập X khi đó $Gen(G) = \{X\} = \{X_1\}$.

Giả sử định lý đúng cho giàn giao G_k chứa k tập, nghĩa là G_k được sinh bởi $Gen(G_k) = \{X_1, X_2,...,X_i\}$ ta phải chứng minh định lý đúng cho G_{k+1} chứa k+1 tập. Thay cho chứng minh ta lấy $Gen(G_{k+1}) = \{X_1, X_2,... X_i, X\}$ với X là tập thứ k+1 vừa được thêm vào.

Cho f là ánh xạ đóng trên U. Gọi G_f là họ các tập đóng của f. Khi đó ta có định lý sau:

Định lý 1.2

G_f là một giàn giao trên U.

Chứng minh:

Giả sử $X, Y \in G_f$, ta phải chứng minh $X \cap Y \in G_f$ tức chứng minh $X \cap Y$ là tập đóng.

Thật vậy vì f là ánh xạ đóng nên theo tính phản xạ ta có:

$$X \cap Y \subseteq f(X \cap Y) \qquad (*)$$

Mặt khác theo tính đơn điệu và vì X, Y là các tập đóng nên ta có:

$$f(X \cap Y) \subseteq f(X) \cap f(Y) = X \cap Y \qquad (**)$$

Từ (*) và (**) ta có $X \cap Y = f(X \cap Y)$ hay $X \cap Y$ là tập đóng.

1.6. LOGIC MỆNH ĐỀ

Logic mệnh đề được xây dựng trên các mệnh đề có giá trị logic đúng hoặc sai.

Một biểu thức logic thường bao gồm các biến mệnh đề, các hằng logic, các dấu đóng, mở ngoặc và các phép toán logic.

Ta thường dùng các biến mệnh đề là p, q, p_1, q_1,... và các phép toán logic:

$\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$, $\neg$ với

Biểu thức $p \wedge q$ ký hiệu p và q (hội của p và q)

Biểu thức $p \vee q$ ký hiệu p hoặc q (tuyển của p và q)

Biểu thức $p \Rightarrow q$ ký hiệu p kéo theo q (p suy ra q)

Biểu thức $p \Leftrightarrow q$ ký hiệu p tương đương q (p kéo theo q và ngược lại)

Biểu thức $\neg p$ ký hiệu phủ định của p (không p)

Gọi giá trị đúng của biến logic là 1, giá trị sai là 0. Khi đó biểu thức $p \wedge q$ có giá trị 1 khi và chỉ khi cả p và q đều là 1. Biểu thức $p \vee q$ có giá trị 0 khi và chỉ khi cả p và q là 0. Biểu thức $p \Rightarrow q$ có giá trị 0 chỉ khi p = 1 và q = 0. Biểu thức $p \Leftrightarrow q$ có giá trị 1 khi p và q giống nhau, ngược lại có giá trị 0. Biểu thức $\neg p$ có giá trị 1 khi p = 0, ngược lại có giá trị 0.

Một biểu thức logic có thể được xây dựng từ nhiều biến logic, các hằng logic, các dấu (,) và thứ tự thực hiện phủ định, hội, tuyển, các dấu đóng, mở ngoặc theo thứ tự.

Giả sử α, β là các biểu thức logic khi đó ta có các tính chất:

(1) $\alpha \wedge \alpha = \alpha$

(2) $\alpha \vee \alpha = \alpha$

(3) $\alpha \wedge (\alpha \vee \beta) = \alpha$

(4) $\alpha \vee (\alpha \wedge \beta) = \alpha$

(5) $(\alpha \vee \gamma) \wedge (\omega \vee \beta) = (\alpha \wedge \omega) \vee (\alpha \wedge \beta) \vee (\gamma \wedge \omega) \vee (\gamma \wedge \beta)$

(6) $(\alpha \wedge \gamma) \vee (\omega \wedge \beta) = (\alpha \vee \omega) \wedge (\alpha \vee \beta) \wedge (\gamma \vee \omega) \wedge (\gamma \vee \beta)$

Để chứng minh các tính chất (1) - (6) ta chỉ cần thay giá trị α lần lượt bằng 1 và 0 vào các đẳng thức và trong mỗi trường hợp ta đều thấy hai vế của các đẳng thức là như nhau.

Lưu ý:

- Để chứng minh đẳng thức logic f(a, b, c,...z) = g(a, b, c,...z) ta chỉ cần chứng minh f và g có giá trị logic như nhau với những bộ giá trị a, b, c,...z như nhau.

- Tối giản biểu thức logic f(a, b, c,...,z) là quá trình dùng các tính chất (1),...,(6)... để rút gọn, tối giản f(a, b,..., z) về dạng đơn giản và thường chứa tuyển của các hội.

Ví dụ 1.5

Tối giản $f = (a \vee b \vee c \vee c \vee b) \wedge (a \vee b \vee c \vee d) \wedge (a \vee b \vee c \vee d) \wedge (a \vee b \vee c \vee d \vee e)$

$$= ((a \vee b \vee c) \wedge (a \vee b \vee c \vee d)) \wedge ((a \vee b \vee c \vee d) \wedge (a \vee b \vee c \vee d \vee e))$$

$$= (a \vee b \vee c \vee d) \wedge (a \vee b \vee c \vee d \vee e)$$

$$= aa \vee ab \vee ac \vee ad \vee ba \vee bb \vee bc \vee bd \vee be \vee ca \vee cb \vee cc \vee cd \vee ce$$

$$\vee\ da \vee db \vee dc \vee dd = ab \vee ac \vee ad \vee bc \vee bd \vee be \vee cd \vee ce.$$

Với $ab = a \wedge b$

CÂU HỎI VÀ BÀI TẬP CHƯƠNG 1

1.1 Cho U = {1, 2, 3, 4, 5}. Tính Card(2^U)); với 2^U là họ các tập con của U.

Một cách tổng quát cho U = {1, 2,..., n}. Tính Card(2^U).

1.2 Cho U = {1, 2, 3, 4, 5}. U có thể có bao nhiêu phân hoạch khác nhau?

Một cách tổng quát nếu U có n phần tử thì U có bao nhiêu phân hoạch khác nhau?

1.3 Cho U = {1, 2, 3, 4, 5, 6, 7, 8, 9} và ba phân hoạch của U như sau:

$P = \{p_1, p_2, p_3\} = \{\{1, 2\}, \{3, 4, 5, 6\}, \{7, 8, 9\}\}$.

$Q = \{q_1, q_2, q_3, q_4\} = \{\{1, 2\}, \{3, 4\}, \{5, 6\}, \{7, 8, 9\}\}$.

$V = \{v_1, v_2, v_3, v_4, v_5, v_6\} = \{\{1\}, \{2\}, \{3, 4\}, \{5, 6\}, \{7\}, \{8, 9\}\}$.

Hãy xây dựng các quan hệ, R_1, R_2, R_3 sao cho $P = U/R_1$; $Q = U/R_2$; $V = U/R_3$.

Chứng minh rằng $R_2 \cup R_3 \subseteq R_1$ và $R_1 \cap R_2 \cap R_3$ là quan hệ tương đương trên U. Quan hệ $R_1 \cup R_2 \cup R_3$ có là quan hệ tương đương không?

1.4 Giả sử U = {Phú, Quí, Giàu, Sang, Hùng, Mạnh, Cường, Hoa, Huệ} là tập các sinh viên trong lớp, với Hoa, Huệ là nữ còn lại là sinh viên nam. Nếu phân theo giới tính ta được phân hoạch U gồm hai nhóm: nhóm nam, nhóm nữ. Hãy liệt kê các phần tử của quan hệ R để U/R đúng bằng phân hoạch nam, nữ.

1.5 Cho U là không gian bất kỳ. R là quan hệ không tương đương trên U. Hãy giải thích tại sao R không chia được U thành các nhóm quan hệ với nhau. Nói cách khác không có phân hoạch U/R.

1.6 Cho R là quan hệ tương đương trên không gian U. Trình bày thuật toán xác định U/R.

1.7 Thử lại thuật toán sau bằng một ví dụ minh họa:

Thuật toán *xác định phân hoạch U/R*

Input Không gian U, quan hệ tương đương R

Output $U/R = \{E_1, E_2,...E_k\}$ phân hoạch tương đương của U

Algorithm

$k = 1; E_k = \phi;$

repeat

Lấy a $\in$ U; {a là đại diện bất kỳ để tạo nhóm chứa a}

$E_k = \{a\}$;

$U = U - \{a\}$;

For each b $\in$ U if (a,b) $\in$ R then

Begin

$E_k = E_k \cup \{b\}$;

$U = U - \{b\}$

End;

$k = k+1$; $E_k = \phi$;

Until $U = \phi$

1.8 Cho R và S là hai quan hệ tương đương trên U. Chứng minh rằng $R \cap S$, $R \cup S$ luôn là quan hệ tương đương trên U.

1.9 Cho $E = \{E_a, E_b, \ldots,\}$ là một phân hoạch của A; với E_a là nhóm chứa a

Cho ánh xạ f: $2^U \rightarrow 2^U$ với $f(X) = \cup\{E_a: E_a \subseteq X\}$.

Chứng minh rằng $f(X) \subseteq X$.

Cho các ánh xạ g, g': $2^U \rightarrow 2^U$

với $g(X) = \cup\{E_a: a \in X\}$; $g'(X) = \cup\{E_a: E_a \cap X \neq \phi\}$.

Chứng minh rằng g, g' là các ánh xạ đóng và g = g'.

1.10 Cho G là giàn giao trên U.

Cho ánh xạ f: G $\rightarrow$ G với $f(X) = \cap\{Y: Y \in G \;\&\; X \subseteq Y\}$.

Chứng minh rằng f là ánh xạ đóng.

1.11 Cho hai tập U, B bất kỳ. $R \subseteq U \times B$ là quan hệ trên U, B. Với mỗi tập X của 2^U ta có tập X* gồm các y của B mà xRy với mọi x của X:

$X^* = \{y \in B: (x,y) \in R \;\forall\; x \in X\}$.

Chứng minh rằng ánh xạ f: $2^U \rightarrow 2^U$ với $f(X) = X^*$ là ánh xạ đóng khi B = U và R là quan hệ tương đương.

1.12 Cho U là tập bất kỳ được gọi là tập biến. F là tập các quy tắc (hay tập công thức) trên U. Giả sử $x \in U$; $X \subseteq U$. Ta nói x *là giải được từ X*, ký hiệu $X \rightarrow x$

nếu sau một số lần sử dụng các quy tắc trong F vào X ta có x. Đặt X^+ là tập các biến giải được từ X.

Gọi ánh xạ g: $2^U \rightarrow 2^U$ mà $g(X) = X^+$. Chứng minh rằng g là ánh xạ đóng

1.13 Cho $A = \{A_1, A_2,..., A_m\}$; $B = \{B_1, B_2,..., B_n\}$; $R \subseteq A \times B$ là quan hệ trên A, B. Để xác định các phần tử của R ta có thể dùng phương pháp liệt kê các cặp (a,b) thuộc R hoặc dùng bảng hai chiều và tại giao điểm của a và b ta ghi số 1 nếu $(a,b) \in R$ ngược lại ghi số 0.

Ví dụ cho $A = \{a, b, c\}$; $B = \{x, y, z, v\}$; $R \subseteq A \times B$ cho trong bảng sau:

	x	y	z	v
a	0	0	1	1
b	0	1	0	0
c	1	1	0	1

Xét $X \subseteq B$. Ta đặt $\rho(X)$ là tập các phần tử của A quan hệ với các phần tử trong X, tức $\rho(X) = \{a \in A: (a,x) \in R \ \forall \ x \in X\}$. Ta gọi *độ thường xuyên* (hay *độ phổ biến*) của X trong A, ký hiệu sp(X) là tỷ số Card(ρ(X))/Card(A).

Vậy sp(X) = Card(ρ(X))/Card(A). Ví dụ nhìn vào bảng trên, độ thường xuyên của $\{x, z\}$ là 0, độ thường xuyên của $\{x, y\}$ trong A là 1/3; $sp(\{x, y\}) = 1/3$ ...

Nếu $U = \{o_1, o_2,..., o_m\}$ là tập các hóa đơn; $B = \{i_1, i_2,..., i_n\}$ là tập hàng. Khi đó sp(X) là độ thường xuyên của tập hàng X trong tập hóa đơn U.

Chứng minh rằng $sp(X \cup Y) \leq sp(X)$ và $sp(X \cup Y) \leq sp(Y)$

1.14 Cho $\alpha, \beta, \gamma, \delta$ là các biến logic. Chứng minh các đẳng thức sau:

(1) $\alpha \vee \beta = \beta \vee \alpha$

(2) $\alpha \wedge \beta = \beta \wedge \alpha$

(3) $\alpha \wedge (\beta \vee \gamma) = (\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$

(4) $\alpha \vee (\beta \wedge \gamma) = (\alpha \vee \beta) \wedge (\alpha \vee \gamma)$

1.15 Cho tập U gồm chín đồ chơi được đánh số từ 1 đến 9 như sau:

$U = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

Trong các đồ chơi có các màu: trắng, đen, đỏ; kích thước: lớn, nhỏ; hình dáng: tam giác, vuông, tròn.

Gọi R_1 là quan hệ cùng màu; R_2 là quan hệ cùng kích thước; R_3 là quan hệ cùng hình dáng.

a. Chứng minh rằng R_1, R_2, R_3 là các quan hệ tương đương.

b. Giả sử chia theo màu ta được ba nhóm trắng, đen, đỏ

$U/R_1 = \{\{1, 2, 3, 4\}, \{5, 6\}, \{7, 8, 9\}\}$.

Chia U theo kích thước ta được hai nhóm lớn, nhỏ

$U/R_2 = \{\{1, 2, 3, 4, 5, 6\}, \{7, 8, 9\}\}$.

Chia theo hình dáng ta được ba nhóm tam giác, vuông, tròn

$U/R_3 = \{\{1, 2, 3\}, \{4, 5, 6\}, \{7, 8, 9\}\}$.

Hãy thực hiện các yêu cầu sau đây:

- Liệt kê R_1, R_2, R_3
- Liệt kê $R_1 \cap R_2$; $R = R_1 \cap R_2 \cap R_3$
- Liệt kê các nhóm của phân hoạch U/R
- Liệt kê các nhóm của $U/R_1 \cap U/R_2 \cap U/R_3$

Chương 2

CƠ SỞ DỮ LIỆU QUAN HỆ

2.1. MỞ ĐẦU

Cơ sở dữ liệu quan hệ là tập các dữ liệu được biểu thị ở dạng bảng nên mô hình cơ sở dữ liệu quan hệ là một mô hình được sử dụng rộng rãi trong đời sống xã hội của mọi tổ chức, cơ quan, xí nghiệp, doanh nghiệp, nơi nào cần quản lý và xử lý các thông tin, nơi đó có dùng các bảng dữ liệu, tức dùng cơ sở dữ liệu quan hệ. Trong chương 4 chúng ta sẽ xét khái niệm hệ tin và vì quan hệ là một hệ tin nên ta nghiên cứu quan hệ như nghiên cứu một lớp các hệ tin. Vậy cơ sở dữ liệu là gì? Và cơ sở dữ liệu quan hệ (CSDLQH) là gì? Ta hiểu *một cơ sở dữ liệu là tập các dữ liệu liên quan đến một bài toán quản lý. Một cơ sở dữ liệu quan hệ là tập các dữ liệu dạng bảng liên quan đến một bài toán quản lý.*

Ví dụ 2.1:

Xét hồ sơ cán bộ của một cơ quan:

TT	MS	TÊN	NS	QUÊ	GT	LƯƠNG
1	01	Huy	1945	Hànội	Nam	300
2	02	Tiến	1950	Hảiphòng	Nam	400
3	03	Lan	1960	Namhà	Nữ	200
4	04	Hiền	1965	Hảidương	Nữ	250

(TT là thứ tự, MS: mã số, NS: năm sinh, GT: giới tính, LƯƠNG là lương của cán bộ tương ứng...).

Ví dụ 2.2:

Xét sổ theo dõi khách của một khách sạn: có MK là mã khách, NR là số phòng, Sốngười là số người, Tiền là số tiền khách hàng phải trả.

MK	NR	Sốngười	Tiền
101	301	2	400
102	302	1	200
103	303	3	600
104	304	2	400
105	304	3	600

Trong hai ví dụ trên tuy quản lý các mảng thông tin (dữ liệu) khác nhau nhưng cả hai đều có chung một đặc thù: dữ liệu được mô tả dưới dạng *bảng*, mỗi bảng có một dòng đầu là dòng các *thuộc tính*.

Trong ví dụ 2.1 các *thuộc tính* là: {TT, ms, tên, ns, quê, gt, lương}.

Trong ví dụ 2.2 các *thuộc tính* là: {MK, NR, Sốngười, Tiền}.

Mỗi thuộc tính có một miền giá trị của nó, ví dụ thuộc tính năm sinh (NS) có miền giá trị là các số nguyên: 1945, 1950, 1960, 1965,... thuộc tính TÊN có miền giá trị là các xâu ký tự (string): Minh, Tiến, Lan, Hiền,...

Trong mỗi ví dụ ở trên mỗi bảng đều có một số phần tử, ví dụ bảng hồ sơ cán bộ của cơ quan có bốn phần tử, bảng theo dõi khách ở khách sạn có năm phần tử, mỗi phần tử là một dòng, một bản ghi. Về sau các bảng dữ liệu được mô tả dưới dạng bảng như vậy sẽ được gọi là các *quan hệ*.

2.2. ĐỊNH NGHĨA QUAN HỆ

Cho tập hữu hạn khác rỗng các phần tử $A = \{A_1, A_2, \ldots A_n\}$ được gọi là tập các *thuộc tính*.

Mỗi A_i của tập A có *miền giá trị (miền trị)* $D(A_i)$.

Định nghĩa 2.1 Quan hệ trên tập thuộc tính A

Mỗi tập con R của tích Descartes (Đề-các) các miền giá trị $D(A_i)$ với i = 1, 2, 3,..., n được gọi là một *quan hệ* trên A.

Vậy R là *quan hệ* trên tập thuộc tính A nếu: $R \subseteq D(A_1) \times D(A_2) \times \ldots \times D(A_n)$.

Từ định nghĩa ta thấy tích Đề-các $D(A_1) \times D(A_2) \ldots \times D(A_n)$ có rất nhiều tập con nên trên A có nhiều quan hệ khác nhau.

Ví dụ 2.3:

Giả sử A = {E, B, C}, D(E) = {0,1}, D(B) = {a, b, c}, D(C) ={x,y}.

Tích Đề-các D(E) × D(B) × D(C) = {(0,a,x), (0,a,y), (0,b,x), (0,b,y), (0,c,x), (0,c,y), (1,a,x), (1,a,y), (1,b,x), (1,b,y), (1,c,x), (1,c,y)}.

Tích D(E) × D(B) × D(C) có 12 phần tử và nó có 2^{12} tập con khác nhau nên trên A = {E,B,C} ta có 2^{12} quan hệ R khác nhau. Ví dụ $R_0 = \varnothing$ là quan hệ rỗng, R_1 = {(0,a,x)}, R_1' = {(0,b,x)} là các quan hệ chứa 1 phần tử, quan hệ R_2 = {(0,a,x), (0,b,x)} là quan hệ chứa 2 phần tử, còn quan hệ R = {(0,a,x), (0,a,y), (0,b,x), (0,b,y), (0,c,x), (0,c,y), (1,a,x), (1,a,y), (1,b,x), (1,b,y), (1,c,x), (1,c,y)} là quan hệ chứa 12 phần tử. Qua ví dụ, ta thấy mỗi phần tử của quan hệ R là một bộ của tích Đề-các D(E) × D(B) × D(C).

Trong CSDL quan hệ, để cho tiện và dễ hình dung với các bài toán quản lý ta viết mỗi quan hệ R trên A dưới dạng bảng. Dòng đầu của bảng là dòng các thuộc tính, các dòng sau của bảng là các bộ của quan hệ. Ví dụ với quan hệ R_0 trên A = {E, B, C} không chứa phần tử nào ta viết:

R_0

E B C hay viết $R_0(E,B,C) = \varnothing$

Hoặc các quan hệ chứa 1 phần tử R_1, R_1' được viết:

R_1			**R_1'**		
E	B	C	E	B	C
0	a	x	0	b	x

Quan hệ chứa 2 phần tử R_2:

R_2

E	B	C
0	a	x
0	b	x

Một cách tổng quát từ định nghĩa ta thấy nếu cho trước tập thuộc tính $A = \{A_1, A_2,..., A_n\}$ thì quan hệ R trên A là một bảng hai chiều, trên cột thứ i là các giá trị của $D(A_i)$, trên mỗi dòng của bảng là bộ n giá trị của các miền giá trị của các thuộc tính $A_1, A_2,... A_n$. Mỗi dòng là một phần tử của quan hệ.

Vậy một quan hệ R trên A là bảng (không có hai dòng giống nhau) có dạng:

Bảng 2.1: Bảng quan hệ R

R

A_1	A_2	...	A_j	...	A_n
a_{11}	a_{12}	...	a_{1j}		a_{1n}
...	...	...	...	...	
a_{i1}	a_{i2}	...	a_{ij}		a_{in}
...	...	...	...	...	...
a_{m1}	a_{m2}	...	a_{mj}	...	a_{mn}

Trong bảng 2.1 mỗi bộ $t_i = (a_{i1}, a_{i2},...a_{ij},...,a_{in})$, với i = 1, 2,..., m; là một phần tử của R và $R = \{t_1, t_2,...t_m\}$.

Lưu ý:

Khi trình bày để cho nhanh chúng ta không kẻ bảng mà viết theo cột, dòng tương ứng.

Hoặc đôi khi ta viết quan hệ R dưới dạng bảng 2.2 như sau:

Bảng 2.2: Bảng quan hệ

R	A_1	A_2	...	A_i	...	A_n
t_1	a_{11}	a_{12}	...	a_{1i}		a_{1n}
...	...	...	...	...	...	
t_i	a_{i1}	a_{i2}	...	a_{ii}		a_{in}
...	...	...	...	...	...	...
t_m	a_{m1}	a_{m2}	...	a_{mi}	...	a_{mn}

Khi đó $R = \{t_1, t_2, \ldots, t_m\}$, $t_i = (a_{i1}, a_{i2}, a_{i3}, \ldots, a_{in})$, $i = 1, 2, \ldots, n$

Để ký hiệu một quan hệ nào đó trên tập thuộc tính $A = \{A_1, A_2, \ldots A_n\}$ ta dùng R hoặc r.

Ký hiệu $R(A_1, A_2, \ldots A_n)$ dùng để chỉ một quan hệ R trên $A = \{A_1, A_2, \ldots A_n\}$, ví dụ cho quan hệ sinh viên sv(MASV, HOTEN, NGAYSINH, QUE, GIOITINH).

Từ định nghĩa và trực quan của quan hệ R trên A ta thấy khi cho tập thuộc tính $A = \{A_1, A_2, \ldots A_n\}$ là đã cho khuôn hoặc *lược đồ* của quan hệ, và thú vị hơn ta có thể coi như đã cho một quan hệ, đó là quan hệ rỗng. Vậy khi nói cho tập thuộc tính $A = \{A_1, \ldots, A_n\}$ ta coi như cho trước *lược đồ quan hệ* (LĐQH). Lược đồ đó có dạng sau:

Bảng 2.3: Lược đồ quan hệ

R

A_1	A_2	...		A_n
		...		

Lưu ý:

- Quan hệ R là một tập hợp nên trong một quan hệ ta coi hai dòng giống nhau chỉ là một phần tử của R. Đồng thời với cách biểu diễn quan hệ như một bảng thì thứ tự trước sau của các dòng (cột) không làm thay đổi bản chất quan hệ.

- Mọi quan hệ R trên tập thuộc tính $A = \{A_1, A_2, \ldots, A_n\}$ lúc nào cũng có thể biểu diễn bằng bảng như 2.1 hoặc bảng 2.2 nhưng điều ngược lại không đúng, nghĩa là có nhiều bảng không phải là quan hệ trên A. Đó là những bảng có các dòng giống nhau.

Ví dụ xét bảng R sau:

R

HOTEN	NS	QUE	KHOA
LAN ANH	1980	Hn	CNTT
LAN ANH	1980	Hn	CNTT
LAN ANH	1980	Hn	CNTT

Đây không phải là quan hệ R trên A = {HOTEN, NS, QUE, KHOA}.

Vì với tư cách quan hệ là một tập hợp, quan hệ R không có 3 phần tử giống nhau, mặc dù trong đời sống thực có thể có rất nhiều LAN ANH cùng sinh năm 1980, cùng quê Hn, cùng khoa CNTT.

Trong thực tế quản lý nếu gặp các trường hợp như trên ta thêm thuộc tính mã cho các sinh viên để phân biệt các sinh viên với nhau.

Ví dụ 2.4:

Xét danh sách sinh viên

Cho A = {MASV, HOTEN, QUE, NS, KHOA} và quan hệ R như sau:

R

MASV	HOTEN	NS	QUE	KHOA
01	LAN ANH	1980	Hn	CNTT
02	LAN ANH	1980	Hn	CNTT
03	LAN ANH	1980	Hn	CNTT

Quan hệ R có 4 phần tử, có 3 sinh viên LAN ANH cùng quê, cùng năm sinh, cùng khoa.

Ví dụ 2.5:

Cho các quan hệ R và S như sau:

Hình 2.1. Quan hệ S và R

Vậy R = $\{t_1, t_2\}$ và t_1 = (sv1, An, 02/01/1986)

t_2 = (sv5, Phu, 02/02/1986)

S = $\{t_1', t_2', t_3', t_4'\}$ và $t_1' = t_1$ = (sv1, An, 02/01/1986)

t_2' = (sv2, Khang, 03/01/1986)

t_3' = (sv3, Thinh, 02/02/1988)

t_4' = (sv4, Vuong, 02/01/1984)

Từ đây ta sẽ ký hiệu t là phần tử của R.

2.3. CÁC PHÉP TOÁN ĐẠI SỐ QUAN HỆ

Như ta đã biết quan hệ là một tập hợp nên trên các quan hệ ta có thể thực hiện các phép toán tập hợp như hợp, giao, hiệu,... các phép toán đó người ta thường gọi là *các phép toán đại số quan hệ (ĐSQH).*

Định nghĩa 2.2 Các quan hệ tương thích

Ta nói hai quan hệ r_1 và r_2 là *tương thích* nếu chúng cùng xác định trên tập thuộc tính A.

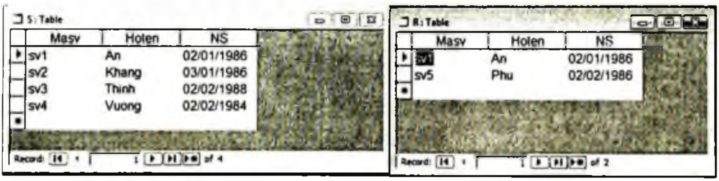
2.3.1. Phép hợp

Định nghĩa 2.3

Hợp của hai quan hệ tương thích r_1 và r_2, ký hiệu $r_1 + r_2$ hoặc $r_1 \cup r_2$ là một quan hệ trên A gồm các phần tử thuộc r_1 hoặc r_2. Tức là: $r_1 \cup r_2 = \{t: t \in r_1 \text{ hoặc } t \in r_2\}$.

Ví dụ 2.6:

Cho hai quan hệ S và R như sau:



S : Table

Masv	Holen	NS
sv1	An	02/01/1986
sv2	Khang	03/01/1986
sv3	Thinh	02/02/1988
sv4	Vuong	02/02/1984

Record: 1 of 4

R : Table

Masv	Holen	NS
sv1	An	02/01/1986
sv5	Phu	02/02/1986

Record: 1 of 2

Hình 2.2. Quan hệ S và R

Khi đó ta có quan hệ R $\cup$ S:

R $\cup$ S

Masv	Hoten	NS
sv1	An	02/01/1986
sv2	Khang	03/01/1986
sv3	Thinh	02/02/1988
sv4	Vuong	02/02/1984
sv5	Phu	02/02/1986

Từ định nghĩa ta thấy ngay rằng:

$\forall\, r_1, r_2$ thì $r_1 \cup r_2 = r_2 \cup r_1$

$\forall\, r$ thì $r \cup r = r$

Một cách tổng quát có thể lấy hợp của m quan hệ tương thích. Cho m quan hệ tương thích $r_1, r_2, r_3, \ldots r_m$.

Hợp của các quan hệ $r_1, r_2, r_3, \ldots r_m$ là một quan hệ $r_1 \cup r_2 \cup \ldots \cup r_m$ gồm các phần tử thuộc r_1 hoặc r_2 hoặc r_3 hoặc... r_m.

Vậy $r_1 \cup r_2 \cup \ldots \cup r_m = \{t: t \in r_1$ hoặc $t \in r_2 \ldots$ hoặc $t \in r_m\}$

2.3.2. Phép giao

Cho $A = \{A_1, A_2, \ldots, A_n\}$, r_1 và r_2 là 2 quan hệ trên A.

Định nghĩa 2.4

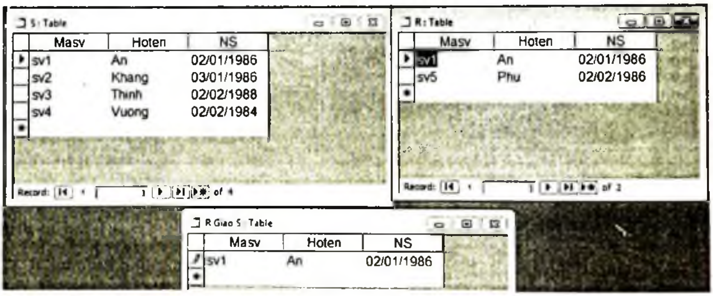
Giao của hai quan hệ r_1 và r_2 ký hiệu: $r_1 * r_2$ hoặc $r_1 \cap r_2$ là một quan hệ trên A gồm các phần tử chung của r_1 và r_2.

Vậy: $r_1 \cap r_2 = \{t: t \in r_1$ và $t \in r_2\}$.

Ví dụ 2.7:

Cho hai quan hệ R và S như trong hình 2.3:

Khi đó $R \cap S$ = R giao S.



S: Table

Masv	Hoten	NS
sv1	An	02/01/1986
sv2	Khang	03/01/1986
sv3	Thinh	02/02/1988
sv4	Vuong	02/02/1984

Record: 1 of 4

R: Table

Masv	Hoten	NS
sv1	An	02/01/1986
sv5	Phu	02/02/1986

Record: 1 of 2

R Giao S: Table

Masv	Hoten	NS
sv1	An	02/01/1986

Hình 2.3. Giao của hai quan hệ S và R

Hoặc cho hai quan hệ r_1 và r_2:

r_1

A_1	B	C	D
a_1	b_1	c_1	d_1
a	b	c	d
a_2	b_2	c_2	d_2
a_3	b_3	c_3	d_3

r_2

A_1	B	C	D
a	b	c	d
a_2	b_2	c_2	d_2
x	y	z	v

Khi đó ta có quan hệ giao:

$r_1 \cap r_2$

a	b	c	d
a	b	c	d
a_2	b_2	c_2	d_2

2.3.3. Phép trừ (hiệu)

Cho hai quan hệ r_1 và r_2 trên A.

Định nghĩa 2.5

Hiệu của r_1 và r_2 ký hiệu $r_1 - r_2$ là một quan hệ trên A gồm các phần tử thuộc r_1 và không thuộc r_2.

Vậy $r_1 - r_2 = \{t: t \in r_1 \text{ và } t \notin r_2\}$.

Ví dụ lấy r_1 và r_2 như sau:

r_1

A_1	B	C	D
a_1	b_1	c_1	d_1
a_3	b_3	c_3	d_3
a_4	b_4	c_4	d_4
a_5	b_5	c_5	d_5
a_6	b_6	c_6	d_6
a_7	b_7	c_7	d_7
a_8	b_8	c_8	d_8
a_9	b_9	c_9	d_9

r_2

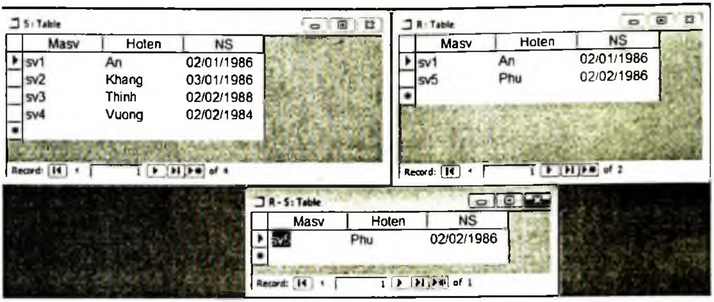
A_1	B	C	D
x	y	z	v
a_3	b_3	c_3	d_3

Khi đó ta có $r_2 - r_1$ là:

$r_2 - r_1$

A_1	B	C	D
x	y	z	v

Hoặc lấy S và R như sau và ta sẽ có R - S tương ứng trong hình 2.4



Hình 2.4. Hiệu của hai quan hệ R và S

2.3.4. Phép chiếu

Cho A = {A_1, A_2,... A_n} và r là một quan hệ trên A, X ⊂ A. Ta xét quan hệ con của quan hệ r chỉ trên tập thuộc tính X, đó là chiếu của r lên X.

Định nghĩa 2.6

Chiếu của r lên tập thuộc tính X, ký hiệu r.X hoặc r[X], là một quan hệ trên lược đồ X gồm các phần tử của r sau khi đã lược bỏ các thuộc tính không thuộc tập X.

Tương tự với r.X, các phần tử của r.X là những phần tử ký hiệu là t.X (hoặc t[X]) chính là chiếu của t lên X. Vậy r.X = {t.X: t ∈ r}, t.X là chiếu của phần tử t lên tập thuộc tính X.

Lưu ý: r.X là thu gọn của r lên tập thuộc tính X.

Ví dụ 2.8:

Cho quan hệ r như sau:

r

A_1	B	C	D	E	F	G
a_1	b_1	c_1	d_1	e_1	f_1	g_1
a_2	b_2	c_2	d_2	e_2	f_2	g_2
a_3	b_3	c_3	d_3	e_3	f_3	g_3
a_4	b_4	c_4	d_4	e_4	f_4	g_4

Giả sử ta có X = {B, C}, Y = {F, G}.

Khi đó ta có hai quan hệ con chiếu của r lên X và Y tương ứng:

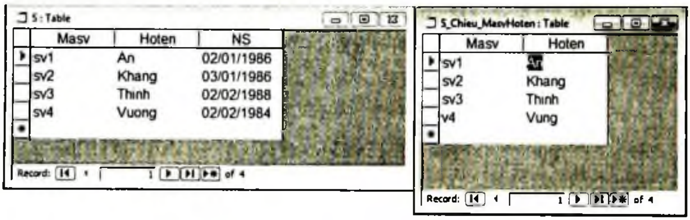
r. X

B	C
b_1	c_1
b_2	c_2
b_3	c_3
b_4	c_4

r. Y

F	G
f_1	g_1
f_2	g_2
f_3	g_3
f_4	g_4

Hoặc cho S như sau và chiếu của S lên Masv và Hoten ta được quan hệ S_MasvHoten như trong hình 2.5.



Hình 2.5. Minh họa chiếu của quan hệ

2.3.5. Phép Tích Đề-các (Descartes)

Tích Đề-các của hai quan hệ ta chỉ xét trên các tập thuộc tính rời nhau.

Cho r_1 là quan hệ trên $A = \{A_1, A_2,\dots A_n\}$,

Cho r_2 là quan hệ trên $A' = \{B_1, B_2,\dots B_m\}$. Với $A \cap A' = \varnothing$.

Định nghĩa 2.7

Tích Đề-các của r_1 và r_2 ký hiệu: $r_1 \times r_2$ là quan hệ trên lược đồ $A \cup A'$ gồm các phần tử tạo ra từ tích Đề-các của hai tập r_1 và r_2.

Vậy quan hệ $r_1 \times r_2$ là quan hệ trên lược đồ:

$$A \cup A' = \{A_1, A_2,\dots A_n, B_1, B_2,\dots B_m\}$$

với $r_1 \times r_2 = \{(t_1, t_2): t_1 \in r_1, t_2 \in r_2\}$.

Ví dụ 2.9:

Cho r_1 và r_2 như sau:

r_1

B	C	D
b_1	c_1	d_1
b_2	c_2	d_2

r_2

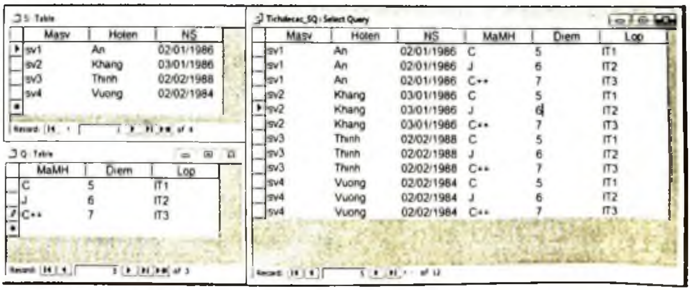
E	F	G
e_1	f_1	g_1
e_2	f_2	g_2
e_3	f_3	g_3
e_4	f_4	g_4
e_5	f_5	g_5

Như vậy r_1 có hai phần tử, r_2 có ba phần tử, tích Đề-các $r_1 \times r_2$ sẽ có sáu phần tử.

$r_1 \times r_2$

B	C	D	E	F	G
b_1	c_1	d_1	e_1	f_1	g_1
b_1	c_1	$đ_1$	e_2	f_2	g_2
b_1	c_1	d_1	e_3	f_3	g_3
b_2	c_2	d_2	e_1	f_1	g_1
b_2	c_2	d_2	e_2	f_2	g_2
b_2	c_2	d_2	e_3	f_3	g_3
b_1	c_1	d_1	e_4	f_4	g_4
b_2	c_2	d_2	e_4	f_4	g_4
b_1	c_1	d_1	e_5	f_5	g_5
b_2	c_2	d_2	e_5	f_5	g_5

hoặc cho 2 quan hệ S có 4 phần tử, Q có 3 phần tử như trong hình 2.6 sau khi lấy tích Đề-các ta được quan hệ Tichdecac_SQ có 12 phần tử.


S : Table

Masv	Hoten	NS
sv1	An	02/01/1986
sv2	Khang	03/01/1986
sv3	Thinh	02/02/1988
sv4	Vuong	02/02/1984

Q : Table

MaMH	Diem	Lop
C	5	IT1
J	6	IT2
C++	7	IT3

Tichdecac_SQ : Select Query

Masv	Hoten	NS	MaMH	Diem	Lop
sv1	An	02/01/1986	C	5	IT1
sv1	An	02/01/1986	J	6	IT2
sv1	An	02/01/1986	C++	7	IT3
sv2	Khang	03/01/1986	C	5	IT1
sv2	Khang	03/01/1986	J	6	IT2
sv2	Khang	03/01/1986	C++	7	IT3
sv3	Thinh	02/02/1988	C	5	IT1
sv3	Thinh	02/02/1988	J	6	IT2
sv3	Thinh	02/02/1988	C++	7	IT3
sv4	Vuong	02/02/1984	C	5	IT1
sv4	Vuong	02/02/1984	J	6	IT2
sv4	Vuong	02/02/1984	C++	7	IT3

Hình 2.6. Minh họa tích Đề-các

2.3.6. Phép nối tự nhiên

Cho r_1 và r_2 là hai quan hệ tương ứng trên A và A'.

Định nghĩa 2.8

Phép nối tự nhiên của r_1 và r_2, ký hiệu r_1 |><| r_2 là quan hệ trên lược đồ $A \cup A'$ gồm các phần tử t mà chiếu của t lên A là phần tử thuộc r_1 còn chiếu của t lên A' là phần tử của r_2.

Vậy r_1 |><| $r_2 = \{t: t.A \in r_1$ và $t.A' \in r_2\}$

Ví dụ 2.10:

Cho r_1 và r_2 là hai quan hệ sau:

$\mathbf{r_1}$

B	C	D
b_1	c_1	d_1
b_2	c_2	d_2
b_3	c_3	d_3
2	3	4

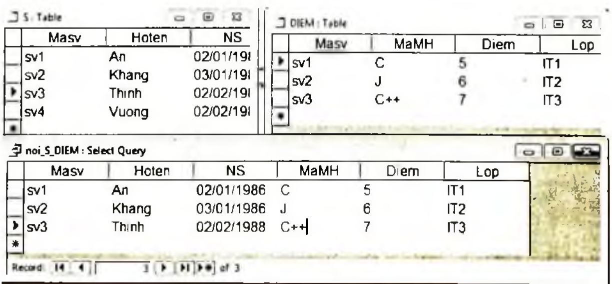
$\mathbf{r_2}$

C	D	E	F
c_1	d_1	e_1	f_1
c_2	d_2	e_2	f_2
x	y	z	v

Nối tự nhiên của r_1 và r_2: $\mathbf{r_1}$ **|><|** $\mathbf{r_2}$

B	C	D	E	F
b_1	c_1	d_1	e_1	f_1
b_2	c_2	d_2	e_2	f_2

Hoặc cho S và DIEM như trong hình 2.7 khi nối ta được bảng noi_S_DIEM



S : Table

Masv	Hoten	NS
sv1	An	02/01/198
sv2	Khang	03/01/198
sv3	Thinh	02/02/198
sv4	Vuong	02/02/198

DIEM : Table

Masv	MaMH	Diem	Lop
sv1	C	5	IT1
sv2	J	6	IT2
sv3	C++	7	IT3

noi_S_DIEM : Select Query

Masv	Hoten	NS	MaMH	Diem	Lop
sv1	An	02/01/1986	C	5	IT1
sv2	Khang	03/01/1986	J	6	IT2
sv3	Thinh	02/02/1988	C++	7	IT3

Record: 3 of 3

Hình 2.7. Minh họa phép nối tự nhiên

Ví dụ 2.11: Xét hai quan hệ cùng tập thuộc tính:

r_1

B	C	D	E
b_1	c_1	d_1	e_1
b_2	c_2	d_2	e_2
b_3	c_3	d_3	e_3

r_2

B	C	D	E
y_1	z_1	w_1	v_1
b_1	c_1	d_1	e_1
y_2	z_2	w_2	v_2

r_1 |><| r_2

B	C	D	E
b_1	c_1	d_1	e_1
b_2	c_2	d_2	e_2
b_3	c_3	d_3	e_3
b_4	c_4	d_4	e_4
b_5	c_5	d_5	e_5
b_6	c_6	d_6	e_6

Vậy trong trường hợp hai tập thuộc tính như nhau thì r_1 |><| $r_2 = r_1 \cap r_2$

Sau đây ta xét ví dụ mà các tập thuộc tính rời nhau.

Cho hai quan hệ r_1 và r_2 như sau:

r_1

B	C	D	E
b_1	c_1	d_1	e_1
b_2	c_2	d_2	e_2
b_3	c_3	d_3	e_3

r_2

F	G	H
f_1	g_1	h_1
f_2	g_2	h_2
f_3	g_3	h_3
x	y	z

Trong trường hợp này ta có r_1|><| r_2 như sau:

r_1 |><| r_2

B	C	D	E	F	G	H
b_1	c_1	d_1	e_1	f_1	g_1	h_1
b_1	c_1	d_1	e_1	f_2	g_2	h_2
b_1	c_1	d_1	e_1	f_3	g_3	h_3
b_1	c_1	d_1	e_1	x	y	z
b_2	c_2	d_2	e_2	f_1	g_1	h_1

b_2	c_2	d_2	e_2	f_2	g_2	h_2
b_2	c_2	d_2	e_2	f_3	g_3	h_3
b_2	c_2	d_2	e_2	x	y	z
b_3	c_3	d_3	e_3	f_1	g_1	h_1
b_3	c_3	d_3	e_3	f_2	g_2	h_2
b_3	c_3	d_3	e_3	f_3	g_3	h_3
b_3	c_3	d_3	e_3	x	y	z

Vậy trong trường hợp $A \cap A' = \varnothing$ thì $r_1 |><| r_2 = r_1 \times r_2$.

Chúng ta cần lưu ý rằng các phép toán tích Đề-các và phép nối là những phép toán ghép các bảng vào với nhau, trong các bài toán quản lý ta thường dùng phép nối để truy xuất dữ liệu.

2.3.7. Phép chia

Cho r là quan hệ trên $A = \{A_1, A_2, \ldots A_n\}$; s là quan hệ trên A' và $A' \subset A$.

Định nghĩa 2.9

Phép chia của quan hệ r cho quan hệ s, ký hiệu $r \div s$ hoặc r/s là quan hệ trên lược đồ A\ A' gồm các phần tử t[A \A'] mà t thuộc r và t[A'] thuộc quan hệ s.

Vậy $r / s = \{t[A \setminus A'] : t \in r \text{ và } t[A'] \in s\}$.

Ví dụ 2.12:

Cho r là quan hệ trên A = {H, B, C, D, E, G}, s là quan hệ trên A' = {H, E, G} như sau:

r

H	B	C	D	E	G
a_1	b_1	c_1	d_1	e_1	g_1
a_2	b_2	c_2	d_2	e_2	g_2
a_3	b_3	c_3	d_3	e_3	g_3
a_4	b_4	c_4	d_4	e_4	g_4

s

H	E	G
a_1	e_1	g_1
x	y	z
a_3	e_3	g_3

Khi đó:

r/s

B	C	D
b_1	c_1	d_1
b_3	c_3	d_3

Từ định nghĩa ta thấy để có thể thực hiện được phép chia r/s thì A' phải là lược đồ con thực sự của A.

2.3.8. Phép chọn

Trong xử lý các CSDL dạng bảng (quan hệ) một phép toán ta thường dùng để truy vấn dữ liệu đó là phép chọn. Phép chọn tức là chọn từ bảng R ra các phần tử thỏa mãn điều kiện E nào đó.

Cho quan hệ r trên A. Cho E là mệnh đề logic trên các thuộc tính của A. Phần tử t thuộc r thỏa mãn điều kiện E ta ký hiệu t(E).

Định nghĩa 2.10

Phép chọn trong quan hệ R theo điều kiện E cho ta một quan hệ trên lược đồ A ký hiệu R(E) hoặc $\sigma_E(R)$ và chứa các phần tử của R thỏa mãn điều kiện E.

Vậy $\sigma_E(R) = \{t(E): t \in R\}$.

Ví dụ 2.13:

Xét hồ sơ kết quả thi của sinh viên.

Giả sử ta có quan hệ r như sau:

r

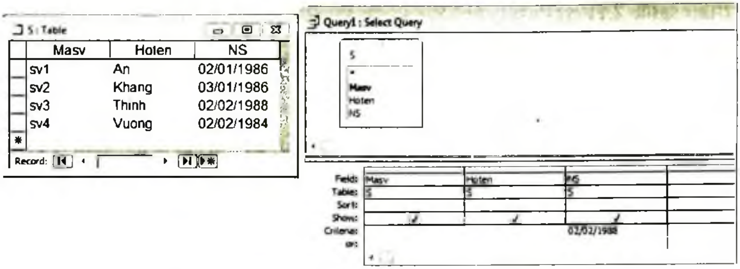
TT	HOTEN	NS	ĐIEMCSDL	ĐIEMFOX
1	Tuấn Anh	1974	7	5
2	Huy Công	1974	8	3
3	Th. Hương	1975	8	9
4	Bình Minh	1976	2	3

Giả sử điều kiện E là sinh viên có ít nhất một điểm kém. Vậy:

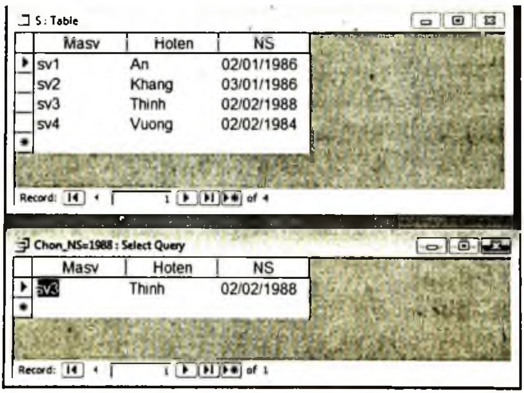
σ_E(R)

TT	HOTEN	NS	ĐIEMCSDL	ĐIEMFOX
2	Huy Công	1974	8	3
4	Bình Minh	1976	2	3

Hoặc cho bảng S như trong màn hình (hình 2.8) và chọn những bộ có năm sinh 2008. Ta được kết quả có trong bảng chon_NS = 2008:



Hình 2.8. Minh họa phép chọn



Hình 2.9. Minh họa kết quả phép chọn

2.3.9. Phép kết nối theo θ

Chúng ta sẽ xét phép kết nối theo toán tử θ, với θ là một toán tử so sánh số học hai ngôi $\{=, <, >, \leq, \geq, \neq\}$.

Cho r và s là hai quan hệ tương ứng trên các lược đồ rời nhau A và A', với $A \cap A' = \varnothing$ và $A = \{A_1,...A_i,..., A_n\}$, $A' = \{B_1,...B_j,...., B_m\}$

Định nghĩa 2.11

Phép nối theo θ của các quan hệ r và s, ký hiệu r $|><|_{i\theta j}$s là một quan hệ trên lược đồ A ∪ A' gồm những bộ thuộc tích Đề-các của r và s sao cho thành phần thứ i của quan hệ r thỏa mãn phép toán θ với thành phần thứ j của quan hệ s.

Vậy kết nối θ: r $|><|_{i\theta j}$s là chọn trong r × s các bộ mà các thành phần thứ i, j của các quan hệ r, s tương ứng thỏa mãn θ, tức là:

$$r\ |><|_{i\theta j}s = \{t \in r \times s: t\ (A_i \theta\ B_j)\}.$$

Đây là phép kết nối gắn với tích Đề-các của r và s.

Ví dụ 2.14:

Giả sử r và s là các quan hệ như sau:

r

H	B	C
1	2	3
4	5	6
7	8	9

s

D	E
3	1
6	2

Khi đó ta có:

r $|><|_{2<1}$s

H	B	C	D	E
1	2	3	3	1
1	2	3	6	2
4	5	6	6	2

Ví dụ 2.15:

Giả sử r và s là các quan hệ:

r

H	B	C
1	2	3
a	b	c
x	y	z

s

D	E	F
1	e	f
a	e	f
5	6	7

Khi đó:

r |><|$_{1=1}$ s

H	B	C	D	E	F
1	2	3	1	e	f
a	b	c	a	e	f

2.3.10. Phép nối nửa

Cho các quan hệ r và s trên các lược đồ A và A' tương ứng.

Định nghĩa 2.12

Nối nửa của các quan hệ r và s, ký hiệu r |>< s là một quan hệ trên lược đồ A gồm các bộ của r |><| s chiếu lên A.

Vậy r |>< s = {t: t $\in$ (r |><| s).A} = {t.A: t $\in$ r |><| s}.

Tức r |>< s = (r |><| s).A

Ví dụ 2.16:

Giả sử r và s là các quan hệ:

r

H	B	C
a	b	c
d	b	c
b	b	f
c	a	d

s

B	C	D
b	c	d
b	c	e
a	d	b

Khi đó ta có:

r |>< s

H	B	C
a	b	c
d	b	c
c	a	d

Chúng ta dễ dàng nhận thấy rằng: Nối nửa là nối tự nhiên xong chiếu lên A.

r |>< s $\neq$ s |>< r.

Lưu ý:

- Hầu hết các phép toán đại số quan hệ đều được cài đặt trong các ngôn ngữ truy vấn quan hệ như QBE, SQL, ORACLE,...

- Cơ sở dữ liệu (CSDL) quan hệ là một cơ sở dữ liệu gồm các quan hệ liên quan đến một bài toán.

Sau đây chúng ta sẽ xét một số ví dụ để minh họa các ứng dụng của các phép toán quan hệ trong các CSDL.

Ví dụ 2.17:

Cho CSDL bán hàng gồm 5 quan hệ sau:

kh(<u>Makh</u>, Tenkh, DS...).

Đây là quan hệ khách hàng gồm các thuộc tính: mã khách hàng (Makh), tên khách hàng (Tenkh), doanh số (DS), ...

nv(<u>Manv</u>, tenv...).

Đây là quan hệ nhân viên gồm các thuộc tính mã nhân viên (Manv), tên nhân viên (Tenv), ...

sp(<u>MASP</u>, TENSP, DVT, NSX, GIA).

Đây là quan hệ sản phẩm gồm các thuộc tính: mã sản phẩm (MASP), tên sản phẩm (TENSP), đơn vị tính (DVT), nơi sản xuất (NSX), giá (GIA).

hd(<u>SOHD</u>, NHD, Makh, Manv, Trigia).

Đây là quan hệ hóa đơn gồm các thuộc tính: số hóa đơn (SOHD), ngày lập hóa đơn (NHD), mã khách hàng (Makh), mã nhân viên lập hóa đơn (Manv), trị giá hóa đơn (Trigia).

cthd(<u>SOHD</u>, <u>Masp</u>, SL).

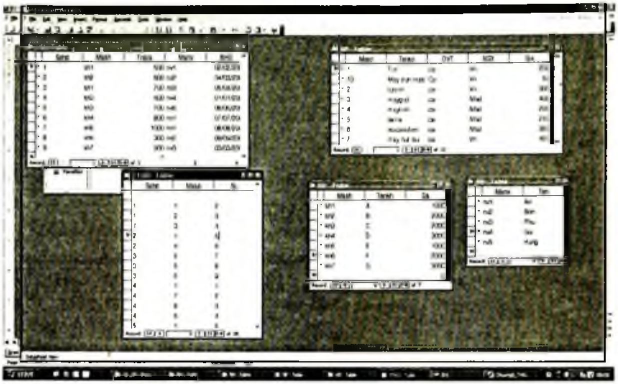
Đây là quan hệ chi tiết hóa đơn gồm các thuộc tính: số hóa đơn (SOHD), mã sản phẩm (Masp), số lượng (SL).

Các thuộc tính khóa được gạch ngang dưới, ví dụ <u>Masp</u>

Bằng các phép toán đại số quan hệ, ta có thể thực hiện các yêu cầu sau:

Lập bảng Masp, Tensp do khách hàng có tên là A mua.

Lập bảng SOHD, TRIGIA do nhân viên có tên là Hoa lập trong ngày 01/01/2009.



Hình 2.10. Minh họa cơ sở dữ liệu bán hàng

Tìm các SOHD đã mua cùng lúc hai mặt hàng có mã 1 và 2.

Lập danh sách Masp, Tensp không bán được.

Lập danh sách Makh, Tenkh đã mua hết tiền (tổng các TRIGIA = DS).

Thực hiện các yêu cầu trên: Muốn lấy tên ông A ta phải vào bảng kh. Muốn biết ông A có mua hàng không phải vào bảng hd tìm (qua Makh của ông A). Muốn biết ông A mua những sản phẩm gì vào bảng cthd (thông qua SOHD). Muốn lấy tên sản phẩm ta phải vào bảng sp (thông qua Masp). Như vậy ta nối tự nhiên bốn bảng kh, hd, cthd, sp, sau đó chọn những bộ chứa họ tên A và chiếu lên Masp, Tensp. Lưu ý quan hệ kh nối với quan hệ hd theo Makh, cthd nối với sp theo Masp, hai kết quả nối với nhau theo SOHD.

Vậy R_1 = kh |><| hd |><| cthd |><| sp

Đặt E là điều kiện: Tenkh = A

Chọn trong R_1 những bộ thỏa mãn điều kiện E cho vào R_2

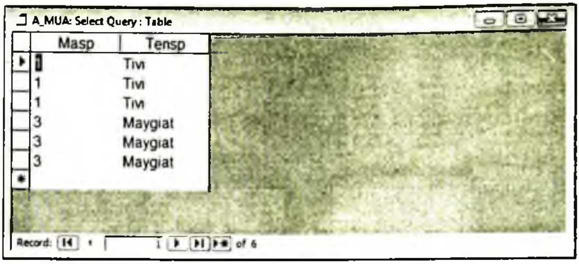
$R_2 = \sigma_E (R_1)$ và kết quả theo yêu cầu có trong R

R = R_2[Masp, Tensp]

Hay viết dưới dạng một biểu thức:

R = (σ_E (kh |><| hd |><| cthd |><| sp))[Masp, Tensp]

= ($\sigma_{HOTENKH= A}$ (kh |><| hd |><| cthd |><| sp))[Masp, Tensp] và kết quả có trong bảng A_MUA (hình 2.11).



A_MUA: Select Query : Table

Masp	Tensp
1	Tivi
1	Tivi
1	Tivi
3	Maygiat
3	Maygiat
3	Maygiat

Record: 1 of 6

Hình 2.11. Minh họa kết quả phép lọc

Lưu ý:

- Hầu hết các câu truy vấn thường phải thực hiện nối - chọn - chiếu. Đầu tiên nối các bảng cần lấy thông tin lại với nhau, sau đó chọn trong bảng đó những bộ thỏa mãn yêu cầu bài toán, cuối cùng chiếu lên những thuộc tính cần lọc.

- Tương tự SOHD, TRIGIA, NHD có trong bảng hd, Tenv có trong bảng nv

Vậy:

- *Nối*

R_1 = hd |><| nv

- *Chọn*

Gọi E là điều kiện: (Tenv = Hoa) and (NHD = 01/01/2007)

$R_2 = \sigma_E(R_1)$ và kết quả có trong R sau:

- *Chiếu*

$R = R_2$[SOHD, Trigia].

Hay R = (σ_E (hd |><| nv))[SOHD, Trigia] với E là điều kiện (Tenv = Hoa) and (NHD = 01/01/2007)

Ta tìm lần lượt các số hóa đơn đã mua 1 và các số hóa đơn đã mua 2, giao của hai tập đó là những số hóa đơn đã mua cả hai mặt hàng 1 và 2.

Vậy $R_1 = \sigma_{MASP=1}$ (cthd)

$R_2 = R_1$[SOHD]

$R_3 = \sigma_{MASP=2}$ (cthd)

$R_4 = R_3$[SOHD] và kết quả có trong R sau $R = R_2 \cap R_4$.

Hay R = ($\sigma_{MASP=1}$ (cthd))[SOHD] $\cap$ ($\sigma_{MASP=2}$ (cthd))[SOHD]

Mã sản phẩm không bán được là những mã sản phẩm có trong bảng sp nhưng không xuất hiện (dù 1 lần) trong các cthd. Vậy ta tìm những mã sản phẩm có trong sp mà không có trong chi tiết hóa đơn, lấy bảng kết quả nối với bảng sản phẩm để có tên sản phẩm và cuối cùng là chiếu.

R_1 = sp[Masp] - cthd[Masp]

R_2 = sp |><| R_1

R = R_2[Masp, Tensp]

Hay R = (sp |><| (sp[Masp] - cthd[Masp]))[Masp, Tensp]

Muốn tính tổng trị giá chúng ta phải có một phép toán số học, phép tính tổng như trong các phép toán quan hệ không có phép toán này. Vậy yêu cầu đặt ra không giải được. Đây là một yếu điểm của các phép toán đại số quan hệ. Để khắc phục các yếu điểm của đại số quan hệ ta dùng các ngôn ngữ quan hệ khác như Access, SQL, Oracle,... trong các ngôn ngữ đó có các phép tính tổng.

Trong các phép toán đại số quan hệ có năm phép toán được suy ra từ năm phép còn lại. Ta sẽ nêu định lý về vấn đề này. Năm phép toán: *hợp, hiệu, tích Đề-các, chọn, chiếu* được gọi là năm *phép toán cơ bản.* Năm phép toán: *Nối tự nhiên, nối theo* θ, *nối nửa, giao, chia* là các *phép toán không cơ bản.*

Định lý 2.1

Các phép toán cơ bản độc lập lẫn nhau.

Các phép toán không cơ bản được biểu thị qua năm phép cơ bản.

Chứng minh:

Ta dễ dàng thấy bằng trực quan rằng một trong năm phép toán cơ bản không thể biểu thị qua bốn phép còn lại.

Ta chứng minh ý hai của định lý 2.1 bằng cách biểu diễn các phép toán không cơ bản qua các phép cơ bản, ví dụ $r \cap s = r - (r - s)$.

2.4. PHỤ THUỘC HÀM

Trong các bảng quan hệ ở phần trước, chúng ta thấy giữa các thuộc tính của quan hệ có một số ràng buộc (phụ thuộc dữ liệu) đóng vai trò quan trọng trong lý thuyết CSDL.

Ví dụ xét tập thuộc tính hay lược đồ quan hệ tuyển sinh vào đại học, với các thuộc tính thứ tự (TT), tên (TÊN), năm sinh (NS), quê (QUÊ), số báo danh (SBD), điểm toán (ĐT), điểm lý (ĐL), điểm hóa (ĐH) ...

A = {TT, TÊN, NS, QUÊ, SBD, ĐT, ĐL, ĐH,...} ta thấy thuộc tính số báo danh (SBD) quyết định duy nhất các thuộc tính khác, nói cách khác nếu biết SBD thì biết được giá trị của các thuộc tính khác như ĐT, ĐL, ĐH,... tức SBD kéo theo các thuộc tính khác. Vậy giữa các tập thuộc tính X, Y của lược đồ quan hệ A có những mối ràng buộc kiểu *kéo theo, xác định duy nhất.*

Để đi sâu hiểu rõ bản chất các mối ràng buộc đó, chúng ta sẽ xét khái niệm phụ thuộc hàm-functional dependence (viết tắt PTH).

2.4.1. Định nghĩa phụ thuộc hàm trên lược đồ A

Cho lược đồ quan hệ $A = \{A_1, A_2, ..., A_n\}$.

Giả sử X, Y là các tập con của A, $X, Y \subset A$.

Định nghĩa 2.15 Phụ thuộc hàm trên lược đồ A

Ta nói Y phụ thuộc hàm vào X (hoặc X xác định phụ thuộc hàm Y) trên lược đồ quan hệ A, ký hiệu $X \to Y$, nếu với mọi quan hệ r trên A và $\forall t, t' \in r$ nếu $t.X = t'.X \Rightarrow t.Y = t'.Y$.

X là tập thuộc tính xác định (determinant), Y là tập thuộc tính phụ thuộc (dependent).

Trong thực tiễn chúng ta thấy rất nhiều lược đồ quan hệ A mà trên nó mọi quan hệ r đều thỏa mãn một phụ thuộc hàm nào đó. Chẳng hạn xét hồ sơ nhân sự của một nước thì số chứng minh thư là một thuộc tính luôn xác định duy nhất các thuộc tính khác. Hoặc nếu chúng ta xét hồ sơ sỹ quan thì số hiệu sỹ quan cũng là thuộc tính xác định các thuộc tính khác (nói cách khác nếu có hai sỹ quan có cùng số hiệu thì hai sỹ quan đó có cùng những tham số khác, tức hai sỹ quan đó là một). Trong các bài toán quản lý nhân sự ta thường thêm thuộc tính mã của đối tượng ID và thuộc tính này luôn xác định duy nhất các thuộc tính khác. Mỗi đối tượng chỉ có một mã ID duy nhất khác với các đối tượng khác.

Như vậy khái niệm phụ thuộc hàm trên lược đồ quan hệ A khẳng định sự ràng buộc mang tính chất nội tại của các tập thuộc tính trong A. Đó là sự ràng buộc của các tập thuộc tính trong A xét với mọi quan hệ r trên A.

Ví dụ 2.18:

Cho lược đồ A = {Manv, Hoten, NS, Que, HS, Luong} với thuộc tính mã nhân viên Manv, Họ tên, năm sinh, quê, Hệ số và Lương tương ứng. Khi đó ta có vì Manv là khóa nên Manv → {Manv, Hoten, NS, Que, HS, Luong}, HS → Luong, Luong → HS. Ta có tập phụ thuộc hàm trên A là F = {Manv → {Manv, Hoten, NS, Que, Hs, Luong}, HS → Luong, Luong → HS} = {Manv → A, HS → Luong, Luong → HS}.

2.4.2. Định nghĩa phụ thuộc hàm trong quan hệ r

Ở trên chúng ta đã định nghĩa khái niệm phụ thuộc hàm trên lược đồ A.

Sau đây là khái niệm phụ thuộc hàm trên một quan hệ r cụ thể của lược đồ A.

Cho lược đồ quan hệ A và X, Y là các tập con của A, r là một quan hệ trên A.

Định nghĩa 2.16 Phụ thuộc hàm trong một quan hệ r

Ta nói X xác định phụ thuộc hàm Y, ký hiệu $X \rightarrow Y$, trong r nếu với mọi t và t' của r mà t, t' bằng nhau trên tập X thì chúng cũng bằng nhau trên tập Y, tức là $\forall$t, t'$\in$ r nếu t.X = t'.X thì t.Y = t'.Y.

Vậy phụ thuộc hàm trên lược đồ A là phụ hàm mà t.X = t'.X thì t.Y = t'.Y thỏa mãn với mọi quan hệ r trên A.

Phụ thuộc hàm trong quan hệ r chỉ đòi hỏi t. X = t'.X thì t.Y = t'.Y thỏa mãn quan hệ r. Hiển nhiên $X \rightarrow Y$ là PTH trên lược đồ A thì nó là PTH thỏa mãn mọi quan hệ r trên A.

Ví dụ xét quan hệ r như sau:

r

H	B	C	D
0	0	0	0
1	1	1	1
0	1	0	1

Rõ ràng trong r thì $H \rightarrow C$ (vì các bộ bằng nhau trong H cũng bằng nhau trong C).

2.4.3. Các tính chất của phụ thuộc hàm

Các tính chất của phụ thuộc hàm ta xét trên lược đồ A. Nếu X, Y, Z và W là những tập thuộc tính của A thì ta có một số tính chất cơ bản của lớp các PTH như sau *(để tiện khi trình bày thay cho tập {H, B, C} ta sẽ viết HBC)*:

Ar1. Tính phản xạ: $X \rightarrow X$, và nếu $Y \subset X$ thì $X \rightarrow Y$

Ar2. Tính mở rộng hai vế: $X \rightarrow Y \Rightarrow XZ \rightarrow YZ$ (mở rộng hai vế tập Z)

Ar3. Tính bắc cầu: $X \rightarrow Y$ và $Y \rightarrow Z \Rightarrow X \rightarrow Z$

4. Tính tựa bắc cầu: $X \rightarrow Y$ và $YZ \rightarrow W \Rightarrow XZ \rightarrow W$

5. Tính mở rộng trái và thu hẹp phải: $X \rightarrow Y \Rightarrow XZ \rightarrow Y - W$

6. Tính cộng đầy đủ: $X \rightarrow Y$ và $Z \rightarrow W \Rightarrow XZ \rightarrow YW$

7. Tính tích lũy: $X \rightarrow Y$ và $Y \rightarrow ZW \Rightarrow X \rightarrow YZW$

Có thể chứng minh các tính chất Ar1, Ar2, Ar3, 4, 5, 6, 7 một cách đơn giản. Giả sử t, t' $\in$ r và r là một quan hệ bất kỳ trên A. Chúng ta lần lượt chứng minh các tính chất trên một cách dễ dàng. Thật vậy:

Tính phản xạ Ar1: Điều này hiển nhiên vì t và t' đã bằng nhau trong tập X thì chúng phải bằng nhau trong mọi tập con của X, nói cách khác $t.X = t'.X \Rightarrow t.X = t'.X$ và $t.Y = t'.Y$ với mọi $Y \subset X$.

Tính bắc cầu Ar3: Giả sử $t.X = t'.X$ theo giả thiết $X \rightarrow Y$ nên ta có $t.Y = t'.Y$ mà $t.Y = t'.Y$ theo giả thiết $Y \rightarrow Z$ ta lại có $t.Z = t'.Z$. Vậy từ $t.X = t'.X$ ta suy ra được $t.Z = t'.Z$, nên $X \rightarrow Z$.

Tính mở rộng hai vế Ar2: Giả sử $t.XZ = t'.XZ$ ta phải chứng minh $t.YZ = t'.YZ$ Từ $t.XZ = t'.XZ$ ta có $t.X = t'.X$ và $t.Z = t'.Z$. Theo giả thiết:

$t.X = t'.X$ thì $t.Y = t'.Y$. Vậy từ $t.XZ = t'.XZ$ ta có $t.Y = t'.Y$ và

$t.Z = t'.Z$ mà $t.Y = t'.Y$ và $t.Z = t'.Z$ thì $t.YZ = t'.YZ$. Vậy $XZ \rightarrow YZ$.

Các tính chất khác có thể chứng minh tương tự. Tuy nhiên ta thấy rằng các tính chất 4, 5, 6, 7 đều có thể suy ra từ các tính chất Ar1, Ar2, Ar3. Trong lý thuyết CSDL, ba tính chất Ar1, Ar2, Ar3 được gọi là *hệ tiên đề Armstrong* (Armstrong là người đầu tiên nêu ba tính chất Ar1,Ar2, Ar3 của các phụ thuộc hàm).

2.4.4. Hệ tiên đề Armstrong và các phép suy dẫn

Hệ ba tính chất {Ar1, Ar2, Ar3} được gọi là *hệ tiên đề Armstrong* của lớp các phụ thuộc hàm PTH trên lược đồ A.

Hệ Armstrong = {Ar1, Ar2, Ar3} là *hệ tiên đề* vì từ đây chúng ta có thể suy ra các tính chất khác. Ví dụ để chứng minh tính chất 6, ta thực hiện:

Từ $X \rightarrow Y$ mở rộng hai vế bằng Z ta có $XZ \rightarrow YZ$ (*)

Từ $Z \rightarrow W$ mở rộng hai vế bằng Y ta có $YZ \rightarrow YW$ (**)

Từ (*) và (**) theo tính bắc cầu ta có $XZ \rightarrow YW$

Hay để chứng minh tính chất 7, ta thực hiện:

Từ $Y \rightarrow ZW$ mở rộng hai vế bằng Y ta có $YY \rightarrow YZW$ mà $YY = Y$ nên ta có $Y \rightarrow YZW$ và theo giả thiết $X \rightarrow Y$ nên theo tính bắc cầu ta có $X \rightarrow YZW$.

a. Phép suy dẫn theo hệ tiên đề

Ta thấy hệ tiên đề Armstrong đóng vai trò *luật sinh* của lớp các PTH.

Thật vậy nếu cho trước tập phụ thuộc hàm F trên lược đồ A thì ta có thể dùng luật suy dẫn trong các tính chất của PTH để nhận được các PTH mới, lớp các PTH nhận được từ các phép suy dẫn như vậy đóng vai trò quan trọng trong lớp các PTH trên lược đồ quan hệ A. Như ta vừa trình bày các tính chất 4, 5, 6, 7 đều được suy dẫn (suy ra) từ hệ tiên đề Armstrong. Thật vậy:

Tính tựa bắc cầu (4) có thể suy ra từ tính mở rộng hai vế và tính bắc cầu vì từ giả thiết $X \rightarrow Y$ mở rộng hai vế bằng Z ta có $XZ \rightarrow YZ$ và vì $YZ \rightarrow W$ theo bắc cầu ta có $XZ \rightarrow W$.

Tính mở rộng trái và thu hẹp phải (5) được suy từ tính phản xạ và bắc cầu vì với mọi X, Y, Z, W theo tính phản xạ ta luôn có $XZ \rightarrow X$ và $Y \rightarrow Y - W$ và từ giả thiết $X \rightarrow Y$ theo tính bắc cầu ta có $XZ \rightarrow Y - W$.

Như vậy ta thấy rằng mọi phụ thuộc hàm f được suy ra từ các tính chất của PTH đều có thể được suy ra từ chỉ ba tính chất của hệ tiên đề Armstrong. Đôi khi thay cho nói phụ thuộc hàm f nhận được từ tập phụ thuộc F dựa vào các luật suy dẫn trong các tính chất của phụ thuộc hàm ta sẽ nói: *f suy dẫn được từ F theo hệ tiên đề Armstrong* (suy dẫn theo hệ tiên đề).

Vậy giả sử F là tập các phụ thuộc hàm và f là một phụ thuộc hàm trên A.

Định nghĩa 2.17 Phụ thuộc hàm suy dẫn được từ hệ tiên đề

Ta nói phụ thuộc hàm *f suy dẫn được theo hệ tiên đề Armstrong* từ tập F, ký hiệu $F \models f$, nếu f có thể nhận được từ F sau một số hữu hạn lần áp dụng hệ tiên đề Armstrong vào các phụ thuộc hàm trong F.

Ví dụ 2.19:

Cho A = {H, B, C, D} là lược đồ quan hệ.

$F = \{H \rightarrow B, H \rightarrow C, H \rightarrow D\}$, f là $H \rightarrow BCD$

Ta thấy ngay f có thể nhận được từ phép cộng đầy đủ tức $F \models H \rightarrow BCD$ và $F \models H \rightarrow HB$ (theo tính phản xạ và cộng đầy đủ) hoặc $F \models H \rightarrow HBCD$ (theo tính cộng đầy đủ),...

b. Phép suy dẫn theo quan hệ

Trên đây chúng ta vừa nêu các phép suy dẫn theo hệ tiên đề.

Sau đây chúng ta sẽ nêu *phép suy dẫn theo quan hệ*.

Cho lược đồ quan hệ $A = \{A_1, A_2, \ldots A_n\}$.

F tập phụ thuộc hàm trên A, f là một phụ thuộc hàm trên A.

Định nghĩa 2.18 Phụ thuộc hàm suy dẫn được theo quan hệ

Ta nói rằng *f suy dẫn được từ tập F theo quan hệ*, ký hiệu F |- f, nếu với mọi quan hệ r trên lược đồ A mà F thỏa mãn quan hệ r thì f cũng thỏa mãn r.

Đây là một phép suy dẫn theo quan hệ, "ở đâu tập F thỏa mãn thì ở đó f thỏa mãn".

Một vấn đề được đặt ra là cho trước A, tập phụ thuộc hàm F, hai luật suy dẫn ở trên có cho ta cùng một tập phụ thuộc hàm suy dẫn {f} không?

Sau đây ta sẽ thấy hai luật suy dẫn là tương đương. Nói cách khác hai luật suy dẫn cho ta cùng một tập kết quả.

Định lý 2.2

Cho tập phụ thuộc hàm F và một phụ thuộc hàm f trên lược đồ A, khi đó ta có:

F |- f khi và chỉ khi F |= f.

Thực chất của việc chứng minh định lý 2.1 là ta phải chứng minh hai ý:

1. Các f suy dẫn được theo hệ tiên đề từ tập F, nghĩa là F |= f thì f là một PTH trên A (tính đúng đắn của luật suy dẫn theo hệ tiên đề), điều này có nghĩa là chúng ta phải chứng minh f thỏa mãn mọi r mà trong đó tập F thỏa mãn.

2. Các phụ thuộc hàm f suy dẫn được theo quan hệ từ F, nghĩa là F |- f. thì f cũng suy dẫn được từ F theo hệ tiên đề (tính đầy đủ của hệ tiên đề Armstrong). Điều này tương đương với việc một phụ thuộc hàm f không suy dẫn được theo hệ tiên đề Armstrong thì cũng không suy dẫn được theo quan hệ, nghĩa là nếu f không suy dẫn được theo hệ tiên đề Armstrong thì tồn tại một quan hệ r mà trong nó tập F thỏa mãn nhưng f không thỏa mãn.

Như vậy để chứng minh định lý 2.2 ta sẽ chứng minh định lý tương đương sau:

Định lý 2.3

Hệ tiên đề Armstrong là đúng đắn và đầy đủ.

Trong quá trình chứng minh ba tính chất Ar1, Ar2, Ar3 của hệ tiên đề Armstrong đồng thời chúng ta đã chỉ ra rằng mọi PTH suy dẫn được từ hệ tiên đề này đều là phụ thuộc hàm thỏa mãn trên lược đồ A, mà trong đó tập PTH F thỏa mãn. Do đó tính đúng đắn đã được chứng minh.

Để chứng minh tính đầy đủ của hệ tiên đề chúng ta sẽ dùng bổ đề sau:

Bổ đề 2.1

Giả sử $X \subseteq A$. Nếu gọi X^+ là tập tất cả các thuộc tính B của A mà $F \models X \rightarrow B$ thì với mọi tập $Y \subseteq A$, $F \models X \rightarrow Y \Leftrightarrow Y \subseteq X^+$.

a. Chứng minh chiều thuận:

Ta có $F \models X \to Y$. Giả sử $Y = \{B,C,D,...\}$ theo tính phản xạ ta có:

$F \models X \to B$, nên $B \in X^+$

$F \models X \to C$, nên $C \in X^+$

$F \models X \to D$, nên $D \in X^+$,...

Vậy $\{B,C,D,...\} = Y \subseteq X^+$

b. Chứng minh chiều ngược:

Ta có $Y \subseteq X^+$. Theo định nghĩa của tập X^+ thì mọi $B \in Y$ ta có:

$F \models X \to B$, vậy theo tính cộng đầy đủ ta có $F \models X \to Y$. Bổ đề đã được chứng minh.

Bây giờ chúng ta sẽ chứng minh tính đầy đủ của hệ tiên đề Armstrong.

Giả sử $f = X \to Y$ là một phụ thuộc hàm trên A không suy dẫn được từ tập F theo hệ tiên đề, tức F not $\models X \to Y$. Ta sẽ xây dựng một quan hệ r trên A mà trong đó tập các phụ thuộc hàm F thỏa mãn nhưng $f = X \to Y$ không thỏa mãn.

Ta xét quan hệ r trên A gồm chỉ hai phần tử t_1 và t_2 như sau:

r

X^+					$A - X^+$				
1	1	1	...	1	0	0	0	...	0
1	1	1	...	1	1	1	1	...	1

Quan hệ r có hai phần tử t_1, t_2. Phần tử t_1 chứa giá trị 1 trong các thuộc tính của X^+ và giá trị 0 trong những thuộc tính $A - X^+$; t_2 chứa toàn giá trị 1 cho mọi thuộc tính.

Ta chứng minh rằng r sẽ thỏa mãn mọi PTH của F. Thật vậy giả sử có một phụ thuộc hàm $W \to V$ của F không thỏa mãn r, thế thì $W \subseteq X^+$, nếu không sẽ vi phạm tính bằng nhau của hai bộ t_1 và t_2 trên W. Hơn nữa V không thể là tập con của X^+, vì nếu không $W \to V$ sẽ thỏa mãn r. Vậy có một thuộc tính B của V không thuộc X^+.

Theo bổ đề 2.1 thì $W \subseteq X^+ \Leftrightarrow F \models X \to W$, mà $W \to V$, nên $W \to B$ và theo tính bắc cầu (vì $X \to W$) thì $X \to B$, tức $F \models X \to B$, hay B thuộc X^+. Điều này là vô lý vì B không thuộc X^+.

Vậy r thỏa mãn mọi PTH của F. Vấn đề còn lại chúng ta phải chứng minh rằng r không thỏa mãn phụ thuộc hàm $f = X \to Y$.

Giả sử $X \rightarrow Y$ thỏa mãn r thế thì X, $Y \subseteq X^+$ nếu không thì vi phạm tính bằng nhau của t_1 và t_2 trên X và Y. Lại sử dụng bổ đề 2.1: $Y \subseteq X^+ \Leftrightarrow F \models X \rightarrow Y$. Điều này vô lý, vì F không suy dẫn được f theo hệ tiên đề. Vậy $X \rightarrow Y$ không thỏa mãn r. Định lý được chứng minh.

2.4.5. Bao đóng F^+ của tập phụ thuộc hàm F

a. Định nghĩa F^+

Trong phần trên chúng ta đã nói đến các phụ thuộc hàm f suy dẫn được từ tập phụ thuộc hàm F cho trước, ta đã có định lý chứng minh các phép suy dẫn theo tiên đề và theo quan hệ là tương đương nên từ nay thay cho nói suy dẫn theo quan hệ hoặc suy dẫn theo tiên đề ta chỉ nói đơn giản là suy dẫn. Tập các phụ thuộc hàm f suy dẫn được từ tập phụ thuộc hàm F ta sẽ gọi là *bao đóng* của tập F và ký hiệu F^+.

Cho lược đồ $A = \{A_1, A_2,... A_n\}$; F là tập phụ thuộc hàm trên A.

Định nghĩa 2.19 Bao đóng của tập phụ thuộc hàm F

Bao đóng của tập phụ thuộc hàm F ký hiệu F^+ là tập tất cả các phụ thuộc hàm f suy dẫn được từ tập F. Vậy $F^+ = \{f: F \models f\}$.

Ví dụ 2.20:

Cho lược đồ A = {B, C, D, E}

Giả sử tập F trên A như sau:

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

Khi đó $F^+ = \{E \rightarrow B, B \rightarrow C, E \rightarrow D, B \rightarrow D, E \rightarrow BD, E \rightarrow BCD, E \rightarrow C, E \rightarrow CD, E \rightarrow BC, B \rightarrow CD,...\}$.

Ta thấy F^+ luôn chứa F.

b. Các tính chất đơn giản của tập F^+

a. *Tính phản xạ*: với mọi tập phụ thuộc hàm F ta luôn có $F \subset F^+$.

b. *Tính đơn điệu*: nếu $F \subseteq G$ thì $F^+ \subseteq G^+$

c. *Tính lũy đẳng*: với mọi tập phụ thuộc hàm F ta luôn có $F^{++} = F^+$.

Phần chứng minh các tính chất a, b, c của bao đóng tập F chúng tôi dành cho các bạn như một bài tập nhỏ.

2.4.6. Bao đóng X^+ của tập thuộc tính X

a. Định nghĩa bao đóng X^+

Cho lược đồ quan hệ $A = \{A_1, A_2,... A_n\}$. Giả sử F là tập phụ thuộc hàm trên A. X là tập con của tập thuộc tính A.

Định nghĩa 2.20 Bao đóng của tập thuộc tính X

Bao đóng của tập thuộc tính X theo F, ký hiệu X^+ (hoặc X_F^+ để chỉ bao đóng lấy theo tập F), là tập tất cả các thuộc tính B của A mà $X \to B$ suy dẫn được từ tập F.

Vậy: $X^+ = \{B: B \in A \text{ và } X \to B \in F^+\}$.

Ta có thể nói bao đóng X^+ là tất cả những thuộc tính B của A mà X kéo theo.

Như vậy bao đóng X^+ của X được xác định qua tập phụ thuộc hàm F, vì thế đôi khi ta ký hiệu X_F^+. Tuy nhiên thông thường ta chỉ có một tập phụ thuộc hàm F, nên thay cho viết X_F^+ ta viết đơn giản X^+.

Ví dụ 2.21:

Giả sử:

$A = \{H, B, C, D, E, G\}$.

$F = \{H \to C, H \to EG, B \to D, G \to E\}$.

$X = \{H, B\}$, $Y = \{C, G, D\}$.

Khi đó ta sẽ có:

$X^+ = \{H, B, C, D, E, G\}$

$Y^+ = \{C, G, D, E\}$.

Tương tự như bao đóng của tập phụ thuộc hàm F^+ tập bao đóng X^+ cũng chứa các phần tử của tập X, tức là $X \subset X^+$.

b. Các tính chất của tập bao đóng X^+

Nếu X, Y là các tập con của tập thuộc tính A thì ta có các tính chất:

1. Tính phản xạ: $X \subseteq X^+$
2. Tính đơn điệu: Nếu $X \subseteq Y$ thì $X^+ \subseteq Y^+$
3. Tính lũy đẳng: $X^{++} = X^+$
4. $(XY)^+ \supseteq X^+Y^+$ (bao đóng của hợp chứa hợp các bao đóng)
5. $(X^+Y)^+ = (XY^+)^+ = (X^+ Y^+)^+ = (XY)^+$
6. $X \to Y \Leftrightarrow Y \subseteq X^+$
7. $X \to Y \Leftrightarrow Y^+ \subseteq X^+$
8. $X \to X^+$ và $X^+ \to X$
9. $X^+ = Y^+ \Leftrightarrow X \to Y$ và $Y \to X$.

Chúng ta thấy bao đóng của tập phụ thuộc hàm F và bao đóng của tập thuộc tính X là những tập liên quan với hệ tiên đề Armstrong. Các tập bao đóng là "kết quả" của các phép suy dẫn dùng luật của tiên đề Armstrong. Sau đây chúng ta sẽ chứng minh các tính chất của tập bao đóng.

Tính chất 1: $X \subseteq X^+$. Thật vậy theo tính phản xạ của hệ tiên đề Armstrong ta có ngay với mọi thuộc tính B của X thì $X \to B$, nên theo định nghĩa của X^+ (X^+ là tập các thuộc tính B mà $X \to B$) thì $X \subseteq X^+$.

Tính chất 2 (tính đơn điệu): Giả sử $X \subseteq Y$ ta phải chứng minh $X^+ \subseteq Y^+$.

Thật vậy lấy $B \in X^+$, theo định nghĩa ta có $X \to B$ mà $X \subseteq Y$ nên ta có $Y \to B$. Vậy $B \in Y^+$.

Tính chất 8: Ta chứng minh tính chất 8 tức là tính chất $\forall X$ thì $X \to X^+$ và $X^+ \to X$.

Theo tính phản xạ vì $X \subseteq X^+$ nên $X^+ \to X$.

Bây giờ ta chứng minh $X \to X^+$.

Theo định nghĩa của tập X^+ ta có $X^+ = \{B: X \to B \in F^+\}$.

Vậy giả sử $X^+ = \{D, B, C,...\}$ khi đó:

$X \to D$

$X \to B$

$X \to C$

...

Theo tính cộng đầy đủ ta có: $X \to \{D, B, C,...\} = X^+$.

Tính chất 3: $X^{++} = X^+$

Rõ ràng theo tính phản xạ ta có ngay $X^+ \subseteq X^{++}$. Bây giờ lấy $B \in X^{++}$ tức là $X^+ \to B$, mà theo tính chất 8 vừa chứng minh thì $X \to X^+$ nên $X \to B$ hay $B \in X^+$. Vậy $X^{++} = X^+$.

Tính chất 4: $(XY)^+ \supseteq X^+Y^+$

Lấy $B \in X^+Y^+$ tức $B \in X^+$ hoặc $B \in Y^+$ tức là $X \to B$ hoặc $Y \to B \Rightarrow XY \to B$. Hay $B \in (XY)^+$.

Tính chất 9: $X^+ = Y^+ \Leftrightarrow X \to Y$ và $Y \to X$

a. Chiều xuôi $\Rightarrow$: ta có $X^+ = Y^+$. Vì $X \to X^+$ và $Y^+ \to Y$ nên $X \to Y$. chứng minh tương tự ta có $Y \to X$.

b. Chiều ngược $\Leftarrow$: Lấy $B \in X^+$ tức là $X \to B$ vì $Y \to X$ nên $Y \to B$ hay $B \in Y^+$. Tương tự lấy $B \in Y^+$ ta chứng minh được $B \in X^+$. Vậy $X^+ = Y^+$.

Tính chất 6: Đã được chứng minh trong bổ đề 2.6.

Tính chất 7: Giả sử $X \to Y$, theo tính chất 6 ta có $Y \subseteq X^+$

Lấy bao đóng ta có $Y^+ \subseteq X^{++}$ mà $X^{++} = X^+$ nên $Y^+ \subseteq X^+$. Ngược lại giả sử $Y^+ \subseteq X^+$ nên $X^+ \to Y^+$. Theo tính chất 8 ta có $X \to X^+$ và $Y^+ \to Y$ nên $X \to Y$.

Tính chất 5: Ánh xạ từ X lên X^+ là ánh xạ đóng nên chúng ta có thể sử dụng kết quả bổ đề 1.3 trong phần 1.4 của chương 1 để chứng minh tính chất 5. Hoặc các bạn có thể chứng minh trực tiếp.

2.4.7. Thuật toán tìm bao đóng X^+, bài toán thành viên

a. Bài toán thành viên

Phần trên chúng ta đã nêu một số khái niệm cơ bản của các tập bao đóng. Ta thấy tập X^+ được xác định thông qua tập F^+. Một vấn đề quan trọng trong lý thuyết về CSDL là cho trước tập F và một phụ thuộc hàm f, có hay không một khẳng định f thuộc F^+? (*bài toán thành viên: f có là thành viên của F^+ không?*)

Để giải *bài toán thành viên* chúng ta có thể dùng tính chất 6 của tập bao đóng X^+ hoặc bổ đề 2.1 đó là tính chất: $X \to Y \in F^+ \Leftrightarrow Y \subset X^+$. Do vậy chỉ cần tính X^+ và so sánh với tập Y ta có ngay câu trả lời là $X \to Y = f$ thuộc F^+ hay không.

b. Thuật toán tìm bao đóng X^+

Thuật toán tìm X^+ của Beeri và Bernstein.

Cho lược đồ $A = \{A_1, A_2,..., A_n\}$; F là tập phụ thuộc hàm trên A, X là tập thuộc tính. Tính X^+ =?

Ta sẽ xây dựng dãy $X^0, X^1,..., X^k$... như sau:

$X^0 = X$

$X^{(i+1)} = X^iZ^i$ với $Z^i = \{B: B \notin X^i$ và $X^i \to B \in F^+\}$, trong đó i = 0, 1, 2,...

Ta có nhận xét rằng: dãy X^0, X^1,... luôn xây dựng được. Z^i là tập thuộc tính mà X^i kéo theo trực tiếp. Dãy X^0, X^1, X^2,... là dãy lồng nhau và tăng dần, tức là $X^0 \subset X^1 \subset$... Vì tập thuộc tính A là hữu hạn nên sau một số hữu hạn bước, thuật toán phải kết thúc. Nói cách khác tồn tại số nguyên k (bé nhất) sao cho:

$X^k = X^{(k+1)} = X^{(k+2)} =$... và $Z^k = Z^{k+1} =... = \phi$

Tập X^k đó chính là tập X^+

Để chứng minh rằng X^k chính là tập X^+ ta sẽ xét ví dụ minh họa thuật toán.

Ví dụ 2.22:

Giả sử A = {H, B, C, D, E, G} và tập F như sau:

F = {HB → C, C → H, BC → D, HCD → B, D → EG, BE → C, CG → BD, CE → HG}, X = {B, D}, X^+ = ?

Đầu tiên ta có X^0 = {B, D}. Để tìm X^1 ta tìm Z^0 gồm những thuộc tính không thuộc X^0 mà X^0 kéo theo trực tiếp. Xét những phụ thuộc hàm trong F có vế trái nằm trong BD. Ta có D → EG thỏa mãn điều kiện đó. Vậy Z^0 = {E, G}, nên X^1 = {B, D, E, G}.

Tiếp tục để tìm X^2 ta tìm những phụ thuộc hàm của F có vế trái nằm trong {B, D, E, G}, đó là D → EG và BE → C. Vậy Z^1 = C; $X^2 = X^1Z^1$ = {B, C, D, E, G}.

Tiếp tục ta có $X^3 = X^2Z^2$ = {H, B, C, D, E, G} = A. Đây là tập X^+

Vậy X^+ = {B, D}$^+$ = {H, B, C, D, E, G} = A.

Thuật toán 2.1 Tính X^+

Input: Lược đồ quan hệ A, tập phụ thuộc hàm F trên A, tập thuộc tính X ⊆ A

Output: Tập X^+

Thuật toán:

```
Begin
   Y: = X;
   repeat
   Z: = ∅;
   for each B in A do
   if (B ∉ Y and Y → B ∈ F⁺) then Z: = Z ∪ B;
   Y: = Y ∪ Z;
   until Z = ∅;
   X⁺ = Y
end;
```

Trong thuật toán này thay cho tính các X^i ta tính Y và tại mỗi bước ta tính Z bằng cách xét các thuộc tính B không thuộc Y mà Y kéo theo B. Thuật toán dừng khi không tìm được thuộc tính mà Y kéo theo.

Định lý 2.4

Trong thuật toán tìm bao đóng X^+, ta có $X^+ = X^k$, với k là số nguyên bé nhất mà $X^k = X^{k+1} = X^{k+2} = \ldots$

Chứng minh:

a. Ta chứng minh $X^+ \subset X^k$. Thật vậy lấy $B \in X^+$. Như ở trên ta đã thấy $X^+ = XZ$ với $Z = \{B: B \notin X \text{ và } X \rightarrow B \in F^+\}$.

Vậy nếu $B \in X$ thì $B \in X^k$ vì $X \subset X^k$; còn nếu $B \in Z$ thì theo định nghĩa của các tập Z^i tồn tại một chỉ số i để $B \in Z^i$ vậy $B \in X^k$ vì với mọi i thì $X^i \subset X^k$.

Trong cả hai trường hợp ta đều có $B \in X^k$ và suy ra $X^+ \subset X^k$.

b. Ta chứng minh $X^k \subset X^+$:

Dễ dàng thấy rằng $X \rightarrow X^1, X^1 \rightarrow X^2, \ldots, X^{k-1} \rightarrow X^k$

$\Rightarrow X \rightarrow X^k \Leftrightarrow X^k \subset X^+$ (theo tính chất 6 của bao đóng).

Ví dụ 2.23:

Cho $A = \{H, B, C, D, E, I\}$.

Tập $F = \{H \rightarrow D, HB \rightarrow E, BI \rightarrow E, CD \rightarrow I, E \rightarrow C\}$.

Tập thuộc tính $X = \{H, E\}$. Tính $X^+ = ?$

Ta có $X^0 = \{H, E\}$

$X^1 = X^0Z^0$ và $Z^0 = \{Y: Y \notin X^0 \text{ và } X^0 \rightarrow Y \in F^+\}$

Vậy $Z^0 = \{D, C\}$ và $X^1 = \{H, E, D, C\}$

$Z^1 = \{Y: Y \notin X^1 \text{ và } X^1 \rightarrow Y \in F^+\}$ và

$Z^1 = \{I\}$ nên $X^2 = \{H, E, D, C, I\}$

$Z^2 = \{Y: Y \notin X^2 \text{ và } X^2 \rightarrow Y \in F^+\} = \varnothing$

Vậy $X^3 = X^2 = X^+ = \{H, E, D, C, I\}$.

Trên đây ta vừa nêu một thuật toán tính bao đóng X^+. Tuy nhiên khi tính X^+ ta nhìn vào tập phụ thuộc hàm và liệt kê lần lượt các thuộc tính mà X kéo theo.

Quay lại ví dụ 2.23; ta có $A = \{H, B, C, D, E, I\}$.

Tập $F = \{H \rightarrow D, HB \rightarrow E, BI \rightarrow E, CD \rightarrow I, E \rightarrow C\}$.

Tập thuộc tính $X = \{H, E\}$. Tính $X^+ = ?$

Vì $X \rightarrow H, X \rightarrow E$ nên $X^+ = \{H, E, \ldots\}$

Vì $H \rightarrow D$ và $E \rightarrow C$ nên $X^+ = \{H, D, E, C, \ldots\}$

Vì CD $\rightarrow$ I nên $X^+ = \{H, E, D, C, I,...\}$

Vì HECDI không kéo thêm thuộc tính nào nên $X^+ = \{H, E, D, C, I\}$

2.5. KHÓA, KHÓA CHÍNH, KHÓA NGOẠI

Trong công tác quản lý các CSDL quan hệ ta đã thấy giữa các tập thuộc tính có các mối ràng buộc kiểu phụ thuộc hàm. Ví dụ, trong hồ sơ nhân sự của cơ quan thuộc tính mã nhân viên (MANV) đóng vai trò "xác định" các thuộc tính khác, tức là có thể dùng thuộc tính MANV như là *khóa*. Hoặc trong lược đồ quản lý tuyển sinh đại học, thuộc tính số báo danh (SBD) đóng vai trò là *khóa*.

Sau đây để cho tiện khi trình bày và sử dụng ta nêu thuật ngữ *sơ đồ quan hệ* (viết tắt là SĐQH) như sau:

Cho lược đồ quan hệ $A = \{A_1, A_2,... A_n\}$; F là tập phụ thuộc hàm trên A.

Định nghĩa 2.21 Sơ đồ quan hệ

Sơ đồ quan hệ là cặp A, F ký hiệu là W = < A, F >.

Vậy nói cho một sơ đồ quan hệ W, thực chất chỉ là cho trước tập thuộc tính A, tập phụ thuộc hàm F trên A và viết chúng thành cặp W = < A, F >.

2.5.1. Định nghĩa khóa của sơ đồ quan hệ

Cho sơ đồ quan hệ W = < A, F >. Trong đó $A = \{A_1, A_2,..., A_n\}$ là tập thuộc tính, F là tập phụ thuộc hàm trên A.

Định nghĩa 2.22 Khóa (key) tối thiểu

Tập thuộc tính $k \subseteq A$ được gọi là *khóa tối thiểu* của W = < A, F > nếu nó thỏa mãn hai điều kiện:

$k^+ = A$ (hay $k \rightarrow A$)

k tối thiểu.

Vậy tập $k \subset A$ được gọi là *khóa tối thiểu* của sơ đồ quan hệ W = < A, F > nếu k là tập tối thiểu kéo theo A.

Tức k là khóa tối thiểu nếu: $k^+ = A$ (hay $k \rightarrow A$) và bớt khỏi k dù một phần tử thì bao đóng của tập còn lại khác A. Vậy tập $k \subset A$ gọi là *khóa tối thiểu* nếu: $k^+ = A$ và $(k - B)^+ \neq A$, với B bất kỳ thuộc k.

Trực quan từ định nghĩa, ta thấy rằng nếu k là một tập thuộc tính mà $k^+ = A$ thì từ k ta có thể bớt dần các phần tử của k, để nhận được tập k tối thiểu có bao đóng bằng A và *đó chính là khóa tối thiểu của sơ đồ quan hệ.*

Lưu ý:

Trong tài liệu này ta chỉ dùng khóa tối thiểu nên từ nay ta nói khóa thay cho nói khóa tối thiểu.

Bổ đề 2.2

Mọi sơ đồ quan hệ W = < A, F > luôn có khóa.

Chứng minh bổ đề 2.2 bằng cách lấy k = A rồi bớt dần các thuộc tính B mà $(k\text{-}B)^+ = A$ để có k tối thiểu.

Về sau các thuộc tính thuộc một khóa nào đó ta gọi là *thuộc tính khóa*, ngược lại thuộc tính không thuộc khóa nào gọi là *thuộc tính không khóa (hoặc thuộc tính thứ cấp)* và ta ký hiệu F_n là tập các thuộc tính thứ cấp của W = < A, F >.

Ví dụ 2.24:

Cho sơ đồ quan hệ W = < A, F > với A = {H, B, C, D, E, G},

F = {HB → C, D → EG, C → H, BE → C, BC → D, CG → BD, HCD → B, CE → HG}. Ta sẽ thấy các tập thuộc tính:

k_1 = {H, B}, k_2 = {B, E}, k_3 = {C, G}, k_4 = {C, E}, k_5 = {C, D}, k_6 = {B, C} đều là các khóa của W và tập các thuộc tính khóa bằng A, và vì vậy $F_n = \varnothing$.

Một sơ đồ quan hệ có thể có nhiều khóa và tập thứ cấp F_n có thể rỗng. Vấn đề còn lại đối với chúng ta là cho trước một sơ đồ quan hệ W = < A, F >, làm thế nào để tìm khóa của nó?

2.5.2. Các thuật toán tìm khóa của W = < A, F >

Ta nhắc lại khóa là tập thuộc tính k tối thiểu mà bao đóng của k đúng bằng A ($k^+ = A$) và nếu bớt khỏi k một phần tử bất kỳ thì bao đóng của nó khác A.

Từ định nghĩa khóa và theo bổ đề 2.2 ta thấy có thể tìm khóa bắt đầu từ tập k = A vì $k^+ = A$ và bớt dần các phần tử của k để nhận được tập bé nhất có bao đóng bằng A.

Cho sơ đồ quan hệ W = < A, F >. Tìm một khóa của W.

Thuật toán 2.2 Tìm 1 khóa của W = < A, F >

Input: W= < A, F >;

Output: k là khóa của W;

Algorithm:

Bước 1: Đặt k = A

Bước 2: Lặp quá trình loại khỏi k các thuộc tính *thừa B* mà $(k - B)^+ = A$

Begin

k: = A;

for each B in k do

if $(k - B)^+ = A$ then k: = k - B

End

Lưu ý:

- Thuật toán 2.2 trên đây cho ta tìm được một khóa của sơ đồ quan hệ W.

- Nếu muốn tìm các khóa khác (nếu có) của sơ đồ quan hệ ta có thể thay đổi thứ tự loại bỏ các phần tử của k.

Ví dụ 2.25.a:

Cho W = < A, F >, với A = {H, B, C, D, E, G, J, I}

F ={HC → B, BI → HCD, HBC → D, J → I, HCE → BCG, CG →HE}

Tìm k = ?

Bước 1: k = A = {H, B, C, D, E, G, J, I}

Bước 2: Lần lượt loại các thuộc tính *thừa* trong k:

Thử loại thuộc tính H: Ta có $\{B, C, D, E, G, J, I\}^+ = A$ (vì CG → HE) nên H được loại và k = {B, C, D, E, G, J, I}

Thử loại phần tử B: Ta có $\{C, D, E, G, H, I\}^+ = A$ (vì CG → HE và HC → B) nên B được loại và k = {C, D, E, G, J, I}

Thử loại phần tử C: Ta có $\{D, E, G, J, I\}^+ \neq A$ nên C không loại được và k = {C, D, E, G, J, I}

Thử loại phần tử D: Ta có $\{C, E, G, J, I\}^+ = A$ (vì CG → HE, HC → B, HBC → D) nên D được loại và k = {C, E, G, J, I}

Thử loại phần tử E: Ta có $\{C, G, J, I\}^+ = A$ (vì CG → HE, HC → B, HBC → D} nên E loại được và k = {C, G, J, I}

Thử loại phần tử G: Ta có $\{C, J, I\}^+ \neq A$ nên G không loại được và k = {C, G, J, I}

Thử loại phần tử J: Ta có $\{C, G, I\}^+ \neq A$ nên J không loại được và k = {C, G, J, I}

Thử loại phần tử I: Ta có $\{C, G, J\}^+ = A$ (vì CG → HE, HC → B, HBC → D, J → I} nên I loại được và k = {C, G, J}. Vậy k = {C, G, J} là một khóa của W.

Thuật toán 2.2 cho ta phương pháp tìm 1 khóa của W = < A, F >. Một vấn đề có tính lý thuyết quan trọng là có thuật toán nào tìm hết các khóa của W = < A, F > không?

Thuật toán 2.3 Thuật toán tìm các khóa của W = < A, F >

Input W = < A, F >

Output $k_1, k_2, \ldots, k_l$ là các khóa của W

Algorithm

Bước1: Xây dựng $P(A) = \{X_1, X_2, \ldots\}$ là họ tất cả các tập con của A

Bước 2: Xác định $M = \{X_i: X_i \in P(A)\ \&\ X_i^+ = A\}$.

Bước 3: Xác định K là họ các tập cực tiểu của M (k là cực tiểu của M nếu nó không chứa phần tử nào của M). Khi đó K là họ khóa của W

Ví dụ 2.25.b:

Cho W = < A, F >, với A = {H, B, C, D}; F ={HC → B, HBC → D}

Bước 1: P(A) = {Φ, H, B, C, D, HB, HC, HD, BC, BD, CD, HBC, HBD, HCD, BCD, HBCD}

Bước 2: M = {HC, HBC, HBCD}

Bước 3: K = {HC}. Vậy W có 1 khóa k = HC.

Ví dụ 2.26:

Cho W = < A, F >, với A = {H, B, C, D}; F ={HC → B, HBC → D, D → HC}

Bước 1: P(A) = {Φ, H, B, C, D, HB, HC, HD, BC, BD, CD, HBC, HBD, HCD, BCD, HBCD}

Bước 2: M = {HC, HBC, HBCD, D}

Bước 3: K = {HC, D}. Vậy W có 2 khóa k_1= HC, k_2 = D.

2.5.3. Các tính chất của khóa

Từ định nghĩa khóa và thuật toán tìm khóa ta có các *kết luận* sau:

Bổ đề 2.3

1. Các thuộc tính không xuất hiện trong cả vế trái và vế phải của tập F phải có trong mọi khóa k của W = < A, F >.

2. Các thuộc tính chỉ xuất hiện bên trái của các phụ thuộc hàm trong F của *W = <A, F > cũng phải thuộc mọi khóa k.*

3. Số lượng các phần tử trong mỗi khóa có thể khác nhau.

4. Họ tất cả các khóa của một W là hệ Sperner (tức không có hai khóa bao nhau)

Trong một số tài liệu các tác giả đã định nghĩa khóa là tập thuộc tính k tối thiểu thỏa mãn tính chất: với mọi cặp t_1, t_2 khác nhau của mọi quan hệ r trên A ta luôn có $t_1.k \neq t_2.k$.

Định nghĩa khác của khóa tối thiểu.

Định nghĩa 2.23 Khóa (key) tối thiểu

Tập thuộc tính $k \subseteq A$ được gọi là khóa của W = < U, F > nếu thỏa mãn hai điều kiện:

Với mọi cặp $t_1, t_2 \in r$ (r bất kỳ trên A) mà $t_1 \neq t_2$ thì $t_1.k \neq t_2.k$

k là tối thiểu.

Định lý 2.5

Hai định nghĩa 2.22 và 2.23 tương đương nhau (nghĩa là từ định nghĩa này suy định nghĩa kia)

Để chứng minh định lý 2.5 ta chỉ cần chứng minh điều kiện (1) của hai định nghĩa tương đương:

a. Nếu k là khóa của sơ đồ quan hệ W = < A, F >, r là quan hệ trên A thì với mọi cặp phần tử khác nhau t_1, t_2 của r luôn có $t_1. k \neq t_2. k$.

Thật vậy:

Giả sử k là khóa của W, t_1, t_2 là hai phần tử khác nhau của r mà $t_1. k = t_2. k$.

Vì k là khóa nên ta có $k \rightarrow A$ và từ $t_1. k = t_2. k$ ta có $t_1. A = t_2. A$ tức $t_1 = t_2$, điều này mâu thuẫn với giả thiết là $t_1 \neq t_2$.

b. Ngược lại nếu k là tập tối thiểu thỏa mãn với mọi quan hệ r trên A và mọi cặp t_1, t_2 của r mà $t_1. k \neq t_2. k$ thì $k \rightarrow A$ nên k là khóa của sơ đồ quan hệ W = < A, F >.

2.5.4. Khóa của một quan hệ R

Chúng ta đã xét khóa và thuật toán tìm khóa của sơ đồ quan hệ W = < A, F >. Bây giờ chúng ta xét khóa của quan hệ R và các thuật toán tìm khóa của quan hệ R. Ta cần lưu ý rằng nếu k là khóa của sơ đồ quan hệ W = < A, F > thì k là khóa của mọi quan hệ R trên A. Cho $A = \{A_1, A_2..., A_n\}$ là lược đồ quan hệ. $R = \{t_1, t_2,..., t_m\}$ là quan hệ trên A.

Định nghĩa 2.24 Khóa của quan hệ R

Tập $k \subseteq A$ được gọi là *khóa* của R nếu:

Mọi cặp t_i và t_j khác nhau của R ta luôn có $t_i.k \neq t_j.k$

k là tối thiểu

Cho $A = \{A_1, A_2..., A_n\}$ là lược đồ quan hệ. $R = \{t_1, t_2,..., t_m\}$ là quan hệ trên A

Ta sẽ định nghĩa ma trận khác nhau D và ma trận giống nhau (bằng nhau) E:

Định nghĩa 2.25 Ma trận khác nhau (ma trận phân biệt) D của quan hệ R

Ma trận vuông m hàng, m cột (m là số các phần tử của R), ký hiệu $D = (d_{ij})$ với $d_{ij} = \{B \in A: t_i.B \neq t_j.B\}$ được gọi là *ma trận khác nhau (*hay *ma trân phân biệt)* của R trên A.

Ta dễ dàng thấy rằng D là ma trận đối xứng qua đường chéo chính.

Định nghĩa 2.26 Ma trận bằng nhau E của quan hệ R

Ma trận vuông m hàng, m cột (m là số các phần tử của R), ký hiệu $E = (e_{ij})$ với $e_{ij} = \{B \in A: t_i.B = t_j.B\}$ được gọi là *ma trận bằng nhau* của R trên A.

Ta dễ dàng thấy rằng E cũng là ma trận đối xứng qua đường chéo chính và với mọi i, j ta luôn có $e_{ij} \cup d_{ij} = A$. Trên đường chéo chính của E luôn là tập thuộc tính A, trên đường chéo chính của D luôn là các tập rỗng ϕ.

Ví dụ 2.27:

Cho quan hệ R gồm 3 phần tử trên A = {I, B, C, G, L, H} như sau:

R

I	B	C	G	L	H
1	2	2	2	3	3
1	2	3	3	3	4
1	1	3	2	2	4

Khi đó ma trận D và E có các dạng:

$$E = \begin{pmatrix} A - - - - - IBL - - - IG \\ IBL - - - - A - - - - IH \\ IG - - - - IH - - - A \end{pmatrix} \quad D = \begin{pmatrix} \phi - - CGH - - BCLH \\ CGH - - \phi - - BCGL \\ BCLH - - BCGL - - \phi \end{pmatrix}$$

Trong ma trận E, A = {I, B, C, G, L, H} là tập tất cả các thuộc tính.

Lưu ý:

- Trong các thuật toán dùng các ma trận D, E các phần tử trên đường chéo chính ta bỏ qua, không sử dụng chúng trong mọi trường hợp và vì D và E đối xứng qua đường chéo chính nên khi xác định các ma trận này ta chỉ cần xác định các d_{ij}, e_{ij} thuộc nửa trên (j >i).

- Sau đây ta sẽ trình bày thuật toán tìm một khóa của R dựa vào E và thuật toán tìm các khóa của R dựa vào D.

Thuật toán 2.4 Thuật toán tìm 1 khóa của R dựa vào ma trận bằng nhau E

Input Quan hệ $R = \{t_1, t_2, ..., t_m\}$ trên $A = \{A_1, A_2, ..., A_n\}$

Output k là một khóa của R

Algorithm

Bước 1. Tính nửa trên của ma trận E, tức tính e_{ij} với j > i

Bước 2. Lọc trong E ra các phần tử cực đại cho vào M (phần tử cực đại của E là phần tử không bị chứa trong phần tử khác của E).

Bước 3. Đặt k = A

Bước 4. Cho i chạy từ 1 đến n nếu $(k - A_i)$ không là tập con của mọi $m \in M$ thì $k = k - A_i$

Kết thúc vòng lặp ở bước 4 ta có k là một khóa của R.

Định lý 2.6

k tìm được theo thuật toán 2.3 là một khóa của R.

Chứng minh:

Để chứng minh k là khóa ta cần lưu ý rằng k là tối thiểu vì trong quá trình tối thiểu hóa k ở bước 4 ta đã kiểm tra cho mọi thuộc tính của A. Bây giờ ta chỉ cần chứng minh nếu $i \neq j$ thì $t_i.k \neq t_j.k$. Điều này hiển nhiên vì nếu $t_i.k = t_j.k$ thì k là tập con của một tập m trong M, điều này vô lý với việc chọn tập k trong bước 4.

Thuật toán 2.5 Thuật toán tìm tất cả các khóa của R dựa vào ma trận khác nhau D

Input Quan hệ $R = \{t_1, t_2, ..., t_m\}$ trên $A = \{A_1, A_2, ..., A_n\}$

Output $K = \{k_1, k_2, ...k_q\}$ là họ khóa của R

Algorithm

Bước 1. Tính nửa trên của ma trận D, tức tính d_{ij} với j > i

Bước 2. Coi mỗi thuộc tính là một biến logic. Đặt $f = \wedge(\vee d_{ij})$; f là hội của các tuyển ($\vee d_{ij}$ là tuyển của các thuộc tính trong d_{ij})

Bước 3. Tối giản f và đưa f về dạng tuyển của các hội $f = k_1 \vee k_2 \vee ... \vee k_q$ với mỗi k_i là một hội của các thuộc tính.

Khi đó mỗi k_i là một khóa của R.

Lưu ý:

Nếu ta coi mỗi thuộc tính A_i của A là một biến logic (miền trị của A_i chỉ có hai giá trị đúng hoặc sai). Khi đó biểu thức $f = \wedge(\vee d_{ij})$ là hội của các tuyển của các d_{ij} được gọi là hàm Bool của R. Mỗi k_i trong bước 3 ở trên có dạng hội các thuộc tính $B \wedge ... \wedge C$ ta coi như $k_i = B...C$

Định lý 2.7

Họ các k_i tìm được trong thuật toán 2.5 là họ khóa của quan hệ R.

Trước khi chứng minh định lý 2.7 ta xét ví dụ

Ví dụ 2.28:

Cho $R = \{t_1, t_2, t_3, t_4, t_5, t_6, t_7\}$ là quan hệ trên A = {I, B, C, G, H} sau:

R

I	B	C	G	H
1	0	1	0	1
1	2	3	3	1
2	2	3	4	2
3	3	2	2	3
4	4	0	5	5
5	5	4	1	1
4	5	5	6	0

Khi đó nửa trên (cả đường chéo) của ma trận D sẽ là

$$\begin{array}{r} d_{11}\, d_{12}\, d_{13}\, d_{14}\, d_{15}\, d_{16}\, d_{17} \\ d_{22}\, d_{23}\, d_{24}\, d_{25}\, d_{26}\, d_{27} \\ d_{33}\, d_{34}\, d_{35}\, d_{36}\, d_{37} \\ d_{44}\, d_{45}\, d_{46}\, d_{47} \\ d_{55}\, d_{56}\, d_{57} \\ d_{66}\, d_{67} \\ d_{77} \end{array} = \begin{array}{ccccccc} \phi & BCG & A & A & A & IBCG & A \\ & \phi & IGH & A & A & IBCG & A \\ & & \phi & A & A & A & A \\ & & & \phi & A & A & A \\ & & & & \phi & A & BCGH \\ & & & & & \phi & ICGH \\ & & & & & & \phi \end{array}$$

Sử dụng các tính chất của tuyển và hội:

$\alpha \wedge \alpha = \alpha$ (1)

$\alpha \vee \alpha = \alpha$ (2)

$\alpha \wedge (\alpha \vee \beta) = \alpha$ (3)

$\alpha \vee (\alpha \wedge \beta) = \alpha$ (4)

$(a \wedge b) \vee (\alpha \wedge \beta) = (a \vee \alpha) \wedge (a \vee \beta) \wedge (b \vee \alpha) \wedge (b \vee \beta)$ (5)

$(a \vee b) \wedge (\alpha \vee \beta) = (a \wedge \alpha) \vee (a \wedge \beta) \vee (b \wedge \alpha) \vee (b \wedge \beta)$ (6)

Khi đó:

$f = \wedge(\vee d_{ij}) = (B \vee C \vee G) \wedge (I \vee G \vee H)$

$= (I \wedge B) \vee (B \wedge H) \vee G = IB \vee BH \vee G = k_1 \vee k_2 \vee k_3.$

Vậy R có 3 khóa $k_1 = IB$, $k_2 = BH$, $k_3 = G$

Chúng ta thấy biểu thức tối giản của $f = \wedge(\vee d_{ij})$ là tuyển của các hội sau khi đã lược bỏ các số hạng tương đương.

Định lý 2.7 sẽ được chứng minh bằng các bước như sau:

Bước 1: Gọi Z là họ của các X_i mà mỗi X_i là một tập các đại diện khác nhau của các d_{ij} và gọi K là họ khóa của R. Ta dễ dàng chứng minh được rằng $K \subseteq Z$.

Bước 2: Sau khi rút gọn $f = \wedge(\vee d_{ij})$ các đại diện của họ Z không là khóa sẽ triệt tiêu.

- Ta xét tập $X \subset A$ gồm các đại diện của các d_{ij}, mỗi d_{ij} lấy 1 thuộc tính, tức:

$X = \{B: B \in d_{ij}; \forall i < j\}$

(ở đây ta xét nửa trên của ma trận D, tức j > i). Ta dễ dàng thấy rằng mọi cặp t_i và t_j khác nhau của quan hệ R ta luôn có $t_i.X \neq t_j.X$.

Thật vậy, vì ta giả sử $\exists$ cặp t_i và t_j khác nhau của R mà: $t_i.X = t_j.X$

Khi đó mọi $B \in X$ thì $t_i.B = t_j.B$ điều này vô lý vì $B \in d_{ij}$

- Bằng cách thay đổi cách chọn thuộc tính B của các d_{ij} ta sẽ có các tập X khác nhau. Gọi $Z = \{X_1, X_2, ...\}$ là họ tất cả các tập X như vậy.

- Gọi $K = \{k_1, k_2, ...\}$ là họ tất cả các khóa của R.

Ta sẽ chứng minh: $K \subset Z$.

Thật vậy, giả sử $\exists$ một khóa $k \in K$ mà $k \notin Z$.

Vì $k \not\in Z$ nên $\exists$ một thuộc tính $B \in k$ sao cho $B \not\in d_{ij}$ ($\forall$ i, j = 1,... m).

Do B không thuộc bất kỳ một tập d_{ij} nào nên: $\forall$ i, j thì $t_i.B = t_j.B$

$$\left.\begin{array}{l} \text{Mà: } B \in k \\ \quad \forall\ i, j = 1,\dots, m \\ \quad t_i.k \neq t_j.k \end{array}\right\} \Rightarrow t_i.(k - B) \neq t_j.(k - B)\ \forall i, j$$

Suy ra, k vi phạm tính tối thiểu của khóa điều này vô lý.

Tức là, ta đã chứng minh được: $K \subset Z$.

• Sử dụng (1), (2), (3) và (4) ta sẽ chứng minh:

$$\wedge(\vee d_{ij}) = \vee(\wedge X_k)\ (X_k \in Z) \tag{7}$$

Vì phép tuyển và phép hội là các phép toán được xây dựng trên các logic mệnh đề nên để chứng minh (7), ta sẽ xét các giá trị đúng hoặc sai của mệnh đề logic.

Nhận xét: $\wedge(\vee d_{ij})$ chỉ có giá trị đúng nếu với mọi $\vee d_{ij}$ có giá trị đúng. Tức không $\exists$ d_{ij} nào mà tất cả các phần tử của nó đều sai ($\vee d_{ij}$ sai nếu tất cả phần tử của nó đều sai)

Xét hai trường hợp:

+ *Trường hợp 1:* $\exists$ một d_{ij} nào đó mà tất cả các phần tử của d_{ij} là sai suy ra $\vee d_{ij}$ có giá trị sai nên $\wedge(\vee d_{ij})$ có giá trị sai.

Vì $\forall X_k \in Z$ thì X_k luôn phải có 1 phần tử của d_{ij}. Mà mọi phần tử của d_{ij} đều sai nên $\forall X_k$ thì $\wedge X_k$ có giá trị sai suy ra $\vee(\wedge X_k)$ có giá trị sai.

Vậy ở trường hợp này $\wedge(\vee d_{ij}) = \vee(\wedge X_k)$

+ *Trường hợp 2:* Không $\exists$ một d_{ij} nào đó mà tất cả các phần tử của nó đều sai.

Suy ra $\forall$ i, j $\exists$ một thuộc tính $A_{ij} \in d_{ij}$ là đúng và $\vee d_{ij}$ có giá trị đúng $\forall$ i, j. Nên $\wedge(\vee d_{ij})$ có giá trị đúng.

Theo cách xây dựng các tập X_k của họ Z thì phải $\exists$ một $X_{k0} \in Z$ mà $X_{k0} = \{A_{ij}\ \forall$ i, j$\}$. Vì X_{k0} gồm toàn bộ phần tử có giá trị đúng nên $\wedge X_{k0}$ có giá trị đúng.

Mà $\vee(\wedge X_k)$ sẽ đúng nếu có ít nhất một X_k nào đó mà $\wedge X_k$ đúng.

Vậy, vì $\wedge X_{k0}$ đúng nên $\vee(\wedge x_k)$ đúng

Suy ra $\wedge(\vee d_{ij}) = \vee(\wedge X_k)$

• Bây giờ ta sẽ chứng minh mọi nhóm $\wedge X_k$ mà X_k không là khóa sẽ triệt tiêu trong biểu thức $\vee(\wedge X_k)$.

Giả sử có $X_{k0} \in Z$ mà X_{k0} không là khóa. Ta sẽ cực tiểu hóa X_{k0} để tìm được khóa $k \subset X_{k0}$. Việc này có thể mô tả bằng thuật toán sau:

Bước 1: $k = X_{k0}$

Bước 2: Lặp quá trình loại phần tử B mà $\forall\ t_i, t_j$ thì: $t_i.(k - B) \neq t_j.(k - B)$

Thuật toán dừng lại khi không tìm được phần tử B như vậy:

Thuật toán trên sẽ đảm bảo k tìm được là khóa mà $k \subset X_{k0}$.

Ta đã chứng minh ở những phần trước nếu $k \in Z$ thì trong biểu thức $\vee(\wedge X_k)$ có nhóm $\wedge k$ và nhóm $\wedge x_{k0} \Rightarrow \vee(\wedge X_k) = T \vee [(\wedge k) \vee (\wedge X_{k0})]$

gọi $X_{k0} - k = \{B_1, B_2,..., B_v\}$. Vì $k \subset X_{k0}$ nên $X_{k0} = k \cup \{B_1, B_2,..., B_v\}$

$\Rightarrow \wedge k \vee (\wedge X_{k0}) = (\wedge k) \vee (\wedge k \wedge \{B_1, B_2,..., B_v\}) = \wedge k$

(Theo tính chất $\alpha \vee (\alpha \wedge \beta) = \alpha$)

$\vee(\wedge X_k) = T \vee (\wedge k)$

Tức là ta đã chứng minh được X_{k0} không khóa đã bị triệt tiêu.

Vậy tính đúng đắn của thuật toán đã được chứng minh.

Lưu ý: Một quan hệ R có thể có nhiều khóa

Định nghĩa 2.27 Khóa chính, khóa ngoại

Khóa chính của quan hệ R là một khóa của R được chọn làm khóa chính. *Khóa ngoại* của R là trường dữ liệu đóng vai trò khóa chính trong quan hệ S khác.

Ví du 2.29:

Xét CSDL trong ví dụ 2.17

kh(<u>Makh</u>, Tenkh, DS...)

nv(<u>Manv</u>, Tnenv...)

sp(<u>Masp</u>, Tensp, DVT, NSX, GIA)

hd(<u>SOHD</u>, NHD, Makh, Manv, Trigia)

cthd(<u>SOHD</u>, <u>Masp</u>, SL)

Khi đó Makh là khóa chính trong quan hệ kh và là khóa ngoại trong quan hệ hd.

Manv là khóa chính trong quan hệ nv và là khóa ngoại trong quan hệ hd.

SOHD là khóa chính trong quan hệ hd và là khóa ngoại trong quan hệ cthd.

Masp là khóa chính trong quan hệ sp và là khóa ngoại trong quan hệ cthd.

2.5.5. Phản khóa của sơ đồ quan hệ W = < A, F>

Cho W = < A, F> là sơ đồ quan hệ. Giả sử $K = \{k_1, k_2, ..., k_l\}$ là họ khóa của W.

Định nghĩa 2.28 Phản khóa của sơ đồ quan hệ

Tập thuộc tính $X \subseteq A$ được gọi là một *phản khóa* của W nếu X không chứa một khóa k_i nào nhưng nếu thêm vào X một thuộc tính B thì XB chứa một khóa k_i nào đó. Ký hiệu K^{-1} là họ các phản khóa của W.

Ví dụ

$W = < HBCD, \{HB \rightarrow C, C \rightarrow H\}>$.

Khi đó họ khóa của W là K = {HBD, CBD} và K^{-1} = {BD}; W chỉ có một phản khóa.

2.6. CÁC DẠNG CHUẨN

Trong công tác quản lý và xử lý các hệ cơ sở dữ liệu, chúng ta thường phải tiến hành "chuẩn hóa" các hệ CSDL. Tức là chúng ta sẽ xét một số dạng đặc biệt mà trong CSDL gọi là các dạng chuẩn NF (Normal Forms) và trong nhiều trường hợp CSDL ở NF này thì tốt hơn ở NF kia. Sự phân loại các sơ đồ quan hệ W = < A, F > theo các NF thực chất là phân loại các sơ đồ quan hệ dựa vào đặc trưng của tập phụ thuộc hàm F. Các dạng chuẩn ta xét đối với các sơ đồ quan hệ. Tất nhiên nếu W = < A, F > là 2NF thì các quan hệ R trên A cũng 2NF. Tương tự W= < A, F > là NF nào thì các quan hệ R trên A là NF đó.

Sau đây chúng ta sẽ xét một số dạng chuẩn quen thuộc của các sơ đồ quan hệ đã được nhiều tác giả quan tâm.

Chúng ta phải nghiên cứu các dạng chuẩn vì chúng ta cảm giác rằng các bảng trong cơ sở dữ liệu ở dạng chuẩn 3NF hay BCNF thường tốt hơn ở chuẩn 1NF, 2NF.

2.6.1. Dạng chuẩn 1: 1NF

Dạng chuẩn 1 (1^{st} Norm Form) ký hiệu là 1NF.

Cho lược đồ quan hệ A, F là tập phụ thuộc hàm trên A.

Khi đó ta có sơ đồ quan hệ W = < A, F >.

Định nghĩa 2.29 Dạng chuẩn 1

W = <A, F> *được gọi là dạng chuẩn 1* (1NF) nếu và chỉ nếu toàn bộ các miền giá trị có mặt trong A đều chỉ chứa các giá trị nguyên tố (giá trị nguyên tố là giá trị không thể tách thành các giá trị khác - giá trị đơn).

Khi W = <A, F> là 1NF thì mọi quan hệ r trên A cũng được gọi tương ứng là *quan hệ 1NF*.

Ví dụ 2.30:

Đặt A = {Masv, Hoten, Que, GT}, Masv (mã sinh viên) là khóa chính khi đó tập phụ thuộc hàm (không tính các phụ thuộc hàm tầm thường):

F = {Masv → A} và W = < A, F > và vì mỗi sinh viên chỉ có một mã duy nhất, một họ tên, một quê và một giới tính nên W là 1NF. Tất nhiên mọi quan hệ r trên A là 1NF.

Lấy A = {Magv, Tenmon, HT} với Magv là mã giáo viên, Tenmon là tên môn học mà giáo viên tham gia dạy, HT là hội trường lên lớp. Khi đó mỗi giáo viên có thể dạy nhiều môn và ở nhiều hội trường khác nhau ví dụ ta xét bảng thông báo R trên A sau:

R

Magv	Tenmon	HT
Gv_1, Gv_2	CSDL, Java	E1, E2
Gv_3, Gv_4	C, C^{++}, Web	E3, E4

Bảng R có thể coi là một quan hệ R trên A (theo nghĩa mở rộng - không đơn trị vì giá trị của mỗi thuộc tính trên mỗi bộ đều không đơn trị).

Đây là một quan hệ không là 1NF vì các thuộc tính có miền giá trị không đơn trị mà đa trị. Vậy quan hệ trên *không là dạng chuẩn 1NF* và tất nhiên W = < A, F > cũng không là 1NF.

Lưu ý:

Tuy nhiên mọi quan hệ R như trên ta luôn đưa về được dạng đơn trị

R

Magv	Tenmon	HT
Gv_1	CSDL	E1
Gv_2	Java	E2
Gv_3	C	E3
Gv_4	C^{++}	E4
Gv_3	Web	E3

Đây là quan hệ đơn trị.

Lưu ý:

Mọi quan hệ đều có thể đưa về dạng đơn trị nên mọi SĐQH W = < A, F> đều có thể coi là chuẩn 1NF. Vậy lớp 1NF chứa tất cả các sơ đồ quan hệ.

2.6.2. Dạng chuẩn 2: 2NF

Định nghĩa 2.30 Định nghĩa phụ thuộc hoàn toàn

Ta nói Y *phụ thuộc hoàn toàn* vào X, nếu trong X không có tập con thực sự X_1 mà $X_1 \rightarrow Y$. Nói cách khác Y phụ thuộc hoàn toàn vào X nếu: $X \rightarrow Y$ và bớt khỏi X dù một thuộc tính B thì tập còn lại X - B không kéo theo Y.

Cho sơ đồ quan hệ W = < A, F >; k là một khóa của W.

Định nghĩa 2.31 Dạng chuẩn 2NF

Ta nói W *là 2NF*, nếu mọi thuộc tính thứ cấp của W phụ thuộc hoàn toàn vào khóa. Nói cách khác W là 2NF nếu: trong W không có phụ thuộc hàm dạng: $X \rightarrow b$ mà $b \in F_n$ và X là tập con thực sự của khóa.

Từ định nghĩa ta thấy ngay là lớp các sơ đồ quan hệ 2NF là lớp con thực sự của lớp 1NF, vì có nhiều sơ đồ quan hệ không là 2NF.

Ví dụ 2.31:

Ta xét hồ sơ nhân sự cơ quan (quan hệ r) như sau:

TT	HOTEN	NS	TĐÔ	QUÊ	GT
01	Tuấn Anh	1960	Đại học	Huế	Nam
02	Lan Anh	1977	Đại học	Hà Nội	Nữ
03	Đình Đông	1945	TS	Vĩnh Phú	Nam
05	Đình Đông	1943	TS	Hà Nội	Nam
06	Công Nụ	1960	TS	Vĩnh Phú	Nam
07	Hoa Huệ	1972	Tr. học	Nghệ An	Nữ

Ta thấy ngay rằng tập có một thuộc tính TT là khóa của quan hệ r vì theo định nghĩa của PTH ta có TT → {HOTEN, NS, TĐÔ, GT, QUÊ}. Vì r là 1NF và tập khóa chỉ có một phần tử nên không thể có phần tử không khóa phụ thuộc hàm vào tập con thực sự của khóa (tập con thực sự của khóa bằng rỗng), vậy r là 2NF. Ta có kết luận sau:

Bổ đề 2.4

W là 2NF nếu mỗi khóa của W chỉ có một thuộc tính.

Ví dụ 2.32:

Quay lại ví dụ 2.23, ta đã thấy khóa của W là:

$K_1 = \{H, B\}, K_2 = \{B, E\}, K_3 = \{C, G\},$

$K_4 = \{C, E\}, K_5 = \{C, D\}, K_6 = \{B, C\}.$

Trong ví dụ này tất cả các phần tử của tập thuộc tính A đều là phần tử khóa, tức là tập các phần tử không khóa bằng rỗng nên không có PTH dạng $B \rightarrow b$ mà b là phần tử thứ cấp. Vậy W là 2NF.

Bổ đề 2.5

W là 2NF nếu tập các thuộc tính không khóa $F_n = \phi$.

Ví dụ 2.33:

Sau đây ta xét một ví dụ W không là 2NF.

Cho W = < A, F >, với A = {G, B, C, D, E, H}

$$F = \{G \rightarrow E, C \rightarrow D, E \rightarrow DH\}.$$

Dễ dàng thấy rằng tập k = {G, B, C} là khóa duy nhất của W, D là thuộc tính không khóa và $C \rightarrow D$, vì C là tập con thực sự của khóa nên *W* không là 2NF.

Hay xét tập thuộc tính A = {Mav, Tenv, Pho, Sonha} gồm các thuộc tính mã vùng, tên vùng, phố, số nhà. Dễ dàng thấy rằng từ mã vùng suy ra tên vùng và từ tên vùng suy ra mã vùng và khi đó tập F = {Mav → Tenv, Tenv → Mav}, sơ đồ quan hệ W = < A, F > là 2NF vì W có hai khóa k_1 = {Mav, Pho, Sonha}, k_2 = {Tenv, Pho, Sonha} và tập thứ cấp $F_n = \varnothing$.

Vậy khi xét một sơ đồ quan hệ W = < A, F > có là 2NF không ta phải tính tất cả các khóa của sơ đồ quan hệ và từ đó suy ra tập thuộc tính thứ cấp F_n, cuối cùng cần xét xem có tập con của khóa kéo theo thuộc tính thứ cấp không?

Ví dụ W = < A, F >; A = {G, B, C, D, E}, F = {$G \rightarrow B, C \rightarrow D$}.

Khi đó mọi khóa k phải chứa E, G, C và ta cũng thấy rằng tập k = {G, C, E} là khóa duy nhất. Vậy các thuộc tính thứ cấp là B, D. Ta thấy trong khóa k có chứa các tập con thực sự G, C và chúng kéo theo các thuộc tính thứ cấp B, D tương ứng, nên W không là 2NF.

Vậy *vi phạm của 2NF là tập con thực sự của khóa kéo theo thuộc tính thứ cấp* và đây là lý do tại sao khi xét dạng 2NF ta phải tìm hết khóa để có tập F_n.

2.6.3. Dạng chuẩn 3: 3NF

Cho sơ đồ quan hệ W = < A, F >.

Định nghĩa 2.32 Dạng chuẩn 3NF

Ta nói W là *3NF* nếu trong W không tồn tại phụ thuộc hàm dạng $X \to b$ mà $b \in F_n$ và $X^+ \neq A$.

Từ định nghĩa ta có nhận xét rằng nếu W là 3NF thì nó là 2NF. Thật vậy:

Giả sử W = < A, F > là 3NF mà không là 2NF. Tức là có tập con X của khóa kéo theo thuộc tính thứ cấp b: $X \to b$ mà $b \in F_n$. Vì X là tập con của khóa nên $X^+ \neq A$ => trong W có $X \to b$ mà $b \in F_n$ và $X^+ \neq A$. Vô lý vì W là 3NF.

Ví dụ 2.34:

a. W không là 3NF, ta lấy ví dụ 2.28. Trong ví dụ này ta thấy D là thuộc tính thứ cấp và $C \to D$, đồng thời $C^+ \neq U$. Vậy W không là 3NF.

b. Trở lại ví dụ 2.23, ta thấy rằng W trong ví dụ này là 3NF, vì tập thuộc tính A ở đây không có thuộc tính không khóa; $F_n = \phi$.

Ví dụ 2.35:

Cho sơ đồ quan hệ W = < A, F >.

với U = {E, B, C, D}; F = {EB → C, D → B, C → EBD}

Ta thấy rằng, các tập k_1 = {E, B}, k_2 = {E, D}, k_3 = {C} là các khóa. Vậy A không có thuộc tính thứ cấp, nên W là 3NF.

Bổ đề 2.6

Mọi sơ đồ quan hệ W = < A, F > không chứa thuộc tính thứ cấp thì W là 3NF.

2.6.4. Dạng chuẩn BCNF: Boyce - Codd Normal Form

Dạng chuẩn Boyce - Codd Normal Form ký hiệu là BCNF.

Cho sơ đồ quan hệ W = < A, F >

Định nghĩa 2.33 Dạng chuẩn BCNF

W *là BCNF* nếu trong W không tồn tại phụ thuộc hàm dạng $X \to b$ với $b \notin X$ và $X^+ \neq A$.

Bổ đề 2.7

Nếu W là BCNF thì W là 3NF.

Thật vậy:

Giả sử W = < A, F > là BCNF và W không là 3NF. Nghĩa là trong W có $X \to b$ mà $b \in F_n$ và $X^+ \neq A$. Vậy trong W có $X \to b$ mà $b \notin X$ và $X^+ \neq A$ điều này vô lý vì W là BCNF.

Ví dụ 2.36:

Cho W = < A, F >, với U = {G, B, C, D}

$$F = \{GB \rightarrow C, C \rightarrow GBD\}$$

Dễ dàng thấy rằng:

Các tập có bao đóng khác A là X = {G} hoặc X = {B} hoặc X = {D} hoặc X = {G, D} hoặc X = {B, D} và trong các tập trên không có phụ thuộc hàm dạng $X \rightarrow b$ với $b \notin X$. Vậy W là BCNF.

Ví dụ 2.37:

Xét một sơ đồ quan hệ W mà W là 3NF nhưng W không là BCNF.

Cho W = < {G, B, C, D}, {G → BCD, BC → DG, B → C}>. Rõ ràng rằng W là 3NF vì W có hai khóa là {G} và {B,C} nên tập không khóa là D và không có tập nào có bao đóng khác A kéo theo thuộc tính thứ cấp D. Ngược lại W không là BCNF vì có phụ thuộc hàm B → C mà B^+ khác A.

Lưu ý:

Nếu W = < A, F> là 1NF, 2NF, 3NF, BCNF,... thì mọi quan hệ r trên A cũng là 1NF, 2NF, 3NF, BCNF,... tương ứng.

Từ định nghĩa suy ra rằng, cho sơ đồ quan hệ W = < A, F >

a. Muốn xét xem W là 2NF hay không ta phải:

- Tính tất cả các khóa k của W và suy ra đồng thời tập thuộc tính thứ cấp F_n
- Xét xem có k'→ x không (k': tập con thực sự của khóa k và x thuộc F_n).

Ví dụ, cho W = < A, F > với A = {H, B, C, D, E, G} và F = {H → BC, C → DE, E → G}. Khi đó ta thấy mọi khóa k phải chứa H và hơn thế nữa tập {H} là khóa nên ta có ngay tập thứ cấp là {B, C, D, E, G} và W là 2NF vì các tập khóa chỉ có một phần tử.

b. Muốn xét xem W là 3NF hay không ta phải:

- Tính tập thuộc tính thứ cấp F_n = {x,...};
- Xét xem có X → x ($x \notin X$ và $X^+ \neq R$).

Ví dụ, W = < A, F >, A = {I, B, C, D, E, G, H}

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

Ta thấy mọi khóa k đều phải chứa H, C, D và tập ba phần tử này là khóa. Vậy ta có tập các thuộc tính thứ cấp {I, B, E, G}. Từ F ta thấy ngay rằng không có tập X mà bao đóng khác A kéo theo thuộc tính thứ cấp. W là 3NF.

c. Muốn xét xem W là BCNF hay không ta phải:

- Xét xem có $X \to a$ với $a \notin X$ và $X^+ \neq A$.

Ví dụ, A = {I, B, C, D, E, G, H}, F = {I → BC, D → E, H → G}. Ta thấy tập con có bao đóng khác A mà kéo theo thuộc tính khác: D → E. Vậy W không là BCNF.

Định lý 2.8

Các lớp dạng chuẩn của các sơ đồ quan hệ có quan hệ lồng nhau, lớp sau nằm trong lớp trước. Nghĩa là ta có: 1NF ⊃ 2NF ⊃ 3NF ⊃ BCNF. Lồng nhau ở đây là thực sự, nghĩa là lớp sau nằm gọn trong lớp trước.

Thật vậy với A = {G, B, C, D} và F_1 = {GB → C, D → B, C → GBD} thì W_1 = <A, F_1> là 3NF nhưng không là BCNF (vì không có tập X mà có bao đóng khác A nhưng lại kéo theo thuộc tính thứ cấp (tập các thuộc tính thứ cấp bằng rỗng) nên W là 3NF. Ngược lại, W_1 không là BCNF vì nó chứa tập X = D có bao đóng khác A và kéo theo B. Còn W_2 = < A, F_2 > với F_2 = {B → D, G → C. C → GBD} là 2NF, nhưng không là 3NF (vì các thuộc tính thứ cấp B, D phụ thuộc hoàn toàn vào khóa nên nó là 2NF, ngược lại có tập X = {B} có bao đóng khác A nhưng kéo theo thuộc tính thứ cấp D nên nó không là 3NF), và rất nhiều sơ đồ quan hệ là 1NF nhưng không là 2NF.

2.7. PHỤ THUỘC ĐA TRỊ VÀ DẠNG CHUẨN 4: 4NF

Lớp các quan hệ chúng ta đã và đang xét rất lớn, một số các quan hệ có ngữ nghĩa (semantic) phức tạp, trong tập các thuộc tính không có phụ thuộc hàm hoặc có các phụ thuộc đặc biệt. Vậy để đi sâu nghiên cứu các đặc thù, tính chất của lớp các quan hệ, chúng ta sẽ trình bày tiếp khái niệm phụ thuộc đa trị.

Ví dụ 2.38:

Ta xét bảng r thông báo chủ và xe như sau:

r

Chủ	Xe	Biển
A	BMW	29F1
A	BMW	29F2
A	BMW	29F3
B	Toyota	29H1
B	Toyota	29H2

Trong quan hệ này không có phụ thuộc hàm giữa Chủ và Biển, nhưng giữa các thuộc tính Chủ, Biển, có mối quan hệ đặc biệt. Ví dụ ta thấy nếu cùng một Chủ thì tráo hai biển bất kỳ cho nhau ta vẫn được một xe hợp lệ của chủ đó. Kiểu phụ thuộc đặc biệt này S. Jajda gọi là phụ thuộc đa trị.

Vậy ta có khái niệm phụ thuộc đa trị trong các lược đồ quan hệ như sau.

2.7.1. Định nghĩa phụ thuộc đa trị

Cho lược đồ quan hệ $A = \{A_1, A_2, \ldots A_n\}$, $n \geq 3$. X và Y là các tập con của A và $Z = A - (X \cup Y)$. Mỗi $t \in r$ ta có thể coi t được ghép từ 3 phần chiếu của t tức là $t = t.Xt.Yt.Z$.

Định nghĩa 2.34 Phụ thuộc đa trị

Ta nói trong lược đồ quan hệ A *xác định phụ thuộc đa trị từ X lên Y* (X xác định đa trị Y), ký hiệu là $X \rightarrow\rightarrow Y$, nếu mọi cặp phần tử t_1, t_2 của r (r là quan hệ bất kỳ trên A) bằng nhau trên tập X, tức:

$$t_1 = t_1.Xt_1.Yt_1.Z \in r;\ t_2 = t_1.Xt_2.Yt_2.Z \in r \qquad (1)$$

và $t_1.X = t_2.X$ thì khi đổi đuôi t_1 và t_2 cho nhau (đổi $t_1.Z$ và $t_2.Z$) chúng ta vẫn nhận được các phần tử thuộc r, tức là:

$$t_1' = t_1.Xt_1.Yt_2.Z \text{ và } t_2' = t_1.Xt_2.Yt_1.Z \qquad (2)$$

cũng thuộc r

Nói cách khác $X \rightarrow\rightarrow Y$ nếu với mọi t_1, t_2 như trong (1) thuộc r và $t_1.X = t_2.X$ thì ta cũng có t_1', t_2' trong (2) thuộc r.

Ví dụ 2.39.a:

Xét quan hệ Chủ - Xe r như trên ta có ngay:

Chủ $\rightarrow\rightarrow$ Xe và Chủ $\rightarrow\rightarrow$ Biển

Ví dụ 2.39.b:

Một cách tổng quát cho lược đồ quan hệ A và hai tập thuộc tính $X, Y \subseteq A$.

Nếu $X \rightarrow Y$ thì $X \rightarrow\rightarrow Y$. Thật vậy giả sử t_1 và t_2 thuộc r và $t_1.X = t_2.X$ vì $X \rightarrow Y$, nên $t_1.Y = t_2.Y$. Khi đó $t_1' = t_1.Xt_1.Yt_2.Z = t_2.Xt_2.Yt_2.Z = t_2$ thuộc r. Tương tự ta có $t_2' = t_1$ thuộc r. Nên $X \rightarrow\rightarrow Y$

2.7.2. Các tính chất của phụ thuộc đa trị

Tính chất d1: Tính bù của phụ thuộc đa trị:

Nếu X, Y, Z là ba tập con rời nhau của lược đồ quan hệ A và $A = X \cup Y \cup Z$ với $A = \{A_1, A_2, \ldots, A_n\}$ thì $X \rightarrow\rightarrow Y$ khi và chỉ khi $X \rightarrow\rightarrow Z$.

Tính chất d2: Tính tăng trưởng:

Nếu $X \rightarrow\rightarrow Y$ và $V \subset W$ thì $XW \rightarrow\rightarrow YV$.

Tính chất d3: Tính bắc cầu:

Nếu $X \rightarrow\rightarrow Y$ và $Y \rightarrow\rightarrow V$ thì $X \rightarrow\rightarrow V \backslash Y$.

Tính chất d4: Tính pha trộn:

Nếu $X \rightarrow Y$ thì $X \rightarrow\rightarrow Y$.

Phụ thuộc hàm là trường hợp riêng của phụ thuộc đa trị.

Tính chất d5: Nếu $X \rightarrow\rightarrow Y$ và $W \rightarrow V$ với $V \subset Y$ và $W \cap Y = \varnothing$ thì $X \rightarrow V$. Và một số tính chất có thể suy dẫn được:

Tính chất 6: (Hợp) $X \rightarrow\rightarrow Y$ và $X \rightarrow\rightarrow Z \Rightarrow X \rightarrow\rightarrow YZ$.

Tính chất 7: (Tựa bắc cầu)

$X \rightarrow\rightarrow Y$ và $YW \rightarrow\rightarrow V \Rightarrow XW \rightarrow V - YW$.

Tính chất 8: (Tựa hợp)

$X \rightarrow\rightarrow Y$ và $XY \rightarrow W$, thì $X \rightarrow W - Y$.

Tính chất 9: Tính phân rã

Nếu $X \rightarrow\rightarrow Y$ và $X \rightarrow\rightarrow V$ thì

$X \rightarrow\rightarrow Y \cap V$;

$X \rightarrow\rightarrow Y - V$;

$X \rightarrow\rightarrow V - Y$.

Về sau ta thường ký hiệu phụ thuộc đa trị là MD (Multivalued Dependencies), ví dụ cho phụ thuộc đa trị ta viết cho MD $X \rightarrow\rightarrow Y$.

Định nghĩa 2.35 Phụ thuộc cơ sở

Một MD $X \rightarrow\rightarrow Y$ trên lược đồ A được gọi là *phụ thuộc cơ sở* nếu $Y \neq \varnothing$ và $X \cap Y = \varnothing$ và $X \cup Y \neq A$.

Sau đây ta sẽ chứng minh một vài tính chất của phụ thuộc đa trị:

Tính chất d1: Ta có $X \rightarrow\rightarrow Y$ và $Z = A - X - Y$.

Giả sử t_1 và t_2 là hai phần tử của r mà $t_1.X = t_2.X$ và $t_1 = t_1.Xt_1.Yt_1.Z$, $t_2 = t_2.Xt_2.Yt_2.Z$ theo giả thiết $X \rightarrow\rightarrow Y$ nên $t_1' = t_1.Xt_1.Yt_2.Z$, và $t_2' = t_2.Xt_2.Yt_1.Z$ cũng thuộc r, vì $t_1.X = t_2.X$ nên đổi $t_1.X$ và $t_2.X$ cho nhau trong các t_1' và t_2' ta vẫn được các phần tử thuộc r, hay $t_1' = t_2.Xt_1.Yt_2.Z$; $t_2' = t_1.Xt_2.Yt_1.Z$ cũng thuộc r. Theo định nghĩa ta có $X \rightarrow\rightarrow Z$. Vậy ta có định nghĩa tương đương:

Định nghĩa 2.36 Phụ thuộc đa trị

$X \rightarrow\rightarrow Y$ nếu t_1 và t_2 mà $t_1.X = t_2.X$ và $t_1 = t_1.Xt_1.Yt_1.Z$ cùng $t_2 = t_2.Xt_2.Yt_2.Z$ thuộc r thì $t_1' = t_1.Xt_2.Yt_1.Z$ và $t_2' = t_2.Xt_1.Yt_2.Z$ cũng thuộc r (thay phần giữa của hai phần tử t_1 và t_2).

Tính chất d2: Ta có $X \rightarrow\rightarrow Y$ và $V \subset W$, ta phải chứng minh: $XW \rightarrow\rightarrow YV$.

Thật vậy, giả sử t_1 và t_2 thuộc r

mà $t_1.\ XW = t_2.\ XW$

và $t_1 = t_1.\ XWt_1.\ YVt_1.\ Z = t_1.\ XWt_1.\ Yt_1.\ Vt_1.\ Z$;

$t_2 = t_2.\ XWt_2.\ YVt_2.\ Z = t_2.\ XWt_2.\ Yt_2.\ Vt_2.\ Z$

và vì $X \rightarrow\rightarrow Y$ nên $t_1' = t_1.XWt_2.Yt_1.Vt_1.Z$ và $t_2' = t_2.XWt_1.Yt_2.Vt_2.Z$ cùng thuộc r (chú ý vì $t_1.XW = t_2.XW$ nên $t_1.X = t_2.X$), mà $V \subset W$ nên $V \subset XW$ và do $t_1.XW = t_2.XW$ nên $t_1.V = t_2.\ V$.

Thay $t_1.V$ và $t_2.V$ vào t_1' và t_2' ta có $t_1' = t_1.XWt_2.Yt_2.Vt_1.Z = t_1.XWt_2.YVt_1Z$ và $t_2' = t_2.XWt_1.Yt_1.Vt_2.Z = t_2.XWt_1.YVt_2.Z$ cùng thuộc r. Vậy $XW \rightarrow\rightarrow YV$.

Tính chất d3:

Ta sẽ chứng minh tính chất 3 bằng phản chứng.

Giả sử kết luận của mệnh đề không thỏa mãn, nghĩa là tồn tại một quan hệ r mà trên đó $X \rightarrow\rightarrow Y$ và $Y \rightarrow\rightarrow V$ thỏa mãn nhưng: $X \rightarrow\rightarrow V \backslash Y$ không thỏa mãn. Nghĩa là trên r tồn tại hai bộ t_1, t_2 mà:

$$t_1.X = t_2.X \text{ và } t_1' = t_1.Xt_2.(V\backslash Y)t_1.\ (R\backslash X\backslash (V\backslash Y)) \notin r$$

Theo giả thiết $X \rightarrow\rightarrow Y$ và từ $t_1.X = t_2.X$ ta có $t_1'' = t_1.Xt_2.Yt_1.(R\backslash X\backslash Y) \in r$. Ta dễ dàng nhận thấy rằng: $t_1''.Y = t_2.Y$ và vì $Y \rightarrow\rightarrow V$ nên ta có thể thay chiếu của V trong t_2 và t_1'' để được các phần tử thuộc r.

Tức là $t_1''' = t_1.Xt_2.Yt_2.Vt_1.(R\backslash X\backslash Y\backslash V)) \in r$.

Ta sẽ chứng minh $t_1' = t_1'''$. Rõ ràng trên tập thuộc tính $X \cup (V\backslash Y)$ thì t_1' và t_1''' bằng nhau. Bây giờ ta chỉ ra rằng t_1' và t_1''' bằng nhau trên phần còn lại, tức trên $R\backslash X\backslash (V\backslash Y)$.

$$\text{Mà } t_1'''.(R\backslash X\backslash (V\backslash Y)) = t_2.Yt_2.(V\cap Y)t_1.(R\backslash X\backslash Y\backslash V)$$

$$= t_2.Yt_1.(R\backslash X\backslash Y\backslash V) = t_1'.(R\backslash (V\backslash Y)).$$

Tương tự ta có thể tiếp tục chứng minh các tính chất còn lại của phụ thuộc đa trị.

2.7.3. Hệ tiên đề của các phụ thuộc FD, MD

Gọi FD là lớp tất cả các phụ thuộc hàm khi đó ta đã có hệ tiên đề Armstrong cho lớp FD:

Ar1 Tính phản xạ: $X \to X$ và nếu $Y \subset X$ thì $X \to Y$.

Ar2 Tính bắc cầu: $X \to Y$ và $Y \to V$ thì $X \to V$.

Ar3 Tính mở rộng hai vế: $X \to Y$ thì $XZ \to YZ$.

Trong lớp MD, hệ năm tính chất d1, d2, d3, d4, d5 gọi là *hệ tiên đề của lớp MD.*

Tổng quát: Hệ tám tiên đề {Ar1, Ar2, Ar3, d1, d2, d3, d4, d5} gọi là hệ tiên đề của các ràng buộc dạng phụ thuộc hàm và phụ thuộc đa trị $FD \cup MD$.

Định lý 2.9

Hệ tiên đề {Ar1, Ar2, Ar3, d1, d2, d3, d4, d5} đúng đắn và đầy đủ trong lớp các ràng buộc $FD \cup MD$.

Từ nay khi nói cho tập ràng buộc Đ ta ngầm chỉ rằng trong Đ có chứa các ràng buộc kiểu phụ thuộc hàm và các phụ thuộc đa trị.

Ví dụ: Đ = $\{H \to B, G \to\to C, E \to F, G \to E\}$.

Tương tự bao đóng của Đ, ký hiệu Đ$^+$ là tập tất cả các ràng buộc được suy dẫn từ Đ. Ví dụ cho Đ = $\{G \to BCE\}$ thì Đ$^+$ = $\{G \to B, G \to C, G \to E, G \to BC, \ldots, G \to\to B, G \to\to C, \ldots\}$. Vậy chúng ta thấy rằng tập Đ có rất ít phần tử nhưng tập Đ$^+$ có thể rất lớn.

2.7.4. Dạng chuẩn 4: 4NF

Cho sơ đồ quan hệ W= < A, Đ >.

Định nghĩa 2.36 Dạng chuẩn 4NF

Ta nói *W là 4NF* nếu mọi MD $X \to\to Y \in$ Đ$^+$ là phụ thuộc cơ sở thì $X^+ = A$. Nói một cách khác W là 4NF nếu mọi phụ thuộc đa trị:

$X \to\to Y \in$ Đ$^+$ mà $Y \neq \varnothing$ và $X \cap Y = \varnothing$ và $X \cup Y \neq A$ thì $X^+ = A$.

Ví dụ 2.40:

W = < A, Đ >, với A = {H, B, C, E, G}, Đ = $\{H \to BCEG\}$ thì W là 4NF vì mọi phụ thuộc đa trị $X \to\to Y \in$ Đ$^+$ đều thỏa mãn định nghĩa của dạng chuẩn 4.

Từ định nghĩa ta có ngay kết luận sau đây:

Bổ đề 2.8

Nếu W là 4NF thì W là BCNF.

Thật vậy giả sử W = < A, Đ > không là BCNF, có nghĩa là trong W có PTH dạng $X \rightarrow B \notin X$ và $X^+ \neq A$, vậy trong Đ có $X \rightarrow\rightarrow B \in Đ^+$ là phụ thuộc cơ sở nhưng $X^+ \neq A$, suy ra vô lý. Vậy W là BCNF.

Muốn xét xem sơ đồ quan hệ W = < A, Đ > có là 4NF hay không ta chỉ cần xét xem trong $Đ^+$ có phụ thuộc cơ sở $X \rightarrow\rightarrow B \in Đ^+$ và $X^+ \neq A$?

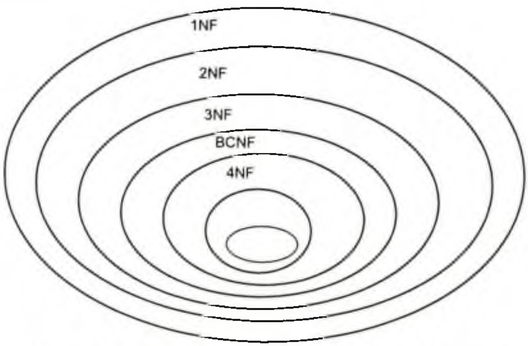
Nếu $Đ^+ = \varnothing$ thì W là 4NF.

Từ định nghĩa chúng ta thấy rằng điều kiện của các dạng chuẩn được thắt dần vào. Đầu tiên là dạng chuẩn1: 1NF là lớp chứa mọi sơ quan hệ. Dạng chuẩn 2 là các sơ đồ quan hệ chắc chắn phải là dạng chuẩn 1 và thêm điều kiện mọi thuộc tính thứ cấp phụ thuộc hoàn toàn vào khóa...

Vậy lớp các sơ đồ quan hệ ở các dạng chuẩn sau lồng trong các dạng chuẩn trước nó.

Định lý 2.10

Trong các lớp của các dạng chuẩn ta có mối quan hệ lồng nhau thực sự như sau: 1NF $\supset$ 2NF $\supset$ 3NF $\supset$ BCNF $\supset$ 4NF $\supset$ ... (lồng nhau thực sự, nghĩa là lớp trước chứa lớp sau thực sự).



Hình 2.12. Sơ đồ biểu thị mối liên hệ của các lớp chuẩn

2.8. RÀNG BUỘC TOÀN VẸN

2.8.1. Định nghĩa Ràng Buộc Toàn Vẹn (RBTV)

Ràng Buộc Toàn Vẹn (RBTV) là gì? Ràng buộc toàn vẹn của một cơ sở dữ liệu là ràng buộc có tính *nhất quán về các giá trị của các thuộc tính* trong CSDL. Ví dụ năm đi làm của một nhân viên không thể bé hơn năm sinh của nhân viên đó. Hoặc trong phần HOTEN của bất kỳ nhân viên nào cũng không thể để trống (HOTEN not null), hay số thí sinh thi trong một phòng thi không thể vượt quá sức chứa của phòng thi đó, các giá trị của thuộc tính khóa chính phải từng đôi một khác nhau, hay giá trị của HOTEN phải thuộc kiểu xâu ký tự, trị của DIEM phải thuộc kiểu số ...

Định nghĩa 2.13 Định nghĩa RBTV

RBTV của một CSDL là ràng buộc về giá trị của các thuộc tính trong CSDL.

Ví dụ miền giá trị của thuộc tính ngày (Day) chỉ có thể từ 1 đến 31, hay thuộc tính GT (giới tính) chỉ có giá trị là 'nam' hoặc 'nữ'...

Lưu ý:

- RBTV là loại ràng buộc nội tại của CSDL, ràng buộc hiển nhiên, tồn tại khách quan. Năm sinh của mỗi nhân viên phải bé hơn năm đi làm của nhân viên ấy, đó là ràng buộc bắt buộc và khách quan.

- Khi thiết kế và thao tác trên CSDL chúng ta phải đảm bảo và tuân thủ các ràng buộc đó. Trong các ngôn ngữ lập trình có một số ràng buộc phải mô tả. Ví dụ HOTEN NOT Null hay khi tạo bảng GIAOVIEN trong SQL có mã số giáo viên (MaGV) là khóa chính, mã học hàm (MaHH) là khóa ngoại, ta mô tả RBTV khóa chính, khóa ngoại như sau:

```
Create table GIAOVIEN
(
MaGV smallint Not null,
MaHH smallint Not null,
Constraint pk_GIAOVIEN primary key (MaGV),
Constraint fk_HOCHAM_GIAOVIEN foreign key (MAHH)
                                        references HOCHAM(MaHH)
)
```

2.8.2. Các loại RBTV

Có nhiều loại RBTV về giá trị, tuy nhiên ta chia chúng thành ba nhóm chính:

(a) RBTV loại 1: RBTV kiểu khóa chính, khóa ngoại:

Trong phần trước ta đã biết khóa chính của quan hệ R là một khóa của R được chọn làm khóa chính. Khóa ngoại của R là trường dữ liệu đóng vai trò khóa chính trong quan hệ S khác. Ví dụ ta có hai quan hệ sinh viên và bảng điểm của sinh viên: SV(SoTT, <u>Masv</u>, Hoten, NS, Que, GT) gồm các thuộc tính số thứ tự, mã sinh viên, họ tên, ngày sinh, quê, giới tính. Quan hệ này có Masv là khóa chính.

Bang_Diem(Masv, diem) gồm hai thuộc tính là mã sinh viên và điểm của sinh viên. Quan hệ Bang_Diem có trường Masv là khóa ngoại vì nó là khóa chính trong SV.

Định nghĩa RBTV loại 1

RBTV loại 1 là kiểu ràng buộc khóa chính, khóa ngoại: mỗi giá trị khóa ngoại phải là một giá trị khóa chính.

Nói cách khác *tập giá trị trong khóa chính chứa tập giá trị trong khóa ngoại.* Ví dụ giá trị mã sinh viên Masv trong bảng điểm phải là một giá trị của Masv trong bảng sinh viên. Hay nói cách khác SV[Masv] $\supseteq$ Bang_Diem[Masv].

(b) RBTV loại 2: RBTV kiểu phụ thuộc hàm:

Định nghĩa RBTV loại 2

RBTV loại 2 là kiểu ràng buộc dạng phụ thuộc hàm: ràng buộc về giá trị của hai tập thuộc tính X và Y, mà X $\rightarrow$ Y.

Ví dụ xét quan hệ R có hai thuộc tính mã vùng và tên vùng R (mã vùng, tên vùng) khi đó ta có RBTV kiểu phụ thuộc hàm: mã vùng $\rightarrow$ tên vùng hay biết tên vùng suy ra mã vùng, tên vùng $\rightarrow$ mã vùng.

(c) *RBTV loại 3:* RBTV loại 3 là kiểu ràng buộc không phải khóa chính, khóa ngoại và không phải kiểu phụ thuộc hàm:

Định nghĩa RBTV loại 3

RBTV loại 3 là kiểu ràng buộc không phải loại 1, không phải loại 2.

Ví dụ giá trị của năm sinh (NS) phải nhỏ thua giá trị năm đi làm (NDL) của nhân viên.

Số tiền các khách hàng đã mua không được vượt quá số tiền (doanh số) mà khách hàng có.

Các trường dữ liệu như họ tên, mã nhân viên của các bảng nhân viên không được để trống... là những RBTV loại 3.

2.8.3. Phương pháp trình bày một RBTV

Khi thiết kế ứng dụng một cơ sở dữ liệu, người thiết kế phải nắm chắc các RBTV có trong CSDL. Để trình bày một RBTV trong *ngôn ngữ lập trình* phải mô tả theo cách viết của ngôn ngữ đó, trong *ngôn ngữ quan hệ* chúng ta phải trình bày các công thức, ràng buộc giá trị bằng các bộ của các quan hệ.

Muốn trình bày được bằng ngôn ngữ quan hệ chúng ta phải nắm chắc cấu trúc của CSDL và các bảng trong CSDL.

a. Phương pháp trình bày một RBTV bằng ngôn ngữ lập trình

Trong phần này ta sẽ minh họa một ngôn ngữ khai báo để đặc tả một RBTV. Ngôn ngữ này giống với SAL (ANSI, 1986) nhưng tổng quát hơn. Ngôn ngữ này cho phép chúng ta mô tả, ghi nhận hoặc tháo bỏ một RBTV. Những RBTV này được định nghĩa, mô tả vào lúc tạo quan hệ hoặc vào một thời điểm nào đó, ngay cả khi quan hệ đã chứa dữ liệu.

Xét CSDL quản lý bán hàng trong ví dụ 2.17.

kh(Makh, Tenkh, DS...)

nv(Manv, Tenv, HS, Luong...)

sp(Masp, Tensp, DVT, NSX, GIA)

hd(SOHD, NHD, Makh, Manv, Trigia)

cthd(SOHD, Masp, SL)

Các ví dụ sau minh họa một số mô tả RBTV

- *RBTV loại 1: Khóa duy nhất*

Cặp (SOHD, Masp) là khóa duy nhất của cthd

(SOHD, Masp) UNIQUE IN cthd

Khóa ngoại

Masp trong cthd là khóa ngoại của Masp trong sp

Masp in cthd REFERENCES Masp IN sp

- *RBTV loại 2: Phụ thuộc hàm*

Manv xác định phụ thuộc hàm Tenv

Manv IN nv DETERMINES Tenv

Phụ thuộc hàm

Hệ số lương của nhân viên (HS) xác định phụ thuộc hàm Luong

HS IN nv DETERMINES Luong

- *RBTV loại 3: Thuộc tính không nhận giá trị null*

Mã nhân viên trong quan hệ nv không nhận giá trị null

Manv not null in nv

Thuộc tính nhận giá trị trong khoảng

Doanh số của mỗi khách hàng nằm trong khoảng 5000 đến 10000

Check on kh (DS >= 5000 and DS <= 10000)

Ràng buộc miền khi xóa

Chỉ với các bộ của kh có DS = 0 mới được xóa

Check on kh when delete (DS = 0)

Ràng buộc giá trị các thuộc tính trong các quan hệ

Tổng tiền (tổng Trigia) của các hóa đơn < 100000

Check on hd (sum(Trigia) < 100000)

b. Phương pháp trình bày một RBTV bằng ngôn ngữ quan hệ

Người thiết kế CSDL cũng như người triển khai ứng dụng phải hiểu rõ bản chất của các bảng trong CSDL cũng như các RBTV của CSDL. Một người không hiểu bản chất cũng như các RBTV của CSDL không thể trình bày, diễn đạt được ràng buộc đó.

Chúng ta thường trình bày RBTV trong *ngôn ngữ quan hệ* bằng ba bước:

Bước1: Nội dung của RBTV. Từ một câu bằng ngôn ngữ tự nhiên ta phải chuyển sang *ngôn ngữ quan hệ.*

Bước 2: Liệt kê các quan hệ tham gia vào RBTV đang xét. Bước này chỉ liệt kê các quan hệ liên đới đến RBTV đang xét.

Bước 3: Lập bảng tầm ảnh hưởng. Đây là bảng có 4 cột. Bảng này cảnh báo, thống kê các ảnh hưởng của các thao tác: *thêm, sửa, xóa* dữ liệu trong các quan hệ tham gia ràng buộc có ảnh hưởng đến RBTV:

Bảng tầm ảnh hưởng

Các quan hệ	Thêm	Sửa	Xóa
R	-		
S		+	+(C)

Trong đó:

R, S là các quan hệ tham gia RBTV.

Dấu - biểu thị thao tác không ảnh hưởng đến RBTV.

Dấu + biểu thị thao tác có ảnh hưởng đến RBTV.

Dấu +(C) biểu thị thao tác ảnh hưởng đến RBTV trên thuộc tính C

Lưu ý: Bảng tầm ảnh hưởng chỉ xét với ràng buộc có trong nội dung ở bước 1. Một CSDL có nhiều RBTV.

Ví du 2.41:

Xét CSDL quản lý bán hàng trong ví dụ 2.17.

kh(<u>Makh</u>, Tenkh, DS, NS, NGN) với NS là năm sinh, NGN là ngày gia nhập

nv(<u>Manv</u>, Tenv, HS, LUONG) với HS là hệ số lương, LUONG là lương

sp(<u>Masp</u>, Tensp, DVT, NSX, GIA)

hd(<u>SOHD</u>, NHD, Makh, Manv, Trigia)

cthd(<u>SOHD</u>, <u>Masp</u>, SL)

Thực hiện (ba bước) các RBTV sau:

RBTV loại 1: Nêu các RBTV khóa chính khóa ngoại của CSDL.

RBTV loại 2: Các nhân viên có cùng hệ số lương HS thì có cùng LUONG.

RBTV loại 3:

- DVT của sản phẩm chỉ có thể là chai, lon, két, thùng.
- Ngày gia nhập (NGN) của khách hàng phải lớn hơn năm sinh (NS) của khách hàng đó.
- Ngày mua hàng của khách hàng phải lớn hơn NGN của khách hàng đó
- Mỗi hóa đơn phải có ít nhất một chi tiết hóa đơn.
- Trigia của một hóa đơn phải bằng tổng tiền (tổng các SL * GIA) của hóa đơn đó.
- DS của một khách hàng phải lớn hơn hoặc bằng tổng các Trigia hóa đơn khách hàng đã mua.

Lưu ý:

- Trong cơ sở dữ liệu này các khóa chính được gạch dưới, khóa ngoại là các khóa chính đó xuất hiện trong các quan hệ khác.
- Ta cần lưu ý rằng trong CSDL trên có các khóa chính và khóa ngoại tương ứng: Makh là khóa chính trong quan hệ kh và nó là khóa ngoại trong quan hệ hd, Manv là khóa chính trong bảng nv và khóa ngoại trong hd, Masp là khóa chính trong sp và nó là khóa ngoại trong cthd, SOHD là khóa chính trong bảng hd và khóa ngoại trong cthd. Để trình bày các ràng buộc toàn vẹn khóa chính khóa ngoại trên ta chỉ cần thực hiện cho một cặp khóa chính, khóa ngoại, các trường hợp khác thực hiện tương tự. Ví dụ thực hiện cho trường hợp Makh

Trình bày các RBTV loại 1

Bước 1: Nội dung RBTV

$\forall t \in hd\ \exists t' \in kh$ sao cho: $t[Makh] = t'[Makh]$

(hoặc làm cách khác $hd[Makh] \subseteq kh[Makh]$)

Bước 2: Các quan hệ tham gia RBTV: hd, kh

Bước 3: Bảng tầm ảnh hưởng của RBTV

Các quan hệ	Thêm	Sửa	Xóa
kh	+(Makh)	+(Makh)	+(Makh)
hd	+(Makh)	+(Makh)	-

Trình bày các RBTV loại 2

Bước 1: Nội dung RBTV

$\forall$t, t' $\in$ nv nếu (t[HS] = t'[HS]) thì (t[LUONG] = t'[LUONG]).

Bước 2: Các quan hệ tham gia RBTV: nv

Bước 3: Bảng tầm ảnh hưởng

Các quan hệ	Thêm	Sửa	Xóa
nv	+(HS, LUONG)	+(HS, LUONG)	-

Trình bày các RBTV loại 3

a. *Bước 1:* Nội dung RBTV

$\forall$t$\in$sp thì (t[DVT] = 'chai') or (t[DVT] = 'lon') or (t[DVT] = 'két') or (t[DVT] = 'thùng').

Bước 2: Các quan hệ tham gia RBTV: sp

Bước 3: Bảng tầm ảnh hưởng của RBTV

Các quan hệ	Thêm	Sửa	Xóa
sp	+(DVT)	+(DVT)	-

b. *Bước 1:* Nội dung RBTV

$\forall$t $\in$kh thì (t[NS] $\leq$ t[NGN])

Bước 2: Các quan hệ tham gia RBTV: kh

Bước 3: Bảng tầm ảnh hưởng của RBTV

Các quan hệ	Thêm	Sửa	Xóa
kh	+(NS,NGN)	+(NS, NGN)	-

c. *Bước 1:* Nội dung RBTV

$\forall$t $\in$ kh & t' $\in$ hd nếu t'[Makh] = t[Makh] thì t[NGN] < t'[NHD]

Bước 2: Các quan hệ tham gia RBTV: Kh, hd

Bước 3: Bảng tầm ảnh hưởng của RBTV

Các quan hệ	Thêm	Sửa	Xóa
kh	-	+(Makh, NGN)	-
hd	+(Makh, NHD)	+(Makh, NHD)	-

d. *Bước 1:* Nội dung RBTV

$\forall t \in hd\ \exists\ t' \in cthd$ sao cho $t'[SOHD] = t[SOHD]$

Bước 2: Các quan hệ tham gia RBTV: hd, cthd

Bước 3: Bảng tầm ảnh hưởng RBTV

Các quan hệ	Thêm	Sửa	Xóa
hd	-	+(SOHD)	+(SOHD)
cthd	+(SOHD)	+(SOHD)	+(SOHD)

e. *Bước 1:* Nội dung RBTV

$\forall t \in hd\ \exists\ t_1, t_2,...,t_m \in cthd$ và $t_1', t_2',...,t_m' \in sp$ sao cho:

$(t[SOHD] = t_i[SOHD])$ and $(t_i[[Masp] = t_i'[Masp])$; i= 1, 2,...,m

và $\sum_{i=1}^{m} t_i[SL] \times t_i'[GIA] \leq t[TRIGIA]$

Bước 2: Các quan hệ tham gia RBTV: hd, cthd

Bước 3: Bảng tầm ảnh hưởng của RBTV

Các quan hệ	Thêm	Sửa	Xóa
hd	-	+(Trigia)	-
cthd	+(SL)	+(SL)	-

f. *Bước 1:* Nội dung RBTV

$\forall t \in kh\ \exists\ t_1, t_2,...,t_m \in hd$ sao cho:

$(t[Makh] = t_i[Makh])$; i= 1, 2,...,m

và $\sum_{i=1}^{m} t_i[TRIGIA] \leq t[DS]$

Bước 2: Các quan hệ tham gia RBTV: kh, hd

Bước 3: Bảng tầm ảnh hưởng

Các quan hệ	Thêm	Sửa	Xóa
kh	-	+(DS, Makh)	+(DS, Makh)
hd	+(Makh, Trigia)	+(Makh, Trigia)	-

Lưu ý:

Bảng tầm ảnh hưởng chỉ xét xem các thao tác thêm, sửa, xóa có ảnh hưởng đến RBTV trong nội dung ở bước 1 không?(và chỉ RBTV đó).

Ví dụ trong câu (f) khi *thêm* một hóa đơn ta chỉ cần chú ý đến Makh, tuy nhiên ta cũng cần chú ý đến SOHD vì mỗi hóa đơn phải có một SOHD duy nhất nhưng RBTV đó không thuộc yêu cầu của câu (f). Tương tự khi *xóa* một khách hàng chúng ta còn phải chú ý các ràng buộc khác vì khách hàng này đã lập và mua nhiều hóa đơn. Tuy nhiên trong câu (f) không yêu cầu về những ràng buộc đó.

2.9. NGÔN NGỮ VẤN TIN SQL

2.9.1. Ngôn ngữ vấn tin

Trong phần trước, chúng ta đã làm quen với các phép toán đại số trên các quan hệ. Trong phần này chúng ta sẽ trình bày tiếp một trong các ngôn ngữ thao tác (Data Manipulation Language - DML) trên các CSDL quan hệ, hay còn gọi là ngôn ngữ vấn tin (query language) để diễn tả các câu lệnh kiểu như:

1. Truy xuất từ CSDL số lượng ghế còn trống trên một chuyến bay.
2. Giảm một số chỗ trống trên chuyến bay 747 vào ngày 03 tháng 9.
3. Tìm tất cả các chuyến bay từ Hà Nội đi Sài Gòn ngày 20 tháng 9.
4. Nhập 100 chỗ từ Hà Nội đi Sài Gòn vào chuyến bay 456 ngày 20 tháng 9.

Với CSDL cài trong máy có quan hệ chuyến bay R (number, data, seats, from, to)

Ngôn ngữ vấn tin là ngôn ngữ dùng để truy xuất dữ liệu

Ví dụ trong ngôn ngữ SQL để diễn tả câu lệnh số 2 ở trên, ta có thể viết:

```
Update R
Set seats = seats - 4
Where number = 747 And Date = '3/9';
```

Hoặc để thực hiện mục 4 ta có thể viết:

```
Insert into R
Values (456, '20/9', 100, 'Hanoi', 'Saigon');
```

Sau đây chúng ta sẽ xét một "đại diện" của các ngôn ngữ vấn tin, đó là ngôn ngữ SQL, một ngôn ngữ được sử dụng khá phổ biến trong các ngôn ngữ lập trình trong các CSDL phân tán trên mạng máy tính.

2.9.2. Câu lệnh tạo quan hệ r

Giả sử ta cần tạo quan hệ r(A_1, A_2,..., A_n) với các thuộc tính A_i có kiểu dữ liệu là K_i. Khi đó ta viết:

CREATE TABLE r(A_1 k_1, A_2 k_2,..., A_n k_n);

Ví dụ 2.42:

Ta cần tạo quan hệ r trên A, với A = {Masv, Hoten, NS, Que} ta viết:

CREATE TABLE r (Masv Char, Hoten char, NS Date, Que char);

2.9.3. Câu lệnh thêm (nhập) dữ liệu vào quan hệ r

Giả sử ta cần nhập bộ dữ liệu t = (v_1, v_2,...,v_n) vào quan hệ r ta viết:

Insert into r

VALUES (v_1, v_2,...,v_n);

Ví dụ 2.43:

Nhập sinh viên có mã số 1000, họ tên là Nguyễn Thị Hoa sinh năm 1986, quê Hà Nội vào bảng r trên ta viết:

Insert into r

Values (1000, 'Nguyễn Thị Hoa', 1986, 'Hà Nội');

2.9.4. Câu lệnh sửa một số bộ của r

Giả sử ta cần sửa lại giá trị các thuộc tính A_1, A_2,..., A_l của các bộ thỏa mãn điều kiện E trong quan hệ r ta viết:

Update r

Set A_1 = v_1, A_2 = v_2,..., A_l = v_l

Where E;

Ví dụ 2.44:

Cộng hai điểm cho những sinh viên thuộc khu vực 4

UPDATE Bangdiem

SET DIEM = DIEM + 2

WHERE KV = 4;

2.9.5. Câu lệnh xóa một số bộ của R thỏa mãn điều kiện E

Giả sử chúng ta cần xóa các bộ trong r thỏa mãn điều kiện E ta viết:

Delete from r

Where E;

Ví dụ 2.45:

Xóa khỏi danh sách lớp các sinh viên có điểm bằng không

Delete from DSLOP

Where DIEM = 0;

2.9.6. Câu lệnh xóa một quan hệ r

Giả sử chúng ta cần xóa quan hệ r và xóa cả lược đồ quan hệ ta viết:

Drop table r;

Ví dụ 2.46:

Ta cần xóa bảng điểm không lưu lại dấu vết

DROP TABLE BANGDIEM;

2.9.7. Câu lệnh Select

Dạng thông dụng nhất của câu vấn tin trong ngôn ngữ SQL là câu lệnh SELECT. Các dạng thường gặp của *SELECT:*

2.9.7.1. Cú pháp của câu lệnh SELECT đơn giản

SELECT r. A,..., s. B

FROM r,..., s

WHERE E;

Trong đó:

r,..., s là các quan hệ.

r. A,..., s. B là chiếu của r,..., s lên A,..., B; E là một biểu thức logic

Thứ tự thực hiện của câu lệnh (đơn giản) là nối các quan hệ có mặt sau from, chọn trong quan hệ kết quả những bộ thỏa mãn điều kiện E, sau đó chiếu lên các thuộc tính A,..., B. Vậy kết quả của câu lệnh Select dạng ngắn là nối - chọn - chiếu.

Ví dụ 2.47:

Xét câu vấn tin "liệt kê tên những khách hàng có số dư âm". Khi đó trong SQL ta có:

SELECT NAME

FROM customers

WHERE BALANCE < 0;

2.9.7.2. Nếu sau from chỉ có một quan hệ r

Nếu sau from chỉ có quan hệ r thì thay cho r.A ta viết A

Ví dụ 2.48:

```
SELECT NAME
FROM customers
Where BALANCE < 0;
```

Trong ví dụ này chỉ có một quan hệ khách hàng customers sau mệnh đề FROM nên không có sự nhầm lẫn nào liên quan đến các thuộc tính.

Nếu muốn lấy mọi thông tin về những khách hàng có số dư âm, chúng ta có thể viết:

```
SELECT NAME, ADDR, BALANCE
FROM customers
WHERE BALANCE < 0;
```

2.9.7.3. Nếu sau from chỉ có một quan hệ r và cần lọc hết tất cả các thuộc tính

Nếu sau from chỉ có quan hệ r và ta cần lọc hết các thuộc tính trong U, khi đó thay cho liệt kê các thuộc tính ta viết dấu *.

Ví dụ 2.49:

```
SELECT *
FROM customers
WHERE BALANCE < 0;\
```

Kết quả của câu lệnh cho bảng tất cả các thuộc tính trong quan hệ customers.

Ví dụ 2.50:

Cho cơ sở dữ liệu quản lý bán hàng:

1. Quan hệ khách hàng

   kh(<u>Makh</u>, Tenkh, DS)

2. Quan hệ nhân viên

   nv(<u>Manv</u>, Ten)

3. Quan hệ sản phẩm

   sp(<u>Masp</u>, Tensp, DVT, NSX, GIA)

4. Quan hệ hóa đơn

   hd(<u>SOHD</u>, NHD, Makh, Manv, Trigia)

5. Quan hệ chi tiết hóa đơn

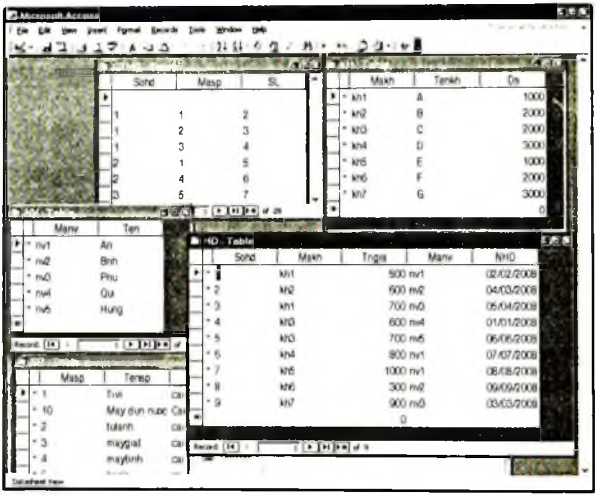
cthd(<u>SOHD, Masp</u>, SL)

Hãy lập bảng Makh, Tenkh, DS có doanh số DS > 100.

Select *

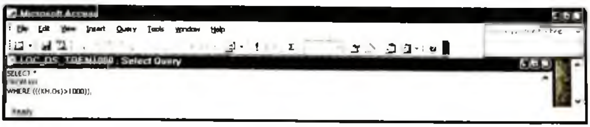
From kh

Where DS > 100;



Hình 2.13. Minh họa kết quả phép lọc

Thực hiện câu lệnh:



Hình 2.14. Câu lệnh lọc

Kết quả câu lệnh có trong hình 2.15.

Hình 2.15. Minh họa kết quả

2.9.7.4. Nếu câu lệnh lọc ra có các bộ trùng nhau dùng DISTINCT

Sau câu lệnh Select đôi khi ta lọc được các bộ lặp nhiều lần, ví dụ lọc những MASP, TENSP mà khách hàng Nguyễn Văn A đã mua, khi đó vì khách hàng A đã mua nhiều hóa đơn nên có thể có những sản phẩm được mua nhiều lần. Để loại trừ các trùng lặp kiểu như vậy ta dùng từ khóa DISTINCT theo sau SELECT, câu lệnh vấn tin khi đó có dạng:

SELECT DISTINCT r. A...., s. B

FROM r,..., s

WHERE E;

Ví dụ 2.51:

Lọc mã sản phẩm, tên sản phẩm mà ông Nguyễn Văn A đã mua

SELECT DISTINCT sp.MASP, sp.TENSP

FROM sp, kh, hd, cthd

WHERE (kh.HOTENKH = 'Nguyễn Văn A') and (kh.MAKH = hd.MAKH)
and (hd. SOHD = cthd.SOHD) and (sp.MASP = cthd.MASP);

2.9.7.5. Biến bộ

Đôi khi chúng ta cần liên hệ đến hai bộ hoặc nhiều hơn hai bộ trong cùng một quan hệ. Để thực hiện việc này chúng ta định nghĩa các biến bộ cho quan hệ sau mệnh đề FROM và dùng biến bộ này làm "bí danh" thay cho quan hệ:

Ví dụ 2.52:

Xét câu "in ra các khách hàng có doanh số nhỏ hơn doanh số của khách hàng Võ Văn Kiệt khi đó câu lệnh sẽ là:

SELECT*

FROM kh

WHERE DS < 50000; (nếu doanh số của Võ Văn Kiệt là 50000).

Tới đây chúng ta cần phải chú ý rằng nếu ta không biết cụ thể doanh số của ông Kiệt thì câu lệnh phải được viết như thế nào? nghĩa là tham chiếu đến doanh số của Kiệt bằng cách nào? Trong trường hợp này ta dùng " biến bộ".

Khi đó câu lệnh có dạng:

SELECT c_1*

FROM kh c_1, kh c_2

WHERE c_1. DS < c_2.DS

AND c_2. Tenkh = 'Võ Văn Kiệt';

Ở đây chúng ta cần nhớ thêm một cách viết của SQL là: các tên đi sau các tên quan hệ là các bí danh của các quan hệ đó, ví dụ kh c_1, kh c_2 thì cả c_1 và c_2 đều là bí danh của quan hệ kh, chúng trở thành các bộ chạy trên quan hệ kh.

2.9.7.6. Mở rộng khả năng dùng where bằng từ khóa like (là)

Nhằm mở rộng thêm khả năng dùng where E ta dùng từ khóa like và hai ký hiệu phụ: %, _

- Từ like ở đây dùng theo nghĩa "là".
- Dấu % được dùng để đại diện cho *các ký tự.*
- Dấu _ dùng để đại diện cho *một ký tự.*

Ví dụ 2.53:

Xét câu vấn tin "in ra các Masp do các khách hàng có họ tên bắt đầu bằng ký tự B đã mua", khi đó ta có câu lệnh:

SELECT cthd.Masp

FROM cthd, hd, kh

WHERE (kh. Tenkh like 'B%') and (kh.Makh = hd.Makh) and

(hd.SOHD = cthd.SOHD);

Ví dụ 2.54:

Xét câu vấn tin "in ra những số hóa đơn nằm trong khoảng 1000 - 1999", khi đó ta có câu lệnh:

SELECT *

FROM hd

WHERE SOHD like '1_ _ _ ';

Tất nhiên ba dấu _ _ _ đại diện cho ba dấu bất kỳ từ 0 đến 9.

SELECT với các hàm gộp

Các hàm gộp cơ bản trong SQL gồm:

SUM(B) tổng các giá trị trong cột B

MIN(B) giá trị min của cột B

MAX(B) giá trị max của cột B

COUNT(B) số lượng các giá trị trong cột B

COUNT(*) số các bản ghi của quan hệ hiện hành

AVG(B) trị trung bình của cột B

Tất cả các hàm gộp đều *phải đi sau Select*

Ví dụ 2.55:

Tính tổng doanh số của các khách hàng nhớ vào TONG

```
SELECT SUM(DS) as TONG
FROM kh;
```

Hoặc tính doanh số trung bình của các khách hàng gia nhập trong quý 1 năm 2008

```
Select AVG(DS) as TB
FROM kh
WHERE NGN BETWEEN 01/01/2008 AND 31/3/2008;
```

2.9.7.7. Gộp theo nhóm

Trong thực tế nhiều khi chúng ta cần gộp theo nhóm, ví dụ tính điểm trung bình theo quê, tính điểm trung bình theo khu vực hoặc tính tổng điểm theo mã sinh viên. Trong SQL cho phép chúng ta thực hiện các phép gộp theo nhóm bằng cách chia các bộ của quan hệ thành các nhóm và thực hiện phép gộp theo các nhóm riêng biệt. Chúng ta sẽ thực hiện các phép gộp qua từ khóa group by (gộp bởi nhóm) và danh sách các thuộc tính cần gộp.

Nghĩa là mệnh đề: GROUP BY $A_1, A_2, \dots, A_k$ chia quan hệ thành các nhóm sao cho hai bộ sẽ cùng một nhóm nếu và chỉ nếu chúng giống nhau ở các thuộc tính $A_1, \dots, A_k$.

Để kết quả của câu vấn tin như thế có nghĩa, tất nhiên các thuộc tính $A_1, \dots, A_k$ phải xuất hiện trong câu lệnh SELECT.

Ví dụ 2.56:

Trong CSDL quản lý điểm sinh viên ta có bảng kết quả gồm mã sinh viên, mã môn học, điểm, lớp, khoa: KQ (Masv, Mamh, Diem, Lop, Khoa).

Hãy thực hiện các yêu cầu sau:

a. Tính điểm trung bình các môn thi của từng sinh viên

b. Tính điểm trung bình theo lớp môn M1

c. Tính điểm trung bình theo khoa môn M2 và chỉ tính cho những khoa có trên 10 sinh viên thi môn M2.

Để thực hiện yêu cầu câu a ta lưu ý trong bảng kết quả có bao nhiêu mã môn học thì một mã sinh viên xuất hiện bấy nhiêu lần. Như vậy điểm trung bình của mỗi sinh viên chính là tổng điểm của sinh viên đó chia cho số lần mã sinh viên xuất hiện.

a.

```
SELECT Masv, AVG(Diem) as TBdiemsv
FROM KQ
GROUP BY Masv;
```

b.

```
SELECT Lop, AVG(Diem) as TBdiemlop
FROM KQ
WHERE Mamh = M1
GROUP BY Lop;
SELECT Khoa, AVG(Diem) as TBkhoa
FROM KQ
WHERE Mamh = M2
GROUP BY Khoa HAVING COUNT(Masv) > 10;
```

Lưu ý: Kết quả của các câu lệnh trên là những bảng hai cột

Ví dụ kết quả của câu c cho ta bảng hai cột là khoa và điểm trung bình môn M2 theo khoa: Khoa, TBkhoa với những khoa có số sinh viên trên 10. Số sinh viên của khoa trên 10 là điều kiện theo nhóm. Trong SQL điều kiện của nhóm đi sau Having.

Vậy gộp theo nhóm có dạng là:

Select

From

Where E

Group by A_1, A_2,..., A_k Having E_1;

Trong đó E là điều kiện toàn cục, E_1 là điều kiện của nhóm

Câu lệnh Select có sắp theo thứ tự tăng hoặc giảm của cột B

Để sắp theo thứ tự tăng (giảm) dần của cột B ta dùng từ khóa ORDER By

Khi ta viết Order by B thì mặc định sắp tăng dần cột B, còn nếu viết Order by B DESC thì cột B được sắp giảm dần.

Vậy một câu lệnh Select đầy đủ của SQL có dạng:

Select...

From...

Where E

Group by A_1, A_2,..., A_k Having E_1

Order by B DESC;

Hoặc

Select...

From...

Where E

Group by A_1, A_2,..., A_k Having E_1

Order by B;

Order by B là mặc định sắp xếp theo thứ tự tăng dần.

2.9.7.8. Select lồng nhau

Trong thực tế nhiều khi ta cần phải dùng các Select lồng nhau để tính một số dữ liệu nhờ vào các dữ liệu khác. Ví dụ cần lọc ra các sinh viên có điểm môn M1 bằng điểm môn M1 của sinh viên sv1000 hoặc in ra các khách hàng có doanh số nhỏ hơn doanh số của khách hàng Võ Văn Kiệt,... Để thực hiện các yêu cầu như vậy ta dùng các Select lồng nhau. Trong SQL ta có thể dùng đến 14 lần lồng nhau của Select, tuy nhiên không mất tính tổng quát ta chỉ xét hai Select lồng nhau. Hai Select lồng nhau có dạng:

Select... From... Where B θ (Select... From... Where...);

Trong đó Select trong dấu ngoặc () là Select trong, Select nằm ngoài dấu ngoặc là Select ngoài. B là một thuộc tính. θ(teta) có thể là một phép toán so sánh hoặc phép toán tập hợp. Nói cách khác θ có thể ∈ {=, <, >, ≤, ≥, ≠} hoặc thuộc {in, not in, all, any}. Dựa vào kết quả của Select trong mà θ thuộc nhóm so sánh hay nhóm các phép toán tập hợp.

- Nếu Select trong lọc ra 1 giá trị thì θ ∈ {=, <, >, ≤, ≥, ≠}

- Nếu Select trong lọc ra nhiều hơn 1 giá trị thì θ ∈ {IN, NOT IN, ANY, ALL}

Ví dụ 2.57:

Lọc ra các sinh viên có điểm môn M1 bằng điểm môn M1 của sinh viên sv1000. Select trong cần tìm ra điểm môn M1 của bạn sinh viên sv1000. Select ngoài lọc ra các sinh viên có điểm bằng điểm vừa lọc được ở Select trong.

```
Select*
From KQ
Where (Mamh = M1) and Diem = (Select Diem from KQ
        Where (Mamh = M1) and (Masv = sv1000));
```

Hoặc lọc ra các sinh viên có điểm môn M1 bằng điểm môn M1 của sinh viên sv1000 hoặc sinh viên sv2000 hoặc sinh viên sv3000. Trong trường hợp này Select trong lọc ra 3 điểm môn M1 của các sinh viên sv1000, sv2000, sv3000.

```
Select*
From KQ
Where (Mamh = M1) and Diem IN (Select Diem from KQ
        Where (Mamh = M1) and ((Masv = sv1000)
        OR (Masv = sv2000) OR (masv = sv3000)));
```

2.9.8. Tính đầy đủ của SQL

Sau đây chúng ta sẽ xét tính đầy đủ của SQL, tức là xét xem khả năng của SQL có thực hiện được tất cả các biểu thức của các phép toán đại số quan hệ. Để thực hiện một biểu thức ta thực hiện theo thứ tự từ các biểu thức con dần ra đến biểu thức toàn bộ. Vậy để xét xem SQL có thực hiện được các biểu thức của đại số quan hệ ta chỉ cần xét lần lượt cho các phép toán cơ bản của đại số quan hệ.

Trong chương 2, mục 2.3 chúng ta đã nói về mối liên hệ giữa các phép toán quan hệ. Bây giờ chúng ta phát biểu lại định lý 2.1 đó như sau:

Định lý 2.1

Năm phép toán quan hệ cơ bản: {hợp, hiệu, tích Đề-các, chọn, chiếu} của đại số quan hệ độc lập với nhau, nghĩa là không một phép nào trong chúng được biểu thị qua các phép còn lại.

Các phép toán khác của đại số quan hệ: {nối tự nhiên, giao, nối nửa, nối theo teta, chia} có thể nhận được từ các phép cơ bản trên.

Vậy để SQL thực hiện được các phép toán đại số quan hệ ta chỉ cần cài đặt cho SQL 5 phép toán cơ bản là {hợp, hiệu, tích Đề-các, chọn, chiếu}.

• *Phép hợp:* Giả sử ta có hai quan hệ r và s có cùng lược đồ $A = \{A_1, \ldots, A_n\}$. Khi đó để tính T = r + s ta viết:

```
INSERT INTO T
SELECT *
FROM r;
```

Tiếp theo là

```
INSERT INTO T
SELECT *
FROM s;
```

• *Phép trừ:* Để tính T = r - s trước tiên chúng ta chèn r vào T như trên. Sau đó dùng câu lệnh xóa như sau:

```
DELETE FROM T
WHERE (A1,..., An) IN
(SELECT * FROM s);
```

• *Tích Đề-các:* Cho r là quan hệ trên A, s là quan hệ trên A'. Điều kiện $A \cap A' = \phi$ và $T = r \times s$, ta có câu lệnh:

```
INSERT INTO T
SELECT r.A1,..., r.An, s.B1,..., s.Bm
FROM r, s;
```

Với $A = \{A_1, A_2, \ldots, A_n\}$ và $A' = \{B_1, B_2, \ldots, B_m\}$.

• *Phép chọn:* T = r(E).

```
INSERT INTO T
```

SELECT *

FROM r

WHERE E;

• *Phép chiếu:* Giả sử X là tập con của R và X = $\{A_1,... A_k\}$

T = r. X, ta có câu lệnh:

INSERT INTO T

SELECT $A_1,..., A_k$

FROM r;

• *Các phép toán còn lại của đại số quan hệ đều có thể nhận được từ các phép toán trên.*

• *Phép giao:* Ta chú ý rằng T = r $\cap$ s = r - (r - s). Các câu lệnh của phép trừ ta dùng như trên.

• *Phép nối tự nhiên* T = r |><| s của r trên A và s trên A'. Ta lưu ý rằng nếu A và A' rời nhau thì nối tự nhiên của r và s là tích Đề-các của r và s. Nếu A và A' chung nhau tập thuộc tính X thì ta chỉ nối những bộ của r và s có giá trị bằng nhau trên X. Vậy câu lệnh để tính T sẽ là:

INSERT INTO T

SELECT $r.A_1,....,r.A_n, s.B_1,..., s.B_m$

FROM r, s

WHERE r.X = s.X

• *Phép chia* T = r ÷ s với r trên A và s trên A', A' là tập con của A. Thực chất kết quả của phép chia r ÷ s là những bộ t.(A-A') mà t $\in$ r và t.A' $\in$ s. Vậy ta có câu lệnh tính T:

INSERT INTO T

SELECT r.(A-A')

FROM r, s

WHERE E; ở đây E là điều kiện t.A' $\in$ s

• *Phép nối theo θ* chọn trong tích Đề-các r×s những bộ thỏa mãn θ

• *Phép nối nửa* T = r |><s. Thực chất là chiếu của r |><| s lên R. Tức là:

r |>< s = (r |><| s). R

2.9.9. Tạo khung nhìn

Trong SQL có một nhóm lệnh đóng vai trò như ngôn ngữ định nghĩa dữ liệu DDL (Data Definition Language), nghĩa là từ một quan hệ, nhóm lệnh này định nghĩa cho ta một quan hệ con, tạo cho ta một khung nhìn (view) có tên mà không tốn thêm bộ nhớ. Cú pháp của câu lệnh tạo khung nhìn:

CREATE VIEW V(A_1,..., A_k) AS Q;

Trong đó V là tên của khung nhìn; A_1,..., A_k là các thuộc tính của nó. Q là câu vấn tin định nghĩa khung nhìn.

Ví dụ 2.58:

Từ quan hệ các nhà cung cấp ta có thể tạo ra một khung nhìn: quan hệ gồm các mặt hàng và giá của chúng do A cung cấp:

```
CREATE VIEW V (ITEM, PRICE) AS
SELECT ITEM, PRICE
FROM supplies
WHERE NAME = 'A';
```

Tương tự như hủy bảng, muốn hủy khung nhìn ta dùng câu lệnh DROP. Ví dụ, DROP V.

Trên đây là những khái niệm cơ bản của ngôn ngữ vấn tin SQL đủ dùng cho các minh họa trong cuốn sách này. Để sử dụng được SQL chúng ta cần tham khảo những qui định cụ thể và sự gắn kết của SQL với các ngôn ngữ chủ khác trong các môi trường cụ thể. Các bạn có thể tìm thấy các phiên bản khác của SQL trong nhiều tài liệu khác hiện đang được bán trên thị trường.

2.10. MINH HỌA ỨNG DỤNG VÀ THAO TÁC TRÊN CSDL QUAN HỆ

Các bước thiết kế và tạo phần mềm ứng dụng một CSDL

Bước 1. Thiết kế các bảng (các quan hệ) có tạo khóa chính

Bước 2. Mô tả RBTV (Relationship) khóa chính-khóa ngoại

Bước 3. Nhập dữ liệu. Chú ý khi nhập dữ liệu phải theo quy tắc khóa chính nhập trước, khóa ngoại nhập sau.

Bước 4. Tạo các truy vấn để phục vụ yêu cầu bài toán

Bước 5. Tạo các giao diện tiện ích phù hợp yêu cầu bài toán ...

Sau đây chúng ta minh họa một số ví dụ thực hiện bốn bước 1, 2, 3, 4 để thiết kế một CSDL.

Ví dụ 1. Thiết kế CSDL quản lý phòng thi gồm 4 bảng:

PT(Sopt, DC, Suchua) đây là bảng phòng thi có 3 thuộc tính là số phòng thi, địa chỉ phòng thi và sức chứa của phòng thi. Sopt là khóa chính.

TS(Sobd, Hoten, Sopt) đây là bảng thí sinh có 3 thuộc tính là số báo danh, họ tên, số phòng thi. Sobd là khóa chính.

MT(Mamt, tenmt, B, N) đây là bảng môn thi có 4 thuộc tính là mã môn thi, tên môn thi, buổi thi, ngày thi. Mamt là khóa chính.

KQ(Sobd, Mamt, Diem, Vangthi) đây là bảng kết quả có 4 thuộc tính là số báo danh, mã môn thi, điểm thi, vắng thi.

Chúng ta sẽ thiết kế trong Access - ngôn ngữ quan hệ đơn giản.

Bước 1. Thiết kế các quan hệ: PT, TS, MT, KQ

Thiết kế bảng PT(Sopt, DC, sSuchua) có:

Sopt khóa chính, kiểu text.

DC kiểu text.

Suchua kiểu number.

Thiết kê bảng TS(Sobd, Hoten, Sopt) có:

Sobd là khóa chính, kiểu text

Hoten kiểu text. Sopt khóa ngoại.

Thiết kế bảng MT(Mamt, tenmt, B, N) có:

Mamt khóa chính, kiểu text

Tenmt và B kiểu text; N kiểu ngày tháng.

Thiết kế bảng KQ(Sobd, Mamt, Diem, Vangthi) có:

Diem kiểu số

Vangthi kiểu text hoặc logic

Hình 2.16. Bảng PT

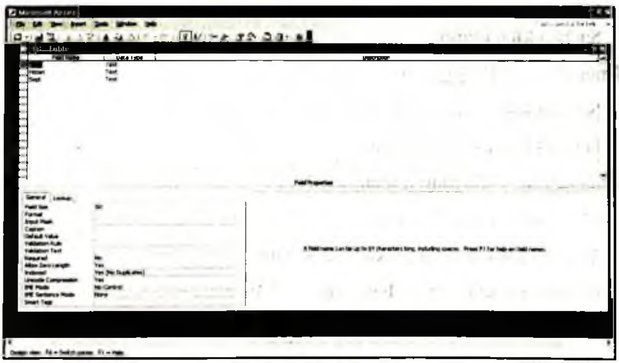
Thiết kế bảng TS có:

Sobd khóa chính, kiểu text

Hoten kiểu text

Sopt kiểu text.

Bảng TS có trong hình 2.17



Hình 2.17. Bảng TS

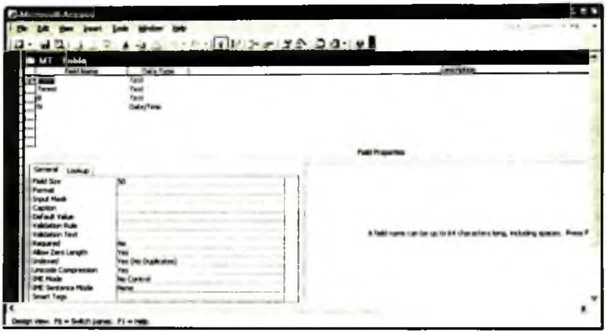
Thiết kế bảng MT có:

Mamt khóa chính, kiểu text

Tenmt kiểu text

B kiểu text

N kiểu ngày tháng



Hình 2.18. Bảng môn thi MT

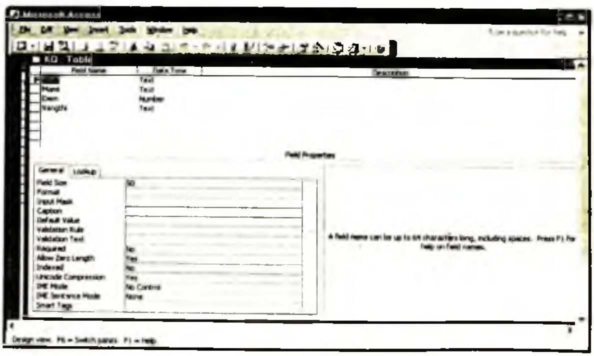
Thiết kế bảng KQ có các thuộc tính:

Sobd kiểu text

Mamt kiểu text

Diem kiểu number.

Lưu ý bảng này có thể tạo khóa chính hoặc không. Trong ví dụ này ta không tạo khóa chính cho bảng này. Ta dùng hai thuộc tính Sobd và Mamt là những khóa ngoại.



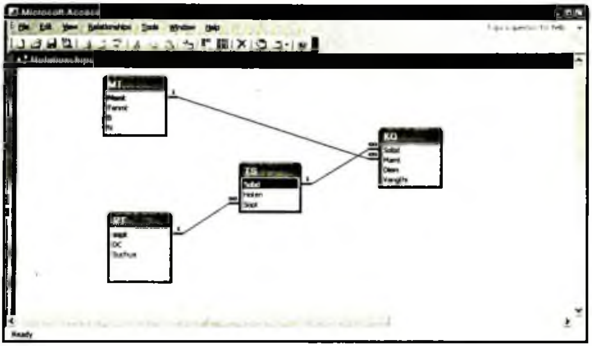
Hình 2.19. Bảng kết quả KQ

Bước 2: Tạo liên kết (Relationship) khóa chính khóa ngoại.

Sự liên kết khóa chính khóa ngoại được mô tả khi thiết kế cơ sở dữ liệu.

Sự liên kết khóa chính khóa ngoại là sự liên kết 1-n (một nhiều). Một giá trị khóa chính có thể xuất hiện nhiều lần trong bảng mà nó đóng vai trò khóa ngoại.

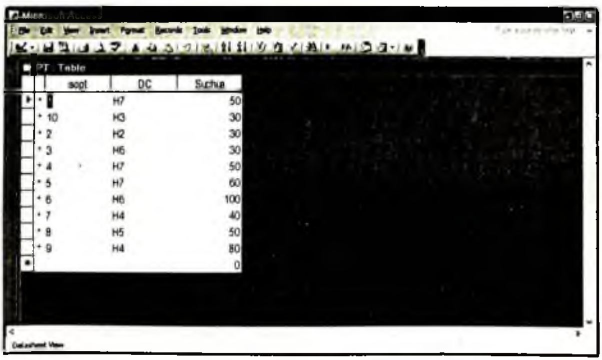
Ví dụ sopt có thể xuất hiện nhiều lần trong bảng TS vì cùng một phòng thi có thể có nhiều thí sinh thi trong đó.



Hình 2.20. Relationship - Liên kết các bảng

Bước 3: Nhập dữ liệu

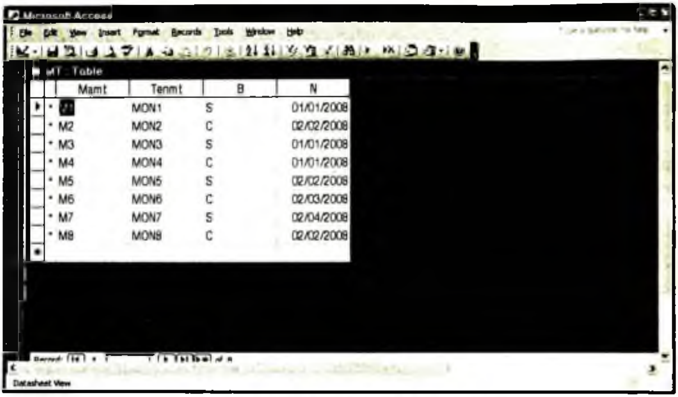
Nhập dữ liệu theo thứ tự PT, MT, TS và cuối cùng là bảng KQ



PT : Table

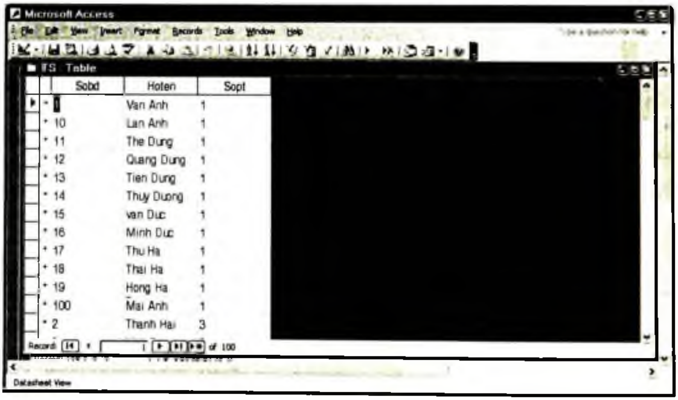
sopt	DC	Suchua
1	H7	50
10	H3	30
2	H2	30
3	H6	30
4	H7	50
5	H7	60
6	H6	100
7	H4	40
8	H5	50
9	H4	80
		0

Hình 2.21. Bảng PT có 10 phòng thi



Mamt	Tenmt	B	N
M1	MON1	S	01/01/2008
M2	MON2	C	02/02/2008
M3	MON3	S	01/01/2008
M4	MON4	C	01/01/2008
M5	MON5	S	02/02/2008
M6	MON6	C	02/03/2008
M7	MON7	S	02/04/2008
M8	MON8	C	02/02/2008

Hình 2.22. Bảng MT có 8 môn thi

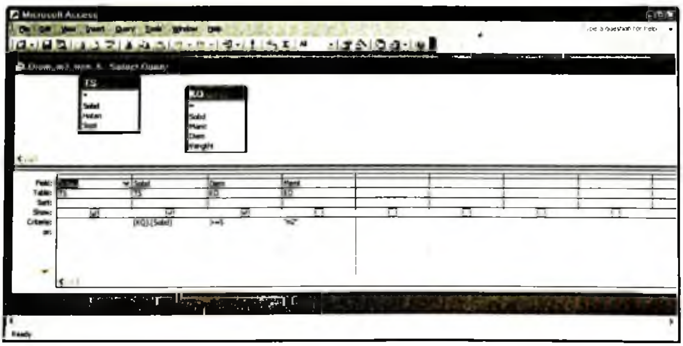


Sobd	Hoten	Sopt
1	Van Anh	1
10	Lan Anh	1
11	The Dung	1
12	Quang Dung	1
13	Tien Dung	1
14	Thuy Duong	1
15	van Duc	1
16	Minh Duc	1
17	Thu Ha	1
18	Thai Ha	1
19	Hong Ha	1
100	Mai Anh	1
2	Thanh Hai	3

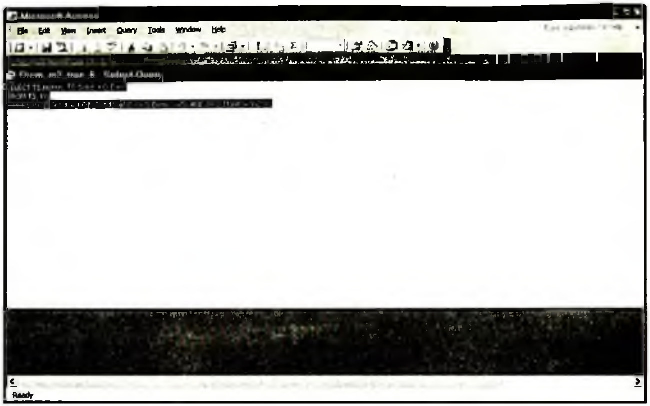
Hình 2.23. Bảng TS có 100 thí sinh

Hình 2.24. Bảng KQ có 800 kết quả

Bước 4: Các truy vấn lọc ra họ tên, số báo danh có điểm môn M2 lớn hơn hoặc bằng 5.

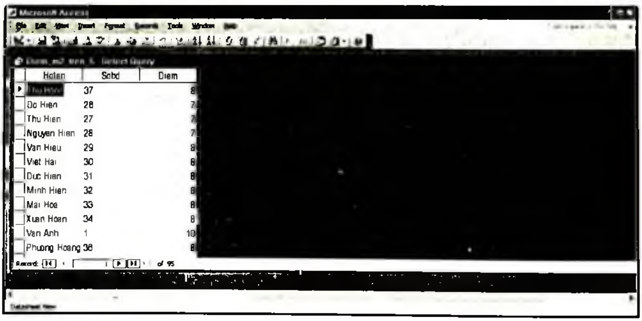


Hình 2.25. Minh họa truy vấn trực quan



Hình 2.26. Minh họa truy vấn bằng SQL

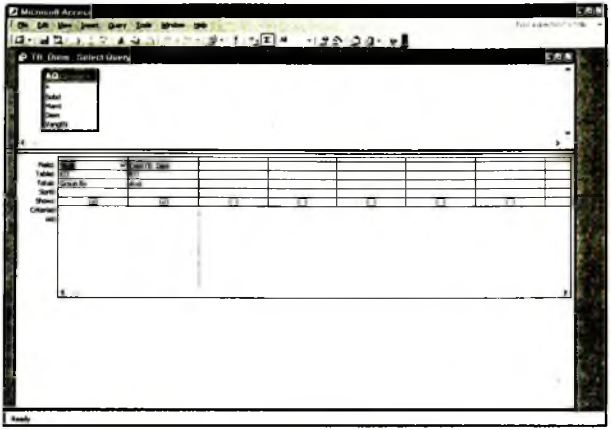
Và kết quả có trong bảng Diem_m2_tren_5



Hình 2.27. Bảng môn m2 trên 5

Lọc ra số báo danh và điểm trung bình các môn thi

Màn hình truy vấn trực quan



Hình 2.28. Truy vấn trực quan

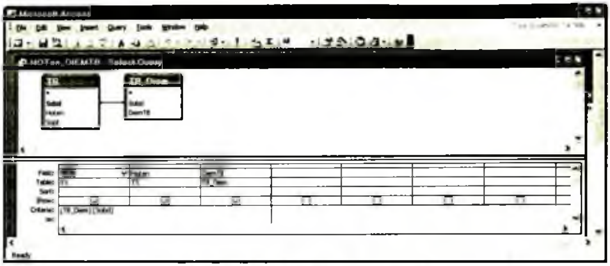
Màn hình truy vấn SQL

Hình 2.29. Câu lệnh SQL

Kết quả có trong màn hình 2.30.

Lọc số báo danh, họ tên và điểm trung bình các môn thi của thí sinh

Để thực hiện câu này ta nối TB_Diem với TS rồi chiếu lên Sobd, Hoten và DiemTB

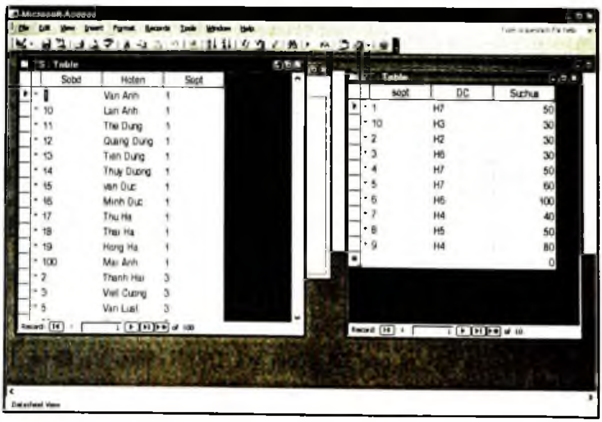


Hình 2.30. Màn hình truy vấn

Truy vấn theo SQL

Hình 2.31. Truy vấn SQL

Bây giờ chúng ta quan sát RBTV: "Với mỗi phòng thi, số thí sinh thi trong đó không được vượt quá sức chứa của phòng thi đó". Ta có 10 phòng thi và 100 thí sinh



Hình 2.32. Kết quả truy vấn

Ứng với mỗi phòng thi có m sinh viên thi trong đó và m $\leq$ sức chứa của phòng thi đó. Ví dụ có 13 thí sinh thi ở phòng số 1 và phòng này có sức chứa là 50.

Nội dung của RBTV là: $\forall t \in PT\ \exists t_1, t_2, ..., t_m \in TS$

sao cho: $t.Sopt = t_i.Sopt$ và $m \leq t.Suchua$; $i = 1, 2, ..., m$

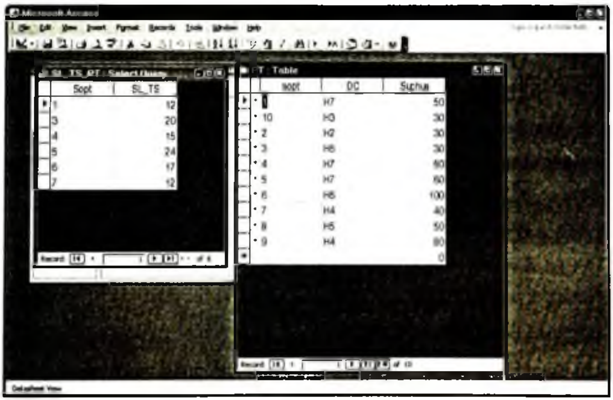
Sau đây là bảng số phòng thi và số lượng thí sinh thi tương ứng.

Sau khi thực hiên câu lệnh:

```
SELECT Sopt, Count(Sobd) AS SL_ TS
FROM TS
GROUP BY Sobd;
```

Ta có kết quả trong bảng SL_TS_PT.

Bên cạnh là bảng các phòng thi, nhìn vào và so sánh ta thấy không phòng thi nào có thí sinh thi vượt quá sức chứa của nó.



Hình 2.33. Minh họa kết quả truy vấn

CÂU HỎI VÀ BÀI TẬP CHƯƠNG 2

2.1 Cho hai quan hệ r và s như sau:

r

E	B	C	D
1	0	0	0
1	1	0	0
1	1	1	0
1	1	1	1

s

E	B	C	D
2	1	1	1
2	2	1	1
1	1	1	0
x	y	z	v

a. Tính r - s và s - r.

b. Tính r + s.

c. Tính r * s.

d. Giả sử X = {E, B, D}, Y = {E, C, D}.

Tính các quan hệ chiếu r. X, r. Y và s. X, s. Y, (r+s). X, (r+s). (X∪ Y).

e. Chứng minh rằng với mọi quan hệ r, s, q thì ta luôn có:

r * s = s * r và r + s = s + r (tính giao hoán)

r * (q + s) = (r * q) + (r * s) (tính kết hợp)

(r + s). X = r. X + s. X

(r * s). X = r. X * s. X

2.2 Cho hai quan hệ r và s như sau:

a. Hai quan hệ giống nhau:

r

E	B	C	D
0	0	0	1
0	0	1	1
0	1	1	1
1	1	1	1

s

E	B	C	D
0	0	0	1
0	0	1	1
0	1	1	1
1	1	1	1

Tính r |><| s, r |>< s, s |>< r.

b. Hai quan hệ hoàn toàn khác nhau trên cùng một lược đồ:

r

E	B	C	D
0	0	0	0
0	0	1	1
0	1	1	1

s

E	B	C	D
a	b	c	d
x	y	z	v

Tính r |><| s.

c. Hai quan hệ có tập các thuộc tính lồng nhau:

r

E	B	C	D
0	1	1	1
x	y	z	v
z	v		
v	z		

s

C	D
1	1
0	0

Tính r |><| s, r |>< s, s |>< r.

d. Hai quan hệ lồng nhau:

r

E	B	C	D
0	0	1	1
1	1	1	1
1	1	0	1

s

C	D
1	1
0	1

Tính r |><| s, Tính r |>< s và tính r |><| $_{2=2}$ s, r |><| $_{1<2}$ s, s|><| $_{1<2}$ r.

e. Cho hai quan hệ r(TT, TÊN, NS, GT) và s(QUÊ, MH, ĐIỂM) sau:

r

TT	TÊN	NS	GT
1	Linh	77	Nữ
2	Quyên	76	Nữ
3	Nam	75	Nam
4	Tuấn	74	Nam

s

QUÊ	MH	ĐIỂM
HN	Anh	18
HF	Hóa	20
SG	Toán	22
VF	Tin học	22

Hãy dùng các thủ thuật nhỏ và sử dụng các phép toán quan hệ để có quan hệ kết quả DS sau:

DS

TT	TÊN	NS	GT	QUÊ	MH	ĐIÊM
1	Linh	77	Nữ	HN	Anh	18
2	Quyên	76	Nữ	HF	Hóa	20
3	Nam	75	Nam	SG	Toán	22
4	Tuấn	74	Nam	VF	Tin học	22

Một cách tổng quát cho hai quan hệ r và s như sau:

r

H	B	C	D
a_1	b_1	c_1	d_1
a_2	b_2	c_2	d_2
a_3	b_3	c_3	d_3
a_4	b_4	c_4	d_4

s

E	F	G
e_1	f_1	g_1
e_2	f_2	g_2
e_3	f_3	g_3
e_4	f_4	g_4

Hãy tìm cách sử dụng các phép toán quan hệ và các thủ thuật để có quan hệ kq:

kq

A	B	C	D	E	F	G
a_1	b_1	c_1	d_1	e_1	f_1	g_1
a_2	b_2	c_2	d_2	e_2	f_2	g_2
a_3	b_3	c_3	d_3	e_3	f_3	g_3
a_4	b_4	c_4	d_4	e_4	f_4	g_4

2.3 Cho hai quan hệ r và s sau:

r

H	B	C
0	0	0
1	1	1
1	1	0

s

D	E
5	6
0	6

Tính tích Đề-các của r và s: r x s.

2.4 Cho hai quan hệ r và s sau:

r

H	B	C	D	E
0	0	0	0	1
0	0	1	1	0
1	1	1	1	1
0	0	0	1	1

s

D	E
0	1
1	1
1	0

Tính r / s.

2.5*

a. Chứng minh rằng: năm (5) phép toán cơ bản của đại số quan hệ hợp, hiệu, Đề-các, chiếu, chọn là độc lập với nhau, nghĩa là không một phép toán nào trong chúng có thể biểu diễn qua các phép còn lại.

b. Chứng minh rằng các phép toán còn lại của đại số quan hệ giao, nối, nối nửa, nối theo θ, phép chia, đều có thể nhận được từ các phép toán cơ bản trên.

2.6 Cho r và s là hai quan hệ trên các lược đồ tương ứng A và A':

$A = \{A_1, A_2,..., A_n\}$, $A' = \{A_1, A_2,..., A_k\}$ với $k < n$.

Giả sử $X = A - A' = \{A_{k+1},..., A_n\}$.

Chứng minh rằng: $r / s = r.\ X - ((r.\ X \times s) - r).\ X$.

2.7

a. Cho lược đồ quan hệ A và tập PTH

$F = \{JB \to E, JG \to I, BE \to I, E \to G, GI \to H\}$ trên A.

Chứng minh rằng $JB \to GH$.

b. Tương tự cho tập PTH F sau:

$F = \{JB \to C, B \to D, CD \to E, CE \to GH, G \to J\}$.

Chứng minh rằng: $JB \to E$, $JB \to G$.

2.8 Cho sơ đồ quan hệ W = < A, F >. Chứng minh (giải thích) rằng: với mọi tập con X bất kỳ của A và mọi phần tử B thuộc tập X thì $X \to B \in F^+$.

Tức là: $\forall\ B \in X \subset A \Rightarrow X \to B \in F^+$.

2.9 Cho sơ đồ quan hệ W = < A, F >, với $A = \{E, B, C, D\}$ và $F = \{E \to B, E \to C\}$. Hãy tìm các PTH suy được từ các quy tắc của PTH trong các ràng buộc sau:

a. $E \to D$.

b. $C \to D$.

c. $EB \to B$.

d. $BC \to E$.

e. $E \to BC$.

2.10 Cho sơ đồ quan hệ W = < A, F >, với A = {E, B, C, D } và F = {$E \to B$, $BC \to D$}. PTH nào trong dãy sau là suy được từ F bằng các quy tắc của PTH:

a. $C \to D$.

b. $E \to D$.

c. $ED \to C$.

d. $BC \to E$.

e. $B \to CD$.

2.11 Cho bảng quan hệ r:

r

E	B	C	D
x	u	x	y
y	x	z	x
z	y	y	y
y	z	w	z

Trong các PTH sau đây PTH nào không thỏa mãn r:

$E \to B$, $E \to C$, $B \to E$, $C \to D$, $D \to C$, $D \to E$.

2.12 Cho hai lược đồ quan hệ A và A', $A \cap A' = X$.

Chứng minh rằng nếu quan hệ r trên tập thuộc tính $A \cup A'$ thỏa mãn $X \to A'$ thì r = r. A |><| r. A'.

2.13 Cho sơ đồ quan hệ W = < A, F >, với A = {I, B, C, D, E, G, H} và tập các PTH F: F = {$I \to D$, $IB \to DE$, $CE \to G$, $E \to H$}. Tính $(IB)^+$.

2.14 Trong thuật toán tìm bao đóng X^+ ta đã xây dựng dãy:

$X^0 \subset X^1 \subset X^2 \ldots \subset X^i \ldots \subset$ với

$X^0 = X$

$X^{i+1} = X^i Z^i$ và $Z^i = \{B \in A: B \notin X^i$ và $X^i \to B \in F^+\}$.

a. Xét giao của các Z^i: $Z^i \cap Z^k = ?$

b. Chứng minh rằng: $\forall$ i $Z^i \subset (X^i)^+$.

2.15

a. Cho sơ đồ quan hệ W = < A, F >, với A = {I, B, C, D, E, H} và F = {I → E, C → D, E → DH}. Chứng minh rằng K = {I,B,C} là khóa duy nhất của W.

b. Cho sơ đồ quan hệ W = < A, F >, với A = {G,B,C,D} và F = {GB → C, D → B, C → GBD}. Tìm họ khóa K và họ phản khóa K^{-1}của W.

c. Cho sơ đồ quan hệ W = < A, F >, với A = {H,B,C,D,E,G} và F = {HB → C, C → H, BC → D, HCD → B, D → EG,BE → C, CG → BD, CE → CG}. Tìm các khóa của W.

2.16 Cho lược đồ quan hệ A = {A_1, A_2,... A_n}. Họ S các tập con của A được gọi là *hệ Sperner* nếu trong S không có hiện tượng lồng nhau, nghĩa là không có $X \subsetneq Y$ với X, Y $\in$ S. Gọi K là họ tất cả các khóa của sơ đồ quan hệ W = < A, F >.

Chứng minh rằng: K là hệ Sperner.

2.17* Cho hệ Sperner không rỗng K trên A. Chứng minh rằng tồn tại một sơ đồ quan hệ W để K là tập các khóa của W. Vậy ta có thể chứng minh mệnh đề K là họ khóa của W = < A, F > khi và chỉ khi K là hệ Sperner.

2.18 Cho sơ đồ quan hệ W = < A, F>. Gọi K^{-1} là họ phản khóa của W.

Chứng minh rằng: K^{-1} là hệ Sperner.

2.19 Giả sử K, K^{-1} là các hệ Sperner trên A. Chứng minh rằng: $\cup K = A - (\cap K^{-1})$

Ở đây $\cup$ K và $\cap$ K^{-1} là hợp và giao của các tập trong K và K^{-1} tương ứng.

2.20 Thuật toán tìm khóa của một quan hệ.

Cho r là quan hệ sau:

r

E	B	C	D
1	1	1	1
2	2	2	2
1	2	3	4

Hãy tìm một khóa và tất cả các khóa của r bằng thuật toán tìm 1 khóa, thuật toán tìm tất cả các khóa của r.

2.21 Ta nói tập phụ thuộc hàm F là *Phủ* của tập phụ thuộc hàm G nếu $F^+ \supset G^+$. Hai tập phụ thuộc hàm F và G là *tương đương*, ký hiệu là F ~ G nếu: $F^+ = G^+$. Chứng minh rằng: F ~ G khi và chỉ khi F phủ G và G phủ F.

2.22 Ký hiệu X_G^+, X_F^+ là các tập bao đóng đối với các tập PTH G và F tương ứng. Chứng minh rằng nếu F ~ G thì $X_G^+ = X_F^+$.

2.23 Cho sơ đồ quan hệ W = < A, F >. Ta nói *W là chính tắc* nếu:

a. Vế phải của mỗi PTH trong F là một thuộc tính.

b. Trong tập phụ thuộc hàm F không có phụ thuộc hàm f (thừa) mà: F - f ~ F.

c. Trong F không có phụ thuộc hàm $X \rightarrow A$

mà có $Z \subset X$ và $(F - (X \rightarrow A)) \cup (Z \rightarrow A) \sim F$. Chứng minh rằng với mọi sơ đồ quan hệ W = < A, F > luôn tồn tại một sơ đồ quan hệ chính tắc W' = < A, G > tương đương với W (hai sơ đồ quan hệ là tương đương nếu các tập phụ thuộc hàm của chúng tương đương).

2.24 Thuật toán tìm phủ chính tắc của một sơ đồ quan hệ.

Input: W = < A, F >, với $F = \{f_1, f_2, \ldots f_m\}$

Output: W' = < A, G > chính tắc và tương đương với W.

Thuật toán:

Bước 1:

$F_0 = F$

$F_i = F_{i-1} - f_i$ nếu $F_{i-1} - f_i$ tương đương với F_{i-1}, ngược lại $F_i = F_{i-1}$,

i = 1, 2,..., m.

Bước 2:

Loại bỏ các thuộc tính thừa trong vế trái của các PTH của F_m.

Chứng minh rằng tập F_m nhận được chính là tập G.

2.25 Cho sơ đồ quan hệ W = < A, F >, với A = {a, b, c, d, e, g} và F = {ab → c, c → a, bc → d, acd → d, d → eg, be → c, cg → bd, ce → ag}.

Dùng thuật toán trên tìm W' = < A, F > chính tắc và tương đương với W.

2.26 Giả sử W = < A, F > là sơ đồ quan hệ, K là họ khóa của W. Đặt M = {k - B: B ∈ k và k là một khóa, tức k ∈ K}.

F_n là tập tất cả các thuộc tính thứ cấp (thuộc tính không khóa).

Đặt $L = \{C^+: C \in M\}$.

Chứng minh rằng khi đó ta có 3 mệnh đề sau là tương đương:

1. W là 2NF.
2. $\forall X \in L$ thì $X \cap F_n = \varnothing$.
3. $\forall X \in L$ và $B \in F_n$ thì $(X - B)^+ = X - B$

2.27 Cho sơ đồ quan hệ W = < A, F >, F_n là tập tất cả các thuộc tính thứ cấp, K là tập các khóa. Đặt $Y = \{k - F_n: k \in K^{-1}\}$. Chứng minh rằng: W là 2NF khi và chỉ khi $\forall X \in Y$ thì $X^+ = X$.

2.28 Cho sơ đồ quan hệ W = < A, F >. Ta nói trong W có *quan hệ bắc cầu* nếu giữa 2 (hoặc nhiều hơn) thuộc tính thứ cấp có ràng buộc phụ thuộc hàm. Chứng minh rằng: W là 3NF nếu trong W không có quan hệ bắc cầu.

2.29 Cho lược đồ quan hệ A = {C, I, D, B, Q, G, L, M} và tập $F = \{C \to IDBQ, D \to B, L \to Q\}$. Xét xem W = < A, F> thuộc dạng chuẩn nào, 2NF, 3NF, BCNF, 4NF?

2.30 Xét xem sơ đồ quan hệ sau đây thuộc dạng chuẩn nào W = < A, F >, với A = {J,B,C,D,E,G,H,I} và $F = \{JC \to B, BI \to JCD, JBC \to D, H \to I, JCE \to BCG, CG \to JE\}$.

2.31 Cho sơ đồ quan hệ W = < A, F >. F_n là các phần tử thứ cấp. Chứng minh rằng: W là 3NF khi và chỉ khi $\forall B \in K^{-1}, a \in F_n$ thì $(B - a)^+ = B - a$

2.32 Cho sơ đồ quan hệ W = < A, F >. Chứng minh rằng W là 3NF khi và chỉ khi với mọi tập thuộc tính $X \neq A$: $X^+ = X$, a là thuộc tính thứ cấp và $a \in X$ thì $(X - a)^+ = X - a$

2.33 Cho sơ đồ quan hệ W = < A, F >. F_n là các thuộc tính thứ cấp. Chứng minh rằng W là BCNF khi và chỉ khi $\forall B \in K^{-1}, a \in B$ thì $(B - a)^+ = (B - a)$.

2.34 Cho lược đồ quan hệ A = {H, B, C, D, E, G}, và tập phụ thuộc hàm $F = \{HB \to C, C \to B, HBD \to E, G \to H\}$.

Xét xem W = < A, F > có là BCNF không?

2.35* Cho quan hệ r, E là ma trận bằng nhau của r.

M là họ cực đại của E. Chứng minh rằng r là BCNF khi và chỉ khi với mọi B thuộc M và a thuộc B thì $(B - a)^+ = B - a$.

2.36* Cho sơ đồ quan hệ W = < A, F >, trong F không có PTH dạng tầm thường ($X \to Y$ mà $Y \subset X$). Chứng minh rằng: W là BCNF khi và chỉ khi $X \to B \in F$ thì $X^+ = A$.

2.37 Cho lược đồ quan hệ $A = \{A_1, A_2,... A_n\}$, $X, Y \subset U$. Chứng minh rằng: Nếu $X \rightarrow Y$ thì $X \rightarrow\rightarrow Y$ (nói cách khác phụ thuộc hàm là trường hợp riêng của MD).

2.38* Ta nói MD $X \rightarrow\rightarrow Y$ trên A là không tầm thường nếu $Y \neq \varnothing$, Y not $\subseteq$ X và $X \cup Y \neq A$.

Cho lược đồ quan hệ A; X, Y, Z là các tập rời nhau và $X \cup Y \cup Z = A$.

Nếu $X \rightarrow Y$ hoặc $X \rightarrow Z$ thì ta có $X \rightarrow\rightarrow Y$ (hoặc Z). Khi đó MD $X \rightarrow\rightarrow Y$ (hoặc Z) ta sẽ gọi là MD kề của phụ thuộc hàm $X \rightarrow Y$ (hoặc Z).

Ta nói rằng phụ thuộc đa trị $X \rightarrow\rightarrow Y$ là thuần nhất nếu nó không tầm thường và không là kề của bất kỳ PTH nào trên A.

Cho sơ đồ quan hệ W = < A, F >. $K = \{K_1, K_2,..., K_m\}$ là tập khóa của W. Chứng minh rằng nếu $Y \cap (\cap K_i) = \varnothing$ và $X \rightarrow\rightarrow (Y - K_i)$ với i = 1, 2,... m thì MD $X \rightarrow\rightarrow Y$ không là một MD thuần nhất.

2.39* Giả sử MD $X \rightarrow\rightarrow Y$ là MD thuần nhất trên A.

$K = \{K_1, K_2,..., K_m\}$ là tập các khóa.

Chứng minh rằng nếu $X \rightarrow\rightarrow (Y - K_i)$, i = 1, 2,... m thì $Y \cap (\cap K_i) \neq \varnothing$.

2.40 Giả sử $X \rightarrow\rightarrow Y$ là MD không tầm thường trên lược đồ A và $K_1, K_2,... K_m$ là tập các khóa mà: $Y - K_i \neq \varnothing$, i = 1, 2,... m. Chứng minh rằng nếu $Y \cap (\cap K_i) = \varnothing$ và $X \rightarrow\rightarrow Y \cap K_1$, i = 1, 2,... m thì $X \rightarrow\rightarrow Y$ là MD không thuần nhất.

2.41* Giả sử $X \rightarrow\rightarrow Y$ là MD thuần nhất trong lược đồ quan hệ A; $K = \{K_1, K_2,..., K_m\}$ là tập các khóa. Chứng minh rằng nếu $X \rightarrow\rightarrow (Y - \cap K_i)$ thì $Y \cap (\cap K_i) \neq \varnothing$.

2.42* Giả sử $X \rightarrow\rightarrow Y$ là MD không tầm thường trên lược đồ quan hệ A. $K_1, K_2,... K_m$ là tập các khóa của W mà $Y - K_i \neq \varnothing$, i = 1, 2... m.

Chứng minh rằng $X (Y \cap K_i) \rightarrow X (Y \cap K_j)$ với $i \neq j$.

2.43* Giả sử $X \rightarrow\rightarrow Y$ là MD không tầm thường trên A.

K là khóa của W, $Y \cap K \neq \varnothing$.

Chứng minh rằng $X (Y \cap K) \rightarrow Y$.

2.44* Giả sử $X \rightarrow\rightarrow Y$ (hoặc Z) là MD thuần nhất trên A.

Chứng minh rằng với khóa k của W thì $k - Y \neq \varnothing$ và $K - Z \neq \varnothing$.

2.45* Chứng minh rằng nếu $X \rightarrow\rightarrow Y$ (hoặc Z) là MD thuần nhất trên A, k là một khóa của W thì k có ít nhất là 3 phần tử, tức là $|k| \geq 3$.

2.46* Chứng minh rằng nếu $X \to\to Y$ là MD không tầm thường thì $X \to Y$ khi và chỉ khi tồn tại khóa k mà $X \to Y \cap k$.

2.47 Cho lược đồ quan hệ A = {B, D, I, O, S, Q}, tập các ràng buộc {$S \to\to D$, $I \to B$, $IS \to Q$, $B \to Q$}. Xét xem W có là 4NF không?

2.48 Cho lược đồ A = {G, B, C, D, E, I} và tập ràng buộc F = {$G \to\to BCD$, $B \to\to GC$, $C \to D$}. W = < A, F > có là 4NF không?

2.49 Gọi A là tập thuộc tính và Đ là tập phụ thuộc (thuộc một loại bất kỳ) trên tập thuộc tính A. Chúng ta hãy định nghĩa SAT (Đ) là tập các quan hệ r trên A sao cho r thỏa mãn mỗi phụ thuộc trong Đ. Hãy chứng minh.

a. SAT ($Đ_1 \cup Đ_2$) = SAT ($Đ_1$) $\cap$ SAT ($Đ_2$)

b. Nếu $Đ_1$ suy diễn được tất cả các phụ thuộc trong $Đ_2$ thì SAT ($Đ_1$) $\supseteq$ SAT ($Đ_2$)

2.50 Gọi F là một tập phụ thuộc với các vế phải chỉ có một thuộc tính.

Chứng minh rằng nếu lược đồ A có một phụ thuộc vi phạm BCNF $X \to B$, trong đó $X \to B$ thuộc F^+ thì tồn tại một phụ thuộc $Y \to C$ trong chính tập F vi phạm dạng BCNF.

2.51 Chứng minh nhận xét sau: Nếu A là một lược đồ quan hệ và $k \subseteq A$ là một khóa của W ứng với tập phụ thuộc F thì k không thể có một vi phạm dạng 3NF ứng với tập phụ thuộc chiếu của F lên k, $\pi_k(F)$.

2.52 Chứng minh rằng không thể có một phụ thuộc được gọi là "*Phụ thuộc hàm gắn kết*" (embedded functional dependency). Nghĩa là nếu $S \subseteq R$ với S, R là các quan hệ trên A và $X \to Y$ đúng trong S thì $X \to Y$ đúng trong R.

2.53 *Phụ thuộc bao hàm đơn ngôi (*unary inclusion dependency*)* $C \subseteq B$ trong đó C, B là các thuộc tính (có thể từ các quan hệ khác nhau) khẳng định rằng trong những giá trị hợp lệ của các quan hệ, mỗi giá trị xuất hiện trong cột của C cũng xuất hiện trong cột của B. Chứng tỏ rằng các tiên đề sau là đúng đắn và đầy đủ đối với các phụ thuộc bao hàm đơn ngôi.

a. $B \subseteq B$ với mọi B

b. Nếu $H \subseteq B$ và $B \subseteq C$ thì $H \subseteq C$

2.54 Giả sử với số chẵn *n* chúng ta có các thuộc tính $A_1, \ldots, A_n$. Cũng giả sử rằng $A_i \subseteq A_{i+1}$ với i lẻ, nghĩa là i = 1,3,..., n-1.

Cuối cùng giả sử rằng với i = 3,5,..., n-1 chúng ta có $A_i \to A_{i+1}$ và $A_1 \to A_n$.

a. Chứng minh rằng nếu các quan hệ được giả định là hữu hạn thì tất cả các phụ thuộc trên có thể đảo ngược lại nghĩa là:

$A_2 \subseteq A_1$, $A_2 \rightarrow A_3$, $A_4 \subseteq A_3$, $A_4 \rightarrow A_5, \ldots$, $A_n \subseteq A_{n-1}$, $A_n \rightarrow A_1$

b. Chứng minh rằng tồn tại những quan hệ vô hạn mà (*a*) không đúng: nghĩa là chúng thỏa mãn tất cả các phụ thuộc đã cho nhưng không thỏa mãn các phụ thuộc đảo ngược.

2.55 Chứng minh rằng nếu Đ chỉ là một tập phụ thuộc hàm thì quan hệ R có dạng BCNF ứng với Đ nếu và chỉ nếu R có dạng 4NF ứng với Đ.

2.56 Chứng minh rằng nếu $X \rightarrow A_1, \ldots, X \rightarrow A_n$ là các phụ thuộc hàm trong một phủ cực tiểu thì sơ đồ $W = < XA_1, \ldots A_n, F >$ có dạng 3NF.

2.57 Cho CSDL quản lý phòng thi gồm các quan hệ sau:

PT(SoPT, DC, Suchua); Quan hệ PT có các thuộc tính số PT, địa chỉ, sức chứa

TS(SBD, Hoten, SoPT,...); Quan hệ TS có số báo danh, họ tên, số phòng thi,...

MON(Mam, Tenm, Buoi, Ngay); Quan hệ môn thi có mã môn, tên môn, buổi và ngày thi

KQ(SBD, Mam, Diem); Quan hệ kết quả có số báo danh, mã môn, điểm.

Trình bày 3 bước các RBTV sau:

Với mỗi phòng thi số thí sinh thi trong đó không được vượt quá sức chứa của nó.

Với mỗi sinh viên số môn sinh viên đó thi không vượt quá 10.

Bằng ĐSQH thực hiện các yêu cầu sau:

Với các thí sinh thi ở phòng thi số 1 hãy lọc ra SBD, Hoten, Diem của buổi chiều ngày 20/10/2008.

Lọc SBD, Hoten của các sinh viên có điểm của 2 môn CSDL và Java đều trên 8

Bằng SQL thực hiện các yêu cầu sau:

Thực hiện bằng SQL hai câu trong phần b.

Lọc SBD, Hoten những sinh viên có điểm CSDL cao hơn điểm CSDL của sinh viên sv200.

Tính điểm trung bình các môn thi của từng sinh viên.

Chương 3

TẬP THÔ VÀ ỨNG DỤNG

3.1. CÁC KHÁI NIỆM VỀ TẬP THÔ

Khái niệm tập thô được GS.TSKH Ba lan Z. Pawlak đưa ra vào năm 1982. Theo một nghĩa nào đó khái niệm tập thô gắn liền với các bài toán phân loại các đối tượng. Xét một không gian các đối tượng U, $P = \{p_1, p_2, \ldots p_k\}$ là một phân hoạch của U, khi đó trong họ các tập con 2^U của U sẽ có một số tập là những *tập rõ*, số còn lại là những *tập thô ứng với phân hoạch P*. Tập thô gắn với các bài toán phân loại, gom cụm các đối tượng. Chính những sự phân loại và gom cụm đó tạo nên phân hoạch P của U. Về mặt trực quan tập thô là tập những đối tượng không phân loại được. Tập rõ là những tập phân loại được. Ví dụ xét U là tập sinh viên trong lớp. Nếu phân loại theo giới tính thì U được chia thành hai nhóm, nhóm nam và nhóm nữ mỗi nhóm là một tập phân loại được, đó là những *tập rõ*. Bây giờ lấy một tập gồm vài bạn nam, vài bạn nữ, tập như vậy không phân loại được, đó là những *tập thô*. Vậy tập thô gắn với một phân hoạch của không gian đang xét U. Để minh họa một ứng dụng của tập thô ta xét không gian U gồm các bệnh nhân trong một bệnh viện. Không gian U được chia theo các nhóm ví dụ như E_1 gồm các bệnh nhân bị bệnh tim, E_2 gồm các bệnh nhân bị bệnh phổi, E_3 gồm các bệnh nhân bị các bệnh còn lại. Như vậy E_1, E_2, E_3 là những tập rõ trong không gian U, đây là những tập *phân loại được, mô tả được*. Bây giờ xét tập X gồm vài người trong E_1, vài người trong E_2 và vài người trong E_3 khi đó tập X là tập *không phân loại được*, tập *không mô tả được* đây là tập thô. Việc nghiên cứu những tập X như vậy là nghiên cứu các tập thô. Vậy ở đâu có phân loại, ở đó có tập thô.

Cho tập U hữu hạn, khác rỗng bất kỳ, U được gọi là tập các đối tượng. $E = \{E_1, E_2, \ldots, E_k\}$ là phân hoạch của U. Trong lý thuyết tập thô các nhóm $E_1, E_2, \ldots, E_k$ được gọi là các *tập sơ cấp* hay các *tập mô tả được*.

Cặp U và phân hoạch E tạo nên không gian được gọi là *không gian xấp xỉ* hay *không gian nền Pawlak:*

Vậy Apr = (U, E) là không gian nền hay không gian xấp xỉ Pawlak.

3.1.1. Xấp xỉ tập hợp

Cho không gian xấp xỉ Apr = (U, E).

Giả sử $X \subseteq U$.

Định nghĩa 3.1 Xấp xỉ của tập X trong không gian Apr = (U, E)

Xấp xỉ trên của X trong Apr = (U, E), ký hiệu X^E (hoặc $\overline{X}(E)$) là hợp của các nhóm E_i có phần tử chung với X hay $X^E = \overline{X}(E)) = \cup\{E_i \in E: E_i \cap X \neq \phi\}$.

Xấp xỉ dưới của X trong Apr = (U, E), ký hiệu X_E (hoặc $\underline{X}(E)$) là hợp của các nhóm E_i mà E_i là tập con của X hay $X_E = \underline{X}(E) = \cup\{E_i \in E: E_i \subseteq X\}$.

Lưu ý:

- Xấp xỉ trên, dưới của X luôn gắn với không gian nền Apr = (U, E) hay gắn với phân hoạch E.

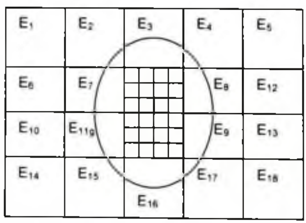
- Xấp xỉ trên của X là hợp của những nhóm E_i giao với X khác rỗng, mà mỗi phần tử x của X thuộc một E_i nào đó nên $X \subseteq X^E$.

- Xấp xỉ dưới của X là hợp của những nhóm E_i mà $E_i \subseteq X$, nên $X_E \subseteq X$. Vậy ta luôn có: $X_E \subseteq X \subseteq X^E$

- Vùng $X^E \setminus X_E$ gọi là vùng biên của X, ký hiệu $BIEN(X) = X^E \setminus X_E$

- Đại lượng $\alpha_X{}^E = card(X_E)/card(X^E)$ là *độ chính xác của xấp xỉ X* ứng với E.

Ví dụ 3.1.a:



Hình 3.1: Minh họa tập xấp xỉ

Giả sử U là hình chữ nhật và phân hoạch của U có 20 nhóm, mỗi nhóm là một hình vuông con trong U, tức $E = \{E_1, E_2, ...\}$

Lấy X là hình elíp.

Khi đó X_E = hình gồm hai nhóm kẻ ô vuông nhỏ trong elíp.

X^E = hợp của các nhóm có phần tử chung với X

$X^E = E_2 \cup E_3 \cup E_4 \cup E_7 \cup E_8 \cup E_9 \cup E_{11} \cup E_{15} \cup E_{16} \cup E_{17} \cup X_E$

$\alpha_X^E = card(X_E)/ card(X^E) = 2/12 = 1/6$ (ở đây phải giả sử các nhóm có số phần tử như nhau).

Ví dụ 3.1.b

Cho U = {1, 2, 3, 4, 5, 6, 7, 8, 9};

$E = \{E_1, E_2, E_3, E_4\} = \{\{1, 2, 3\}, \{4, 5\}, \{6, 7\}, \{8, 9\}\}$.

Lấy X, Y là các tập con của U và X = {1, 2, 3}; Y = {1, 4}.

Khi đó: $X^E = X_E = \{1, 2, 3\}$;

$Y^E = \{1, 2, 3\} \cup \{4, 5\} = \{1, 2, 3, 4, 5\}$ và $Y_E = \phi$.

Bổ đề 3.1

Cho không gian Apr = (U, E).

a. Khi đó với mọi tập X mà X là hợp của một số nhóm E_i nào đó, tức $X = \cup E_i$ ta luôn có $X_E = X^E = X$.

b. Với mọi $X \subseteq U$ khác rỗng, X^E luôn khác rỗng và X^E là hợp của một số nhóm E_i nào đó, tức là $X^E = \cup E_i$.

c. Với mọi $X \subseteq U$ khác rỗng, X_E có thể rỗng hoặc khác rỗng, nếu X_E khác rỗng thì X_E là hợp của một số nhóm E_i nào đó, tức là $X_E = \cup E_i$.

Chứng minh:

a. Theo định nghĩa ta có $X_E = \cup\{E_i \in E: E_i \subseteq X\}$; nếu $X = \cup E_i$ thì mọi E_i trong $\cup E_i$ và chỉ những E_i đó thỏa mãn $E_i \subseteq X$ nên $X_E = \cup E_i = X$.

Tương tự theo định nghĩa ta có $X^E = \cup\{E_i \in E: E_i \cap X \neq \phi\}$; nếu $X = \cup E_i$ thì mọi E_i trong $\cup E_i$ và chỉ những E_i đó thỏa mãn $E_i \cap X \neq \phi$ nên $X^E = \cup E_i = X$

Hai mệnh đề b và c của bổ đề ta dễ dàng suy ra từ định nghĩa của X^E và X_E.

Chúng ta cũng dễ dàng chứng minh được một số tính chất đơn giản của các tập xấp xỉ trên, dưới của X trong không gian Apr = (U, E):

Bổ đề 3.2

Giả sử cho Apr =(U, E) là không gian xấp xỉ, X, Y $\subseteq$ U.

Với $E = \{E_1, E_2, ..., E_k\}$. Gọi $-X = U \setminus X$ là phần bù của X

Khi đó xấp xỉ dưới (Lower approximation) của X trong Apr = (U, E) thỏa mãn các tính chất [16]:

L0. $X_E = -(-X)^E$

L1. $X_E \subseteq X$

L2. $\phi_E = \phi$

L3. $U_E = U$

L4. Nếu $X \subseteq Y$ thì $X_E \subseteq Y_E$

L5. $(X \cup Y)_E \supseteq (X)_E \cup (Y)_E$

L6. $(X \cap Y)_E = (X)_E \cap (Y)_E$

L7. $X \subseteq (X^E)_E$

L8. $(X_E)_E = X_E$

L9. $(X^E)_E = X^E$

Và xấp xỉ trên (Upper approximation) của X trong Apr = (U, E) thỏa mãn các tính chất:

U0. $X^E = -(-X)_E$

U1. $\Phi^E = \Phi$

U2. $U^E = U$

U3. $X \subseteq (X^E)$

U4. $X \subseteq Y$ thì $X^E \subseteq Y^E$

U5. $(X \cup Y)^E = (X)^E \cup (Y)^E$

U6. $(X \cap Y)^E \subseteq (X)^E \cap (Y)^E$

U7. $(X_E)^E \subseteq X$

U8. $X^{EE} = X^E$

U9. $(X_E)^E = X_E$

Ngoài ra ta dễ dàng thấy một vài tính chất của xấp xỉ trên dưới như sau:

T1. Nếu $X = E_1 \cup E_2 \cup \ldots \cup E_i$ thì $X^E = X_E = X$ tức là khi X là hợp của một số nhóm thì xấp xỉ dưới và xấp xỉ trên của X bằng nhau và bằng X và khi đó $X^{EE\ldots E} = X_{EE\ldots E} = X$; với $(X^E)^E = X^{EE}$

T2. $(-X)_E = U \setminus (X^E) = -(X^E)$ với $-X = U \setminus X$ là phần bù của X

T3. $(-X)^E = U \setminus (X_E) = -(X_E)$

T4. $(X^E \cup Y_E)^E = (X^E \cup Y_E)$

T5. $(X^E \cup Y_E)_E = (X^E \cup Y_E)$

Chứng minh

L0. Chứng minh $X_E = -(-X)^E$ tức là chứng minh $-(X_E) = (-X)^E$

Thật vậy giả sử $X_E = E_1 \cup E_2 \cup ... \cup E_i$ thì bù của X_E là:

$$-X_E = E_{i+1} \cup E_{i+2} \cup ... \cup E_k$$

trong $-X_E$ có thể có những nhóm có phần tử chung với X, có những nhóm nằm ngoài X và đó chính là xấp xỉ trên của -X, tức $-X_E = E_{i+1} \cup E_{i+2} \cup ... \cup E_k = -(X)^E$.

Tương tự chứng minh tính chất U0. Thay cho chứng minh $X^E = -(-X)_E$ ta chứng minh $-(X^E) = (-X)_E$

Thật vậy giả sử $X^E = E_1 \cup E_2 \cup ... \cup E_i$ thì bù của X^E là:

$$-X^E = E_{i+1} \cup E_{i+2} \cup ... \cup E_k$$

trong $-X^E$ các $E_{i+1}, E_{i+2}, ... E_k$ đều được chứa trong -X

nên $(-X)_E = E_{i+1} \cup E_{i+2} \cup ... \cup E_k = -(X)^E$

Ví dụ xét X trong hình 3.1 ta có $(-X)^E = \bigcup_{i=1}^{18} E_i$ và xấp xỉ dưới của X; X_E là hai hình vuông kẻ trong X tức là $X_E = U - (-X)^E = -(-X)^E$.

$X^E = E_2 \cup E_3 \cup E_4 \cup E_7 \cup E_8 \cup E_9 \cup E_{11} \cup E_{15} \cup E_{16} \cup E_{17} \cup X_E$

$= U - (E_1 \cup E_5 \cup E_6 \cup E_{10} \cup E_{14} \cup E_{12} \cup E_{13} \cup E_{18}) = U - (-X)_E = -(-X)_E$.

Các tính chất L1, L2, L3, U1, U2, U3 suy trực tiếp từ định nghĩa.

L4, U4. Ta chứng minh tính đơn điệu của xấp xỉ trên và dưới

Giả sử $X \subseteq Y$ ta phải chứng minh $X^E \subseteq Y^E$. Lấy $x \in X^E$ suy ra $x \in E_i$ nào đó mà $E_i \subseteq X$ và vì $X \subseteq Y$ nên $E_i \subseteq Y$ tức $x \in Y_E$.

Tương tự lấy $x \in X^E$ suy ra $x \in E_i$ nào đó mà E_i có phần tử chung với X và vì $X \subseteq Y$ nên E_i có phần tử chung với Y, suy ra $x \in Y^E$.

L5, U5. Ta chứng minh các tính chất liên quan đến xấp xỉ trên, dưới của hợp $X \cup Y$.

Trước tiên ta chứng minh tính chất L5: $(X \cup Y)_E \supseteq (X)_E \cup (Y)_E$

Vì $(X \cup Y) \supseteq X$ và $(X \cup Y) \supseteq Y$ nên theo tính chất đơn điệu L4 ta có:

$(X \cup Y)_E \supseteq (X)_E$; $(X \cup Y)_E \supseteq (Y)_E$.

Từ đó suy ra: $(X \cup Y)_E \supseteq (X)_E \cup (Y)_E$

Tương tự chứng minh tính chất U5: $(X \cup Y)^E = (X)^E \cup (Y)^E$.

Theo tính chất đơn điệu U4 ta có:

$(X \cup Y)^E \supseteq (X)^E$

$(X \cup Y)^E \supseteq (Y)^E$ nên lấy hợp hai vế ta có $(X \cup Y)^E \supseteq (X)^E \cup (Y)^E$

Bây giờ ta chứng minh $X^E \cup Y^E \supseteq (X \cup Y)^E$, lấy $x \in (X \cup Y)^E$ suy ra $x \in E_i$ nào đó mà E_i có phần tử chung với $X \cup Y$ suy ra E_i hoặc có phần tử chung với X và khi đó $x \in X^E$ hoặc E_i có phần tử chung với Y suy ra $x \in Y^E$ suy ra $x \in X^E \cup Y^E$.

L6, U6. Chứng minh các tính chất xấp xỉ trên, dưới của giao hai tập $X \cap Y$.

Chứng minh L6: $(X \cap Y)_E = (X)_E \cap (Y)_E$

Vì $X \cap Y \subseteq X$ và $X \cap Y \subseteq Y$ nên theo tính đơn điệu L4 ta có:

$(X \cap Y)_E \subseteq (X)_E$; $(X \cap Y)_E \subseteq (Y)_E$.

Từ đây suy ra ta được một vế $(X \cap Y)_E \subseteq (X)_E \cap (Y)_E$.

Ta chứng minh vế còn lại $(X)^E \cap (Y)^E \subseteq (X \cap Y)^E$

Lấy $x \in (X)^E \cap (Y)^E$ suy ra x thuộc E_i mà E_i có phần tử chung với X và Y như vậy E_i có phần tử chung với $X \cap Y$ nên $x \in (X \cap Y)^E$

Chứng minh U6: $(X \cap Y)^E \subseteq (X)^E \cap (Y)^E$. Tính chất này suy trực tiếp từ tính chất U4.

Các tính chất L7, L8, L9, U7, U8, U9 suy trực tiếp từ định nghĩa và kết quả trong bổ đề 3.1.

Bổ đề 3.3

Từ các tính chất U3, U4, U8 ta thấy ánh xạ $\phi: 2^U \to 2^U$ với $\phi(X) = X^E$ là ánh xạ đóng và ta có đẳng thức $(X \cup Y)^E = (X^E \cup Y)^E = (X \cup Y^E)^E = (X^E \cup Y^E)^E$.

3.1.2. Định nghĩa tập thô, tập rõ theo xấp xỉ

Cho không gian xấp xỉ Apr = (U,E); $X \subseteq U$.

Định nghĩa 3.2 Định nghĩa *tập thô, tập rõ* theo xấp xỉ

Tập $X \in 2^U$ được gọi là *thô* trong không gian Apr = (U, E) (hay X là *thô* ứng với phân hoạch E) nếu $X^E \neq X_E$.

Tập $X \in 2^U$ được gọi là *rõ* trong không gian Apr = (U, E) (hay X là *rõ* ứng với phân hoạch E) nếu $X^E = X_E$.

Hoặc:

Tập X được gọi là *thô* trong Apr = (U, E) nếu $\alpha_X^E < 1$.

Tập X được gọi là *rõ* trong Apr = (U, E) nếu $\alpha_X^E = 1$.

Quay lại ví dụ 3.3.b ta thấy X là tập rõ ứng với phân hoạch E, Y là tập thô ứng với phân hoạch E.

Lưu ý:

- Trong định nghĩa tập thô X luôn có cụm từ ứng với phân hoạch E. Vì E là phân hoạch của U trong không gian Apr = (U, E) nên đôi khi thay cho nói X thô ứng với phân hoạch E, ta nói X thô trong không gian Apr. Ta luôn nhớ rằng bản thân các tập X, X_E, X^E là những tập con của U.

- Khái niệm tập thô luôn gắn với phân hoạch E. Cho trước hai không gian Apr1 = (U, E) và Apr2 = (U, P), X có thể thô trong không gian này nhưng lại rõ trong không gian kia.

Ví dụ 3.2.a:

Cho Apr1 = (U, E);

với U = {1, 2, 3, 4, 5, 6, 7, 8, 9};

E = $\{E_1, E_2, E_3, E_4\}$ = {{1, 2, 3}, {4, 5}, {6, 7}, {8, 9}}.

Apr2 = (U, P); với P = $\{p_1, p_2, p_3, p_4\}$ = {{1, 4}, {2, 3}, {5, 6, 7}, {8, 9}}.

Lấy X, Y là các tập con của U và X = {1, 2, 3}; Y = {1, 4} thì như đã xét ở trên X là rõ trong Apr1, Y là thô trong Ap1.

Trong không gian Apr2 ta có $X^P = \{1,4\} \cup \{2,3\} = \{1, 2, 3, 4\}$ và $X_P = \{2, 3\}$ vậy X thô trong Apr2 và $Y^P = \{1,4\} = Y_P$ nên Y tập rõ trong Apr2.

Ví dụ 3.2.b:

Xét tập 7 sinh viên có các thuộc tính năm sinh, quê như bảng sau:

Masv	NS	QUE
1	82	HN
2	82	HN
3	82	HP
4	82	HP
5	84	HP
6	84	HT
7	84	HT

Xét không gian Apr1 = (U, E); U = {1, 2, 3, 4, 5, 6, 7} và theo tiêu chí cùng năm sinh thì U được chia thành hai nhóm:

E = $\{E_1, E_2\}$ = {{1, 2, 3, 4}, {5, 6, 7}}. Theo tiêu chí cùng quê thì U được phân hoạch thành 3 nhóm P = $[p_1, p_2, p_3\}$ = {{1, 2}, {3, 4, 5}, {6, 7}}.

Bây giờ lấy tập X = {1, 2}. Khi đó theo phân hoạch P ta có $X^P = X_P$ nên X rõ ứng với P.

Ngược lại theo phân hoạch E thì $X^E = \{1, 2, 3, 4\}$ và $X_E = \phi$ vậy $X^E \neq X_E$ nên X thô ứng với E.

Trong chương 1 ta đã biết rằng mọi phân hoạch $E = \{E_1,...,E_k\}$ luôn có quan hệ tương đương R để U/R = E và ngược lại với mọi quan hệ tương đương R trên U thì U/R là phân hoạch E nên các phần sau thay cho E đôi khi ta có thể viết R.

Ví dụ xét quan hệ R là quan hệ đồng dư khi chia các phần tử của U = {1, 2, 3, 4, 5, 6, 7, 8, 9} cho 3. Khi đó $U/R = \{E_1, E_2, E_3\} = \{3, 6, 9\}, \{1, 4, 7\}, \{2, 5, 8\}\}\}$.

Lấy X = {1, 2, 3}.

Ta có: $X^R = X^E = \{1, 4, 7\} \cup \{2, 5, 8,\} \cup \{3, 6, 9\} = U$.

$X_R = X_E = \phi$. Vậy X là tập thô ứng với E (hay ứng với R).

Lấy $Y = E_1 = \{1, 4, 7\}$.

Ta có: $Y^R = Y_R = E_1$. Vậy Y là tập rõ ứng với E (hay ứng với R).

Lấy $Z = E_1 \cup E_2 = \{1, 4, 7\} \cup \{2, 5, 8,\}$.

Ta cũng có $Z^R = Z_R = \{1, 4, 7\} \cup \{2, 5, 8,\}$, Z cũng là tập rõ.

3.1.3. Định nghĩa tập thô, tập rõ theo tập hợp

Cho không gian Apr = (U, E); $E = \{E_1, E_2,..., E_k\}$ là phân hoạch của U thường được gọi là họ các tập *sơ cấp*.

Ký hiệu 2^U là họ các tập con của U.

Từ các tính chất của xấp xỉ ở trên ta dễ dàng thấy rằng trong họ các tập con của U thì tập rỗng ϕ, hợp của các tập sơ cấp E_i là những tập rõ. Những tập X không biểu diễn được qua hợp của các tập sơ cấp là những tập thô.

Cho Apr = (U, E).

Đặt $RO = \{\phi, E_1, E_2,..., E_k, \cup E_i, U\}$, trong đó $\cup E_i$ là hợp của các nhóm E_i nào đó.

Ta dễ dàng thấy rằng $(2^U, \cup, \cap)$ với 2^U là họ các tập con của U là một đại số và $(RO, \cup, \cap)$ cũng là một đại số.

Nói chung RO là họ con thực sự của 2^U. Ví dụ xét không gian xấp xỉ

Apr = (U, E) = ({1,2,3}, E) và $E_1 = \{1,2\}$; $E_2 = \{3\}$.

Khi đó:

$RO = \{\phi, \{1,2\}, \{3\}, \{1,2,3\}\}$.

$2^U = \{\phi, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$.

2^U - RO = $\{\{1\}, \{2\}, \{1,3\}, \{2,3\}\}$.

Cho không gian Apr = (U, E). Trong họ 2^U *chỉ có các tập rõ và các tập thô* ứng với phân hoạch E.

Đặt RO = $\{\phi, E_1, E_2, \ldots, E_k, \cup E_i, U\}$, trong đó $\cup E_i$ là hợp của các nhóm E_i nào đó.

THO = $2^U \setminus$ RO

Định nghĩa 3.3 Định nghĩa tập thô, tập rõ theo tập hợp

Cho không gian Apr = (U, E)

$X \in 2^U$ là tập *rõ* trong Apr = (U, E) nếu $X \in \{\phi, E_1, E_2, \ldots, E_k, \cup E_i, U\}$ = RO

$X \in 2^U$ là tập *thô* trong Apr = (U, E) nếu $X \in$ THO = $2^U \setminus$ RO

Trong đó $\cup E_i$ là hợp của một số nhóm E_i nào đó.

Bổ đề 3.4

Cho không gian Apr = (U, E); với E có k nhóm E = $\{E_1, E_2, \ldots, E_k\}$.

Khi đó số các tập rõ trong không gian nền Apr = (U, E) là card(RO) = 2^k. Số các tập thô trong không gian nền Apr = (U, E) là card(THO) = card(2^U) - 2^k.

Chứng minh:

Ta có họ các tập rõ trong Apr là:

RO = $\{\phi, E_1, E_2, \ldots, E_k, E_1 \cup E_2, E_1 \cup E_3, \ldots, U\}$.

Số tập không phần tử ϕ là $C_n^0 = 1$, số tập gồm một tập $E_1, E_2, \ldots, E_k$ là C_n^1, số tập là hợp hai tập là $C_n^2, \ldots$ nên Card(RO) = $\sum_{i=0}^{k} C_n^i = 2^k$.

Trong đó C_n^i là tổ chập i của n.

Lưu ý:

(RO, $\cup$, $\cap$) là một đại số

(THO, $\cup$, $\cap$) không là một đại số vì $\cup$, $\cap$ của hai tập trong THO có thể không thuộc THO. Nói cách khác hợp, giao của hai tập thô có thể thô hoặc rõ.

RO là một giàn giao trong U.

Nếu X, Y là những tập rõ thì $X \cap Y$, $X \cup Y$, $X \setminus Y$ là những tập rõ

Nếu X, Y là những tập thô thì $X \cap Y$, $X \cup Y$, $X \setminus Y$ là những tập thô hoặc rõ

3.1.4. Định nghĩa tập thô, tập rõ theo hàm thuộc thô

Cho không gian Apr = (U, E); trong đó E = $\{E_1, E_2, ..., E_k\}$ là phân hoạch của U. Ký hiệu E_o là nhóm của E chứa đối tượng o.

Bây giờ chúng ta sẽ dùng hàm thuộc thô để định nghĩa tập thô, tập rõ.

Giả sử $X \subseteq U$.

Xét hàm $\mu_X^E : U \to [0,1]$.

Định nghĩa 3.4 Định nghĩa hàm thuộc thô

Cho không gian Apr = (U, E); $X \in 2^U$.

Ứng với X và E ta xây dựng hàm thuộc thô $\mu_X^E : U \to [0, 1]$ với $\forall o \in U$

$$\mu_X^E(o) = \frac{card(X \cap E_o)}{card(E_o)}$$

Trong đó card(E_o) là số phần tử của nhóm chứa o.

Nhận xét rằng với mọi $o \in U$ thì E_o có thể nằm gọn trong X

khi đó $E_o \cap X = E_o$ và $\mu_X^E(u) = 1$ với mọi u thuộc E_o, hoặc E_o có phần chung với X nhưng không nằm trong X khi đó $E_o \cap X$ là tập con thực sự của E_o và do đó $0 < \mu_X^E(u) < 1$ với mọi u thuộc E_o, hoặc E_o nằm ngoài X. Khi đó $E_o \cap X = \phi$ và $\mu_X^E(u) = 0$ với mọi u thuộc E_o.

Ví dụ U = {1, 2, 3, 4, 5, 6, 7, 8, 9}; E = {{1, 2, 3}, {4, 5}, {6, 7, 8, 9}}.

Giả sử X = {1, 4, 5}. Khi đó:

$\mu_X^E(1) = Card(E_1 \cap X)/ Card(E_1) = \mu_X^E(2) = \mu_X^E(3) = 1/3,$

$\mu_X^E(4) = \mu_X^E(5) = 1, \mu_X^E(6) = \mu_X^E(7) = \mu_X^E(8) = \mu_X^E(9) = 0;$

Giá trị $\mu_X^E(u)$ ta sẽ gọi là *giá trị thuộc thô của u* ứng với X và E.

Vậy giá trị của hàm thuộc thô $\mu_X^E(u)$ trên mọi nhóm E_i nằm gọn trong X hay trên X_E và là 0 trên mọi nhóm E_j nằm ngoài X hay trên $U \setminus X^E$ và $\mu_X^E(u)$ có thể có giá trị giữa 0 và 1 trên các nhóm có cả phần trong X và phần ngoài X hay trên BIEN(X) = $X^E \setminus X_E$.

Vậy tập Values = $\{\mu_X^E(a) : a \in U\}$ có thể dùng để xét tập X là thô hay rõ ứng với phân hoạch E.

Ví dụ 3.3:

Xét tập X như trong hình 3.1: khi đó μ_X^E có giá trị 0 trong các nhóm 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,12, 13, 14, 15, 16. μ_X^E có giá trị 1 trong hai nhóm kẻ ô vuông nhỏ và μ_X^E có giá trị giữa 0 và 1 trong các nhóm a, b, c, d, e, f, g, h, i, k.

Ví dụ 3.4:

Lấy U = {1, 2, 3, 4, 5, 6, 7, 8, 9}. R là quan hệ đồng dư khi chia cho 3.

U/R = {E_1. E_2, E_3} = {{1, 4, 7}, {5, 8, 2}, {3, 6, 9}}.

Lấy X = {1, 2}

Khi đó ta có Values = {1/3, 1/3, 0, 1/3, 1/3, 0, 1/3, 1/3, 0}.

Tập X là thô ứng với phân hoạch E = U/R.

Nếu lấy X = {1, 4, 7}. Khi đó tập Values = {1, 0} nên X là tập rõ ứng với E.

Bổ đề 3.5

Cho không gian Apr = (U, E). X, Y $\subseteq$ U.

Từ định nghĩa hàm thuộc thô ta luôn có:

1. $\mu_X^E(a) = 1$ nếu $a \in X_E$
2. $\mu_X^E(a) = 0$ nếu $a \in U - X^E$
3. $\mu_{(X \cup Y)}{}^E(a) \geq \max(\mu_X^E(a), \mu_X^E(a))$, với mọi $a \in U$
4. $\mu_{(X \cap Y)}{}^E(a) \leq \min(\mu_X^E(a), \mu_X^E(a))$, với mọi $a \in U$
5. $0 < \mu_X^E(a)$ với mọi $a \in X^E - X_E$
6. $\mu^E_{-X}(a) = 1 - \mu_X^E(a)$, với mọi $a \in U$

Định nghĩa 3.5 Định nghĩa tập thô, tập rõ theo hàm thuộc thô

Cho không gian Apr = (U, E)

$X \in 2^U$ là tập *rõ* trong Apr = (U, E) nếu tập giá trị của hàm thuộc thô μ_X^E *không chứa* giá trị khác 0 và 1.

$X \in 2^U$ là tập *thô* trong Apr = (U, E) nếu tập giá trị của hàm thuộc thô μ_X^E *có chứa* giá trị khác 0 và khác 1.

3.1.5. Sự tương đương của ba định nghĩa tập thô, tập rõ

Chúng ta dễ dàng chứng minh rằng các định nghĩa 3.2, 3.3, 3.5 về tập thô, tập rõ trong các mục 3.1.2, 3.1.3, 3.1.4 là tương đương. Để làm sáng tỏ điều này ta chứng minh định lý cả ba cách tiếp cận tập thô, tập rõ ở trên đều cho chúng ta cùng họ các tập rõ, tập thô ứng với E như nhau.

Định lý 3.1

Cho không gian Apr = (U, E); với $E = \{E_1, E_2, ..., E_k\}$ là phân hoạch của U. Ba định nghĩa 3.2, 3.3, 3.5 là tương đương theo nghĩa nếu $X \subseteq U$ là thô hay rõ trong định nghĩa này thì nó cũng là thô hay rõ tương ứng trong hai định nghĩa còn lại.

Để chứng minh định lý 3.1 ta sẽ chứng minh hai bổ đề 3.6 và 3.7.

Bổ đề 3.6 Cho không gian Apr = (U, E); với $E = \{E_1, E_2, ..., E_k\}$

Hai định nghĩa sau là tương đương

(1) Tập $X \in 2^U$ là tập *rõ* trong Apr = (U, E) nếu $X^E = X_E$

Tập $X \in 2^U$ là tập *thô* trong Apr = (U, E) nếu $X^E \neq X_E$

(2) Tập $X \in 2^U$ là tập *rõ* trong Apr = (U, E) nếu $X \in \{\phi, E_1, E_2, ..., E_k, \cup E_i, U\}$

Tập $X \in 2^U$ là tập *thô* trong Apr = (U, E) nếu $X \notin \{\phi, E_1, E_2, ..., E_k, \cup E_i, U\}$.

Chứng minh

(1) => (2)

Lấy $X \in 2^U$ là *tập rõ*, tức $X^E = X_E$.

Khi đó nếu $X = \phi$ thì $\phi^E = \phi = \phi_E$ nên $X \in \{\phi, E_1, E_2, ..., E_k, \cup E_i, U\}$.

Nếu $X = E_i$ thì $X^E = X_E$ nên $X \in \{\phi, E_1, E_2, ..., E_k, \cup E_i, U\}$

Nếu $X = \cup E_i$ thì $X^E = X_E$ nên $X \in \{\phi, E_1, E_2, ..., E_k, \cup E_i, U\}$.

Nếu $X = U$ thì $X^E = X_E$ nên $X \in \{\phi, E_1, E_2, ..., E_k, \cup E_i, U\}$.

Vậy X là tập rõ ứng với E theo định nghĩa (2)

Lấy $X \in 2^U$ là *tập thô*, tức $X^E \neq X_E$

Khi đó chắc chắn $X \notin \{\phi, E_1, E_2, ..., E_k, \cup E_i, U\}$ vì nếu X thuộc một trong các tập rõ $\{\phi, E_1, E_2, ..., E_k, \cup E_i, U\}$ thì $X^E = X_E$. Vậy X là tập thô ứng với E theo định nghĩa (2).

(2) => (1)

Lấy $X \in 2^U$ là *tập rõ* ứng với E theo định nghĩa (2),

tức $X \in \{\phi, E_1, E_2,..., E_k, \cup E_i, U\}$. Dễ dàng thấy rằng tất cả các tập X như vậy luôn có $X^E = X_E$ nên X là tập rõ theo định nghĩa (1).

Lấy $X \in 2^U$ là *tập thô* ứng với E theo định nghĩa (2),

tức $X \notin \{\phi, E_1, E_2,..., E_k, \cup E_i, U\}$. Vì $X \notin \{\phi, E_1, E_2,..., E_k, \cup E_i, U\}$ nên $X \neq \phi$ và X phải chứa ít nhất một tập con thực sự của một E_i nào đó, nếu không X là hợp của một số nhóm nào đó. Khi đó xấp xỉ trên X^E của X có chứa E_i và xấp xỉ dưới X_E của X không chứa E_i. Vậy $X^E \neq X_E$ nên X thô theo định nghĩa (1).

Ví dụ 3.5:

Cho Apr = (U, P)

với $U = \{1, 2, 3, 4, 5\}$; $P = \{p_1, p_2, p_3\} = \{\{1, 2\}, \{3, 4\}, \{5\}\}$.

Khi đó ta dễ dàng liệt kê các tập thô, tập rõ ứng với P

$RO = \{\phi, p_1, p_2, p_3, \cup p_i, U\} = \{\phi, \{1, 2\}, \{3, 4\}, \{5\}, \{1, 2, 3, 4\}, \{1, 2, 5\}, \{3, 4, 5\}, \{1, 2, 3, 4, 5\}\}$.

$2^U = \{\phi, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \{2, 3\}, \{2, 4\}, \{3, 4\}, \{3, 5\}, \{4, 5\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 5\}, \{1, 3, 4\}, \{1, 4, 5\}, \{2, 3, 4\}, \{2, 3, 5\}, \{2, 4, 5\}, \{3, 4, 5\}, \{1, 3, 5\}, \{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \{1, 2, 4, 5\}, \{1, 3, 4, 5\}, \{2, 3, 4, 5\}, \{1, 2, 3, 4, 5\}$.

$THO = 2^U \setminus RO = \{\{1\}, \{2\}, \{3\}, \{4\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \{2, 3\}, \{2, 4\}, \{3, 5\}, \{4, 5\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}, \{1, 4, 5\}, \{2, 3, 4\}, \{2, 3, 5\}, \{2, 4, 5\}, \{1, 3, 5\}, \{1, 2, 3, 5\}, \{1, 2, 4, 5\}, \{1, 3, 4, 5\}, \{2, 3, 4, 5\}\}$

Bổ đề 3.7

Cho không gian Apr = (U, E); với $E = \{E_1, E_2,..., E_k\}$

Hai định nghĩa sau là tương đương

(1) Tập $X \in 2^U$ là tập *rõ* trong Apr = (U, E) nếu $X^E = X_E$

Tập $X \in 2^U$ là tập *thô* trong Apr = (U, E) nếu $X^E \neq X_E$

(3) Tập $X \in 2^U$ là tập *rõ* trong Apr = (U, E) nếu tập giá trị của hàm thuộc thô $\mu_X^E(o)$ không chứa giá trị khác 0 và 1

Tập $X \in 2^U$ là tập *thô* trong Apr = (U, E) nếu tập giá trị của hàm thuộc thô μ_X^E có chứa giá trị khác 0 và khác 1.

Chứng minh

(1) => (3)

Lấy $X \in 2^U$ là *tập rõ*, tức $X^E = X_E$.

Khi đó nếu $X^E \setminus X_E = \phi$ tức vùng biên giữa X^E và X_E là ϕ, BIEN(X) = $X^E \setminus X_E = \phi$, với mọi $o \in X_P$ ta có $p_o \cap X = p_o$ nên $\mu_x^E(o) = 1$ và $o \notin X^P$ ta có $p_o \cap X = \phi$ nên $\mu_x^E(o) = 0$ với mọi o nằm ngoài $X^E = X_E$. Vậy tập giá trị $\mu_x^E(o)$ của hàm thuộc thô không chứa giá trị khác 0 và 1 nên X rõ theo định nghĩa (3).

Lấy $X \in 2^U$ là *tập thô*, tức $X^E \neq X_E$

Khi đó vùng biên giữa X^E và X_E khác ϕ, tức BIEN(X) = $X^E \setminus X_E \neq \phi$. Ta dễ dàng thấy rằng với mọi $o \in X_E$ thì $E_o \cap X = E_o$ nên $\mu_x^E(o) = 1$. Với mọi $o \notin X^E$ ta có $E_o \cap X = \phi$ nên $\mu_x^E(o) = 0$. Với mọi $o \in$ BIEN(X) thì $E_o \cap X$ khác rỗng và là tập con thực sự của E_o nên giá trị khác 0 và khác 1. Vậy X là tập thô ứng với E theo định nghĩa (3).

(3) => (1)

Lấy $X \in 2^U$ là *tập rõ* ứng với E theo định nghĩa (3), tức $\mu_x^E(o)$ không chứa giá trị khác 0 và 1. Tức là $E_o \cap X$ bằng E_o hoặc bằng ϕ. Mà $E_o \cap X = E_o$ chỉ khi $E_o \subseteq X$, hay $E_o \subseteq X_E$. $E_o \cap X = \phi$ chỉ khi E_o nằm ngoài X^E. Như vậy giữa X_E và X^E hàm thuộc thô không có giá trị nào nên BIEN(X) = $X^E \setminus X_E = \phi$. Vậy $X^E = X_E$ nên X là tập rõ theo định nghĩa (1).

Lấy $X \in 2^U$ là *tập thô* ứng với P theo định nghĩa (3), tức $\mu_x^E(o)$ có chứa giá trị khác 0 và 1. Ta dễ dàng thấy rằng với mọi o mà $E_o \subseteq X_E$ thì $E_o \cap X = E_o$ nên $\mu_x^E(o) = 1$. Với mọi o mà E_o nằm ngoài X^P thì $E_o \cap X = \phi$ nên $\mu_x^E(o) = 0$. Vậy $\mu_x^E(o)$ có giá trị khác 0 và 1 chỉ có thể có với những o trên vùng BIEN(X) = $X^E \setminus X_E$. Vậy BIEN(X) phải khác rỗng hay $X^E \neq X_E$. X là thô theo định nghĩa (1).

Lưu ý:

- Cho không gian Apr = (U, E). Thay cho nói tập $X \in 2^U$ là tập thô (tập rõ) ứng với E, đôi khi ta nói X là tập thô (tập rõ) trong không gian Apr.

- Khi cho trước không gian Apr = (U, E) và $X \in 2^U$ ta nói X là tập thô (tập rõ) có nghĩa X là tập thô (tập rõ) trong không gian Apr = (U, E).

- Như vậy khi ta nói X là tập thô (tập rõ) hàm chỉ trong một không gian Apr = (U, E).

Ví dụ 3.6:

Cho Apr = (U, E)

với U = {1, 2, 3, 4, 5}; E = {E_1, E_2, E_3} = {{1, 2}, {3, 4}, {5}}.

Khi đó ta dễ dàng liệt kê các tập thô, tập rõ ứng với P theo định nghĩa hàm thuộc thô $\mu_X^E(o) = \dfrac{card(X \cap E_o)}{card(E_o)}$

RO = {ϕ, {1, 2}, {3, 4}, {5}, {1, 2, 3, 4}, {1, 2, 5}, {3, 4, 5}, {1, 2, 3, 4, 5}}.

2^U = {ϕ, {1}, {2}, {3}, {4}, {5}, {1, 2}, {1, 3}, {1, 4}, {1, 5}, {2, 3}, {2, 4}, {3, 4}, {3, 5}, {4, 5}, {1, 2, 3}, {1, 2, 4}, {1, 2, 5}, {1, 3, 4}, {1, 4, 5}, {2, 3, 4}, {2, 3, 5}, {2, 4, 5}, {3, 4, 5}, {1, 3, 5}, {1, 2, 3, 4}, {1, 2, 3, 5}, {1, 2, 4, 5}, {1, 3, 4, 5}, {2, 3, 4, 5}, {1, 2, 3, 4, 5}.

THO = {{1}, {2}, {3}, {4}, {1, 3}, {1, 4}, {1, 5}, {2, 3}, {2, 4}, {3, 5}, {4, 5}, {1, 2, 3}, {1, 2, 4}, {1, 3, 4}, {1, 4, 5}, {2, 3, 4}, {2, 3, 5}, {2, 4, 5}, {1, 3, 5}, {1, 2, 3, 5}, {1, 2, 4, 5}, {1, 3, 4, 5}, {2, 3, 4, 5}}

3.2. CÁC PHÉP TOÁN TẬP HỢP TRÊN CÁC TẬP THÔ, TẬP RÕ

3.2.1. Các phép toán tập hợp trên các tập rõ

Cho không gian Apr = (U, E) và X, Y $\subseteq$ U.

Trong phần này ta sẽ xét bài toán nếu X, Y là những tập rõ trong không gian Apr thì hợp, giao, hiệu,... của X, Y có là tập rõ không?

Bổ đề 3.8

Cho không gian Apr = (U, E); X $\subseteq$ U.

X là tập rõ khi và chỉ khi X = ϕ hoặc X = $\cup$ E_i

Chứng minh:

Theo định nghĩa của họ các tập rõ RO thì X = ϕ hoặc X = $\cup$ E_i là tập rõ. Ngược lại nếu X là tập rõ thì X phải là tập rỗng hoặc X là hợp của một số nhóm E_i nào đó, tức X = $\cup$ E_i nếu không X chứa tập con thực sự của một nhóm E_j nào đó và X thô.

Bổ đề 3.9

Cho không gian Apr = (U, E); X, Y $\subseteq$ U.

a. Nếu X, Y $\in$ RO thì $X \cup Y$ là tập rõ, tức $X \cup Y \in$ RO

b. Nếu X, Y $\in$ RO thì $X \cap Y$ là tập rõ, tức $X \cap Y \in$ RO.

c. Nếu X, Y $\in$ RO thì X \ Y là tập rõ, tức X \ Y $\in$ RO.

d. Nếu X $\in$ RO thì -X (phần bù của X) là tập rõ, tức -X $\in$ RO.

Chứng minh:

Theo bổ đề 3.7 vì X, Y là những tập rõ nên X = $\cup$ E_i và Y = $\cup$ E_j nên

a. Hợp X $\cup$ Y = $\cup$ E_i $\cup$ E_j là tập rõ.

b. Giao X $\cap$ Y = $\cup$ E_i $\cap$ $\cup$ E_j = ϕ hoặc bằng hợp của các nhóm nào đó trong E cũng là tập rõ.

c. Hiệu X \ Y = $\cup$ E_i \ $\cup$ E_j = ϕ hoặc bằng hợp của các nhóm nào đó trong E cũng là tập rõ.

d. Phần bù của X là - X = U \ X = U \ $\cup$ E_i = $\cup$ E_j là tập rõ.

Lưu ý:

- Hợp, giao, hiệu, phần bù của các tập rõ là một tập rõ.
- Tập con thực sự của tập rõ có thể là tập rõ hoặc tập thô.

Ví dụ 3.7:

Xét không gian xấp xỉ Apr = (U,E) = ({1,2,3}, E) = {{1,2,3},{{1,2},{3}}.

Khi đó:

RO = {ϕ, {1,2}, {3}, {1,2,3}} là họ các tập rõ.

2^U = {ϕ,{1},{2},{3},{1,2},{1,3},{2,3},{1,2,3}}.

THO = 2^U - RO = {{1},{2},{1,3},{2,3}} là họ các tập thô.

Lấy X = {1, 2}, Y = {3} là những tập rõ.

Khi đó: X $\cup$ Y = {1, 2, 3} = U là tập rõ. X $\cap$ Y = ϕ là tập rõ.

Hiệu X \ Y = X, phần bù -X = U\X = {3} = Y là những tập rõ.

Xét tập rõ Z = X $\cup$ Y = {1, 2, 3} = U. Khi đó tập con thực sự {1} của Z là tập thô. Tập con thực sự {3} của Z là tập rõ.

3.2.2. Các phép toán tập hợp trên các tập thô

Cho không gian Apr = (U, E) và X, Y $\subseteq$ U.

Trong phần này ta sẽ xét bài toán nếu X, Y là những tập thô trong không gian Apr thì hợp, giao, hiệu ... của X, Y có là tập thô không?

Bổ đề 3.10

Cho không gian Apr = (U, E) và X $\subseteq$ U.

X là tập thô khi và chỉ khi X chứa tập con thực sự (khác rỗng, khác E_i) của một nhóm E_i nào đó.

Lưu ý:

Ta nói X chứa tập con thực sự của E_i có nghĩa là X chỉ chứa tập con thực sự khác rỗng của E_i, để tránh trường hợp hiểu nhầm nếu X chứa E_i thì X chứa mọi tập con của E_i.

Chứng minh:

Giả sử X là tập thô khi đó X phải chứa tập con thực sự khác rỗng của một nhóm E_j nào đó, nếu không $X = \cup E_i$ là tập rõ.

Bây giờ giả sử X chứa Y là tập con thực sự khác rỗng của E_j khi đó $Z = E_j \setminus Y$ là tập con thực sự khác rỗng của E_j. Do đó xấp xỉ trên X^E của X có chứa E_j; còn xấp xỉ dưới X_E của X không chứa E_j nên $X^E \neq X_E$, vậy X là tập thô.

Bổ đề 3.11

Cho không gian Apr = (U, E); X, Y $\subseteq$ U.

a. Nếu X, Y $\in$ THO thì $X \cup Y$ là tập thô hoặc tập rõ.

b. Nếu X, Y $\in$ THO thì $X \cap Y$ là tập thô hoặc tập rõ.

c. Nếu X, Y $\in$ THO thì $X \setminus Y$ là tập thô hoặc tập rõ.

d. Nếu X $\in$ THO thì -X (phần bù của X) là tập thô.

Chứng minh:

d. Nếu X là thô thì X chứa tập con thực sự của nhóm E_i và phần còn lại của E_i thuộc -X. Vậy -X chứa tập con thực sự của E_i nên -X là tập thô.

Thay cho chứng minh các kết luận a, b, c bổ đề 3.10 ta xét ví dụ minh họa.

Ví dụ 3.8:

Xét không gian xấp xỉ Apr = (U,E) = ({1,2,3}, E) = {{1,2,3},{{1,2},{3}}.

Khi đó:

RO = {ϕ, {1,2}, {3}, {1,2,3}} là họ các tập rõ.

2^U = {ϕ,{1},{2},{3},{1,2},{1,3},{2,3},{1,2,3}}.

THO = 2^U - RO = {{1},{2},{1,3},{2,3}} là họ các tập thô.

Lấy X = {1, 3}, Y = {2, 3} là những tập thô.

Khi đó:

$X \cup Y$ = {1, 2, 3} = U là tập rõ.

$X \cap Y = \{3\}$ là tập rõ.

Hiệu $X \setminus Y = \{1\}$ là tập thô, phần bù $-X = U \setminus X = \{2\}$ là tập thô.

Nếu lấy $X' = \{1\}$, $Y' = \{1, 3\}$ là những tập thô.

Khi đó:

$X' \cup Y' = \{1, 3\}$ là tập thô. $X \cap Y = \{1\}$ là tập thô.

Hiệu $X' \setminus Y' = \{\}$ là tập rõ.

Các phần bù $-X' = \{2,3\}$; $-Y' = \{2\}$ là những tập thô.

Nếu $X, Y \in$ THO thì $X \cup Y$, $X \cap Y$, $X - Y$ có thể thô hoặc rõ. Hợp, giao, hiệu của các tập thô có thể là tập thô hoặc tập rõ.

Lưu ý:

- Các tập X, Y thô không nhất thiết $X \cup Y$, giao $X \cap Y$, hiệu $X - Y$ phải thô.
- Phần bù -X của tập thô X là tập thô.
- Tập con thực sự của một tập thô có thể là tập thô hoặc tập rõ.

Ví dụ 3.9:

Xét không gian xấp xỉ $Apr = (U,E) = (\{1,2,3\}, E) = \{\{1,2,3\},\{\{1,2\},\{3\}\}$.

Xét các tập thô X, Y, Z với $X = \{1,3\}$, $Y = \{2\}$, $Z = \{2,3\}$.

Khi đó: $X \cup Y$ là tập rõ, $Y \cup Z$ là tập thô.

$X \cap Y = \phi$ là tập rõ, $Y \cap Z = \{2\}$ là tập thô.

$Z \setminus Y = \{3\}$ là tập rõ, $Y \setminus X = \{2\}$ là tập thô.

Phần bù $-X = U \setminus X = \{2\}$; $-Y = U \setminus Y = \{1, 3\}$; $-Z = U \setminus Z = \{1\}$ là những tập thô.

Bổ đề 3.12

Cho không gian xấp xỉ $Apr = (U, E)$; với $E = \{E_1, E_2, \dots E_k\}$. Khi đó họ các tập rõ: $RO = \{\phi, E_1, E_2, \dots, E_k, \cup E_i, U\}$ trong Apr là một topo.

3.3. KHÔNG GIAN NỀN VÀ ỨNG DỤNG CỦA NÓ

3.3.1. Khái niệm không gian nền

Cho hai không gian nền

$Apr1 = (U, E)$

$Apr2 = (U, P)$;

Với $U = \{o_1, o_2, ..., o_m\}$ là tập đối tượng, $E = \{E_1, E_2, ..., E_k\}$, $P = \{p_1, p_2, ..., p_l\}$ là các phân hoạch của U.

Lưu ý: Trong phần này ta chỉ xét các không gian nền cùng tập đối tượng U.

Định nghĩa 3.6

Ta nói Apr1 = (U, E) *mịn hơn* Apr2 = (U, P) (hay Apr2 *thô hơn* Apr1) nếu số tập rõ trong Apr1 lớn hơn số tập rõ trong Apr2.

Ví dụ 3.10

Cho $U = \{1, 2, 3, 4, 5\}$; $E = \{E_1, E_2, E_3\} = \{\{1\}, \{2, 3, 4\}, \{5\}\}$.

$P = \{p_1, p_2\} = \{\{1, 2, 3, 4\}, \{5\}\}$.

Khi đó số tập rõ trong Apr1 là 8, số tập rõ trong Apr2 là 4. Vậy Apr1 = (U, E) mịn hơn Apr2 = (U, P).

Ví dụ 3.11

Xét hệ tin dạng quan hệ gồm 10 người có 2 thuộc tính như sau:

U	Tỉnh	Huyện
T_1	Nghệ An	Nam Đàn
T_2	Nghệ An	Nam Đàn
T_3	Nghệ An	Thanh Chương
T_4	Nghệ An	Thanh Chương
T_5	Hà Tĩnh	Thạch Hà
T_6	Hà Tĩnh	Đức Thọ
T_7	Hà Tĩnh	Đức Thọ
T_8	Hà Nội	Từ Liêm
T_9	Hà Nội	Từ Liêm
T_{10}	Hà Nội	Sóc Sơn

$U = \{T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_9, T_{10}\}$

Phân hoạch U theo Tỉnh ta có Apr1 = (U, E)

$E = \{E_1, E_2, E_3\}$

$E_1 = \{T_1, T_2, T_3, T_4\}$

$E_2 = \{T_5, T_6, T_7\}$

$E_3 = \{T_8, T_9, T_{10}\}$

Bây giờ phân hoạch U theo Huyện ta có Apr2 = (U, P)

$P = \{p_1, p_2, p_3, p_4, p_5, p_6\}$

$p_1 = \{T_1, T_2\}$

$p_2 = \{T_3, T_4\}$

$p_3 = \{T_5\}$

$p_4 = \{T_6, T_7\}$

$p_5 = \{T_8, T_9\}$

$p_6 = \{T_{10}\}$

Apr2 mịn hơn Apr1 hay Apr1 thô hơn Apr2.

Lưu ý: Nếu chỉ có hai không gian Apr1 và Apr2 thì không gian mịn hơn ta gọi là không gian mịn, không gian kia gọi là không gian thô.

Định lý 3.2

Cho Apr1 = (U, P); Apr2 = (U, E). Giả sử rằng Apr1 mịn hơn Apr2. Khi đó:

a. Mọi tập $X \in 2^U$ là tập rõ trong Apr2 thì X là tập rõ trong Apr1.

b. Mọi tập $X \in 2^U$ là thô trong Apr1 thì X là thô trong Apr2.

Chứng minh

a. Giả sử X là rõ trong không gian thô Apr2, khi đó xét định nghĩa thô, rõ theo tập hợp ta có X chỉ có thể là tập ϕ, E_i, $\cup E_i$, U. Nếu X = ϕ thì X là rõ trong Apr1. Nếu X = U thì X là rõ trong Apr1. Nếu X = E_i hay X = $\cup E_i$ thì X = $\cup p_j$ vậy X là tập rõ trong Apr1.

b. Giả sử X là tập thô trong không gian mịn Apr1, X phải chứa tập con thực sự của một nhóm p_i nào đó, vì P là phân hoạch con của E, nên $p_i \subseteq E_j$ và như vậy X chứa tập con thực sự của E_j nào đó, X là tập thô trong Apr2 = (U, E).

Lưu ý:

- Qua định lý 3.2 ta thấy tập rõ trong không gian thô là tập rõ trong không gian mịn, tập thô trong không gian mịn là thô trong không gian thô.

- Qua ví dụ 3.10 ta thấy Apr = (U, P) cho ta một cách nhìn tổng thể về tập các đối tượng U. Từ quan điểm này ta có thể coi Apr = (U, P) như là một tri thức của con người đối với tập đối tượng U.

- Ta có tri thức (quan sát được) về U trong Apr1 = (U, P) nhiều hơn trong Apr2 = (U, E) nếu Apr1 mịn hơn Apr2.

Định lý 3.3

Cho hai không gian Apr1 = (U, P) và Apr2 = (U, E). Nếu P << E (tức P là phân hoạch con của E) thì Apr1 mịn hơn Apr2.

Chứng minh định lý 3.3 hoàn toàn suy từ định nghĩa 3.6.

3.3.2. Các ứng dụng không gian nền

Cho không gian nền Apr = (U, E).

Giả sử U là hữu hạn ánh xạ $\varphi: 2^U \to R^+$; với R^+ là tập số thực không âm, thỏa mãn ba điều kiện:

(1) $\varphi(\phi) = 0$

(2) $\varphi(X) > 0$ nếu X khác rỗng

(3) φ đơn điệu, tức là nếu $X \subseteq Y$ thì $\varphi(X) \leq \varphi(Y)$ được gọi là *hàm độ đo* trên U

Ứng với $X \in 2^U$ ta xây dựng hàm thuộc thô $\mu_X^E : U \to [0, 1]$ như sau:

$\mu_X^E(o) = \dfrac{\varphi(X \cap E_o)}{\varphi(E_o)}$; trong đó E_o là nhóm chứa đối tượng o.

Chúng ta dễ dàng chỉ ra rằng X là rõ nếu tập trị của hàm μ_X^E không chứa trị khác 0 và 1. X là tập thô nếu tập trị của hàm μ_X^E có chứa trị khác 0 và khác 1. (Xem bài tập 3.9).

Cho hai không gian nền Apr1 = (U, E); Apr2 = (U, P).

Đặt $h(E_i) = \dfrac{card(\{o \in E_i : \varphi(o) = 1\})}{card(E_i)}$.

Ứng với $X \in 2^U$ ta xây dựng hai hàm thuộc thô $\mu_X^P, \mu_X^E : U \to [0, 1]$ như sau:

$$\mu_X^P(o) = \frac{\varphi(X \cap P_o)}{\varphi(P_o)} \text{ và } \mu_X^E(o) = \frac{\varphi(X \cap E_o)}{\varphi(E_o)};$$

trong đó P_o, E_o là các nhóm chứa đối tượng o.

Định lý 3.4

Cho hai không gian Apr1 = (U, E); Với $E = \{E_1, E_2, ..., E_k\}$

Apr2 = (U, P); Với $P = \{p_1, p_2, ..., p_l\}$

$\varphi: 2^U \to R^+$ là hàm độ đo trên U

Khi đó Apr2 mịn hơn Apr1 nếu tồn tại tập $X \in 2^U$ sao cho:

$$\sum_{i=1}^{l} card(p_i)\left|\mu_X^P(o) - h(p_i)\right| \leq \sum_{i=1}^{k} card(E_i)\left|\mu_X^E(o) - h(E_i)\right|$$

Lưu ý: Việc làm mịn không gian Apr = (U, E) là làm tăng các tập rõ, giảm các tập thô trong không gian Apr = (U, E).

Ví dụ 3.11

Xét lại hệ tin trong ví dụ 10 gồm 10 người có 2 thuộc tính như sau:

U	Tỉnh	Huyện
T_1	Nghệ An	Nam Đàn
T_2	Nghệ An	Nam Đàn
T_3	Nghệ An	Thanh Chương
T_4	Nghệ An	Thanh Chương
T_5	Hà Tĩnh	Thạch Hà
T_6	Hà Tĩnh	Đức Thọ
T_7	Hà Tĩnh	Đức Thọ
T_8	Hà Nội	Từ Liêm
T_9	Hà Nội	Từ Liêm
T_{10}	Hà Nội	Sóc Sơn

$U = \{T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8, T_9, T_{10}\}$

Nếu chỉ phân hoạch U theo Tỉnh ta có Apr1 = (U, E)

$E = \{E_1, E_2, E_3\}$

$E_1 = \{T_1, T_2, T_3, T_4\}$

$E_2 = \{T_5, T_6, T_7\}$

$E_3 = \{T_8, T_9, T_{10}\}$.

Mỗi tập con thực sự của E_1 hay E_2, hay E_3 là những tập thô. Số các tập thô, tập rõ trong Apr1 là card(RO) = card($\{\phi, E_1, E_2, E_3, E_1 \cup E_2, E_1 \cup E_3, E_2 \cup E_3, E_1 \cup E_2 \cup E_3\}$) = 2^3 = 8. Card(THO) = 2^{10} - 8 = 1016 tập thô.

Bây giờ phân hoạch U theo Tỉnh và Huyện ta có không gian mịn hơn Apr2 = (U, P)

$P = \{p_1, p_2, p_3, p_4, p_5, p_6\}$

$p_1 = \{T_1, T_2\}$

$p_2 = \{T_3, T_4\}$

$p_3 = \{T_5\}$

$p_4 = \{T_6, T_7\}$

$p_5 = \{T_8, T_9\}$

$p_6 = \{T_{10}\}$

Các tập p_1, p_2, p_3, p_4, p_5, p_6 là những tập thô trong không gian thô Apr1, nhưng lại rõ trong không gian mịn Apr2.

Số các tập rõ trong Apr2 là:

$$\text{card(RO)} = \text{Card}(\{\phi, p_1, p_2, p_3, p_4, p_5, p_6, p_1 \cup p_2, p_1 \cup p_3, p_1 \cup p_4, p_1 \cup p_5, p_1 \cup p_6, p_2 \cup p_3, p_2 \cup p_4, p_2 \cup p_5, p_2 \cup p_6, \ldots, U\}) = 2^6 = 64.$$

Số tập thô là $2^{10} - 64 = 960$.

Số tập rõ tăng lên và số tập thô giảm xuống.

Nếu tiếp tục làm mịn Apr2 bằng cách phân rã U theo Tỉnh, Huyện, Xã ta sẽ giảm được các tập thô và tăng các tập rõ.

Lưu ý: Một vấn đề được đặt ra là nên làm mịn không gian nền Apr = (U, E) đến mức độ nào? bởi vì muốn làm mịn Apr ta lại phải thêm các thuộc tính vào cho tập đối tượng U. Thêm các thuộc tính cho các đối tượng U lại tốn không gian lưu trữ U. Nhiều vấn đề truy cập trên U lại phức tạp. Hiện tại bài toán ngược đang được nhiều tác giả quan tâm đó là giảm các thuộc tính của U mà vẫn giữ nguyên số các tập thô, tập rõ trong 2^U. Đó là vấn đề rút gọn hệ tin được xét trong chương 4.

3.4. PHỦ VÀ TẬP THÔ

Trong các phần trên chúng ta đã xét các khái niệm liên quan đến tập thô '*kinh điển*'. Tức tập thô trong không gian Apr = (U, P) gắn với phân hoạch P hoặc quan hệ tương đương R.

Chúng ta thấy rằng trong đời sống thực tồn tại nhiều mối quan hệ không tương đương giữa tập các đối tượng thuộc U. Cũng trong thế giới thực có nhiều bài toán, nhiều vấn đề phân loại không thể tách rạch ròi các đối tượng trong U, giữa các nhóm phân loại các đối tượng có những phần tử chung nhau. Ví dụ có nhiều bệnh nhân lúc nhập viện dựa vào các triệu chứng, bệnh nhân có thể điều trị ở khoa nọ hoặc khoa kia.

Trong phần này chúng ta nêu và xét mở rộng khái niệm tập thô khi bỏ điều kiện rời nhau của các nhóm trong phân hoạch P hay bỏ điều kiện bắc cầu của quan hệ tương đương R. Đây là các khái niệm mở rộng để xét các tập thô theo phủ hay tập thô dung sai đã và đang được nhiều tác giả quan tâm trong thời gian gần đây.

3.4.1. Phủ và phân hoạch

a. Phân hoạch

Cho tập đối tượng $O = \{o_1, o_2, \ldots, o_m\}$

Họ các tập con của O, $P = \{p_1, \ldots, p_k\}$ được gọi là phân hoạch của O nếu P thỏa mãn 3 điều kiện:

(1) $p_i \neq \phi$ với mọi i

(2) $p_i \cap p_j = \phi$ với $i \neq j$

(3) $O = \bigcup_{i=1}^{k} p_i$

b. Phủ

Họ các tập con $C = \{C_1, C_2, ..., C_k\}$ của O được gọi là phủ của O nếu:

(i) $C_i \neq \phi$

(ii) $O = \bigcup_{i=1}^{k} C_i$

Như vậy phân hoạch của O là một phủ của O. Mọi tính chất kết quả đúng cho phủ thì đúng cho phân hoạch. Tất nhiên có nhiều tính chất có thể đúng cho phân hoạch mà không đúng cho phủ.

Ví dụ 3.12:

Cho $O = \{1, 2, 3, 4, 5\}$

$C = \{C_1, C_2, C_3\} = \{\{1, 2\}, \{2, 3\}, \{3, 4, 5\}\}$ là một phủ của O.

3.4.2. Tập thô theo phủ

Cho $O = \{o_1, o_2, ... o_m\}$; $C = \{C_1, C_2, ... C_k\}$ là phủ của O

Không gian PHU = (O, C) được gọi là không gian phủ

Giả sử $X \subseteq O$.

Xấp xỉ của tập X

Xấp xỉ trên của X trong không gian PHU = (O, C),

ký hiệu $X^C = \cup\{C_i : C_i \cap X \neq \phi\}$

Xấp xỉ dưới của X trong không gian PHU = (O, C),

ký hiệu $X_C = \cup\{C_i : C_i \subseteq X\}$

Định nghĩa 3.6 Tập thô trong không gian PHU theo xấp xỉ

Tập X được gọi là *rõ* trong không gian PHU = (O, C) nếu $X^C = X_C$

Tập X được gọi là *thô* trong không gian PHU = (O, C) nếu $X^C \neq X_C$

Định nghĩa 3.7 Tập thô trong không gian PHU theo tập hợp

Tập X được gọi là *rõ* trong không gian PHU = (O, C) nếu $X \in \{\phi, C_1, C_2, \ldots, C_k, \cup C_i, O\}$; trong đó $\cup C_i$ là hợp của các C_i nào đó.

Tập X được gọi là *thô* trong không gian PHU = (O, C)

nếu $X \notin \{\phi, C_1, C_2, \ldots, C_k, \cup C_i, O\}$;

Định nghĩa 3.8 Tập thô trong không gian PHU theo hàm thuộc

Cho $O = \{o_1, o_2, \ldots o_m\}$; $C = \{C_1, C_2, \ldots C_k\}$ là phủ của O

Không gian PHU = (O, C) được gọi là không gian phủ

Giả sử $X \subseteq O$.

Ứng với X và C ta xây dựng hàm thuộc $\mu_x^C : O \to [0.1]$ như sau:

$$\text{với mọi } o \in O;\ \mu_x^C(o) = \min\left\{\frac{\text{card}(C_i(o) \cap X)}{\text{card}(C_i(o))}\right\}$$

trong đó $C_i(o)$ là nhóm C_i chứa đối tượng o, min theo các C_i khác nhau.

Tập X được gọi là *rõ* trong không gian PHU = (O, C) nếu tập giá trị của hàm thuộc μ_x^C chỉ chứa trị 0 hoặc 1.

Tập X được gọi là *thô* trong không gian PHU = (O, C) nếu tập giá trị của hàm thuộc μ_x^C có chứa trị khác 0 và khác 1.

Ví dụ 3.13:

Cho $O = \{1, 2, 3, 4, 5\}$

$C = \{C_1, C_2, C_3\} = \{\{1, 2\}, \{2, 3\}, \{3, 4, 5\}\}$ là một phủ của O.

Xét xem $X = \{1, 2\}$ và $Y = \{1, 3\}$ là thô hay rõ theo cả ba định nghĩa 3.6; 3.7; 3.8.

Theo định nghĩa 3.6 ta có $X^C = \{1, 2, 3\}$; $X_C = \{1, 2\}$; vậy X thô

Theo định nghĩa 3.7 ta có $X = C_1$ nên X là rõ

Theo định nghĩa 3.8 ta có tập giá trị của hàm thuộc μ_x^C là:

$\mu_x^C(1) = 1,\ \mu_x^C(2),$

$\mu_x^C(3) = 0,\ \mu_x^C(4) = 0,\ \mu_x^C(5) = 0$; X là tập rõ.

Tương tự xét tập Y

Theo định nghĩa 3.6 ta có $Y^C = \{1, 2, 3, 4, 5\}$; $Y_C = \{\}$; vậy Y thô

Theo định nghĩa 3.7 ta có Y không thuộc tập rõ nên Y thô

Theo định nghĩa 3.8 ta có tập giá trị của hàm thuộc μ_Y^C là:

$\mu_Y^C(1) = 1/3$, $\mu_Y^C(2) = 0$, $\mu_Y^C(3) = 1/3$,

$\mu_Y^C(4) = 1/3$, $\mu_Y^C(5) = 1/3$; Y là tập thô.

Lưu ý: Trong không gian PHU = (O, C) ba định nghĩa 3.6; 3.7; 3.8 không tương đương.

3.5. TẬP THÔ DUNG SAI (TOLERANCE ROUGH SET MODEL-TRSM)

Cho $U = \{o_1, o_2, \ldots o_m\}$;

a. Quan hệ tương đương

Định nghĩa 3.9 Quan hệ $R \subseteq U \times U$ được gọi là quan hệ tương đương trên U nếu R thỏa mãn ba điều kiện:

(*) Phản xạ: $\forall o \in U$ thì $(o, o) \in R$

(**) Đối xứng: $\forall o, o' \in U$ nếu $(o, o') \in R$ thì $(o', o) \in R$

(***) Bắc cầu: $\forall o, o', o'' \in U$ nếu $(o, o') \in R$ & $(o', o'') \in R$ thì $(o, o'') \in R$

b. Quan hệ dung sai (Tolerance relation)

Định nghĩa 3.10 Quan hệ $R \subseteq U \times U$ được gọi là quan hệ dung sai trên U nếu R thỏa mãn hai điều kiện:

(*) Phản xạ: $\forall o \in U$ thì $(o, o) \in R$

(**) Đối xứng: $\forall o, o' \in U$ nếu $(o, o') \in R$ thì $(o', o) \in R$.

Như vậy quan hệ tương đương là quan hệ dung sai. Mọi tính chất kết quả đúng cho quan hệ dung sai thì đúng cho quan hệ tương đương.

Ví dụ 3.14:

Xét tập U = {1, 2, 3, 4, 5, 6, 7};

Quan hệ $R \subseteq U \times U$ với $(o, o') \in R$ nếu $|o-o'| \le 1$ là quan hệ dung sai vì R thỏa mãn hai điều kiện:

Phản xạ: $\forall o \in U$ thì $(o, o) \in R$

Đối xứng: $\forall o, o' \in U$ nếu $(o, o') \in R$ thì $(o', o) \in R$. Tuy nhiên R không thỏa mãn tính bắc cầu vì $|1-2'| \le 1$ và $|2-3| \le 1$ nhưng $|1-3| > 1$.

Ví dụ 3.15:

Cho bảng điểm ba môn toán lý hóa của các sinh viên

U	T	L	H
o_1	2	2	1
o_2	4	4	2
o_3	6	3	3
o_4	8	9	4
o_5	8	9	5
o_6	9	8	8
o_7	3	8	7
o_8	5	7	8
o_9	8	7	9

Trên U ta xây dựng các quan hệ dung sai như sau:

$R_T \subseteq U \times U$ với o R_To' nếu $|o(T) - o'(T)| \leq 1$. Với o(T) là điểm toán của o

$R_L \subseteq U \times U$ với o R_Lo' nếu $|o(L) - o'(L)| \leq 1$.

$R_H \subseteq U \times U$ với o R_Ho' nếu $|o(H) - o'(H)| \leq 1$.

Khi đó rõ ràng R_T, R_L, R_H là các quan hệ dung sai chỉ thỏa mãn hai điều kiện phản xạ, đối xứng và không bắc cầu.

Nhìn vào bảng ta thấy rằng:

R_T chia U thành các nhóm rời nhau:

$C_1 = \{o_1, o_7, o_2, o_8, o_3\}$, $C_2 = \{o_4, o_5, o_6, o_9\}$

hoặc các nhóm không rời nhau:

$C_1' = \{o_1, o_7, o_2\}$, $C_2' = \{o_2, o_8, o_3\}$, $C_3 = \{o_4, o_5\}$, $C_4 = \{o_5, o_6, o_9\}$.

Vậy quan hệ dung sai có thể cho chúng ta nhiều phủ $C = \{C_1, C_2, ..., C_k\}$ khác nhau.

Lưu ý: Mọi quan hệ dung sai R trên tập đối tượng U luôn chia U thành phủ.

- Nếu đặt R(x) là nhóm các đối tượng quan hệ với x thì ta dễ dàng chỉ ra rằng $y \in R(x)$ khi và chỉ khi $x \in R(y)$.

- Ta dễ dàng chỉ ra rằng quan hệ $R \subseteq U \times U$ mà o R_Ho' <=> oR_To' & oR_Lo' & oR_Ho' là quan hệ dung sai trên U.

Định lý 3.5

Mọi quan hệ dung sai R trên U thì $C = \{R(x): x \in U\}$ là một phủ của U

Mọi phủ $C = \{C_1, C_2, ..., C_k\}$ của U thì quan hệ

$$R = C_1 \times C_1 \cup C_2 \times C_2 \cup ... C_k \times C_k$$ là quan hệ dung sai trên U.

Ta dễ dàng chỉ ra rằng quan hệ dung sai xây dựng như trong định lý 3.5 là không bắc cầu.

Ví dụ 3.16:

Cho U = {1, 2, 3, 4, 5}; C = {C_1, C_2} = {{1, 2, 3}, {3, 4, 5}}.

Khi đó R = $C_1 \times C_1 \cup C_2 \times C_2$ = {(1,1), (2,2), (3,3), (1,2), (2,1), (1,3), (3,1), (2,3), (3,2)} ∪ {(3,3), (4,4), (5,5), (3,4), (4,3), (3,5), (5,3), (4,5), (5,4)}. Rõ ràng R không bắc cầu vì (1,3) ∈ R và (3,4) ∈ R nhưng (1,4) ∉ R.

Lưu ý: Từ định lý 3.5 ta thấy tiếp cận tập thô theo phủ và theo quan hệ dung sai là hai hướng của một vấn đề.

c. Không gian tập thô dung sai (Tolerance Rough Set Model-TRSM)

$\Re = (U, I, \upsilon, P)$:

U: Tập các đối tượng.

I: $U \rightarrow 2^U$ - Hàm không chắc chắn (uncertainty function)

υ: $2^U \times 2^U \rightarrow [0,1]$ - độ mập mờ (vague inclusion)

P: $I(U) \rightarrow \{0,1\}$ - Hàm cấu trúc (structurality function)

Giả sử đối tượng x được nhận biết bằng hàm thông tin Inf(x)

I: $U \rightarrow 2^U$: hàm không chắc chắn xác định I(x) là một lớp dung sai (tolerance class) của đối tượng x.

υ: $2^U \times 2^U \rightarrow [0,1]$: hàm mập mờ đánh giá mức độ bao hàm của các tập hợp, cụ thể nó đánh giá độ bao hàm của lớp dung sai I(x) trong tập thô X.

P: $I(U) \rightarrow \{0,1\}$: Hàm cấu trúc phân lớp tập I(x) của đối tượng $x \in U$ vào một trong hai tập con là tập có cấu trúc với P(I(x)) = 1 và tập không có cấu trúc với P(I(x)) = 0

Ví dụ 3.17:

Sau đây ta xét một ví dụ biểu diễn văn bản trong TRSM

Cho U = {t_1, t_2,..., t_n} là tập các thuật ngữ.

D = {d_1, d_2,..., d_m} là tập m văn bản.

$\theta \in R^+$ là một số thực không âm, được gọi là *ngưỡng*.

F: $U \times U \rightarrow R^+$ với F(t,t') là số văn bản của tập D mà t và t' cùng xuất hiện.

Đặt: I: $U \rightarrow 2^U$ như sau: $I(t) = \{t' : F(t,t') \geq \theta\} \cup \{t\}$

Hàm tính độ mập mờ:

$$v(X, Y) = \frac{|X \cap Y|}{|X|}$$

Ta có thể coi tất cả các lớp dung sai của mỗi t đều là các tập con có cấu trúc, nghĩa là $P(I(t)) = 1$ với $\forall\ t \in U$

Vậy ta có không gian tập thô dung sai: $\Re = (U, I, \nu, P)$.

* Các xấp xỉ trên và dưới của tập $X \subseteq U$ trong không gian $\Re$ vừa xác định được tính theo các công thức:

Xấp xỉ dưới: $L(\Re, X) = \{t \in U: \upsilon(I(t), X) = 1\}$

Xấp xỉ trên: $U(\Re, X) = \{t \in U: \upsilon(I(t), X) > 0\}$

Giả sử cho tập gồm 31 thuật ngữ như trong bảng dưới đây. Với ngưỡng $\theta = 2$ ta tính được các lớp dung sai ứng với các từ t như sau:

$I(t_1) = \{t_1, t_2, t_3, t_{16}\}$

$I(t_2) = \{t_1, t_2, t_4, t_5, t_{26}\}$

$I(t_3) = \{t_3\}$

$I(t_4) = \{t_2, t_4\}$

$I(t_5) = \{t_1, t_2, t_3\}$

$I(t_6) = \{t_6, t_7\}$ ...

STT	Từ khóa	Xấp xỉ dưới	Xấp xỉ trên
d_1	t_1, t_2, t_3, t_4, t_5	t_3, t_4, t_5	$t_1, t_2, t_3, t_4, t_5, t_{16}, t_{26}$
d_2	t_6, t_7, t_8, t_9	t_6, t_7, t_8, t_9	t_6, t_7, t_8, t_9
d_3	$t_5, t_1, t_{10}, t_{11}, t_{12}$	t_5, t_{10}, t_{11}	$t_1, t_{24}, t_4, t_5, t_{10}, t_{11}, t_{16}, t_{26}$
d_4	$t_6, t_7, t_{12}, t_{13}, t_{14}$	$t_6, t_7, t_{12}, t_{13}, t_{14}$	$t_6, t_7, t_{12}, t_{13}, t_{14}$
d_5	t_2, t_{15}, t_4	t_4, t_{15}	$t_1, t_2, t_4, t_5, t_{15}, t_{26}$
d_6	$t_1, t_{16}, t_{17}, t_{18}, t_{19}, t_{20}$	$t_{16}, t_{17}, t_{18}, t_{19}, t_{20}$	$t_1, t_2, t_5, t_{16}, t_{17}, t_{18}, t_{19}, t_{20}$
d_7	$t_{21}, t_{22}, t_{23}, t_{24}, t_{25}$	$t_{21}, t_{22}, t_{23}, t_{24}, t_{25}$	$t_{21}, t_{22}, t_{23}, t_{24}, t_{25}$
d_8	$t_2, t_{12}, t_{26}, t_{27}$	t_{12}, t_{26}, t_{27}	$t_1, t_2, t_4, t_5, t_{12}, t_{26}, t_{27}$
d_9	t_{26}, t_2, t_{28}	t_{26}, t_{28}	$t_1, t_4, t_5, t_{26}, t_{28}$
d_{10}	$t_1, t_{16}, t_{21}, t_{26}, t_{29}, t_{30}, t_{31}$	$t_{16}, t_{21}, t_{26}, t_{29}, t_{30}, t_{31}$	$t_1, t_2, t_5, t_{16}, t_{21}, t_{26}, t_{29}, t_{30}, t_{31}$

CÂU HỎI VÀ BÀI TẬP CHƯƠNG 3

3.1 Cho X là tập thô nào đó trong không gian Apr = (U, E). Tập con B của tập thô X là tập như thế nào? có thô không? cho ví dụ minh họa.

3.2 Cho U là tập gồm 9 sinh viên.

U = {Phú, Quí, Giàu, Sang, Hùng, Mạnh, Cường, Hoa, Huệ}. Gọi X là tập thô với tập giá trị của hàm thuộc thô như sau:

X = {(Phú, 0.3), (Quí, 0.1), (Giàu,0.5), (Sang,0.9), (Hùng,0.0), (Mạnh,0.2), (Cường,0.4), (Hoa,0.3), (Huệ,1.0)};

Gọi Y là tập thô với các giá trị thô như sau:

Y = {(Phú, 0.3), (Quí, 0.1), (Giàu,0.5), (Sang,0.9), (Hùng,0.0), (Mạnh,0.2), (Cường,0.4), (Hoa,0.3), (Huệ,1.0)}.

Hãy nêu ý nghĩa và xác định giá trị thuộc thô của các phần tử trong các tập $X \cup Y$ và $X \cap Y$.

3.3 Đặt $U = M \cup V$; trong đó M là nhóm sinh viên Mỹ, V là nhóm các sinh viên Việt Nam. Gọi X là tập các sinh viên sẽ trở thành thiên tài. Theo bạn tập X là thô hay rõ theo phân hoạch P = {M, V}?

3.4 Giải thích ý nghĩa của các hàm thuộc thô trong tích Đề-các và chiếu sau:

Tích Đề-các

Giả sử $X_1, X_2, ... X_n$ là các tập con trong các không gian tương ứng $U_1, U_2, ..., U_n$.

Tích Đề-các của $X_1, X_2, ..., X_n$ là tập $X = X_1 \times X_2 ... \times X_n$ trong không gian $U_1 \times U_2 ... \times U_n$ với hàm thuộc thô được xác định như sau:

$$\mu_X(a) = \min(\mu_X(a_1), \mu_X(a_2), ..., \mu_X(a_n)); \text{ cho mọi } a = (a_1, a_2, ..., a_n)$$

Phép chiếu

Giả sử X là tập con trong không gian $U_1 \times U_2$ chiếu của tập X lên U_1 là tập X_1 trong U_1 với hàm thuộc thô $\mu_{X1}(a_1) = \max\{\mu_X(a_1, a_2 : a_2 \in U_2\}$.

Một cách tổng quát giả sử $X = X_1 \times X_2 ... \times X_n$ là tập trong không gian $U_1 \times U_2 ... \times U_n$ với hàm thuộc $\mu_X(a)$ được xác định cho mọi $a = (a_1, a_2, ... a_n)$

Hãy xác định tập X_i là chiếu của X lên U_i. Cho ví dụ.

3.5 Cho tập sinh viên A = {Phú, Quí, Giàu, sang, Hùng, Mạnh, Cường, Hoa, Huệ}. Trong đó Hoa và Huệ là những sinh viên nữ, còn lại là những sinh viên nam.

Cho X = {Phú, Quí}, Y = {Hoa, Huệ}, Z là tập những người sẽ thành đạt. Hãy cho biết và giải thích các mệnh đề sau đây mệnh đề nào đúng:

a. Tập X là tập rõ

b. Tập X là tập thô

c. Tập X là tập rõ theo phân hoạch giới tính

d. Tập X là tập thô theo phân hoạch giới tính

e. Tập Y là tập rõ

f. Tập Y là tập rõ theo phân hoạch giới tính

g. Tập $X \cup Y$ là tập rõ

h. Tập $X \cup Y$ là tập rõ theo phân hoạch giới tính

i. Tập $X \cup Y$ là tập thô theo phân hoạch giới tính

j. Z là tập thô

k. Z là tập rõ

3.6 Cho ví dụ không gian U, phân hoạch E và các tập X, Y, Z mà:

a. X là tập con thực sự của một nhóm E_i của E.

b. $Y = \cup E_i$ là hợp của các nhóm của E

c. Z là tập chứa mỗi nhóm một đại diện

Xét xem X, Y, Z là thô hay rõ theo phân hoạch E bằng cả ba định nghĩa 3.6, 3.7, 3.8.

3.7 Cho quan hệ sinh viên sv như sau:

sv

TT	TÊN	NS	QUÊ	GT	Màutóc	CAO
1	Phú	1985	HN	Nam	Đen	1.70^m
2	Quí	1985	HN	Nam	Đỏ	1.65^m
3	Giàu	1985	HN	Nam	Đen	1.70^m
4	Sang	1985	HN	Nam	Đỏ	1.65^m
5	Hùng	1986	HN	Nam	Đen	1.70^m
6	Mạnh	1986	HP	Nam	Đỏ	1.65^m
7	Cường	1986	HP	Nam	Đen	1.70^m
8	Hoa	1986	HP	Nữ	Đỏ	1.65^m
9	Huệ	1986	HP	Nữ	Trắng	1.65^m

Cho X = {Phú, Quí}, Y = {Hoa, Huệ}, Z = {Cường, Hoa}, MO là tập những sinh viên sẽ thành 'Đại Gia'. Gọi R là quan hệ cùng giới tính, S là quan hệ cùng quê, R' là quan hệ cùng năm sinh, S' là quan hệ cùng chiều cao. Xét xem X, Y, Z thô theo quan hệ nào, rõ theo quan hệ nào trong không gian sv = {Phú, Quí,..., Huệ}?

3.8 Xét không gian các bệnh nhân của một tỉnh được điều trị tại một bệnh viện đa khoa. Mỗi khoa của bệnh viện sẽ điều trị cho một số bệnh và triệu chứng nhất định. Mỗi ngày có X bệnh nhân nhập viện, Y bệnh nhân ra viện. Xét xem X, Y là thô hay rõ theo phân hoạch khoa?

3.9 Cho không gian nền Apr = (U, E).

Giả sử U là hữu hạn và đo được bởi *hàm độ đo* $\varphi: 2^U \rightarrow R^+$; với R^+ là tập số thực không âm, thỏa mãn các điều kiện:

(1) $\varphi(\phi) = 0$

(2) $\varphi(X) > 0$ nếu X khác rỗng

(3) φ đơn điệu, tức là nếu $X \subseteq Y$ thì $\varphi(X) \leq \varphi(Y)$

Ứng với $X \in 2^U$ ta xây dựng hàm thuộc thô $\mu_X^E : U \rightarrow [0, 1]$ như sau:

$$\mu_X^E(o) = \frac{\varphi(X \cap E_o)}{\varphi(E_o)}$$; trong đó E_o là nhóm chứa đối tượng o.

a. Chứng minh rằng X là rõ trong Apr nếu giá trị của hàm μ_X^E không chứa trị khác 0 hoặc 1. X là thô trong Apr nếu hàm μ_X^E có chứa trị khác 0 và 1.

b. Cho hai không gian Apr1 = (U, E) và Apr2 = (U, P). Giả sử Apr1 mịn hơn Apr2.

Ứng với $X \in 2^U$ ta xây dựng hai hàm thuộc thô $\mu_X^P, \mu_X^E : U \rightarrow [0, 1]$ như sau:

$$\mu_X^P(o) = \frac{\varphi(X \cap P_o)}{\varphi(P_o)}$$

$$\mu_X^E(o) = \frac{\varphi(X \cap E_o)}{\varphi(E_o)}$$; trong đó P_o, E_o là các nhóm chứa đối tượng o

Chứng minh rằng nếu X là thô trong Apr1 thì số trị của μ_X^E nhiều hơn số trị của μ_X^P, tức $card(\{\mu_X^E(o)\}) > card(\{\mu_X^P(0)\})$; với mọi o thuộc U.

3.10 Cho tập U gồm chín đồ chơi được đánh số từ 1 đến 9 như sau:

U = {1, 2, 3, 4, 5, 6, 7, 8, 9}

Trong các đồ chơi có các màu: trắng, đen, đỏ; kích thước: lớn, nhỏ; hình dáng: tam giác, vuông, tròn.

Gọi R1 là quan hệ cùng màu; R2 là quan hệ cùng kích thước; R3 là quan hệ cùng hình dáng.

a. Chứng minh rằng R1, R2, R3 là các quan hệ tương đương.

b. Giả sử chia theo màu ta được ba nhóm trắng, đen, đỏ

$U/R1 = \{\{1, 2, 3, 4\}, \{5, 6\}, \{7, 8, 9\}\}$.

Chia U theo kích thước ta được hai nhóm lớn, nhỏ

$U/R2 = \{\{1, 2, 3, 4, 5, 6\}, \{7, 8, 9\}\}$.

Chia theo hình dáng ta được ba nhóm tam giác, vuông, tròn

$U/R3 = \{\{1, 2, 3\}, \{4, 5, 6\}, \{7, 8, 9\}\}$.

Hãy thực hiện các yêu cầu sau đây:

- Liệt kê (hoặc cho biết chính xác số tập hợp) các tập rõ của không gian $Apr = (U, U/R1)$

- Liệt kê (hoặc cho biết chính xác số tập hợp) các tập rõ của không gian $Apr = (U, U/(R1 \cap R2 \cap R3))$

Chương 4

HỆ TIN VÀ CÁC ỨNG DỤNG

4.1. ĐỊNH NGHĨA HỆ TIN

Hệ tin (Information System) được Z. Pawlak định nghĩa lần đầu tiên trong [18].

Hệ tin là một bảng thông tin của một tập đối tượng U (Universe). Mỗi đối tượng $o \in U$ được đặc trưng bởi tập thuộc tính $A = \{A_1, A_2, ..., A_n\}$. Ví dụ xét hệ tin gồm 5 sinh viên:

Bảng 4.1. Bảng hệ tin

	Masv	Hoten	NS	Que
o_1	1	An	86	Hà Nội
o_2	2	Bình	86	Hà Nội
o_3	3	Sinh	86	Hà Nội
o_4	4	Phú	87	Hải Phòng
o_5	5	Quí	87	Hải Phòng

Bảng 4.1 cho ta biết thông tin về Hoten, năm sinh, quê quán của 5 sinh viên.

Trong bảng này ta có $U = \{o_1, o_2, o_3, o_4, o_5\}$

Tập thuộc tính A = {Masv, Hoten, NS, Que}

Định nghĩa 4.1 Hệ tin

Hệ tin là bộ bốn thành phần $S = (U, A, V, f)$.

Trong đó:

* $U = \{o_1, o_2, ... o_m\}$, $m \geq 1$, được gọi là *tập các đối tượng, hay vũ trụ (Universe).*

* $A = \{A_1, A_2, ... A_n\}$, $n \geq 1$ là *tập các thuộc tính*, mỗi thuộc tính $a \in A$ có miền trị (Domain(a)) là V_a;

* $V = \cup V_a$ là miền trị của các thuộc tính;

* Hàm $f: U \times A \rightarrow V$ thỏa mãn điều kiện $f(o, a) \in V_a$ được gọi là *hàm thông tin.*

Từ định nghĩa ta nhận xét rằng một hệ tin hoàn toàn được xác định nếu:

(1) Tập đối tượng $U = \{o_1, o_2, ..., o_m\}$, với $m \geq 1$, xác định.

(2) Tập các thuộc tính $A = \{A_1, A_2, ..., A_n\}$, $n \geq 1$, xác định.

(3) Tập các miền trị $V = \cup V_a$ xác định.

(4) Ánh xạ f: $U \times A \rightarrow V$ xác định và $f(o,a) \in V_a$.

Lưu ý:

Đôi khi để cho tiện ta có thể viết hệ tin S = (U, A). Với V, f coi như xác định.

Trong nhiều tài liệu các tác giả thường viết a(o) = v thay cho f(o, a) = v.

Ví dụ 4.1:

Cho $U = \{o_1, o_2, o_3, o_4, o_5\}$ tập các đối tượng.

A = {Manv, NS, QUE, GT}.

Domain(Manv) = {nv1, nv2, nv3, nv4, nv5}.

Domain(NS) ={50, 60,70, 80, 82, 83, 84}.

Domain(QUE) = {HN, HT, HD, HB, LS}

Domain(GT) = {Nam, Nu}.

V = {nv1, nv2, nv3, nv4, nv5, 50, 60, 70, 80, 82, 83, 84, HN, HT, HB, HD LS, Nam, Nu}. Hàm f cho trong bảng 4.2:

Bảng 4.2. Bảng hệ tin

	Manv	NS	GT	QUE
o_1	nv1	50	Nu	HN
o_2	nv2	70	Nu	HN
o_3	nv3	82	Nam	HT
o_4	nv4	83	Nam	HT
o_5	nv5	84	Nam	HB

$f(o_1, Manv) = nv1$, $f(o_2, Manv) = nv2$,

$f(o_3, Manv) = nv3$, $f(o_4, Manv) = nv4$, $f(o_5, Manv) = nv5$.

$f(o_1, NS) = 50$, $f(o_2, NS) = 70$, $f(o_3, NS) = 82$, $f(o_4, NS) = 83$, $f(o_5, NS) = 84$.

$f(o_1, GT) = Nu$, $f(o_2, GT) = Nu$,

$f(o_3, GT) = Nam$, $f(o_4, GT) = Nam$, $f(o_5, GT) = Nam$.

$f(o_1,QUE) = HN$, $f(o_2,QUE) = HN$,

$f(o_3,QUE) = HT$, $f(o_4,QUE) = HT$, $f(o_5, QUE) = HB$.

Từ định nghĩa hệ tin ta dễ dàng suy ra các bổ đề 4.1 và 4.2 sau đây:

Bổ đề 4.1

Mọi hệ tin S = (U, A) đều được biểu diễn dưới dạng bảng:

Bảng 4.3. Bảng hệ tin tổng quát

	A_1	A_2	...	A_j	...	A_n
o_1	a_{11}	a_{12}	...	a_{1j}		a_{1n}
...	...	...	...	...	...	
o_i	a_{i1}	a_{i2}	...	a_{ij}		a_{in}
...	...	...	...	...	...	...
o_m	a_{m1}	a_{m2}	...	a_{mj}	...	a_{mn}

Mọi bảng dạng bảng 4.3 đều xác định hệ tin S = (U, A, V, f).

Với U = $\{o_1, o_2, \ldots, o_m\}$ và A = $\{A_1, A_2, A_3, \ldots, A_m\}$; các giá trị trên giao của dòng o_i với cột A_j là giá trị của hàm thông tin $f(o_i, A_j) = a_{ij}$. V là tập các giá trị có trong bảng.

4.2. HỆ QUYẾT ĐỊNH, HỆ KHAI THÁC DỮ LIỆU, QUAN HỆ R

4.2.1. Quan hệ R trên tập thuộc tính A là một Hệ tin

Cho A = $\{A_1, A_2, \ldots, A_n\}$ là tập thuộc tính.

R = $\{t_1, t_2, \ldots, t_m\}$ là quan hệ trên A.

Trong chương 2 ta đã biết mọi quan hệ R đều có thể biểu diễn dạng bảng:

Bảng 4.4. Bảng hệ tin dạng quan hệ

r	A_1	A_2	...	A_j	...	A_n
t_1	a_{11}	a_{12}	...	a_{1j}		a_{1n}
...	...	...	...	...	...	
t_i	a_{i1}	a_{i2}	...	a_{ij}		a_{in}
...	...	...	...	...	...	...
t_m	a_{m1}	a_{m2}	...	a_{mj}	...	a_{mn}

Vậy quan hệ R trên A là một Hệ tin S = (U, A, V, f).

Trong đó U = R = $\{t_1, t_2, \ldots t_m\}$; A = $\{A_1, A_2, \ldots, A_n\}$ là tập thuộc tính; V và f cho trong bảng.

Bổ đề 4.2

Mọi quan hệ R trên A, với R = $\{t_1, t_2, \ldots, t_m\}$; khi đó S = (R, A, V, f) là một hệ tin, với $f(t_i, A_j) = t_i.A_j$

Ví dụ cho quan hệ r như sau:

r

	Manv	NS	GT	QUE
t_1	nv1	50	Nu	HN
t_2	nv2	70	Nu	HN
t_3	nv3	82	Nam	HT
t_4	nv4	83	Nam	HT
t_5	nv5	84	Nam	HB

Khi đó ta có hệ tin hoàn toàn xác định S = (U, A, V, f) với U = $\{t_1, t_2, ..., t_5\}$, A = $\{A_1, A_2, A_3, A_4\}$ = {Manv, NS, GT, QUE} và miền giá trị các thuộc tính V được cho như trong bảng.

Lưu ý: Mọi quan hệ r trên A luôn biểu thị được dạng bảng, nhưng có nhiều bảng không phải là một quan hệ r trên A và vì vậy có nhiều hệ tin không phải là quan hệ. Ví dụ xét hệ tin S = (U, A, V, f) và hàm f được xác định như sau:

Ví dụ 4.2:

	HOTEN	NS	QUE
o_1	T. ANH	82	HN
o_2	T. ANH	82	HN
o_3	T. ANH	82	HN
o_4	T. ANH	82	HN
o_5	T. ANH	82	HN

Đây là hệ tin mà các đối tượng giống nhau (bất khả phân biệt) trên tập thuộc tính A.

Theo định nghĩa quan hệ r trên A những bộ giống nhau trên toàn bộ tập thuộc tính A phải được coi là một phần tử của quan hệ. Vậy hệ tin trên chỉ được biểu thị bằng quan hệ r trên tập thuộc tính A = {HOTEN, NS, QUE} là quan hệ chỉ 1 phần tử:

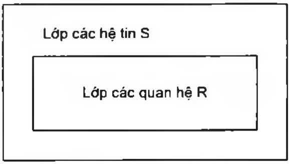
r

HOTEN	NS	QUE
T. ANH	82	HN

Lưu ý:

- Mọi quan hệ là một hệ tin. Lớp các quan hệ {R} là lớp con thực sự của lớp các hệ tin {S}.

- Quan hệ là một hệ tin nên mọi khái niệm, tính chất, kết quả đúng trong hệ tin đều có thể dùng trong các quan hệ.



Hình 4.1. Lớp các hệ tin

Ví dụ 4.3:

r	HOTEN	NS	GT	QUE
t_1	An	45	Nữ	Hà Nội
t_2	An	45	Nam	Hà Nội
t_3	Bính	80	Nam	Thái Nguyên
t_4	Bích	60	Nam	Tuyên Quang
t_5	Ngọc	60	Nữ	Lạng Sơn
t_6	Lan	82	Nữ	Thanh Hóa
t_7	Hương	82	Nữ	Nghệ An

Đây là hệ tin và là một quan hệ.

4.2.2. Hệ quyết định là một Hệ tin

Định nghĩa 4.2 Hệ quyết định

Hệ quyết định là hệ tin S = (U, A, V, f) với $A = C \cup D$ trong đó C được gọi là tập *thuộc tính điều kiện*, D là tập *thuộc tính quyết định* và $C \cap D = \phi$.

Vậy hệ quyết định là hệ $S = (U, C \cup D, V, f)$.

Thông thường tập thuộc tính quyết định và tập thuộc tính điều kiện rời nhau, tức là $C \cap D = \varnothing$ và *giá trị các thuộc tính điều kiện quyết định giá trị thuộc tính quyết định.*

Ví dụ 4.4 cho biết điểm thi 3 môn Toán (T), Lý (L), Hóa (H) của các thí sinh quyết định (d) thí sinh đó đậu hay trượt cuộc thi:

Ví dụ 4.4:

THISINH

SBD	HOTEN	T	L	H	d
1	A	4	5	4	Trượt
2	B	6	9	9	Đậu
3	C	2	3	4	Trượt
4	E	8	8	10	Đậu
5	F	9	9	9	Đậu.

Hệ quyết định trong ví dụ 4.4 có: tập 5 đối tượng U = {1, 2, 3, 4, 5}; Tập thuộc tính A = {HOTEN, T, L, H, d} với tập thuộc tính điều kiện C = {HOTEN, T, L, H} và tập thuộc tính quyết định D = {d} có 1 thuộc tính.

Lưu ý:

- Trong hệ quyết định thông thường tập thuộc tính quyết định chứa ít thuộc tính và miền trị của các thuộc tính quyết định chỉ có vài giá trị.

- Trong ví dụ 4.5 ta xét hệ tin các ứng viên thi người đẹp có thuộc tính quyết định (hoa hậu) có hai giá trị 0 (trượt) và 1 (hoa hậu). Các thuộc tính điều kiện như: cao, nặng, vòng1(V1), vòng2(V2), vòng3(V3), thuộc tính quyết định D = {Hoa hậu}.

Ví dụ 4.5:

THISINH_ĐEP

SBD	Cao	Nặng	V1	V2	V3	Hoa hậu
1	1.6	55	85	58	88	0
2	1.7	55	88	60	90	1
3	1.75	58	88	60	88	1
4	1.55	55	70	70	70	0
5	1.50	50	60	70	70	0

4.2.3. Hệ khai thác dữ liệu là một Hệ tin

Định nghĩa 4.3 Định nghĩa hệ khai thác thông tin (hệ khai thác dữ liệu)

Hệ khai thác thông tin (KTTT) là hệ tin S = (U, A, V, f).

Trong đó tập U = {o_1, o_2,..., o_m} được gọi là tập các hóa đơn (hay tập các giao dịch).

Tập A = $\{i_1, i_2,..., i_n\}$ được gọi là tập các mặt hàng (hay tập các mục).

Tập V = {0, 1}.

Giá trị $f(o_j, i_k) = 1$ cho ta biết hóa đơn o_j chứa mặt hàng i_k và $f(o_j, i_k) = 0$ nếu hóa đơn o_j không chứa mặt hàng i_k.

Ví dụ 4.6:

	i_1	i_2	i_3	i_4	i_5	i_6	i_7
o_1	1	1	1	1	1	1	1
o_2	1	1	1	0	0	1	0
o_3	1	1	0	1	0	0	0
o_4	1	0	0	0	1	0	0
o_5	1	0	0	0	0	0	0

Lưu ý: Các hệ tin như trong ví dụ 4.6 được gọi là hệ khai thác thông tin vì từ hệ thống này ta có thể khai thác được các thông tin như mặt hàng i_1 bán tốt nhất vì nó xuất hiện trong tất cả các hóa đơn. Mặt hàng i_7 bán kém nhất vì nó chỉ xuất hiện một lần trong các hóa đơn.

Định nghĩa 4.4 Độ phổ biến (hay độ thường xuyên) của tập hàng s

Cho hệ khai thác dữ liệu (KTDL) = (O, I, V, f); $s \subseteq I$.

Độ phổ biến của tập mặt hàng s, ký hiệu sp(s) là tỷ số giữa số lần xuất hiện trong các hóa đơn của tập s trên tất cả các hóa đơn. Gọi m là số các hóa đơn khi đó ta có: *Sp(s) = (số lần xuất hiện của s)/m.*

Ví dụ xét dữ liệu trong ví dụ 4.6 ta có:

$sp(\{i_1\}) = 5/5 = 1$; $sp(\{i_2,i_3\}) = 2/5$;

$sp(\{i_2,i_3,i_4\}) = 1/5$; $sp(\{i_7\}) = 1/5$.

Vậy với mọi tập hàng $s \subseteq I$ thì $0 \leq sp(s) \leq 1$ và mọi số $\varepsilon \in (0,1]$ chia họ các tập con của I thành hai phần. Một phần gồm các tập s mà $sp(s) < \varepsilon$ và phần kia gồm các tập s mà $sp(s) \geq \varepsilon$. Trong khai thác dữ liệu họ $\{s \subseteq I: sp(s) \geq \varepsilon\}$ được gọi là các *tập phổ biến với ngưỡng* ε, *gọi tắt là các tập phổ biến.* Gọi FS là họ các tập s mà $sp(s) \geq \varepsilon$. Một bài toán quan trọng trong khai thác dữ liệu là tìm các thuật toán để tính FS.

Lưu ý:

Trong khai thác dữ liệu để tìm các tập phổ biến FS người ta dùng nguyên lý Apriori như sau:

- Nếu s là tập phổ biến với ngưỡng minsupp thì mọi tập con của s là tập phổ biến.

- Nếu s không là tập phổ biến tức sp(s) < minsupp thì mọi tập cha của s là tập không phổ biến.

Định nghĩa 4.5.a Luật kết hợp

Cho hệ KTDL = (O, I, V, f); X, Y $\subseteq$ I.

Luật kết hợp của X và Y ký hiệu X $\rightarrow$ Y là luật chỉ khả năng xuất hiện của Y khi X xuất hiện.

Trong khai thác dữ liệu *khả năng xuất hiện* của Y khi X xuất hiện được đo bằng *độ tin cậy của luật*.

Ví dụ nếu lấy X = $\{i_1, i_2\}$ và Y = X thì khả năng xuất hiện của Y khi X xuất hiện là 100% và luật X $\rightarrow$ Y có độ tin cậy 100%. Như vậy mỗi luật kết hợp có độ tin cậy CF(X $\rightarrow$ Y).

Định nghĩa 4.5.b Định nghĩa độ tin cậy của luật kết hợp

Cho hệ KTDL = (O, I, V, f); X, Y $\subseteq$ I & $X \cap Y = \phi$

Độ tin cậy của luật X $\rightarrow$ Y, ký hiệu CF(X $\rightarrow$ Y) là đại lượng chỉ khả năng xuất hiện của Y khi X xuất hiện. Độ tin cậy được tính theo công thức:

$$CF(X \rightarrow Y) = sp(X \cup Y)/sp(X).$$

Ví dụ Xét dữ liệu trong ví dụ 4.6 ta có:

$$\begin{aligned} CF(\{i_1, i_2\} \rightarrow \{i_3, i_4, i_5\}) &= sp(\{i_1, i_2, i_3, i_4, i_5\})/sp(\{i_1, i_2\}) \\ &= 1/3.\ CF(\{i_1\} \rightarrow \{i_2\}) \\ &= sp(\{i_1, i_2\})/\ sp(\{i_1\}) = (3/5)/(5/5) = 3/5 \end{aligned}$$

4.3. QUAN HỆ BẤT KHẢ PHÂN BIỆT TRONG HỆ TIN

Cho hệ tin S = (U, A); t, t' $\in$ U; X $\subseteq$ A

Trong tập các đối tượng U ta xét quan hệ IND(X) $\subseteq$ U $\times$ U được gọi là *quan hệ bất khả phân biệt (INDiscernibility)* hay *quan hệ giống nhau* trên tập thuộc tính X.

Định nghĩa 4.6 Quan hệ bất khả phân biệt (giống nhau) trên tập thuộc tính X

Quan hệ IND(X) $\subseteq$ U $\times$ U được gọi là *quan hệ bất khả phân biệt (INDiscernibility)* trên tập thuộc tính X, nếu hai đối tượng bất kỳ t, t' $\in$ U thì (t, t') $\in$ IND(X) khi và chỉ khi f(t, a) = f(t', a) $\forall$ a $\in$X.

Lưu ý: Hai đối tượng t và t' có thể giống nhau trên tập thuộc tính A, tức t IND(A) t' khi t và t' giống nhau trên mọi thuộc tính a của A và khi đó ta nói t và t' giống nhau trên A.

Ví du 4.7:

Cho hệ tin như sau:

(a)

O	E	B	C	D
o_1	a_1	b_1	c_1	d_1
o_2	a_2	b_2	c_2	d_2
o_3	1	2	3	3
o_4	1	2	3	3
o_5	1	2	4	4
o_6	2	3	3	2

Khi đó tập các đối tượng $\{o_3, o_4, o_5\}$ giống nhau trên $\{E, B\}$

Hoặc trong hệ tin

(b)

r	HOTEN	NS	GT	QUE
t_1	N.V.An	1980	Nữ	Hanội
t_2	N.Bính	1930	Nam	Hanội
t_3	Trần Minh	1960	Nam	Hanam
t_4	Trần Minh	1960	Nam	Hanam

Khi đó $r = \{t_1, t_2, t_3, t_4\}$ hai đối tượng t_1 và t_2 giống nhau trên thuộc tính QUE, hay $(t_1, t_2) \in$ IND(QUE). Hai đối tượng t_3 và t_4 giống nhau trên toàn bộ tập thuộc tính A, như vậy t_3 và t_4 giống nhau hay nói cách khác t_3 và t_4 bất khả phân biệt trên A, tức $(t_3, t_4) \in$ IND(A).

Bổ đề 4.3

Cho hệ tin S = (U, A, V, f). $X \subseteq A$.

Quan hệ IND(X) là quan hệ tương đương trên U

Khi đó U/ IND(X) là phân hoạch tương đương của U, ký hiệu PART(X). Vậy phân hoạch tương đương PART(X) là: PART(X) = U/IND(X) = $\{[o]_{IND(X)}: o \in U\}$; $[o]_{IND(X)}$ là nhóm các đối tượng quan hệ IND(X) với o.

Cặp U và PART(X) tạo nên không gian Pawlak: Apr(X) = (U, PART(X)).

Ví du 4.8:

Xét hệ tin (a) trong ví dụ 4.7, với A = {E, B, C, D}, $U = \{o_1, o_2, o_3, o_4, o_5, o_6\}$.

Lấy X ={E, B}.

Khi đó PART(X) = U/ IND(X) = $\{E_1, E_2, E_3, E_4\}$

$= \{\{o_1\}, \{o_2\}, \{o_3, o_4, o_5\}, \{o_6\}\}$.

Tức là phân hoạch PART(X) có 4 nhóm:

$E_1 = \{o_1\}$

$E_2 = \{o_2\}$

$E_3 = \{o_3, o_4, o_5\}$

$E_4 = \{o_6\}$.

Bổ đề 4.4

Cho hệ tin S = (U, A). X $\subseteq$ A, o $\in$ U. Khi đó:

a. $IND(X) = \bigcap_{x \in X} IND(x)$

b. Với mọi tập thuộc tính X, Y ta có: $[o]_{IND(X)} \subseteq [o]_{IND(Y)} \Leftrightarrow IND(X) \subseteq IND(Y)$.

c. Với mọi tập thuộc tính X, Y nếu $X \subseteq Y \Rightarrow IND(Y) \subseteq IND(X)$

Chứng minh:

Kết luận a là hiển nhiên vì hai đối tượng o và o' giống nhau trên tập thuộc tính X khi và chỉ khi chúng giống nhau trên từng thuộc tính x của X. Tương tự b, c cũng hiển nhiên.

Hoàn toàn bằng trực quan chúng ta có kết luận sau:

Bổ đề 4.5

a. Với mọi tập thuộc tính X, Y của U ta luôn có:

$IND(X \cup Y) \subseteq IND(X)$ & $IND(X \cup Y) \subseteq IND(Y)$

b. Cho r là quan hệ trên A, hay cho hệ tin S = (r, A); X, Y $\subseteq$ A.

Khi đó r thỏa mãn phụ thuộc hàm $X \rightarrow Y$ khi và chỉ khi $IND(X) \subseteq IND(Y)$.

Sau đây ta sẽ định nghĩa *không gian dương* hay *vùng dương* của tập đối tượng liên quan hai tập thuộc tính X, Y.

Cho hệ tin S = (U, A, V, f). X, Y $\subseteq$ A. Khi đó PART(X) = U/IND(X), PART(Y) = U/IND(Y) là các phân hoạch tương đương của U.

4.4. VÙNG DƯƠNG

Cho hệ tin S = (U, A, V, f); X, Y $\subset$ A;

$PART(X) = \{p_1, p_2,...,p_k\}$;

$PART(Y) = \{q_1, q_2,..., q_l\}$ là các phân hoạch của U.

Đặt $Apr(X) = (U, PART(X))$

$Apr(Y) = (U, PART(Y))$ là các không gian Pawlak tương ứng với X và Y.

Định nghĩa 4.7 Vùng dương (hay không gian dương)

Vùng dương (POSitive region) của hai tập thuộc tính X và Y, ký hiệu là POS(X,Y) là hợp của các nhóm p thuộc PART(X) mà p được chứa trong nhóm q nào đó của PART(Y) hay $POS(X,Y) = \cup\{p \in PART(X): \exists\, q \in PART(Y) \text{ sao cho } p \subseteq q\}$.

Lưu ý:

Giả sử $PART(X) = P = \{p_1, p_2,..., p_k\}$, $PART(Y) = Q = \{q_1, q_2,..., q_n\}$ là các phân hoạch của U ứng với X, Y.

- Từ định nghĩa ta dễ dàng suy ra rằng $POS(X,Y) = \bigcup_{i=1}^{n} q_{iP}$; tức vùng dương POS(X,Y) là hợp của các xấp xỉ dưới của các q_i ứng với phân hoạch P.

- Vùng dương trong U gắn với hai tập thuộc tính X, Y nên đôi khi thay cho nói vùng dương của hai tập thuộc tính X, Y ta nói vùng dương của X, Y.

Ví dụ 4.9:

Cho hệ tin S = (U, A, V, f) như sau:

U	H	B	C	D
o_1	a	b	c	d
o_2	a	b	c_1	d_1
o_3	a	b	c_2	d_2
o_4	a_1	b_1	c_3	d_3
o_5	a_1	b_1	c_4	d_4
o_6	a_2	b_2	c_5	d_5
o_7	a_2	b_2	c_5	d_5
o_8	a_2	b_2	c_5	d_5
o_9	a_3	b_3	c_5	d_5

Khi đó nếu lấy X = {H, B} và Y = {C, D} ta có:

$PART(X) = \{p_1, p_2, p_3, p_4\} = \{\{o_1, o_2, o_3\}, \{o_4, o_5\}, \{o_6, o_7, o_8\}, \{o_9\}\}$.

$PART(Y) = \{q_1, q_2, q_3, q_4, q_5, q_6\}$

$= \{\{o_1\}, \{o_2\}, \{o_3\}, \{o_4\}, \{o_5\}, \{o_6, o_7, o_8, o_9\}\}$.

Vậy trong phân hoạch PART(X) chỉ có hai nhóm p_3, p_4 là các tập con của nhóm q_6 trong PART(Y) nên:

$$POS(X,Y) = p_3 \cup p_4 = \{o_6, o_7, o_8\} \cup \{o_9\} = \{o_6, o_7, o_8, o_9\}.$$

Và tương tự:

$$POS(Y,X) = q_1 \cup q_2 \cup q_3 \cup q_4 \cup q_5$$
$$= \{o_1\} \cup \{o_2\} \cup \{o_3\} \cup \{o_4\} \cup \{o_5\} = \{o_1, o_2, o_3, o_4, o_5\}$$

Lưu ý: POS(X,Y) hoặc chứa trọn một số nhóm của PART(X) hoặc không chứa nhóm nào của PART(X). Nói cách khác nếu POS(X,Y) đã chứa một phần tử nào đó của một nhóm trong PART(X) thì POS(X,Y) phải chứa hết các phần tử của nhóm đó.

Ví dụ 4.10:

Xét quan hệ r(A) trên tập thuộc tính A = {B, C, D}.

Với r gồm 8 phần tử $r = \{t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8\}$.

r

D	B	C
1	0	2
1	1	3
3	1	3
3	1	2
3	1	3
4	1	5
4	2	5
4	3	5

Khi đó ta có:

IND(D) chia r thành 3 nhóm và $PART(D) = \{\{t_1,t_2\}, \{t_3,t_4,t_5\}, \{t_6,t_7,t_8\}\}$.

IND(B) chia r thành 3 nhóm và $PART(B) = \{\{t_1\}, \{t_2,t_3,t_4,t_5, t_6\}, \{t_7,t_8\}\}$.

IND(C) chia r thành 3 nhóm và $PART(C) = \{\{t_1,t_4\}, \{t_2,t_3,t_5\}, \{t_6,t_7,t_8\}\}$.

IND(BC) chia r thành 6 nhóm

và $PART(BC) = \{\{t_1\}, \{t_2, t_3, t_5\}, \{t_4\}, \{t_6\}, \{t_7\}, \{t_8\}\}$.

IND(DC) chia r thành 5 nhóm

và $PART(DC) = \{\{t_1\}, \{t_2\}, \{t_3, t_5\}, \{t_4\}, \{t_6, t_7, t_8\}\}$.

Khi đó:

$POS(D,B) = \{t_3,t_4,t_5\}$, đây là nhóm thứ 2 của PART(D

POS(D,C) = $\{t_6,t_7,t_8\}$, đây là nhóm thứ 3 của PART(D).

POS(D,BC) = ϕ, không chứa bộ nào của bất kỳ nhóm nào của PART(D).

POS(DC,BC) = $\{t_1, t_2, t_3, t_5, t_4\}$ chứa 4 nhóm của PART(DC).

Ta thấy POS(D,B) $\subseteq$ POS(DC,BC).

Bổ đề 4.6

Cho hệ tin S = (U, A); X, Y, Z $\subseteq$ A. Khi đó ta có:

a. card(PART(X)) $\leq$ card(PART(XY)) & card(PART(Y)) $\leq$ card(PART(XY))

b. card(POS(X,Y)) $\leq$ card(POS(XZ,YZ)

Chứng minh:

a. PART(X) = U/IND(X) = $\{[t]_{IND(X)}: t \in U\}$

PART(XY) = U/IND(XY) = $\{[t]_{IND(XY)}: t \in U\}$.

Ta nhận xét rằng mọi cặp t, t' bất khả phân biệt trên XY thì bất khả phân biệt trên X. Như vậy mọi nhóm của PART(XY) là một nhóm con của PART(X) và mỗi nhóm $[t]_{IND(X)}$ của PART(X) có chứa ít nhất một nhóm con $[t]_{IND(XY)}$ của PART(XY). Như vậy số nhóm của PART(XY) nhiều hơn hoặc bằng số nhóm của PART(X).

Tức card(PART(X)) $\leq$ card(PART(XY)).

Tương tự ta có card(PART(Y)) $\leq$ card(PART(XY))

b. Lấy t $\in$ POS(X,Y) khi đó ta có $[t]_{IND(X)} \subseteq [t]_{IND(Y)}$ <=> IND(X) $\subseteq$ IND(Y) => IND(XZ) $\subseteq$ IND(YZ) <=> $[t]_{IND(XZ)} \subseteq [t]_{IND(YZ)}$ => t $\in$ POS(XZ,YZ).

Ví dụ xét hệ tin S = (U, A) trong bảng sau:

Với U = {1, 2, 3, 4, 5}; A = {X, Y, X, QUE}.

U	X	Y	Z	QUE
1	a	82	Nữ	HN
2	a	82	Nữ	HN
3	b	83	Nam	HP
4	b	84	Nam	HP
5	b	84	Nữ	HP

Khi đó:

PART(X) = $\{p_1, p_2\}$ = {{1, 2}, {3, 4, 5}}. PART(X) có 2 nhóm.

PART(XY) = $\{q_1, q_2, q_3\}$ = {{1, 2}, {3}, {4, 5}}. PART(XY) có 3 nhóm.

Vậy POS(X,Y) = {1, 2}.

PART(XZ) = {{1,2}, {3},{4, 5}}.

PART(YZ) = {{1,2}, {3}, {4}, {5}}. Vậy POS(XZ,YZ) = {1, 2, 3}

Bổ đề 4.7

- Cho hệ tin S = (r, A); X, Y $\subseteq$ A

- Nếu r thỏa mãn phụ thuộc hàm X $\rightarrow$ Y thì mỗi nhóm p_i trong PART(X) nằm gọn trong một nhóm q_j nào đó của PART(Y).

Chứng minh:

Giả sử p là một nhóm thuộc PART(X), như vậy p gồm những bộ bằng nhau trong X và vì r thỏa mãn X $\rightarrow$ Y nên các bộ đó cũng bằng nhau trong Y. Vậy các bộ đó cùng nằm trong một nhóm q của PART(Y).

Bổ đề 4.8

- Cho hệ tin S = (r, A); X, Y $\subseteq$ A

- Với mọi cặp X, Y coi POS(X,Y) là một quan hệ con của r. Khi đó ta có:

X $\rightarrow$ Y thỏa mãn trong POS(X,Y)

Nói cách khác POS(X,Y) thỏa mãn X $\rightarrow$Y.

Chứng minh:

Giả sử có hai bộ t và t'của POS(X,Y) mà t.X = t'.X thì hai bộ đó thuộc cùng một nhóm u của PART(X), theo định nghĩa của POS(X,Y) thì có một nhóm v của PART(Y) chứa u, do đó t và t' thuộc cùng nhóm v và vì vậy t.Y = t'.Y. Kết quả được chứng minh.

Bổ đề 4.9

Cho hệ tin S = (r, A); X, Y $\subseteq$ A

Nếu POS(X, Y) = r $\Leftrightarrow$ r thỏa mãn X $\rightarrow$ Y

Chứng minh suy trực tiếp từ định nghĩa và bổ đề 4.8.

Bổ đề 4.10

Cho r(U).

Với mọi X, Y, Z $\subseteq$ U ta luôn có POS(X, Y) $\subseteq$ P(XZ,YZ).

Chứng minh:

Giả sử t $\in$ POS(X,Y). Lấy t làm đại diện ta có $[t]_{IND(X)} \subseteq [t]_{IND(Y)}$.

Ta phải chứng minh t $\in$ POS(XZ,YZ), tức ta phải chứng minh $[t]_{IND(XZ)} \subseteq [t]_{IND(YZ)}$ hay theo bổ đề 4.4b ta phải chứng minh IND(XZ) $\subseteq$ IND(YZ).

Cũng theo bổ đề 4.4 ta thấy IND(XZ) = IND(X) $\cap$ IND(Z)

và IND(YZ) = IND(Y) $\cap$ IND(Z).

Theo giả thiết ta có $[t]_{IND(X)} \subseteq [t]_{IND(Y)} \Leftrightarrow$ IND(X) $\subseteq$ IND(Y).

Vậy IND(XZ) $\subseteq$ INDYZ). Bổ đề được chứng minh.

Bổ đề 4.11

Cho r là quan hệ trên U; X, Y,Z $\subseteq$ U

Nếu POS(X,Y) = r thì POS(Y,Z) $\subseteq$ POS(X, Z).

Chứng minh:

Lấy t $\in$ POS(Y, Z) $\Rightarrow$ $[t]_{IND(Y)} \in$ PART(Y) và $[t]_{IND(Z)} \in$ PART(Z) và tất nhiên $[t]_{IND(Y)} \subseteq [t]_{IND(Z)}$. Mặt khác vì POS(X,Y) = r nên r thỏa mãn phụ thuộc hàm X $\rightarrow$ Y, tức IND(X) $\subseteq$ IND(Y).

Vậy $[t]_{IND(X)} \subseteq [t]_{IND(Y)}$ và vì $[t]_{IND(Y)} \subseteq [t]_{IND(Z)}$ nên $[t]_{IND(X)} \subseteq [t]_{IND(Z)}$, điều này có nghĩa là t $\in$ POS(X, Z).

Bổ đề 4.12

Nếu POS(X,Y) = r thì POS(Z,X) $\subseteq$ POS(Z,Y) với mọi X, Y,Z $\subseteq$ U.

Chứng minh:

Tương tự như chứng minh bổ đề 4.11 lấy t $\in$ POS(Z, X) và lấy t làm đại diện ta có $[t]_{IND(Z)} \subseteq [t]_{IND(X)}$ và vì POS(X,Y) = r nên r thỏa mãn phụ thuộc hàm X $\rightarrow$ Y, tức IND(X) $\subseteq$ IND(Y) => $[t]_{IND(X)} \subseteq [t]_{IND(Y)}$ => $[t]_{IND(Z)} \subseteq [t]_{IND(Y)}$ điều này tương đương với t $\in$ POS(Z, Y).

Sau đây chúng ta sẽ xét khái niệm về phụ thuộc hàm trong hệ tin. Đây là khái niệm mở rộng của khái niệm phụ thuộc hàm trong chương 2. Trong chương 2 ta đã định nghĩa r thỏa mãn phụ thuộc hàm X $\rightarrow$ Y nếu mọi cặp t_1, t_2 $\in$ r mà $t_1.X = t_2.X$ thì $t_1.Y = t_2.Y$. Bây giờ ta nới rộng điều kiện t_1, t_2 có thể bằng nhau trên X nhưng không bằng nhau trên Y. Khi đó ta nói r thỏa mãn phụ thuộc hàm X $\rightarrow$ Y ở mức độ k nào đó và ký hiệu $X \xrightarrow{k} Y$. Tất nhiên k $\in$ [0,1]. Như vậy để xác định phụ thuộc hàm giữa X và Y trong quan hệ r ta phải xác định độ phụ thuộc k của Y vào X.

Để làm rõ điều này ta xét khái niệm phụ thuộc hàm trong hệ tin.

4.5. PHỤ THUỘC HÀM TRONG HỆ TIN

Cho hệ tin S = (U, A, V, f); X, Y $\subseteq$ A.

Khi đó ta có POS(X,Y) và đặt k = card(POS(X,Y))/ card(U).

Định nghĩa 4.8 Phụ thuộc hàm trong Hệ tin S

Với mọi cặp X, Y $\subseteq$ A.

Ta nói Y *phụ thuộc hàm vào* X (hoặc X xác định *phụ thuộc hàm* Y) *mức độ k*, ký hiệu $X \xrightarrow{k} Y$; Trong đó k = card(POS(X,Y))/ card(U).

Lưu ý:

k = card(POS(X,Y))/card(U) là độ phụ thuộc của Y vào X.

Mọi cặp X, Y ta luôn có $X \xrightarrow{k} Y$ với k = card(POS(X,Y))/card(U) và $0 \le k \le 1$.

Khi k = 1 thay cho viết $X \xrightarrow{1} Y$ ta chỉ viết $X \to Y$.

Định nghĩa 4.9 Độ phụ thuộc

Giá trị k = card(POS(X,Y))/card(U) được gọi là *độ phụ thuộc* của Y vào X

Ví dụ 4.11:

Xét X và Y như trong ví dụ sau:

Cho hệ tin S như sau:

U	H	B	C	D
o_1	a	b	c	d
o_2	a	b	c_1	d_1
o_3	a	b	c_2	d_2
o_4	a_1	b_1	c_3	d_3
o_5	a_1	b_1	c_4	d_4
o_6	a_2	b_2	c_5	d_5
o_7	a_2	b_2	c_5	d_5
o_8	a_2	b_2	c_5	d_5
o_9	a_3	b_3	c_5	d_5

Khi đó nếu lấy X = {H, B} và Y = {C, D} ta có:

PART(X) = $\{E_1, E_2, E_3, E_4\} = \{\{o_1, o_2, o_3\}, \{o_4, o_5\}, \{o_6, o_7, o_8\}, \{o_9\}\}$.

PART(Y) = $\{E_1', E_2', E_3', E_4', E_5', E_6'\}$

$=\{\{o_1\}, \{o_2\}, \{o_3\}, \{o_4\}, \{o_5\}, \{o_6, o_7, o_8, o_9\}\}$.

Vậy trong phân hoạch PART(X) chỉ có hai nhóm E_3, E_4 là các tập con của nhóm E_6' trong PART(Y) nên:

POS(X,Y) = $E_3 \cup E_4 = \{o_6, o_7, o_8\} \cup \{o_9\} = \{o_6, o_7, o_8, o_9\}$.

ta có:

$k = card(POS(X,Y))/card(U) = 4/9$, vậy $X \xrightarrow{4/9} Y$.

Lưu ý:

- Khi $k = 1$ ta có $X \rightarrow Y$. Đây là phụ thuộc hàm bình thường.

- Khi $k = 0$, X và Y nói chung không phụ thuộc hay độ phụ thuộc của Y vào X là 0.

Bổ đề 4.13

Với mọi cặp số hữu tỷ k, k' thuộc đoạn [0,1], mọi tập X, Y $\subseteq$ A, ta luôn xây dựng được các quan hệ R, S trên A sao cho:

$X \xrightarrow{k} Y$ thỏa mãn R

$X \xrightarrow{k'} Y$ thỏa mãn S

$X \xrightarrow{k''} Y$ thỏa mãn $R \cup S$, với $0 < k'' < 1$. Giả thiết các miền trị của các thuộc tính trong A không hạn chế.

Chứng minh:

Để chứng minh bổ đề 4.13 ta chỉ cần xây dựng các quan hệ R và S trên A ứng với cặp độ phụ thuộc k, k' của Y vào X.

Giả sử $k = m_1/n_1$, $k' = m_2/n_2$;

Với m_1, n_1, m_2, n_2 là những số tự nhiên, và $m_1 < n_1$ và $m_2 < n_2$.

Không mất tính tổng quát ta luôn có thể giả sử rằng $m_1 < m_2$ và $m_1 > 1$.

Trước khi xây dựng R và S ta cần lưu ý rằng, nếu trong R mà mỗi nhóm của phân hoạch PART(X) có nhiều hơn 1 bộ (có ít nhất hai bộ bằng nhau trên X) và mỗi nhóm của phân hoạch PART(Y) chỉ chứa đúng 1 bộ (tất cả các bộ đều khác nhau trên Y từng đôi một) thì $POS(X,Y) = \phi$.

Tất nhiên trong những trường hợp như vậy độ phụ thuộc của Y vào X luôn bằng 0. Tức là $k = card(POS(X,Y))/card(r) = 0$.

Bây giờ ta xây dựng R và S:

Khi xây dựng các quan hệ R và S ta luôn coi X, Y là các nhóm thuộc tính và các giá trị x_i, y_i cũng là những nhóm giá trị. Ta sẽ xây dựng quan hệ R có n_1 phần tử, quan hệ S có n_2 phần tử và ta có thể xây dựng để R và S không có phần tử chung.

R			**S**		
X	Y	X	Y		
x_1	y_1	t_1	x_1	y'_1	t'_1
x_2	y_2	t_2	x_2	y'_2	t'_2
...			...		
x_{m1}	y_{m1}	t_{m1}	x_{m1}	y'_{m1}	t'_{m1}
x	y_{m1+1}	t_{m1+1}	x_{m1+1}	y'_{m1+1}	t'_{m1+1}
...			...		
x	y_{n1}	t_{n1}	x_{m2}	y'_{m2}	t'_{m2}
			a	y'_{m2+1}	t'_{m2+1}
			...		
			a	y'_{n2}	t'_{n2}

Với các giá trị trên Y của R và S từng đôi một khác nhau, các giá trị trên X trong R từ x_1 đến x_{m1} từng đôi một khác nhau. Các giá trị trên X trong S từ x_1 đến x_{m2} từng đôi một khác nhau, từ x_1 đến x_{m1} trong R và S giống nhau, a và x khác các x_1, $x_2,\ldots, x_{m2}$ và $a \neq x$.

Với cách xây dựng như vậy khi đó trong R ta có:

$PART(X) = \{\{t_1\},\ldots,\{t_{m1}\}, \{t_{m1+1},\ldots,t_{n1}\}\}$

$PART(Y) = \{\{t_1\},\ldots,\{t_{m1}\}, \{t_{m1+1}\},\ldots\{t_{n1}\}\}$.

Mỗi nhóm của phân hoạch PART(Y) có đúng 1 bộ.

Khi đó $POS(X,Y) = \{t_1,\ldots t_{m1}\}$ và $k = card(POS(X,Y))/card(r) = m_1/n_1$.

Trong S ta có:

$PART(X) = \{\{t'_1\},\ldots,\{t'_{m2}\}, \{t'_{m2+1},\ldots,t'_{n2}\}\}$

$PART(Y) = \{\{t'_1\},\ldots,\{t'_{m2}\}, \{t'_{m2+1}\},\ldots\{t'_{n2}\}\}$.

Mỗi nhóm của phân hoạch PART(Y) trong S có đúng 1 bộ.

Khi đó $POS(X,Y) = \{t'_1,\ldots, t'_{m2}\}$ và $k' = card(POS(X,Y))/card(s) = m_2/n_2$

Trong $R \cup S$ ta có:

$PART(X) = \{\{t_1,t'_1\},\ldots,\{t_{m1},t'_{m1}\}, \{t'_{m1+1}\},\ldots,\{t'_{m2}\},$
$\{t_{m1+1},\ldots,t_{n1}\}, \{t'_{m2+1},\ldots, t'_{n2}\}\}$.

Có $m_2 - m_1$ nhóm chứa đúng một bộ từ t'_{m1+1} đến t'_{m2}: $\{t'_{m1+1}\},\ldots,\{t'_{m2}\}$

Các nhóm khác của PART(X) có ít nhất là hai bộ.

PART(Y) có n_1+n_2 nhóm, mỗi nhóm có đúng một bộ.

Khi đó trong $R \cup S$ ta có:

$POS(X,Y) = \{t'_{m1+1},..., t'_{m2}\}$

và $K'' = card(POS(X,Y))/card(R \cup S) = (m_2 - m_1)/(n_1+n_2)$.

Ta dễ dàng thấy rằng $0 < k'' < 1$.

Ví dụ cho k = 3/5 và k' = 4/7. Khi đó ta xây dựng R và S như sau:

R

X	Y	
1	5	t_1
2	6	t_2
3	7	t_3
4	8	t_4
4	9	t_5

S

X	Y	
1	11	t_1'
2	12	t_2'
3	13	t_3'
5	14	t_4'
6	15	t_5'
6	16	t_6'
6	17	t_7'

R ∪ S

X	Y	
1	5	t_1
2	6	t_2
3	7	t_3
4	8	t_4
4	9	t_5
1	11	t_1'
2	12	t_2'
3	13	t_3'
5	14	t_4'
6	15	t_5'
6	16	t_6'
6	17	t_7'

Khi đó trong R rõ ràng:

$R/IND(X) = \{\{t_1\},\{t_2\},\{t_3\},\{t_4, t_5\}\}$

$R/IND(Y) = \{\{t_1\},\{t_2\},\{t_3\},\{t_4\},\{t_5\}\}$

$POS(X,Y) = \{t_1, t_2, t_3\}$

$k = 3/5.$

Trong S có:

$S/IND(X) = \{\{t_1'\},\{t_2'\},\{t_3'\},\{t_4'\}, \{t_5', t_6', t_7'\}\}$

$S/IND(Y) = \{\{t_1'\},\{t_2'\},\{t_3'\},\{t_4'\},\{t_5'\}, \{t_6'\}, \{t_7'\}\}$

$POS(X,Y) = \{t_1', t_2', t_3', t_4'\}$

$k' = 4/7.$

Trong $R \cup S$:

$(R \cup S)/IND(X) = \{\{t_1, t_1'\}, \{t_2, t_2'\}, \{t_3, t_3'\}, \{t_4, t_5\}, \{t_4'\}, \{t_5', t_6', t_7'\}\}$

$(R \cup S)/IND(Y) = \{\{t_1\},\{t_1'\},\{t_2\},\{t_2'\},\{t_3\},\{t_3'\},\{t_4\},\{t_5\},\{t_4'\},\{t_5'\}, \{t_6'\},t_7'\}\}$

$POS(X,Y) = \{t_4'\}$

$K'' = 1/12$

Bổ đề 4.14

Với mọi cặp số hữu tỷ khác 0, $k = m_1/n_1$, $k'= m_1/n_2$ thuộc đoạn [0,1], mọi tập $X, Y \subseteq A$, ta luôn xây dựng được các quan hệ R, S trên A sao cho:

$X \xrightarrow{k} Y$ thỏa mãn R

$X \xrightarrow{k'} Y$ thỏa mãn S

$X \xrightarrow{0} Y$ thỏa mãn $R \cup S$. Với điều kiện các miền trị của các thuộc tính không có hạn chế.

Chứng minh:

Để chứng minh bổ đề ta sẽ xây dựng các quan hệ R và S trên A ứng với cặp k, k' như đã xây dựng trong bổ đề 4.13. Chú ý vì $m_1 = m_2$ nên ta có R và S như sau:

r				s		
X	Y			X	Y	
x_1	y_1	t_1		x_1	y'_1	t'_1
x_2	y_2	t_2		x_2	y'_2	t'_2
...				...		

x_{m1}	y_{m1}	t_{m1}	x_{m1}	y'_{m1}	t'_{m1}
x	y_{m1+1}	t_{m1+1}	a	y'_{m2+1}	t'_{m2+1}
...			...		
x	y_{n1}	t_{n1}	a	y'_{n2}	t'_{n2}

Với các giá trị trên X,Y được lấy như trong bổ đề 4.13.

Khi đó trong R ta có:

$PART(X) = \{\{t_1\},...,\{t_{m1}\}, \{t_{m1+1},...,t_{n1}\}\}$

$PART(Y) = \{\{t_1\},...,\{t_{m1}\}, \{t_{m1+1}\},...\{t_n1\}\}$.

Mỗi nhóm của phân hoạch PART(Y) có đúng 1 bộ.

Khi đó $POS(X,Y) = \{t_1,.. t_{m1}\}$ và $k = card(POS(X,Y))/card(R) = m_1/n_1$.

Trong S ta có:

$PART(X) = \{\{t'_1\},...,\{t'_{m2}\}, \{t'_{m2+1},...,t'_{n2}\}\}$, lưu ý $m_2 = m_1$.

$PART(Y) = \{\{t'_1\},...,\{t'_{m2}\}, \{t'_{m2+1}\},...\{t'_{n2}\}\}$.

Mỗi nhóm của phân hoạch PART(Y) có đúng 1 bộ.

Khi đó $POS(X,Y) = \{t'_1,..., t'_{m2}\}$ và $k' = card(POS(X,Y))/card(s) = m_2/n_2$

Trong $R \cup S$ ta có:

$PART(X) = \{\{t_1,t'_1\},...,\{t_{m1},t'_{m1}\}, \{t_{m1+1},...,t_{n1}\}, \{t'_{m2+1},..., t'_{n2}\}\}$. Các nhóm chứa ít nhất 2 bộ.

PART(Y) có n_1+n_2 nhóm, mỗi nhóm có đúng một bộ.

Khi đó trong $R \cup S$ ta có:

$POS(X,Y) = \varnothing$ và $K'' = Card(POS(X,Y))/ Card(R \cup S) = 0$.

Bổ đề 4.15

Với mọi cặp số hữu tỷ k, k' thuộc đoạn [0,1], mọi tập X, Y $\subseteq$ U, ta luôn xây dựng được các quan hệ R, S trên A sao cho:

$X \xrightarrow{k} Y$ thỏa mãn R

$X \xrightarrow{k'} Y$ thỏa mãn S

$X \xrightarrow{1} Y$ thỏa mãn $R \cap S$.

Chứng minh:

Để chứng minh bổ đề ta chỉ cần xây dựng các quan hệ R và S trên A ứng với cặp độ phụ thuộc k, k' của Y vào X.

Giả sử $k = m_1/n_1$, $k' = m_2/n_2$; $m_1 < n_1$ và $m_2 < n_2$.

Không mất tính tổng quát ta luôn có thể giả sử rằng $m_1 < m_2$ và $m_1 > 1$.

Trước khi xây dựng R và S ta cần lưu ý và nhắc lại rằng, nếu trong R mà mỗi nhóm của phân hoạch PART(X) có nhiều hơn 1 bộ (có ít nhất hai bộ bằng nhau trên X) và mỗi nhóm của phân hoạch PART(Y) chỉ chứa đúng 1 bộ (tất cả các bộ đều khác nhau trên Y từng đôi một) thì POS(X,Y) = ∅. Tất nhiên trong những trường hợp như vậy độ phụ thuộc của Y vào X luôn bằng 0. Tức là:

$k = card(POS(X,Y))/card(R) = 0.$

Bây giờ ta xây dựng R và S:

R			**S**		
X	Y		X	Y	
x_1	y_1	t_1	x_1	y'_1	t'_1
x_2	y_2	t_2	x_2	y'_2	t'_2
...			...		
x_{m1}	y_{m1}	t_{m1}	x_{m1}	y'_{m1}	t'_{m1}
x	y_{m1+1}	t_{m1+1}	x_{m1+1}	y'_{m1+1}	t'_{m1+1}
...			...		
x	y_{n1}	t_{n1}	x_{m2}	y'_{m2}	t'_{m2}
			a	y'_{m2+1}	t'_{m2+1}
			...		
			a	y'_{n2}	t'_{n2}

Với các giá trị trên Y của R từng đôi một khác nhau, các giá trị trên X trong R từ x_1 đến x_{m1} từng đôi một khác nhau.

Các giá trị trên X trong S từ x_1 đến x_{m2} từng đôi một khác nhau, từ x_1 đến x_{m1} trong R và S giống nhau, a và x khác các $x_1, x_2, \ldots, x_{m2}$.

Các giá trị trên Y trong S từng đôi một khác nhau và $y'_1 = y_1$, $y'_2 = y_2, \ldots, y'_{m1} = y_{m1}$.

Với cách xây dựng như vậy khi đó:

Trong R ta có:

$PART(X) = \{\{t_1\}, \ldots, \{t_{m1}\}, \{t_{m1+1}, \ldots, t_{n1}\}\}$

$PART(Y) = \{\{t_1\}, \ldots, \{t_{m1}\}, \{t_{m1+1}\}, \ldots \{t_{n1}\}\}.$

Mỗi nhóm của phân hoạch PART(Y) có đung 1 bộ.

Khi đó $POS(X,Y) = \{t_1, \ldots t_{m1}\}$ và $k = card(POS(X,Y))/card(R) = m_1/n_1$.

Trong S ta có:

PART(X) = {{t'$_1$},...,{t'$_{m2}$}, {t'$_{m2+1}$,...,t'$_{n2}$}}

PART(Y) = {{t'$_1$},...,{t'$_{m2}$}, {t'$_{m2+1}$},...{t'$_{n2}$}}.

Mỗi nhóm của phân hoạch PART(Y) có đúng 1 bộ.

Khi đó POS(X,Y) = {t'$_1$,..., t'$_{m2}$} và k' = card(POS(X,Y))/card(S) = m_2/n_2

Trong R ∩ S:

X	Y	
x_1	y_1	t_1
x_2	y_2	t_2
...		
x_{m1}	y_{m1}	t_{m1}

PART(X) = {{t_1},...,{t_{m1}}} có m_1 nhóm, mỗi nhóm chứa đúng 1 bộ.

PART(Y) = PART(X).

Khi đó trong R ∩ S ta có:

POS(X,Y) = {t_1, t_2,...t_{m1}} và k''= card(POS(X,Y))/card(R ∩ S) = m_1/m_1 = 1.

Bổ đề 4.16

Với mọi cặp số hữu tỷ khác 0, k = m_1/n_1, k'= m_1/n_2 thuộc đoạn [0,1], mọi tập X, Y ⊆ U, ta luôn xây dựng được các quan hệ R, S trên A sao cho:

$X \xrightarrow{k} Y$ thỏa mãn R

$X \xrightarrow{k'} Y$ thỏa mãn S

$X \xrightarrow{0} Y$ thỏa mãn R ∩ S.

Chứng minh:

Để chứng minh bổ đề ta sẽ xây dựng các quan hệ R và S trên A ứng với cặp k, k' như đã xây dựng trong bổ đề 4.13. Chú ý vì $m_1 = m_2$ nên ta có R và S như sau:

R

X	Y	
x_1	y_1	t_1
x_2	y_2	t_2
...		

S

X	Y	
x_1	y'$_1$	t'$_1$
x_2	y'$_2$	t'$_2$
...		

x_{m1}	y_{m1}	t_{m1}	x_{m1}	y'_{m1}	t'_{m1}
x	y_{m1+1}	t_{m1+1}	a	y'_{m2+1}	t'_{m2+1}
...			...		
x	y_{n1}	t_{n1}	a	y'_{n2}	t'_{n2}

Với các giá trị trên X,Y được lấy như trong bổ đề 4.13.

Khi đó trong R ta có:

PART(X) = $\{\{t_1\},...,\{t_{m1}\}, \{t_{m1+1},...,t_{n1}\}\}$

PART(Y) = $\{\{t_1\},...,\{t_{m1}\}, \{t_{m1+1}\},...\{t_n1\}\}$.

Mỗi nhóm của phân hoạch PART(Y) có đúng 1 bộ.

Khi đó POS(X,Y) = $\{t_1,...t_{m1}\}$ và k = card(POS(X,Y))/card(R) = m_1/n_1.

Trong S ta có:

PART(X) = $\{\{t'_1\},...,\{t'_{m2}\}, \{t'_{m2+1},...,t'_{n2}\}\}$, lưu ý $m_2 = m_1$.

PART(Y) = $\{\{t'_1\},...,\{t'_{m2}\}, \{t'_{m2+1}\},...\{t'_{n2}\}\}$.

Mỗi nhóm của phân hoạch PART(Y) có đúng 1 bộ.

Khi đó POS(X,Y) = $\{t'_1,..., t'_{m2}\}$ và k' = card(POS(X,Y))/card(S) = m_2/n_2

Trong $R \cap S$:

PART(X) = $\{\{t_1,t'_1\},...,\{t_{m1},t'_{m1}\}, \{t_{m1+1},...,t_{n1}\}, \{t'_{m2+1},..., t'_{n2}\}\}$.

Các nhóm chứa ít nhất 2 bộ.

PART(Y) có n_1+n_2 nhóm, mỗi nhóm có đúng một bộ.

Khi đó trong $R \cap S$ ta có:

POS(X,Y) = $\varnothing$ và K''= card(POS(X,Y))/card($R \cap S$) = 0.

Bổ đề 4.17

Nếu $X \xrightarrow{k} Y$ thì $XZ \xrightarrow{k'} YZ$ với $k \le k'$ và X, Y, Z $\subseteq$ A.

Chứng minh:

Theo bổ đề 4.6 ta có:

$$card(\text{POS}(X,Y)) \le card(\text{POS}(XZ,YZ))$$

$$\Rightarrow k = \frac{\text{card(POS}(X,Y))}{\text{card}(r)} \le \frac{\text{card(POS}(XZ,YZ))}{\text{card}(r)} = k'$$

Bổ đề 4.18

Nếu $X \rightarrow Y$ và $Y \xrightarrow{k} Z$ thì $X \xrightarrow{k'} Z$ với $k \le k'$

Chứng minh:

Vì $X \rightarrow Y$ nên IND(X) $\subseteq$ IND(Y), điều này có nghĩa là mỗi nhóm của *PART*(*X*) nằm gọn trong một nhóm của *PART*(*Y*), tức là $\text{POS}(X, Y) = r$. Theo bổ đề 4.11 ta có:

$$k = \frac{\text{card}(\text{POS}(Y, Z))}{\text{card}(r)} \le \frac{\text{card}(\text{POS}(X, Z))}{\text{card}(r)} = k'$$

Bổ đề 4.19

Nếu $X \xrightarrow{k} Y$ và $Y \rightarrow Z$ thì $X \xrightarrow{k'} Z$ với $k \le k'$

Chứng minh:

Ta sẽ chứng minh POS(X,Y) $\subseteq$ POS(X,Z)

lấy $t \in P(X,Y) \Rightarrow [t]]_{IND(X)} \subseteq [t]_{IND(Y)}$

vì $Y \rightarrow Z \Rightarrow [t]_{IND(Y)} \subseteq [t]_{IND(Z)} \Rightarrow [t]_{IND(X)} \subseteq [t]_{IND(Z)} \Rightarrow t \in \text{POS}(X,Z)$.

Vậy k = card(POS(X,Y))/ card(r) $\le$ card(POS(X,Z))/card(r) = k'

Bổ đề 4.20

Nếu $X \xrightarrow{k} YZ$ thì $X \xrightarrow{k'} Y$ và $X \xrightarrow{k''} Z$ với $k \le k', k''$.

Chứng minh:

Vì vai trò của *Y* và *Z* như nhau nên ta chỉ cần chứng minh cho *Y*. Tức ta phải chứng minh:

K = card(POS(X,YZ))/card(r) $\le$ card(POS(X,Y))/card(r) = k'

$\Leftrightarrow \text{POS}(X, YZ) \subseteq \text{POS}(X, Y)$

Lấy $t \in \text{POS}(X, YZ)$ ta có: $[t]_{IND(X)} \subseteq [t]_{IND(YZ)}$ vì IND(YZ) là quan hệ con của IND(Y) nên $[t]_{IND(YZ)} \subseteq [t]_{IND(Y)} \Rightarrow [t]I_{ND(X)} \subseteq [t]_{IND(Y)} \Rightarrow t \in \text{POS}(X,Y)$ (đpcm)

4.6. ĐỊNH NGHĨA PHỤ THUỘC HÀM THEO IND(X), IND(Y)

Cho quan hệ R trên tập thuộc tính $A = \{A_1, A_2, \ldots, A_n\}$; $X, Y \subset A$.

Định nghĩa 4.7 Phụ thuộc hàm trong R

Ta nói X xác định phụ thuộc hàm Y (Y phụ thuộc hàm vào X) trong R, ký hiệu $X \rightarrow Y$ nếu IND(X) $\subseteq$ IND(Y).

Khi đó ta cũng nói R thỏa mãn phụ thuộc hàm $X \rightarrow Y$.

Chúng ta dễ dàng chứng minh rằng định nghĩa 4.7 và định nghĩa 2.14 trong chương 2 về phụ thuộc hàm trong một quan hệ R là tương đương. Để chứng minh điều này ta nhắc lại định nghĩa 2.14 trong chương 2:

Định nghĩa 2.14 trong chương 2 nói rằng:

Ta nói X xác định phụ thuộc hàm Y, ký hiệu $X \rightarrow Y$ trong R nếu với mọi t và t' của R mà t, t' bằng nhau trên tập X thì chúng cũng bằng nhau trên tập Y, tức là $\forall t, t' \in R$ nếu t. X = t'. X thì t. Y = t'. Y.

Định lý 4.1

Hai định nghĩa 4.7 và 2.14 là tương đương. Tức từ định nghĩa này suy ra định nghĩa kia và ngược lại.

Chứng minh:

Cho quan hệ R trên tập thuộc tính A, viết R(A)

+ Định nghĩa 2.14 ⇒ Định nghĩa 4.7

Ta có R thỏa mãn phụ thuộc hàm $X \rightarrow Y$ theo Định nghĩa 2.14 nghĩa là mọi t, t' của R nếu t.X = t'.X thì t.Y = t'.Y => nếu t IND(X) t' thì t IND(Y) t'.

Vậy IND(X) $\subseteq$ IND(Y) và khi đó R thỏa mãn phụ thuộc hàm $X \rightarrow Y$ theo định nghĩa 4.7.

+ Định nghĩa 4.7 => Định nghĩa 2.14

Ta có R thỏa mãn $X \rightarrow Y$ theo định nghĩa 4.7 tức là IND(X) $\subseteq$ IND(Y).

Lấy t, t' thuộc R và giả sử t.X = t'.X ta phải chứng minh t.Y = t'.Y

Vì t.X = t'.X nên t IND(X) t'

và vì IND(X) $\subseteq$ IND(Y) nên t IND(Y) t' => t.Y = t'.Y. Bổ đề được chứng minh.

4.7. RÚT GỌN HỆ TIN

Cho hệ tin S = (U, A, V, f); U là vũ trụ, A là tập thuộc tính.

Khi đó PART(A) = $\{E_1, E_2, ..., E_k\}$.

Sau đây ta sẽ xét quá trình rút gọn tập thuộc tính A bằng cách bỏ khỏi A những *thuộc tính thừa*. Thuộc tính $B \in A$ được gọi là *thừa* nếu bỏ B khỏi A không làm thay đổi phân hoạch PART(A). Nói cách khác *thuộc tính B được gọi là thừa* trong A nếu PART(A) = PART(A-B) = $\{E_1, E_2, ..., E_k\}$.

Thực hiện lặp quá trình loại bỏ các thuộc tính thừa của A ta sẽ nhận được tập tối thiểu k mà PART(k) = PART(A). Tập k như vậy được gọi là một rút gọn (reduct) của A.

Ví dụ 4.12:

Cho hệ tin S = (U, A, V, f); A = {Massv, Hoten}; U = {o_1, o_2, o_3, o_4} với hàm thông tin như trong bảng sau:

U	Masv	Hoten
o_1	Sv1	Khang
o_2	Sv2	Khang
o_3	Sv3	Thịnh
o_4	Sv4	Vượng

Khi đó ta có:

PART(A) = {E_1, E_2, E_3, E_4} = {{o_1}, {o_2}, {o_3}, {o_4}}

và dễ dàng thấy rằng thuộc tính Hoten là thừa trong A

vì PART(A) = PART(A - Hoten) = {E_1, E_2, E_3, E_4}. Như vậy khái niệm thừa của một thuộc tính B ở đây ta xét theo quan điểm phân loại tập đối tượng U theo những thông tin trong tập thuộc tính A.

Định nghĩa 4.10 Rút gọn hệ tin

Cho hệ tin S = (U, A). Khi đó ta có PART(A) = {E_1, E_2,..., E_k}

Tập thuộc tính reduct $\subseteq$ A được gọi là *rút gọn của A* (đôi khi ta còn gọi là *rút gọn của hệ tin S*) nếu reduct là tập tối thiểu mà PART(reduct) = PART(A).

Nói cách khác reduct $\subseteq$ A được gọi là *rút gọn của A* nếu:

(1) PART(A) = PART(reduct).

(2) reduct tối thiểu

Lưu ý:

reduct tối thiểu theo nghĩa nếu bớt khỏi reduct một thuộc tính thì PART(reduct) $\neq$ PART(A). Reduct là tập bé nhất có thể không làm thay đổi phân hoạch trên U.

Ví dụ 4.13:

Xét ví dụ 4.9 thì reduct của hệ tin là Masv. Vì Masv chỉ có 1 thuộc tính nên tối thiểu và tất nhiên PART(Masv) = PART(A).

Ví dụ 4.14:

Xét hệ tin S = (U, A); với U = R = {t_1, t_2, t_3, t_4, t_5, t_6, t_7} là quan hệ trên A = {Q, B, C, D, E, H, I, J, L, M, N} và hàm thông tin được cho trong bảng sau:

R

	Q	B	C	D	E	H	I	J	L	M	N
t_1	0	1	2	3	3	4	2	1	2	3	4
t_2	0	0	0	0	3	3	3	3	4	0	3
t_3	1	2	1	1	2	2	4	2	6	1	1
t_4	4	4	4	5	5	5	5	5	5	5	5
t_5	1	3	3	3	4	7	6	7	8	9	0
t_6	2	5	6	4	5	6	8	6	7	8	9
t_7	2	6	5	2	2	1	7	8	3	4	4

Khi đó ta có các reduct của hệ tin là:

$reduct_1$ = B; $reduct_2$ = C; $reduct_3$ = H; $reduct_4$ = I; $reduct_5$ = J;

$reduct_6$ = L; $Reduct_7$ = M; $reduct_8$ = QD;

$reduct_9$ = QE; $reduct_{10}$ = DE; $reduct_{11}$ = NQ;...

Định lý 4.2

Cho hệ tin dạng quan hệ S = (r, A) khi đó reduct $\subseteq$ A là một rút gọn của A khi và chỉ khi reduct là khóa (tối thiểu) của quan hệ r.

Chứng minh:

a. Ta chứng minh một reduct là một khóa.

Giả sử r = $\{t_1, t_2, ..., t_m\}$; k $\subseteq$ A là một rút gọn của A. Vì S là hệ tin dạng quan hệ nên không có hai phần tử của r giống nhau trên toàn bộ A, nói cách khác $t_i.A \neq t_j.A$ $\forall$i khác j. Tức là PART(A) = $\{\{t_1\}, \{t_2\}, ..., \{t_m\}\}$; mỗi nhóm là một phần tử của r. Vì k là rút gọn của A nên k là tập tối thiểu giữ nguyên phân hoạch của r, nghĩa là k là tập tối thiểu của A mà $t_i.k \neq t_j.k$. Vậy k là khóa của r.

b. Ta chứng minh khóa k của r là một rút gọn của A.

Vì k là khóa của r nên k là tập tối thiểu mà $t_i.k \neq t_j.k$ nên k là tập tối thiểu mà r/IND(k) = $\{\{t_1\}, \{t_2\}, ..., \{t_m\}\}$ = r/IND(A). Vậy k là một rút gọn của A.

Từ định lý 4.2 ta có thuật toán tìm một rút gọn, tất cả các rút gọn của hệ tin S = (U, A) như sau:

Thuật toán 4.1 Tìm các rút gọn của hệ tin S = (U, A)

a. Trường hợp S là hệ tin dạng quan hệ, các đối tượng từng đôi một khác nhau trên A, khi đó thuật toán tìm một rút gọn, tìm tất cả các rút gọn được tiến hành như các thuật toán tìm khóa dựa vào ma trận bằng nhau E, ma trận khác nhau D đã xét trong quan hệ r.

b. Trường hợp S là hệ tin dạng không phải quan hệ; tức có các đối tượng giống nhau trên tập thuộc tính A. Khi đó thực hiện hai bước:

Bước 1: Trong mỗi nhóm đối tượng giống nhau trên A chọn ra một đại diện

Bước 2: Gọi r là tập các đại diện như vậy và thực hiện như trong phần a của thuật toán 4.1.

Ví dụ 4.15:

Cho hệ tin dạng quan hệ như sau:

R

H	B	C	D
1	1	1	2
1	2	3	2
1	3	3	2
2	2	2	2
2	3	3	3
3	3	3	3
1	2	1	3

Vậy R = $\{t_1, t_2, t_3, t_4, t_5, t_6, t_7\}$ = U là tập đối tượng, không có hai đối tượng giống nhau trên A = {E, B, C, D}.

Tìm một rút gọn bằng thuật toán tìm khóa 2.3. Tìm tất cả các rút gọn bằng thuật toán 2.4.

a. Dùng thuật toán 2.3 tìm một rút gọn của A:

- Tính nửa trên của ma trận bằng nhau E

HD	HD	D	ϕ	ϕ	HC
	HCD	BD	C	C	HB
		D	BC	BC	H
			H	ϕ	B
				BCD	D
					D

- Họ các tập cực đại của E là M = {HCD, BCD, HB}

- Đặt k = HBCD

- Lần lượt thử loại bỏ các thuộc tính trong k:

+ Thử loại H tập còn lại BCD là tập con của một tập trong M nên H không bỏ được và k = HBCD.

+ Thử bỏ B tập còn lại HCD là tập con của tập trong M nên B không bỏ được và k = ABCD.

+ Thử loại C tập còn lại ABD không là tập con của tập nào trong M nên k = HBD.

+ Thử loại D tập còn lại HB là tập con của một tập trong M nên D không loại được và k = HBD là một rút gọn của A.

b. Dùng thuật toán 2.4 để tìm tất cả các rút gọn của A:

- Tính nửa trên của ma trận khác nhau D

BC	BC	HBC	HBCD	HBCD	BD
	B	HC	HBD	HBD	CD
		HBC	HD	HD	BCD
			BCD	HBCD	HCD
				H	HBC
					HBC

- Đặt $f = \wedge(\vee d_{ij})$ và rút gọn ta được $f = B \wedge H \wedge (C \vee D) = HBC \vee HBD$. Trong đó $H \wedge B \wedge C$ viết thành HBC.

- Vậy A có hai rút gọn là $reduct_1$ = HBC; $reduct_2$ = HBD.

Ví dụ 4.16:

Cho hệ tin dạng không phải quan hệ (có các đối tượng giống nhau trên A) như sau:

	H	B	C	D
o_1	1	1	1	2
o_2	1	2	3	2
o_3	1	3	3	2
o_4	2	2	2	2
o_5	2	3	3	3
o_6	3	3	3	3
o_7	1	2	1	3
o_8	1	2	1	3
o_9	1	2	1	3

Vậy $U = \{o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8, o_9\}$ là tập các đối tượng; $A = \{H, B, C, D\}$ là tập các thuộc tính, có ba đối tượng $\{o_7, o_8, o_9\}$ giống nhau trên $A = \{H, B, C, D\}$.

Chúng ta cần lưu ý rằng trong phân hoạch $PART(A) = U/IND(A) = \{\{o_1\}, \{o_2\}, \{o_3\}, \{o_4\}, \{o_5\}, \{o_6\}, \{o_7, o_8, o_9\}\}$ các đối tượng giống nhau trên A thuộc cùng một nhóm.

Theo thuật toán 4.1.b ta chọn trong nhóm các đối tượng $\{o_7, o_8, o_9\}$ giống nhau trên U, đối tượng o_7 làm đại diện. Khi đó xét hệ tin dạng quan hệ $S = (r, U)$; với $r = \{o_1, o_2, o_3, o_4, o_5, o_6, o_7\}$.

Tiếp tục dùng các thuật toán 2.3 ta được một rút gọn là HBD; Dùng thuật toán 2.4 ta được hai rút gọn là HBC và HBD.

Lưu ý:

Việc rút gọn các thuộc tính có thể thực hiện trên một tập thuộc tính X bất kỳ.

Khi đó thuộc tính x của X được gọi là thừa trong X nếu $U/IND(X) = U/IND(X-x)$.

4.8. TẬP THÔ TRONG HỆ TIN

Cho hệ tin $S = (U, A)$; $X \subseteq A$

Đặt $P = U/IND(X) = \{p_1, p_2, \ldots, p_k\}$ là phân hoạch các đối tượng giống nhau trên tập thuộc tính X

Khi đó họ các tập rõ $RO = \{\phi, p_1, \ldots, p_k, \cup p_i, U\}$.

Họ các tập thô $THO = 2^U - RO$ là các tập thô ứng với P.

Ví dụ 4.17:

Cho hệ tin như trong ví dụ 4.16 và xét $X = \{H, B\}$.

Xét xem tập $\Omega = \{o_1, o_2, o_3\}$ thô hay rõ ứng với phân hoạch PART(X).

Ta có hai cách để xét tập $\Omega = \{o_1, o_2, o_3\}$ thô hay rõ ứng với phân hoạch PART(X)

Cách 1: Liệt kê hết các tập rõ ứng với phân hoạch PART(X) rồi xét xem $\Omega = \{o_1, o_2, o_3\}$ có là tập rõ không, nếu không nó là tập thô.

Ta có:

$PART(X) = \{\{o_1\}, \{o_2, o_7, o_8, o_9\}, \{o_3\}, \{o_4\}, \{o_5\}, \{o_6\}\}$

$= \{p_1, p_2, p_3, p_4, p_5, p_6\} = P$

$RO = \{\phi, p_1, \ldots, p_k, \cup p_i, U\}$

$= \{\phi, \{o_1\}, \{o_2, o_7, o_8, o_9\}, \{o_3\}, \{o_4\}, \{o_5\}, \{o_6\}, \{\{o_1, o_2, o_7, o_8, o_9\}, o_1, o_3\}, \{o_1, o_4\}, \{o_1, o_5\}, \{o_1, o_6\}, \{o_2, o_7, o_8, o_9, o_3\}, \{o_2, o_7, o_8, o_9 o_4\}, \{o_2, o_7, o_8, o_9 o_5\}, \{o_2, o_7, o_8, o_9 o_6\}, \{o_3, o_4\}, \{o_3 o_5\}, \{o_3 o_6\}, \{o_4, o_5\}, \{o_4 o_6\}, \{o_5, o_6\}, \ldots, U\}$.

Ta dễ dàng thấy rằng trong RO không có tập $\Omega = \{o_1, o_2, o_3\}$ vì tập $\Omega = \{o_1, o_2, o_3\}$ không thể biểu thị qua hợp các tập trong PART(X).

Vậy $\Omega = \{o_1, o_2, o_3\}$ là tập thô.

Cách 2: Tính xấp xỉ trên dưới của $\Omega = \{o_1, o_2, o_3\}$ và so sánh chúng với nhau.

$\Omega^P = \{o_1, o_2, o_3, o_7, o_8, o_9\}$ và $\Omega_P = \{o_1, o_3\}$.

Vậy $\Omega = \{o_1, o_2, o_3\}$ là thô ứng với PART(X) vì xấp xỉ trên khác xấp xỉ dưới.

4.9. CÁC BÀI TOÁN ỨNG DỤNG CỦA HỆ TIN

4.9.1. Sử dụng luật kết hợp để tìm qui luật những sinh viên ở khu vực nào thì khả năng đậu đại học là cao nhất

Cho hệ quyết định $S = (U, A \cup d)$; $U = \{o_1, o_2, ..., o_m\}$ là các ứng viên thi đại học; $A = \{T, L, H, kv1, kv2, kv3\}$ là tập thuộc tính gồm các thuộc tính điểm toán (T), điểm lý (L), điểm hóa (H), thuộc các khu vực 1, 2, 3 (kv1, kv2, kv3), thuộc tính d là thuộc tính quyết định. Nếu d = 1 thì ứng viên là đậu, d = 0 ứng viên không đậu. Xét bảng kết quả của m sinh viên thi đại học:

Bảng_Diem: Kết quả thi đại học

	T	L	H	kv1	kv2	kv3	d
o_1	7	5	5	0	1	0	0
o_2	8	9	9	1	0	0	1
...	...	...	...				...
o_m	7	7	7	0	0	1	1

Chú ý rằng mỗi ứng viên chỉ có thể thuộc một khu vực trong ba khu vực. Trong cột khu vực một ta ghi 1 nếu ứng viên thuộc khu vực một, ngược lại ghi không. Tương tự cho các khu vực hai, ba.

Bây giờ xét hệ khai thác dữ liệu trên các thuộc tính kv1, kv2, kv3, d:

Bảng_D: Kết quả thi đại học của khu vực

	kv1	kv2	kv3	d
o_1	0	1	0	0
o_2	1	0	0	1
...				...
o_m	0	0	1	1

Khi đó khả năng đậu đại học của các khu vực được tính dựa vào độ tin cậy của các luật kết hợp:

$CF(kv1 \to d)$; $CF(kv2 \to d)$; $CF(kv3 \to d)$

4.9.2. Sử dụng luật kết hợp để tìm qui luật những môi trường sống thế nào thì khả năng sống thọ của con người là cao

Cho hệ tin từ cuộc điều tra về môi trường và tuổi thọ của con người tương ứng. Từ hệ tin này ta có thể dùng luật kết hợp để tìm độ phụ thuộc của tuổi thọ vào môi trường sống.

Giả sử ta có hệ tin S = (U, A); U = $\{o_1, o_2, \ldots, o_m\}$ là tập các ứng viên được khảo sát.

A = {K, Q, N, M, T, tho} gồm các thuộc tính K - môi trường không khí; Q - quyền con người; N - nền văn hóa; M - mức sống, T - tinh thần, tho là tuổi thọ.

Bảng_DT: Kết quả điều tra

	K	Q	N	M	T	tho
o_1	tốt	bình đẳng	cao	0	1	0
o_2	0	0	cao	cao	0	1
...	...	...	...			...
o_m	tốt	0	0	0	0	1

Mỗi thuộc tính của A chỉ có hai giá trị là tốt không tốt hay cao không cao, bình đẳng không bình đẳng.

Trong cột tho ta ghi là 1 nếu đối tượng quan sát thọ cao, ngược lại ghi 0.

Ta đưa bảng điều tra về hệ khai thác dữ liệu (Bảng DT: Hệ khai thác dữ liệu):

Mỗi thuộc tính điều kiện có hai giá trị 0 và 1: K là không khí có giá trị 1 nếu môi trường không khí tốt, ngược lại có giá trị 0. Q có giá trị 1 nếu con người có quyền bình đẳng, ngược lại có giá trị 0...

Bảng_DT: Hệ khai thác dữ liệu

	K	Q	N	M	T	tho
o_1	1	1	1	0	1	0
o_2	0	0	1	1	0	1
...	...	...	...			...
o_m	1	0	0	0	0	1

Bảng là hệ khai thác dữ liệu ta có thể dùng độ tin cậy CF(s $\rightarrow$ tho) để tính độ phụ thuộc của tuổi thọ vào môi trường s; s là tập con của A.

CÂU HỎI VÀ BÀI TẬP CHƯƠNG 4

4.1 Cho hệ khai thác dữ liệu như sau:

Bảng_BT4.1: Hệ khai thác dữ liệu

	i_1	i_2	i_3	i_4	i_5
o_1	1	1	1	0	1
o_2	1	0	1	1	1
o_3	1	1	1	1	0
o_4	1	1	0	1	1
o_5	1	1	1	1	0
o_6	1	1	1	0	0
o_7	1	1	0	0	0
o_8	1	0	0	0	0

Đặt $S = \{i_1, i_3\}$, $S' = \{i_4, i_5\}$; $X = \{o_2, o_3, o_4\}$

a. Tính độ tin cậy $CF(S \rightarrow S')$.

b. Tính k trong $S \xrightarrow{k} S'$.

c. So sánh độ tin cậy $CF(S \rightarrow S')$ và độ phụ thuộc k trong $S \xrightarrow{k} S'$. Hai giá trị này có bắt buộc bằng nhau không vì sao?

d. Xét xem X là thô hay rõ ứng với PART(S). Liệt kê tất cả tập rõ ứng với PART(S).

e. Xét xem X là thô hay rõ ứng với PART(S'). Liệt kê tất cả tập rõ ứng với PART(S').

4.2 Trong quan hệ và lược đồ quan hệ ta có hệ tiên đề Armstrong. Trong hệ tin ta có hệ tiên đề Armstrong không? các tính chất của Armstrong có đúng cho phụ thuộc hàm trong hệ tin không?

4.3 Tìm các rút gọn của hệ tin trong bài **4.1**

4.4 Cho hệ tin S = (U, A, V, f). Nhân của hệ tin S, ký hiệu Core(S) là giao của các rút gọn của S: Core(S) = $\cap$ reduct.

a. Cho ví dụ hệ tin S = (U, A, V, f) để Core(S) = ϕ

b. Cho ví dụ hệ tin S = (U, A, V, f) để Core(S) $\neq$ ϕ

c. Cho hệ tin như *bảng_BT4.4*: Tính các rút gọn của S

d. Tính Core(S)

e. Tính PART(CD), PART(DE)

f. Liệt kê các tập rõ ứng với các phân hoạch PART(CD), PART(DE).

g. Tính độ phụ thuộc k trong các phụ thuộc hàm sau:

$CD \xrightarrow{k} DE$; $B \xrightarrow{k} H$; $B \xrightarrow{k} DE$; $D \xrightarrow{k} E$; $H \xrightarrow{k} D$

Bảng_BT4.4

	B	C	D	E	H
o_1	1	1	1	0	8
o_2	2	0	1	1	7
o_3	3	1	2	1	6
o_4	4	1	0	1	5
o_5	5	2	1	2	4
o_6	6	1	3	0	3
o_7	7	2	2	0	2
o_8	8	0	0	0	1

4.5 Cho hệ tin như trong bảng Bang_BT4.4 của bài 4.4. Hãy tìm các rút gọn của hệ tin.

4.6 Cho bảng kết quả thi đại học của các sinh viên như sau:

Bảng_BT4.6: Bảng điểm kết quả thi đại học

	T	L	H	kv1	kv2	kv3	d
o1	7	5	5	0	1	0	0
o2	8	9	9	1	0	0	1
o3	8	8	8	1	0	0	1
o4	7	7	7	1	0	0	1
o5	3	4	5	0	0	1	0
o6	2	5	5	0	1	0	0
o7	7	7	7	0	0	1	1
o7	3	4	6	0	0	1	0

a. Tính xem khả năng đậu đại học của khu vực nào cao nhất?

b. Nếu qui định loại giỏi là $(T+L+H)/3 \geq 8$;

Khá là $(T+L+H)/3 \geq 7$ và $(T+L+H)/3 < 8$

Trung bình là $(T+L+H)/3 \geq 5$ và $(T+L+H)/3 < 7$

Kém là $(T+L+H)/3 < 5$.

Hãy tính các khả năng đậu đại học của từng loại học lực, dựa vào bảng điểm trên.

4.7 Cho hệ quyết định S = (U, A $\cup$d) như trong Bảng_Diem bài 4.6. Hãy thực hiện các yêu cầu sau:

a. Tính k trong phụ thuộc hàm $T \xrightarrow{k} d$

b. Tính k trong phụ thuộc hàm $\{T, L, H\} \xrightarrow{k} d$

4.8 Cho hệ KTDL = (O, I, V, f).

Chứng minh rằng nếu s và s' là những tập phổ biến thì s $\cap$ s', s \ s' là những tập phổ biến.

4.9 Cho hệ KTDL = (O, I, V, f).

Chứng minh rằng họ các tập phổ biến FS với ngưỡng minsupp là một giàn giao trong I.

4.10 Cho hệ tin S = (U, A); X,Y $\subseteq$ A.

Ta định nghĩa phụ thuộc hàm trong hệ tin như sau:

$X \rightarrow Y$ nếu $IND(X) \subseteq IND(Y)$.

Chứng minh rằng các phụ thuộc hàm trong hệ tin thỏa mãn 3 tính chất của hệ tiên đề Armstrong.

KÝ HIỆU (CHỮ VIẾT TẮT)

$A = \{A_1, A_2, \ldots A_n\}$	Tập thuộc tính hay lược đồ quan hệ
$Apr = (U, E)$	Không gian xấp xỉ Pawlak
CSDL	Cơ sở dữ liệu
F	Tập phụ thuộc hàm
$W = \langle A, F \rangle$	Sơ đồ quan hệ
$W \dashrightarrow \{W_1, W_2, \ldots, W_k\}$	Phân rã W thành các sơ đồ con $W_1, W_2, \ldots, W_k$
$\pi_{U_i}(F)$	Chiếu của tập phụ thuộc hàm F lên tập thuộc tính U_i
$R(A_1, A_2, \ldots A_n)$	Quan hệ R trên tập thuộc tính $\{A_1, A_2, \ldots A_n\}$
$\delta_E(R)$	Chọn trong R những bộ thỏa mãn E
$R \cup S$	Hợp của R và S
$R \cap S$	Giao của R và S
$R \times S$	Tích Đề-các của R và S
$R|><|S$	Nối tự nhiên của R và S
$R|>< S$	Nối nửa của R và S
$R[U_i]$	Chiếu của quan hệ R lên tập thuộc tính U_i
R/S	Phép chia của quan hệ R cho quan hệ S
$R|><|_{\theta_j}S$	Nối theo têta của R và S
$R \setminus S$	Hiệu của quan hệ R với quan hệ S
$U = \{o_1, o_2, \ldots o_m\}$	Tập đối tượng
$E = \{E_1, E_2, \ldots E_k\}$	Phân hoạch của U
hoặc $P = \{P_1, P_2, \ldots P_k\}$	
U/R	Phân hoạch tương đương của U được tạo ra từ quan hệ tương đương R
X^E	Xấp xỉ trên của tập đối tượng X ứng với phân hoạh E
X_E	Xấp xỉ dưới của X ứng với phân hoạch E

$\mu_X^E : U \to [0,1]$	Hàm thuộc thô của tập X ứng với phân hoạch E
S = (U, A, V, f)	Hệ tin
S = (U, A)	Hệ tin mà tập giá trị thuộc tính V và hàm thông tin f đã xác định.
IND(X)	Quan hệ bất khả phân biệt trên tập thuộc tính X
CF(X $\to$ Y)	Độ tin cậy của luật kết hợp X $\to$ Y
$X \xrightarrow{k} Y$	Y phụ thuộc hàm vào X với độ phụ thuộc k
PTH	Phụ thuộc hàm

TÀI LIỆU THAM KHẢO

[1]. J. Ulman, *Nguyên lý các hệ cơ sở dữ liệu và cơ sở tri thức. Tập 1, 2.* NXB Thống kê, 1998.

[2]. M. Tamer Özsu, Patrick Valduriez, *Nguyên lý các hệ cơ sở dữ liệu phân tán. Tập 1, 2.* NXB Thống kê, 2000.

[3]. Nguyễn Xuân Huy, *Các phụ thuộc logic trong cơ sở dữ liệu.* NXB Thống kê, 2006

[4]. Vũ Đức Thi, *Cơ sở dữ liệu - kiến thức và thực hành.* NXB Thống kê, 1997.

[5]. Đỗ Trung Tuấn, *Cơ sở dữ liệu.* NXB Giáo dục 1998.

[6]. Hồ Thuần, Nguyễn Quang Vinh, Nguyễn Xuân Huy dịch. *Nhập môn các hệ cơ sở dữ liệu.* NXB Thống kê, 1986.

[7]. Nguyễn Bá Tường. *Nhập môn cơ sở dữ liệu phân tán.* NXB Khoa học và Kỹ thuật, 2005.

[8]. Nguyễn Bá Tường. *Cơ sở dữ liệu - Lý thuyết và thực hành.* NXB Khoa học và Kỹ thuật, 2005.

[9]. Lê Tiến Vương. *Nhập môn cơ sở dữ liệu quan hệ.* NXB Khoa học và Kỹ thuật, 1997.

[10]. Nguyễn Văn Hoàng. *Tự học Microsoft SQL Server 7.0.* NXB Thống kê, 2000.

[11]. G. Coulouris, J. Dollimore, Tim Kindberg. *Distributed Systems - Concepts and Design.* NXB Addison –Wesley Publishing Company, 1994.

[12]. David W. Embley. *Object Database Development-concepts and principles.* NXB Addison Wesley Longman, Inc.1997.

[13]. Gary J. Nutt. *Centralized and distributed operating systems.* NXB Prentice - Hall International, Inc.

[14]. J. Martin. *Organizacja Baz Danych.* NXB Khoa học Quốc gia Warszawa, 1983.

[15]. Han Yin Shi, *A classification study of Rough Sets generalization.* A thesis Master of science in Lakehead University 2004.

[16]. Han Yin Shi, *A classification study of Rough Sets generalization Thander Bay*, Ontario 2004.

[17]. Z. Pawlak, *Rough Sets - Theorycal Aspect of Reasoning about Data*, Kluwer Akademic publishers 1991.

[18]. Z. Pawlak, *Systemy informacyjne- Podstawy teoretyczne*. Warszawa 1983

MỤC LỤC

CƠ SỞ DỮ LIỆU
QUAN HỆ & ỨNG DỤNG

Chịu trách nhiệm xuất bản
NGUYỄN THỊ THU HÀ

Biên tập: NGÔ MỸ HẠNH
TRỊNH THU CHÂU
Trình bày sách: BÙI CHÂU LOAN
Sửa bản in: TRỊNH THU CHÂU
Thiết kế bìa: TRẦN HỒNG MINH

NHÀ XUẤT BẢN THÔNG TIN VÀ TRUYỀN THÔNG
Trụ sở: 18 Nguyễn Du, TP. Hà Nội
ĐT Biên tập: 04.35772143, 04.35772145 ĐT Phát hành: 04.35772138
E-mail: nxb.tttt@mic.gov.vn Fax: 04.35772037
Website: www.nxbthongtintruyenthong.vn
Chi nhánh TP. Hồ Chí Minh: 8A đường D2, P25, Quận Bình Thạnh, TP. Hồ Chí Minh
Điện thoại: 08.35127750, 08.35127751 Fax: 08.35127751
E-mail: cnsg.nxbtttt@ mic.gov.vn
Chi nhánh TP. Đà Nẵng: 42 Trần Quốc Toản, Quận Hải Châu, TP. Đà Nẵng
Điện thoại: 0511.3897467 Fax: 0511.3843359
E-mail: cndn.nxbtttt@ mic.gov.vn

In 1.000 bản, khổ 19x27cm tại Công ty In Hải Nam
Số đăng ký kế hoạch xuất bản: 88-2011/CXB/16-03/TTTT
Số quyết định xuất bản: 51/QĐ-NXB TTTT ngày 17 tháng 3 năm 2011
In xong và nộp lưu chiểu tháng 03 năm 2011.