

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC QUY NHƠN**

ThS. Phạm Văn Phú

**TÀI LIỆU GIẢNG DẠY
NHẬP MÔN CƠ SỞ DỮ LIỆU**

(TRÌNH ĐỘ ĐẠI HỌC, NGÀNH CÔNG NGHỆ THÔNG TIN)

Quy Nhơn, 6/2016

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC QUY NHƠN**

ThS. Phạm Văn Phú

**TÀI LIỆU GIẢNG DẠY
NHẬP MÔN CƠ SỞ DỮ LIỆU**
(TRÌNH ĐỘ ĐẠI HỌC, NGÀNH CÔNG NGHỆ THÔNG TIN)

SỐ TÍN CHỈ: 3
(LÝ THUYẾT: 45, THỰC HÀNH : 0, THẢO LUẬN: 0)

Quy Nhơn, 6/2016

Mục lục

Chương 1: CÁC KHÁI NIỆM VÀ MÔ HÌNH DỮ LIỆU CƠ BẢN.....	5
MỞ ĐẦU	5
§1 CÁC KHÁI NIỆM CƠ BẢN.....	6
§2 CÁC MÔ HÌNH DỮ LIỆU CƠ BẢN	7
§3 CÁC KHÁI NIỆM CƠ BẢN CỦA MÔ HÌNH QUAN HỆ	10
Chương 2: ĐẠI SỐ QUAN HỆ VÀ CÀI ĐẶT ĐẠI SỐ QUAN HỆ	13
MỞ ĐẦU	13
§1 PHÉP CHỌN	14
§2 PHÉP CHIẾU	16
§3 PHÉP KẾT NỐI TỰ NHIÊN.....	18
§4 PHÉP CHIA QUAN HỆ	20
§5 QUAN HỆ TƯƠNG THÍCH.....	21
BÀI TẬP CHƯƠNG 2.....	23
Chương 3: PHỤ THUỘC HÀM.....	25
MỞ ĐẦU	25
§1 PHỤ THUỘC HÀM VÀ CÁC TÍNH CHẤT CƠ BẢN	26
§2 HỆ TIỀN ĐỀ ARMSTRONG CHO CÁC PHỤ THUỘC HÀM	29
§3 BAO ĐÓNG CỦA TẬP PHỤ THUỘC HÀM.....	30
§4 BAO ĐÓNG CỦA TẬP THUỘC TÍNH	31
BÀI TẬP CHƯƠNG 3	33
Chương 4: KHOÁ CỦA LƯỢC ĐỒ QUAN HỆ VÀ HỆ SPERNER.....	35
§1 KHOÁ VÀ CÁC TÍNH CHẤT CƠ BẢN.....	35
§2 HỆ SPERNER VÀ TẬP KHOÁ CỦA LƯỢC ĐỒ QUAN HỆ.....	37
BÀI TẬP CHƯƠNG 4.....	38
Chương 5: PHỦ CỦA TẬP PHỤ THUỘC HÀM.....	39
MỞ ĐẦU	39
§1 PHỦ CỦA TẬP PHỤ THUỘC HÀM.....	40
§2 CÁC KẾT QUẢ QUAN TRỌNG VỀ PHỦ CỦA TẬP PHỤ THUỘC HÀM ..	42
§3 CÀI ĐẶT THUẬT TOÁN TÌM PHỦ CỦA TẬP PHỤ THUỘC HÀM	43

BÀI TẬP CHƯƠNG 5.....	45
Chương 6: CHUẨN HÓA.....	46
MỞ ĐẦU	46
§1 PHÉP TÁCH KHÔNG TỔN THẤT THÔNG TIN.....	47
§2 CÁC DẠNG CHUẨN CỦA QUAN HỆ	47
§3 CÁC THUẬT TOÁN CHUẨN HÓA	49
BÀI TẬP CHƯƠNG 6.....	51
TÀI LIỆU THAM KHẢO	52

Chương 1: CÁC KHÁI NIỆM VÀ MÔ HÌNH DỮ LIỆU CƠ BẢN

MỞ ĐẦU

Cơ sở dữ liệu (CSDL) là lĩnh vực của tin học nhằm nghiên cứu các cơ chế, nguyên lý và phương pháp tổ chức dữ liệu trên các thiết bị nhớ bên ngoài của máy tính, nhằm phục vụ cho việc khai thác dữ liệu trong các hệ thống tin học ứng dụng.

Trong số 3 mô hình cơ bản - 3 cách tiếp cận cho việc tổ chức và khai thác các CSDL là mô hình phân cấp, mô hình mạng và mô hình quan hệ thì mô hình quan hệ được quan tâm hơn cả vì mô hình này được xây dựng trên một cơ sở toán học chặt chẽ lý thuyết về các quan hệ và có hình ảnh trực quan gần với các quan niệm thông thường của người dùng cuối.

§1 CÁC KHÁI NIỆM CƠ BẢN

1.1. CSDL

CSDL là một tập các dữ liệu về các đối tượng cần được quản lí, được lưu trữ đồng thời trên các vật mang tin (các thiết bị nhớ ngoài) của máy tính điện tử và được quản lí theo một cơ chế thống nhất gọi là hệ quản trị cơ sở dữ liệu nhằm thực hiện 3 chức năng sau đây một cách tối ưu:

- i) Mô tả dữ liệu.
- ii) Cập nhật dữ liệu.
- iii) Tìm kiếm dữ liệu.

1.2. HỆ QUẢN TRỊ CSDL

Hệ QTCSDL là một hệ thống phần mềm (các chương trình) giúp cho người sử dụng khai thác các CSDL theo 3 chức năng nói trên.

Chú ý: Các CSDL là đối tượng quản lí của các Hệ QTCSDL, chúng được tạo lập và lưu trữ trên các thiết bị nhớ ngoài của máy tính.

1.3. TẠI SAO CẦN CÓ CSDL?

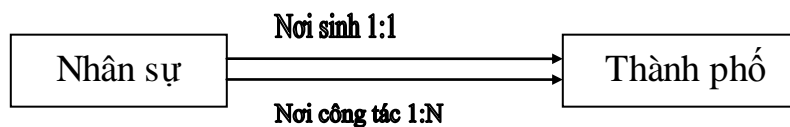
CSDL là bộ phận không thể thiếu được trong các hệ lưu trữ và tìm kiếm thông tin, các hệ thống quản lí kinh tế các ngành, các cấp, các hệ thống quản lí kho tàng, tư liệu, các hệ thống phục vụ công cộng như ngân hàng, bán vé máy bay, các phương tiện giao thông, các hệ thống thiết kế tự động,...

§2 CÁC MÔ HÌNH DỮ LIỆU CƠ BẢN

2.1. MÔ HÌNH MẠNG

Mô hình mạng được xây dựng trên các tập dữ liệu và các quan hệ.

Từ các tập dữ liệu rời nhau sau đó liên kết lại với nhau thành một mạng liên thông.



Tập dữ liệu được tạo ra từ những dữ liệu cùng một kiểu, gọi là bản ghi.

Mỗi bản ghi được tạo bởi các trường. Theo hình trên ta có hai tập dữ liệu là: Nhân sự và Thành phố, với các bản ghi tương ứng như sau:

- Mỗi phần tử của tập Nhân sự được mô tả qua 6 trường (6 thuộc tính):
Mã CB, Họ và tên, Năm sinh, Lương, Số con, Chức vụ.
- Mỗi phần tử của tập Thành phố được mô tả qua 5 thuộc tính:
Mã thành phố, Tên gọi, Diện tích, Dân số, Khoảng cách đến thủ đô.

Quan hệ xác lập một tương quan - ánh xạ giữa 2 tập dữ liệu.

Theo hình trên ta có quan hệ:

Nơi sinh: Nhân sự → Thành phố

Mỗi nhân sự cụ thể của tập Nhân sự có một nơi sinh cụ thể.

Giữa hai tập trên còn có thể có quan hệ thứ hai, chẳng hạn:

Nơi đến công tác: Cho biết cán bộ X đến làm việc tại những thành phố nào.

Các quan hệ được phân loại theo kiểu ánh xạ, ví dụ:

Nơi sinh là ánh xạ đơn trị, được kí hiệu là: 1-1 (mỗi người có một nơi sinh),

còn nơi đến công tác là một ánh xạ không đơn trị, kí hiệu là: 1-N (một người có thể tổ chức chuyển đi công tác ở nhiều thành phố).

Để tạo lập một CSDL theo mô hình mạng chúng ta cần các thao tác sau:

- Tạo lập một tập: Bao gồm việc khai báo tên tập và mô tả các thuộc tính của tập.
- Thiết lập một quan hệ giữa hai tập: Bao gồm việc khai báo tên quan hệ, tên tập nguồn, tên tập đích và kiểu quan hệ.

Chú ý:

Mô hình mạng có những ưu, nhược điểm sau:

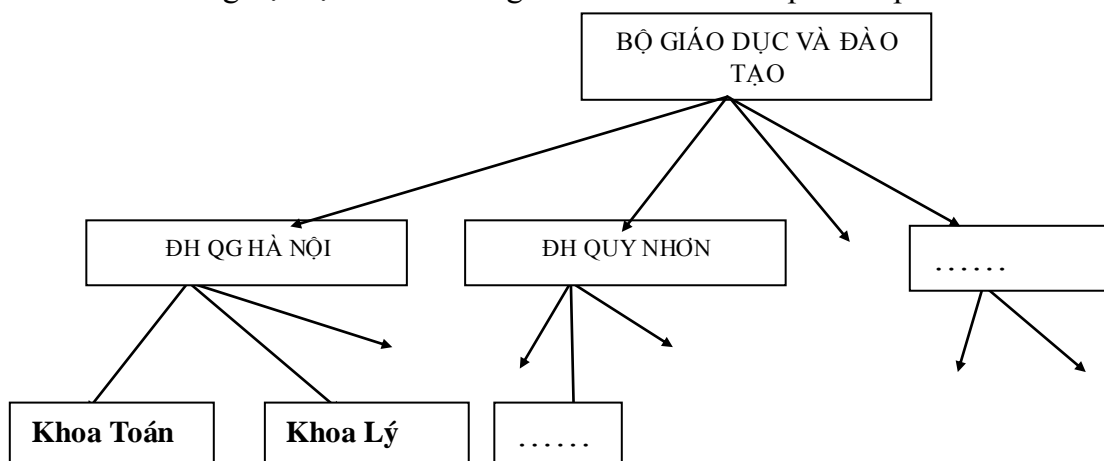
- Ưu điểm: Mang lại cho ta một thông tin của sự hiểu biết, để bổ sung các lớp đối tượng (tập dữ liệu) và các quan hệ.
- Nhược điểm: Phức tạp, luôn đòi hỏi một hình ảnh trực quan, dễ sinh ra nhọc nhằn bởi vì dễ sinh ra quan hệ giữa các đối tượng.

2.2. MÔ HÌNH PHÂN CẤP

Đây là một trường hợp riêng của mô hình mạng, trong đó khái niệm tập được giữ nguyên còn khái niệm quan hệ được giới hạn ở kiểu phân cấp.

Giữa 2 tập (nếu có) không quá một quan hệ và quan hệ này tuân thủ trật tự trên dưới.

Loại mô hình phân cấp khá phù hợp với những hình thức tổ chức phân cấp trong xã hội. Ví dụ một cơ sở dữ liệu quản lý chất lượng học tập của sinh viên các trường đại học và cao đẳng có thể có cấu trúc phân cấp sau:



Một mô hình phân cấp thường gặp trong các hệ thống máy tính là mô hình quản lý thư mục.

Đặc điểm nổi bật trong các thủ tục truy nhập tới một đối tượng trong mô hình phân cấp là đường dẫn - đường đi từ gốc (đỉnh đầu tiên) tới phần tử cần xét trong cây phân cấp.

Chú ý:

Mô hình phân cấp có những ưu, nhược điểm sau:

- Ưu điểm: Thể hiện được nguyên lý nhìn theo từng mức và muốn chế ngự độ phức tạp nên “chia để trị”.
- Nhược điểm: Theo nguyên tắc này chỉ lên xuống 1 mức nên không nhìn thấy những phần tử nằm ở những nhánh khác.

2.3. MÔ HÌNH QUAN HỆ

Mô hình này được E.F Codd đề xuất năm 1970.

Theo mô hình này: Một CSDL quan hệ được tạo lập từ các quan hệ có hình ảnh trực quan là các bảng. Mỗi bảng bao gồm các cột được gọi là thuộc tính và các dòng được gọi là bộ.

Ví dụ: Giả sử ta có cơ sở dữ liệu THỰC TẬP, nhằm lưu trữ thông tin về đợt thực tập của sinh viên, được tạo từ 3 quan hệ sau đây:

1. Quan hệ Sinh viên (SV)

Mã sinh viên	Họ và tên	Năm sinh	Quê quán
...

2. Quan hệ Đề tài (ĐT)

Mã đề tài	Tên đề tài	Chủ nhiệm	Kinh phí
....

3. Quan hệ Sinh viên đề tài (SVĐT)

Mã sinh viên	Mã đề tài	Nơi thực tập	Kết quả
....

2.4. CÁC ĐẶC ĐIỂM CỦA MÔ HÌNH QUAN HỆ

- Có cơ sở toán học chặt chẽ, cho phép áp dụng rộng rãi các công cụ đại số và logic.
- Khá tự nhiên, gần với quan niệm thông thường của người sử dụng.
- Ngôn ngữ thao tác trong sáng và có khả năng tổ hợp cao.
- Dễ đảm bảo tính an toàn dữ liệu, có thể đặt mật khẩu truy nhập ở nhiều mức: mức quan hệ, mức thuộc tính, mức bộ, mức thuộc tính - bộ.
- Dễ cập nhật tới các đơn vị dữ liệu.
- Dễ đảm bảo tính độc lập dữ liệu.

Chú ý:

Khi xây dựng các mô hình dữ liệu cần phân biệt các thành phần cơ bản sau:

- Thực thể: là đối tượng có trong thực tế mà chúng ta cần mô tả các đặc trưng của nó.
- Thuộc tính: là tính chất thuộc thực thể.
- Ràng buộc: là các mối quan hệ logic của các thực thể.

§3 CÁC KHÁI NIỆM CƠ BẢN CỦA MÔ HÌNH QUAN HỆ

3.1. QUAN HỆ

Cho tập hữu hạn $U = \{A_1, A_2, \dots, A_n\} \neq \emptyset$ ($n \geq 1$). Các phần tử của U được gọi là thuộc tính. Ứng với mỗi thuộc tính $A_i, i=1, n$ ta đặt tương ứng một tập d_i và gọi là miền trị của thuộc tính A_i , kí hiệu: $d_i = \text{dom}(A_i)$. Đặt

$$D = \bigcup_{i=1}^n \text{dom}(A_i)$$

Khi đó ta có

3.1.1. Định nghĩa

Quan hệ R với tập thuộc tính U (k/h: $R(U)$) là một tập các ánh xạ $t: U \rightarrow D$ sao cho với mỗi $A_i \in U$ ta phải có $t(A_i) \in \text{dom}(A_i), i=1, n$.

Các ánh xạ này được gọi là bộ của quan hệ R .

3.1.2. Định nghĩa

Quan hệ $R(U)$ là một bộ phận của tích Đề Các những $\text{dom}(A_i), i=1..n$

(tức: $R(U) \subseteq \prod_{i=1}^n \text{dom}(A_i)$).

3.1.3. Chú ý

- Hai định nghĩa trên là tương đương.
- Về mặt trực quan, một quan hệ có thể biểu diễn như một bảng 2 chiều gồm các cột là các thuộc tính và các dòng là các bộ và bảng phải thoả mãn 2 tính chất sau:
 - Hữu hạn tại mọi thời điểm.
 - Thứ tự của các hàng và cột là không quan trọng.
- Vì quan hệ là tập hợp nên trong quan hệ không có 2 bộ giống nhau hoàn toàn.

3.2. CƠ SỞ DỮ LIỆU QUAN HỆ

CSDLQH là một tập các quan hệ biến thiên theo thời gian và được quản lý bằng một cơ chế thống nhất gọi là hệ quản trị cơ sở dữ liệu quan hệ.

3.3. HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU QUAN HỆ

Hệ QTCSDLQH là hệ thống phần mềm giúp chúng ta quản xuyên toàn bộ việc tạo lập và khai thác các cơ sở dữ liệu quan hệ.

3.4. RÀNG BUỘC DỮ LIỆU

Ràng buộc dữ liệu là những quy định mà dữ liệu trong một cơ sở dữ liệu phải thỏa mãn.

Mục đích của việc đặt ra các ràng buộc dữ liệu là nhằm đảm bảo cho dữ liệu trong cơ sở dữ liệu phản ánh đúng thế giới hiện thực.

Các ràng buộc dữ liệu thường được mô tả vào lúc tạo lập quan hệ và thuộc tính.

Có các loại ràng buộc dữ liệu cơ bản như:

- Ràng buộc về kiểu dữ liệu: Đây là loại ràng buộc thấp nhất và cũng là loại ràng buộc tối thiểu, bắt buộc đối với hầu hết các CSDL.

Ví dụ: Khai báo $\text{dom}(\text{HOTEN}) = \text{C}(25)$; $\text{dom}(\text{NAMSINH}) = \text{N}(4)$;

- Ràng buộc giải tích: Đây là loại ràng buộc giữa các thuộc tính với nhau và thể hiện qua các biểu thức tính toán.

Ví dụ: $\text{DTB} = (\text{TOAN} * 3 + \text{TIN} * 2) / 5$

- Ràng buộc logic: Ở mức độ tổng quát các ràng buộc loại này được mô tả dưới dạng các biểu thức logic.

Ví dụ: Mã sinh viên \rightarrow Họ tên, năm sinh, quê quán;

$\text{DTB} \rightarrow \text{XEPLOAI}; \dots$

3.5. CÁC KÝ HIỆU TRUYỀN THÔNG

Theo truyền thông của lý thuyết CSDL, chúng ta chấp nhận các quy định sau đây:

- Các thuộc tính được kí hiệu bằng các chữ Latinh hoa đầu bảng: A, B, C,...
- Tập thuộc tính được kí hiệu bằng các chữ Latinh hoa cuối bảng: X, Y, Z,...
- Các thuộc tính trong một tập được liệt kê như một xâu kí tự, không có các dấu biểu diễn tập, chẳng hạn: $U = \text{ABC}$ thay vì viết: $U = \{A, B, C\}$.
- XY biểu diễn cho hợp của 2 tập thuộc tính X và Y.
- Các bộ trong quan hệ được biểu diễn bằng các chữ Latinh thường: s, t, u,...
- Với $t \in R(U)$, $A \in U$. Kí hiệu: $t.A$ là giá trị của bộ t tại thuộc tính A.
- Với $t \in R(U)$, $X \subseteq U$. Kí hiệu: $t.X$ là hạn chế/ thu hẹp của ánh xạ t trên X.
(tức: $t.X$ là bộ con của bộ t, chỉ lấy những giá trị trên X)

3.6. VÍ DỤ THỰC TẾ VỀ CƠ SỞ DỮ LIỆU QUAN HỆ

Giả sử ta có cơ sở dữ liệu quan hệ THỰC TẬP, nhằm lưu trữ thông tin về đợt thực tập của sinh viên, được tạo từ 3 quan hệ sau đây:

1. Quan hệ Sinh viên (SV)

Mã sinh viên	Họ và tên	Năm sinh	Quê quán
...

Quan hệ SV lưu trữ thông tin về sinh viên, bao gồm các thuộc tính sau:

Mã sinh viên (MSV); Họ và tên (HT); Năm sinh (NS); Quê quán (QQ)

2. Quan hệ Đề tài (ĐT)

Mã đề tài	Tên đề tài	Chủ nhiệm	Kinh phí
....

Quan hệ ĐT lưu trữ thông tin về các đề tài do nhà trường quản lý, bao gồm các thuộc tính sau:

Mã đề tài (MĐT); Tên đề tài (TĐT); Chủ nhiệm đề tài (CN); Kinh phí cấp cho đề tài (KP) và giả sử đơn vị tính của cột Kinh phí là triệu đồng.

3. Quan hệ Sinh viên đề tài (SVĐT)

Mã sinh viên	Mã đề tài	Nơi thực tập	Kết quả
....

Quan hệ SVĐT lưu trữ thông tin về việc thực tập của sinh viên: ai (MSV) tham gia đề tài nào (MĐT) thực tập ở đâu (NTT) và kết quả thực tập đạt bao nhiêu điểm (KQ).

Và giả sử rằng các quan hệ của CSDLQH nói trên thỏa mãn các ràng buộc sau đây:

- Mỗi sinh viên chỉ tham gia một đề tài.
- Mỗi đề tài có thể được triển khai ở nhiều thành phố khác nhau.
- Chủ nhiệm đề tài chỉ đảm nhiệm duy nhất đề tài đó.

Chương 2: ĐẠI SỐ QUAN HỆ VÀ CÀI ĐẶT ĐẠI SỐ QUAN HỆ

MỞ ĐẦU

- Đại số quan hệ là một bộ đôi sắp thứ tự gồm B và P (kí hiệu $\alpha = (B,P)$), ở đây B là tập các quan hệ trong một CSDLQH cho trước, còn P là tập các phép toán thao tác trên các quan hệ của B, gồm có các phép toán cơ bản sau:

- | | |
|-------------------------|--------------|
| - Phép chọn | kí hiệu: () |
| - Phép chiếu | [] |
| - Phép kết nối tự nhiên | * |
| - Phép tích Đề Các | x |
| - Phép hợp | + |
| - Phép giao | . |
| - Phép hiệu | - |
| - Phép chia | ÷ |

- Tập hợp các phép toán của đại số quan hệ sẽ tạo thành một cơ chế truy nhập linh hoạt, nhờ có các phép toán này mà chúng ta có thể trích dữ liệu từ một hay nhiều quan hệ cho trước để tạo thành một quan hệ mới nhằm trả lời cho một câu hỏi tìm kiếm đặt ra.
- Đại số quan hệ là một trong những ngôn ngữ con truy nhập dữ liệu (còn được xem như một ngôn ngữ hỏi).
- Đại số quan hệ là cơ sở toán học để cài đặt cú pháp SQL và cho phép diễn đạt các thao tác xử lý quan hệ.

Chú ý:

- Những phép toán nói trên còn được gọi là các phép toán đại số quan hệ/ phép toán quan hệ.
- SQL (Structured Query Language): là một ngôn ngữ hỏi có cấu trúc.

§1 PHÉP CHỌN

2.1.1. Định nghĩa Cho quan hệ $R(U)$. Xét một bộ $t \in R(U)$ và một biểu thức logic E phát biểu trên U (E còn gọi là điều kiện chọn). Khi đó:

Bộ t thỏa E (kh: $t(E)$) \Leftrightarrow Thay mọi xuất hiện của thuộc tính A trong E bởi giá trị $t.A$ thì ta thu được một mệnh đề logic đúng.

2.1.2. Ví dụ Giả sử xét quan hệ SV , với bộ $t = (2, \text{Hỗ Quý}, 1985, \text{Hà Nội}) \in SV$ và xét điều kiện chọn $E = \text{“tuổi đời hiện nay trên 26 và có quê quán ở Huế”}$.

Hỏi rằng: Bộ t này có thỏa E hay không?

Đs: Bộ t không thỏa E

2.1.3. Định nghĩa Cho quan hệ $R(U)$ và một biểu thức logic E phát biểu trên U .

Phép chọn quan hệ R theo E (k/h: $R(E)$) cho kết quả là một quan hệ P có tập thuộc tính là U và có các bộ của R thỏa E .

Nói gọn ta có: $R(E) = P(U) = \{t \in R \mid t(E)\}$

2.1.4. Ví dụ

a) Cho quan hệ R như sau:

A	B	C
a_1	x	2
a_2	y	5
a_1	m	4
a_3	x	1

và điều kiện chọn $E = \text{“}A \neq a_1 \vee C - 3 > 0\text{”}$

Khi đó ta có: $R(E) = P$

A	B	C
a_2	y	5
a_1	m	4
a_3	x	1

b) Cho biết thông tin nói về những đề tài có kinh phí trên 5 triệu và dưới 12 triệu?

$$P = \text{ĐT}(\text{KP} > 5 \wedge \text{KP} < 12)$$

c) Cho biết thông tin nói về việc thực tập của những sinh viên đạt kết quả ít nhất là 9 và thực tập tại Hà Nội?

$$P = \text{SVĐT}(\text{KQ} \geq 9 \wedge \text{NTT} = \text{'Hà Nội'})$$

Cài đặt thuật toán thực hiện phép chọn:

(Sử dụng ngôn ngữ ALGOL/ ngôn ngữ tựa Pascal để mô tả cài đặt)

Algorithm Selection:

Input: $R(U)$, E .

Output: Tính $R(E) = ?$

Actions:

 Create(P, U);

 For each tuple t in R do

 If $t(E)$ then

 Add(P, t);

 End if;

 End for;

 Return P ;

End.

§2 PHÉP CHIẾU

2.2.1. Định nghĩa Cho quan hệ $R(U)$ và tập thuộc tính $X \subseteq U$.

Phép chiếu quan hệ R trên X (k/h: $R[X]$) cho kết quả là quan hệ P có tập thuộc tính X và chứa các bộ của R hạn chế trên X .

Nói gọn ta có: $R[X] = P(X) = \{t.X \mid t \in R\}$

2.2.2. Ví dụ

a) Cho quan hệ R như sau:

A	B	C
a_1	x	5
a_2	y	7
a_1	x	4
a_2	y	3

Khi đó ta có: $R[AB] = P$

A	B
a_1	x
a_2	y

b) Cho biết họ tên và quê quán của các sinh viên?

$P = SV[HT, QQ]$

c) Cho biết mã của những sinh viên thực tập tại Bình Định và đạt kết quả cao nhất là 8?

$P = SV\Join T(NTT = \text{'Bình Định'} \wedge KQ \leq 8)[MSV]$

Cài đặt thuật toán thực hiện phép chiếu:

(Sử dụng ngôn ngữ ALGOL/ ngôn ngữ tựa Pascal để mô tả cài đặt)

Algorithm Projection:

Input: $R(U)$, $X \subseteq U$.

Output: Tính $R[X] = ?$

Actions:

 Create(P , X);

 For each tuple t in R do

 If $\neg(t.X \text{ in } P)$ then

 Add(P , $t.X$);

 End if;

 End for;

 Return P ;

End.

§3 PHÉP KẾT NỐI TỰ NHIÊN

2.3.1. Định nghĩa Cho 2 quan hệ $R(U)$ và $S(V)$ với $U \cap V \neq \emptyset$.

Phép kết nối tự nhiên 2 quan hệ R và S (k/h: $R \bowtie S$) cho kết quả là quan hệ P có tập thuộc tính UV và chứa các bộ được xác định như sau:

$$R \bowtie S = P(UV) = \{t \mid t.U \in R \wedge t.V \in S\}$$

2.3.2. Ví dụ

a) Cho 2 quan hệ sau

R		S		
A	B	B	C	E
a ₁	x	x	4	n
a ₂	y	y	5	z
a ₃	x	y	2	m
		z	3	n

Khi đó ta có: $R \bowtie S = P$

A	B	C	E
a ₁	x	4	n
a ₂	y	5	z
a ₂	y	2	m
a ₃	x	4	n

b) Cho biết họ tên và kết quả thực tập của những sinh viên thực tập tại quê nhà?

$$P = (SV \bowtie SV\Delta T)(QQ = NTT)[HT, KQ]$$

c) Cho biết mã của những sinh viên có tuổi đời hiện nay dưới 22 và thực tập đạt kết quả trên 9?

$$P = (SV \bowtie SV\Delta T)(2012 - NS < 22 \wedge KQ > 9)[MSV]$$

2.3.3. Chú ý

- Vì quan hệ là tập hợp nên trong quan hệ không chứa 2 bộ giống nhau hoàn toàn.

- Khi sử dụng đồng thời 2 phép toán chọn và chiếu, thông thường nên “chọn trước rồi chiếu sau”.

- Nếu 2 quan hệ $R(U)$ và $S(V)$ có $U \cap V = \emptyset$ thì ta có phép toán Tích Đề các được định nghĩa như sau:

$$R \times S = P(UV) = \{(a,b) \mid a \in R \wedge b \in S\}$$

Cài đặt thuật toán thực hiện phép kết nối tự nhiên 2 quan hệ:

(Sử dụng ngôn ngữ ALGOL/ ngôn ngữ tựa Pascal để mô tả cài đặt)

Algorithm Natural Join:

Input: $R(U)$, $S(V)$, $U \cap V \neq \emptyset$.

Output: Tính: $R * S = ?$

Actions:

 Create(P , UV);

$X := U \cap V$;

 For each tuple t in R do

 For each tuple s in S do

 If $t.X = s.X$ then

 Add(P , $(t, s.V \setminus X)$);

 End if;

 End for;

 End for;

 Return P ;

End.

§4 PHÉP CHIA QUAN HỆ

2.4.1. Định nghĩa Cho 2 quan hệ $R(U)$ và $S(V)$ với $V \subset U$. Đặt $X = U \setminus V$.

Phép chia quan hệ R cho quan hệ S (k/h: $R \div S$) cho kết quả là quan hệ P có tập thuộc tính là X và có các bộ của P được xác định như sau:

$$R \div S = P(X) = \{t.X \mid t \in R \wedge \forall s \in S: (t.X, s) \in R\}$$

2.4.2. Ví dụ Cho 2 quan hệ sau:

R

S

Khi đó ta có: $R \div S = P$

A	B
1	a
1	b
1	c
2	c
2	d
3	a
3	b
4	a

B
a
b

A
1
3

§5 QUAN HỆ TƯƠNG THÍCH

2.5.1. Định nghĩa Hai quan hệ $R(U)$ và $S(V)$ được gọi là tương thích với nhau khi và chỉ khi $U = V$.

Cho 2 quan hệ tương thích $R(U)$ và $S(U)$.

Khi đó ta có các phép toán trên 2 quan hệ tương thích này như sau:

- Phép hợp: $R(U) + S(U) = P(U) = \{ t \mid t \in R \vee t \in S \}$
- Phép giao: $R(U) \cdot S(U) = P(U) = \{ t \mid t \in R \wedge t \in S \}$
- Phép trừ: $R(U) - S(U) = P(U) = \{ t \mid t \in R \wedge t \notin S \}$

Cài đặt thuật toán thực hiện các phép toán trên 2 quan hệ tương thích:

(Sử dụng ngôn ngữ ALGOL/ ngôn ngữ tựa Pascal để mô tả cài đặt)

- **Phép hợp:**

Input: $R(U), S(U)$

Output: Tính: $R + S = ?$

Actions:

```
Create(P, U);
P := R;
For each tuple s in S do
    If !(s in P) then
        Add(P, s);
    End if;
End for;
Return P;
```

End.

- **Phép giao:**

Input: $R(U), S(U)$

Output: Tính: $R.S = ?$

Actions:

 Create(P, U);

 For each tuple t in R do

 If (t in S) then

 Add(P, t);

 End if;

 End for;

 Return P;

End.

- **Phép trừ:**

Input: $R(U), S(U)$

Output: Tính: $R - S = ?$

Actions:

 Create(P, U);

 For each tuple t in R do

 If $!(t \text{ in } S)$ then

 Add(P, t)

 End if;

 End for;

 Return P;

End.

2.5.2. Chú ý

- Độ ưu tiên theo thứ tự từ cao đến thấp của các phép toán quan hệ trên đây như sau: (chọn, chiếu); (kết nối tự nhiên, tích Đề các, giao, chia); (hợp, trừ).

- Dãy các phép toán có cùng độ ưu tiên được thực hiện lần lượt từ trái sang phải. Nếu biểu thức quan hệ có chứa các cặp dấu ngoặc () thì các biểu thức con trong các cặp dấu ngoặc được thực hiện trước nhất.

BÀI TẬP CHƯƠNG 2

Câu I. Xét cơ sở dữ liệu quan hệ THUCTAP đã biết.

Hãy dùng các phép toán đại số quan hệ để trả lời các câu hỏi sau:

1) Cho biết họ tên và quê quán của những sinh viên không đi thực tập?

(yêu cầu: sử dụng phép trừ)

2) Cho biết mã của những đề tài có kinh phí trên 15 triệu và mã của những đề tài có kinh phí dưới 8 triệu?

(yêu cầu: giải theo 2 cách khác nhau và có sử dụng phép hợp)

3) Cho biết mã của những sinh viên có tuổi đời trên 20 và điểm thực tập dưới 6?
(yêu cầu: giải theo 2 cách khác nhau và có sử dụng phép giao)

4) Cho biết mã của những đề tài có nơi thực tập ít ra là giống như đề tài có mã 01? (yêu cầu: sử dụng phép chia)

Câu II. Viết thuật toán thực hiện (trực tiếp/ gián tiếp) các biểu thức quan hệ sau:

1) $R(E)[X]$

2) $(R * S)(E)$

3) $(R * S)[X]$

4) $(R * S)(E)[X]$

Câu III.

1) Cho 2 quan hệ $R(U)$ và $S(V)$. Chứng minh: $R * S = S * R$

2) Cho 2 quan hệ $R(U)$ và $S(U)$. Chứng minh: $R - (R - S) = R.S$

3) Cho 2 quan hệ $R(UV)$ và $S(V)$.

Chứng minh: $R \div S = R[U] - (R[U] * S - R)[U]$

5) Cho quan hệ $R(U)$ và các tập con X_1, X_2, \dots, X_k của U thỏa:

$X_1 \cup X_2 \dots \cup X_k = U$ và $X_i \cap X_j \neq \emptyset$ ($i \neq j$).

a/ Chứng minh: $R \subseteq R[X_1] * R[X_2] * \dots * R[X_k]$ (1)

b/ Cho ví dụ thực tế để khẳng định chiều ngược lại của (1) không xảy ra?

Câu IV: Cho 2 quan hệ:

R

A	B	C
a ₁	2	c ₁
a ₂	3	c ₂
a ₃	4	c ₃

S

A	B	C	D
a ₁	1	c ₁	d ₁
a ₁	2	c ₂	d ₁
a ₃	3	c ₂	d ₂
a ₄	4	c ₃	d ₁

Tính các biểu thức quan hệ sau:

1) $R[BC] * S(5 - B > 2) = ?$

2) $(R * S)(A = a_1 \vee C \neq c_2) = ?$

3) $R(A \neq a_2 \wedge B > 3)[AC] * S = ?$

Chương 3: PHỤ THUỘC HÀM

MỞ ĐẦU

- Phụ thuộc hàm thực chất là một loại ràng buộc dữ liệu kiểu logic, có dạng: $X \rightarrow Y$, với X, Y là các tập thuộc tính của quan hệ R .

Ví dụ: Xét quan hệ SV trong cơ sở dữ liệu quan hệ THUCTAP.

Ta có phụ thuộc hàm: $MSV \rightarrow \{HT, NS, QQ\}$

- Việc nghiên cứu phụ thuộc hàm nhằm mục đích: đảm bảo tính nhất quán dữ liệu trong cơ sở dữ liệu quan hệ.
- Phụ thuộc hàm có tầm quan trọng rất lớn trong việc phân tích và thiết kế mô hình dữ liệu, các hệ cơ sở dữ liệu.

§1 PHỤ THUỘC HÀM VÀ CÁC TÍNH CHẤT CƠ BẢN

3.1.1. Định nghĩa Cho quan hệ $R(U)$ và $X, Y \subseteq U$. Ta nói:

- X xác định hàm Y (hay Y phụ thuộc hàm vào X) \Leftrightarrow tồn tại một ánh xạ $f: X \rightarrow Y$ và kí hiệu: $X \rightarrow Y$
- Quan hệ R thỏa phụ thuộc hàm $X \rightarrow Y$ (k/h: $R(X \rightarrow Y)$) \Leftrightarrow Với mọi bộ u, v lấy ra bất kỳ trong quan hệ R , nếu 2 bộ này giống nhau trên X thì 2 bộ này cũng giống nhau trên Y .

Nói gọn ta có: $R(X \rightarrow Y) \Leftrightarrow \forall u, v \in R : u.X = v.X \Rightarrow u.Y = v.Y$

3.1.2. Ví dụ: Cho quan hệ HOCTAP như sau:

Lớp	Môn	Thầy
12	Toán	A
11	Toán	B
10	Toán	A
12	Địa	C
11	Địa	C
10	Sử	D

Khi đó ta thấy quan hệ này thỏa các phụ thuộc hàm:

$\text{Thầy} \rightarrow \text{Môn}; \{ \text{Lớp}, \text{Môn} \} \rightarrow \text{Thầy}$

3.1.3. Nhận xét

- Nếu quan hệ R có không quá 1 bộ thì quan hệ này luôn luôn thỏa mọi phụ thuộc hàm.
- Nếu quan hệ R thỏa $X \rightarrow Y$ thì mọi quan hệ con của R cũng thỏa $X \rightarrow Y$.

3.1.4. Các tính chất cơ bản của phụ thuộc hàm

Cho quan hệ $R(U)$ và với mọi tập thuộc tính $X, Y, Z, W \subseteq U$. Ta có:

1. Tính chất phản xạ:

Nếu $X \supseteq Y$ thì $X \rightarrow Y$

Chú ý: $X \rightarrow Y$ trong trường hợp này còn gọi là phụ thuộc hàm tầm thường, phụ thuộc hàm này luôn thỏa với mọi quan hệ R cho trước!

2. Tính chất gia tăng:

Nếu $X \rightarrow Y$ thì $XZ \rightarrow YZ$

3. Tính chất bắc cầu:

Nếu $\begin{cases} X \rightarrow Y \\ Y \rightarrow Z \end{cases}$ thì $X \rightarrow Z$

Chú ý: 3 tính chất trên đây dễ dàng suy ra từ định nghĩa. Từ 3 tính chất cơ bản này ta có thể chứng minh được các tính chất khác của phụ thuộc hàm sau đây:

4. Tính chất cộng tính:

Nếu $\begin{cases} X \rightarrow Y \\ Z \rightarrow W \end{cases}$ thì $XZ \rightarrow YW$

Chúng minh:
$$\left. \begin{array}{l} X \rightarrow Y \xRightarrow{t/c2} XZ \rightarrow YZ \\ Z \rightarrow W \xRightarrow{t/c2} YZ \rightarrow YW \end{array} \right\} \xRightarrow{t/c3} XZ \rightarrow YW$$

5. Tính chất tăng giảm:

Nếu $X \rightarrow Y$ thì $XZ \rightarrow Y \setminus W$

Chúng minh:

ta có $\left. \begin{array}{l} X \rightarrow Y \Rightarrow XZ \rightarrow YZ \\ YZ \supseteq Y \Rightarrow YZ \rightarrow Y \end{array} \right\} \Rightarrow XZ \rightarrow Y (1).$

Vì $Y \supseteq Y \setminus W \Rightarrow Y \rightarrow Y \setminus W (2)$

Từ (1) và (2) $\Rightarrow XZ \rightarrow Y \setminus W$

6. Tính chất phản xạ chặt:

$\forall X \subseteq U$, ta có: $X \rightarrow X$ (hiển nhiên)

7. Tính chất chiếu:

$X \rightarrow YZ \Rightarrow \begin{cases} X \rightarrow Y \\ X \rightarrow Z \end{cases}$ (hiển nhiên)

8. Tính chất tựa bắc cầu:

Nếu $\begin{cases} X \rightarrow Y \\ YZ \rightarrow W \end{cases}$ thì $XZ \rightarrow W$

Chứng minh: ta có $X \rightarrow Y \Rightarrow \begin{cases} XZ \rightarrow YZ \\ YZ \rightarrow W \end{cases} \Rightarrow XZ \rightarrow W$

§2 HỆ TIÊN ĐỀ ARMSTRONG CHO CÁC PHỤ THUỘC HÀM

3.2.1. Định nghĩa Hệ tiên đề Armstrong cho các phụ thuộc hàm (k/h: A^0) bao gồm 3 tính chất cơ bản của phụ thuộc hàm: tính chất phản xạ, tính chất gia tăng và tính chất bắc cầu.

Vậy $A^0 = \{t/c \text{ phản xạ, } t/c \text{ gia tăng, } t/c \text{ bắc cầu}\}$

3.2.2. Định lí Hệ tiên đề Armstrong là đúng và đầy đủ. (tự c/m)

3.2.3. Chú ý

- Hệ tiên đề là những tính chất đầu tiên được chọn và thừa nhận sao cho từ đó ta có thể suy ra được những tính chất khác còn lại.
- Hệ tiên đề bao giờ cũng thoả mãn các tính chất sau đây:

+ Tính phi mâu thuẫn: Từ hệ tiên đề ta không thể suy ra được 2 mệnh đề mâu thuẫn nhau.

+ Tính đúng đắn: Nếu từ hệ tiên đề này ta suy ra được tính chất (mệnh đề) P thì tính chất P này cũng phải đúng theo định nghĩa.

+ Tính đầy đủ: Nếu từ định nghĩa ta suy ra được 1 tính chất Q nào đó thì tính chất Q này cũng phải được suy ra từ hệ tiên đề.

- Hệ tiên đề Armstrong ra đời năm 1974. Cho đến nay có nhiều hệ tiên đề khác về phụ thuộc hàm như:

Hệ tiên đề $B^0 = \{\text{tính chất phản xạ, tính chất tựa bắc cầu}\}$

Hệ tiên đề $D^0 = \{\text{tính chất phản xạ chặt, tính chất bắc cầu, tính chất cộng tính, tính chất tăng giảm}\}, \dots$

Và các hệ tiên đề này đều tương đương với hệ tiên đề Armstrong!

§3 BAO ĐÓNG CỦA TẬP PHỤ THUỘC HÀM

3.3.1. Định nghĩa Cho tập phụ thuộc hàm F và một phụ thuộc hàm f .

Ta nói: f được suy dẫn bởi F (k/h: $F \rightarrow f$) \Leftrightarrow Với mọi quan hệ R , nếu R thỏa F thì R cũng thỏa f .

3.3.2. Chú ý

- Ta nói R thỏa F có nghĩa là R thỏa mọi phụ thuộc hàm f trong F .

- Nói gọn: $F \rightarrow f \Leftrightarrow \forall R, R(F) \Rightarrow R(f)$

$$R(F) \Leftrightarrow R(g), \forall g \in F$$

$$R(g) \Leftrightarrow \forall u, v \in R : u.left(g) = v.left(g) \Rightarrow u.right(g) = v.right(g)$$

3.3.3. Định nghĩa Bao đóng của tập phụ thuộc hàm F (k/h: F^*) là tập hợp các phụ thuộc hàm f sao cho f được suy dẫn bởi F .

Nói gọn ta có: $F^* = \{ f \mid F \rightarrow f \}$.

3.3.4. Ví dụ Giả sử $U = ABC$ và $F = \{ A \rightarrow B, B \rightarrow C \}$

Khi đó: $F^* = \{ A \rightarrow B; B \rightarrow C; A \rightarrow C; AB \rightarrow BC; AC \rightarrow BC; AB \rightarrow AC; \dots \}$

Các tính chất cơ bản của bao đóng tập phụ thuộc hàm:

1. Tính phản xạ: $F \subseteq F^*$
2. Tính đơn điệu: $F \subseteq G \Rightarrow F^* \subseteq G^*$
3. Tính lũy đẳng: $(F^*)^* = F^*$

Chú ý:

- Bài toán tìm F^* trong trường hợp tổng quát là một bài toán khó (có độ phức tạp tính toán là hàm mũ - Vì nếu tập U có n phần tử thì số tập con của U là 2^n)

- **Bài toán thành viên**

“Cho tập phụ thuộc hàm F và một phụ thuộc hàm f . Hỏi rằng: $f \in F^*$?”

- Để giải quyết bài toán thành viên này chúng ta sử dụng định lý quan trọng trong §4.

§4 BAO ĐÓNG CỦA TẬP THUỘC TÍNH

3.4.1. Định nghĩa Lược đồ quan hệ α là một bộ sắp thứ tự gồm U và F

(k/h: $\alpha = (U, F)$), với U là tập thuộc tính của quan hệ R và F là tập các phụ thuộc hàm trong R.

3.4.2. Ví dụ Cho quan hệ KETQUATHI như sau:

Mã môn thi	Mã sinh viên	Họ tên	Tuổi	Quê quán	Điểm thi
....

Khi đó ta có lược đồ quan hệ là: $\alpha = (U, F)$. Với:

$U = \{\text{Mã môn thi, Mã sinh viên, Họ tên, Tuổi, Quê quán, Điểm thi}\}$

$F = \{\text{Mã sinh viên} \rightarrow \text{Họ tên, Tuổi, Quê quán};$

$\text{Mã môn thi, Mã sinh viên} \rightarrow \text{Điểm thi}\}$

3.4.3. Định nghĩa Cho lược đồ quan hệ $\alpha = (U, F)$ và tập thuộc tính $X \subseteq U$.

Bao đóng của tập thuộc tính X đối với tập phụ thuộc hàm F (k/h: X_F^+) là tập hợp các thuộc tính A trong U sao cho: $X \rightarrow A \in F^*$

Nói gọn ta có: $X_F^+ = \{A \in U \mid X \rightarrow A \in F^*\}$

Thuật toán tìm bao đóng của tập thuộc tính:

Input: $\alpha = (U, F)$, U là tập hữu hạn, $X \subseteq U$.

Output: Tính: $X_F^+ = ?$

Actions:

B1: Xây dựng một dãy các tập bao nhau: $X^0 \subseteq X^1 \subseteq \dots \subseteq X^k \subseteq \dots$ theo quy luật sau:

Đặt $X^0 = X$.

Nếu tìm được X^k ($k \geq 0$) thì tập X^{k+1} được tính như sau:

$$X^{k+1} = X^k \cup Z \text{ với } \begin{cases} Y \rightarrow Z \in F \\ Y \subseteq X^k \end{cases}$$

B2: Nếu tìm được một chỉ số nguyên dương k đầu tiên sao cho: $X^{k+1} = X^k$ thì dừng thuật toán và kết luận: $X_F^+ = X^k$.

Chú ý:

- Thuật toán trên đây luôn đảm bảo tính dừng (vì U là tập hữu hạn)
- Thuật toán trên được xây dựng và chứng minh bởi Beeri và Bernstein.

Các tính chất cơ bản của bao đóng tập thuộc tính:

1. Tính phản xạ: $X \subseteq X_F^+, \forall X \subseteq U$
2. Tính đơn điệu: $X \subseteq Y \Rightarrow X_F^+ \subseteq Y_F^+, \forall X, Y \subseteq U$
3. Tính lũy đẳng: $(X_F^+)_F^+ = X_F^+, \forall X \subseteq U$

3.4.4. Định lí quan trọng

$$X \rightarrow Y \in F^* \Leftrightarrow Y \subseteq X_F^+$$

Chứng minh:

(\Rightarrow) Giả sử $Y = B_1 B_2 \dots B_k$ với $B_i \in U$.

Theo tính chất chiếu của phụ thuộc hàm, ta có: $R(X \rightarrow B_1 B_2 \dots B_k) \Rightarrow R(X \rightarrow B_i), \forall i = 1, k$.

Từ đó suy ra: $B_i \in X_F^+, \forall i = 1, k$. Vậy: $Y \subseteq X_F^+$

(\Leftarrow) Giả sử $Y = B_1 B_2 \dots B_k$ với $B_i \in U$.

Vì $Y \subseteq X_F^+$ nên ta có: $B_i \in X_F^+, \forall i = 1, k$. Suy ra: $X \rightarrow B_i \in F^*, \forall i = 1, k$.

Theo tính chất cộng tính của phụ thuộc hàm, ta có: $X \rightarrow B_1 B_2 \dots B_k \in F^*$

Hay $X \rightarrow Y \in F^*$.

BÀI TẬP CHƯƠNG 3

Câu I. Cho quan hệ BANHANG như sau:

Ngày bán hàng	Mã hàng	Tên hàng	Đơn giá	Số lượng	Tổng	Thanh toán
21/04/2016	M1	Radio	100000	1	1000000	600000
21/04/2016	M3	Ti vi	400000	2	1000000	600000
21/04/2016	M6	Xe đạp	100000	1	1000000	600000
22/04/2016	M2	Máy giặt	750000	2	1500000	600000
23/04/2016	M1	Radio	100000	9	1500000	900000
23/04/2016	M4	Video	150000	4	1500000	900000

Hãy xác định tập phụ thuộc hàm trong quan hệ trên?

Câu II.

1. Sử dụng 3 tính chất cơ bản của bao đóng tập thuộc tính, chứng minh:

a) $X^+Y^+ \subseteq (XY)^+$ (1)

b) $(X^+Y)^+ = (XY)^+$

2. Tìm một phản ví dụ để chứng tỏ chiều ngược lại của (1) không xảy ra.

3. Chứng minh:

a) Các tính chất cơ bản của bao đóng tập phụ thuộc hàm.

b) Các hệ tiên đề sau đây tương đương với nhau: A^0, B^0, D^0

Câu III.

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

$F = \{N \rightarrow P, M \rightarrow MQ, MN \rightarrow QR, M \rightarrow OP, PQ \rightarrow MNQ, N \rightarrow OR, R \rightarrow MQ, N \rightarrow O\}$

a) Chứng minh rằng: Từ F , theo hệ tiên đề Armstrong ta có: $M \rightarrow PR$

b) Tính: $(KY)^+ \cap (X^+ \setminus Y)^+$ (nêu chi tiết cách tính)

(Biết rằng: $K^+ = U$, $X = PQ$, $Y = NOP$)

c) Kiểm tra xem: $PQ \rightarrow MOS \in F^*$? $MS \rightarrow PQR \in F^*$?

Câu IV.

Dùng ngôn ngữ tựa Pascal để cài đặt thuật toán:

a) Kiểm tra quan hệ R cho trước có thỏa: $X \rightarrow Y$

b) Tìm X_F^+

Chương 4: KHOÁ CỦA LƯỢC ĐỒ QUAN HỆ

§1 KHOÁ VÀ CÁC TÍNH CHẤT CƠ BẢN

4.1.1. Định nghĩa Cho lược đồ quan hệ $\alpha = (U, F)$ và tập thuộc tính $K \subseteq U$. Khi đó:

- K được gọi là siêu khoá của $\alpha \Leftrightarrow K_F^+ = U$ (hay: $K \rightarrow U$)
- K được gọi là Khoá của $\alpha \Leftrightarrow \begin{cases} K_F^+ = U \\ \forall X \subset K : X_F^+ \neq U \end{cases}$

4.1.2. Ví dụ Cho lược đồ quan hệ $\alpha = (U, F)$ với:

$$U = MNOPQ, F = \{MN \rightarrow OPQ, N \rightarrow OQ, O \rightarrow NP, MO \rightarrow NPQ\}$$

- a) Tìm ít nhất 3 siêu khoá? Đs: U, MNO, MO, \dots
- b) Tìm một khoá của α ? Đs: MN
- c) Tìm toàn bộ các khoá của α Đs: MN, MO .

4.1.3. Nhận xét

- Khoá là siêu khoá nhưng siêu khoá chưa phải là khoá.
- Khoá là siêu khoá nhỏ nhất.
- Một lược đồ quan hệ cho trước bao giờ cũng có ít nhất một siêu khoá (trường hợp xấu nhất thì siêu khoá chính là U)
- Một lược đồ quan hệ cho trước bao giờ cũng có ít nhất một khoá (trường hợp xấu nhất khoá là U)
- Nếu K_1 và K_2 là khoá của α thì ta có: $\begin{cases} K_1 \not\subset K_2 \\ K_2 \not\subset K_1 \end{cases}$

4.1.4. Các tính chất cơ bản của khoá

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

Khi đó ta chứng minh được các tính chất sau:

1. Nếu thuộc tính A không xuất hiện trong mọi vế phải của các phụ thuộc hàm trong F thì A luôn luôn có mặt trong mọi khoá.
2. Nếu $M = U \setminus \cup \text{right}(f)$ (với $f \in F$) và $M^+ = U$ thì lược đồ quan hệ đã cho chỉ có duy nhất một khoá là M.

Thuật toán tìm 1 khoá của lược đồ quan hệ:

Algorithm Key:

Input: $\begin{cases} \alpha = (U, F) \\ U = A_1, \dots, A_n (n \geq 1) \\ F = \{L_i \rightarrow R_i \mid i = m; L_i, R_i \subseteq U\} \end{cases}$

Output: Tìm 1 khoá K của α ?

Actions:

K:=U;

for i:=1 to n do

if $(K \setminus A_i)^+_F = U$ then

K:= K \setminus A_i;

end if;

end for;

return K;

End.

4.1.5. Nhận xét

- Thuật toán trên được xây dựng dựa trên ý tưởng sau đây:

Vì khoá là siêu khoá nhỏ nhất, do đó để tìm một khoá của α đầu tiên ta xuất phát từ một siêu khoá U cho trước, rồi tiến hành loại bỏ dần các thuộc tính trong U để tìm ra siêu khoá nhỏ nhất - Đó chính là khoá cần tìm.

- Nếu cần tìm nhiều khoá, ta chỉ cần thay đổi trật tự loại bỏ các phần tử của K.

4.1.6. Chú ý

Bài toán tìm toàn bộ các khoá của lược đồ quan hệ α cho trước trong trường hợp tổng quát là một bài toán khó (vì phải duyệt toàn bộ các tập con của U nên có độ phức tạp tính toán là hàm mũ - $O(2^n)$, với n là số thuộc tính trong U).

§2 HỆ SPERNER VÀ TẬP KHOÁ CỦA LƯỢC ĐỒ QUAN HỆ

4.2.1. Định nghĩa Cho lược đồ quan hệ $\alpha = (U, F)$. Khi đó:

Họ $s = \{X_i \subseteq U \mid X_i \text{ và } X_j (\forall i \neq j) \text{ từng đôi một không bao nhau}\}$ được gọi là hệ Sperner.

4.2.2. Ví dụ Giả sử $U = ABCDE$. Khi đó ta có các hệ Sperner sau đây:

$s_1 = \{A, B, C, D, E\} \rightarrow$ Hệ Sperner có 5 phần tử.

$s_2 = \{AB, AC, AD, AE, BC, BD, BE, CD, CE, DE\} \rightarrow$ Hệ Sperner có 10 phần tử.

$s_3 = \{ABC, ABD, ABE, ACD, ADE, BCD, BCE\} \rightarrow$ Hệ Sperner có 7 phần tử.

4.2.3. Nhận xét

- Với mọi tổ hợp chập i của n phần tử ($1 \leq i \leq n$) đều tạo thành một hệ Sperner và hệ này có đúng C_n^i phần tử.
- Lược đồ quan hệ α có tập thuộc tính U gồm n phần tử thì số khoá tối đa của lược đồ quan hệ này là C_n^{ndiv2} .

4.2.4. Định lý Sperner

Cho một hệ Sperner s trên U . Khi đó luôn luôn tồn tại một lược đồ quan hệ α sao cho $K_\alpha = s$. (với K_α là tập toàn bộ các khoá của α).

4.2.5. Chú ý

Lược đồ quan hệ α tồn tại trong định lý trên đây được Sperner xây dựng như sau: $\alpha = \langle U, F \rangle$ với U đã cho và $F = \{X \rightarrow U \setminus X \mid X \in s\}$.

4.2.6. Ví dụ Giả sử $U = ABCDE$. Hãy xây dựng một lược đồ quan hệ trên U có đúng 5 khoá và chỉ ra các khoá này?

+ Xây dựng lược đồ quan hệ $\alpha = \langle U, F \rangle$ như sau: $U = ABCDE$ và

$F = \{A \rightarrow BCDE, B \rightarrow ACDE, C \rightarrow ABDE, D \rightarrow ABCE, E \rightarrow ABCD\}$

+ Theo định lý Sperner trên đây, lược đồ quan hệ này có đúng 5 khoá và ta có:

$K_\alpha = \{A, B, C, D, E\}$ (vì $\{A, B, C, D, E\}$ là một hệ sperner có 5 phần tử)

Chú ý: Ta có thể xây dựng lược đồ quan hệ này có đúng 5 khoá nhưng tiến hành theo một cách khác như sau:

$\alpha = \langle U, F \rangle$ với $U = ABCDE$ và

$F = \{A \rightarrow B, B \rightarrow C, C \rightarrow D, D \rightarrow E, E \rightarrow A\}$. Ta có: $K_\alpha = \{A, B, C, D, E\}$.

BÀI TẬP CHƯƠNG 4

Câu I.

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

Chứng minh rằng: α có một khóa duy nhất khi và chỉ khi $M^+ = U$

(với M là giao của các khóa của α)

2. Cho lược đồ quan hệ $\alpha = (U, F)$ và K là một khóa bất kỳ của α .

Chứng minh rằng: Với mọi $X \subset K$ ta có: $X^+ \cap K = X$

3. Chứng minh định lý Sperner.

4. Chứng minh rằng: Mọi tổ hợp chập i của n phần tử ($1 \leq i \leq n$) đều tạo thành một hệ Sperner và hệ này có đúng C_n^i phần tử.

5. Chứng minh rằng: Lược đồ quan hệ α có tập thuộc tính U gồm n phần tử thì số khóa tối đa của lược đồ quan hệ này là C_n^{ndiv2} .

6. Chứng minh rằng: Tập toàn thể các khóa của α tạo thành một hệ Sperner.

Câu II.

Cho tập thuộc tính $U = MNOPQR$.

Hãy xây dựng một lược đồ quan hệ trên U có đúng:

1. 20 khóa và chỉ rõ 20 khóa này.
2. một khóa duy nhất và chỉ rõ khóa này.
3. 4 khóa và chỉ rõ 4 khóa này.

Câu III.

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

$F = \{M \rightarrow MOP, PQ \rightarrow MNQ, N \rightarrow OR, R \rightarrow MQ, N \rightarrow NO, M \rightarrow MQ, PQ \rightarrow Q, N \rightarrow P, MN \rightarrow NQR\}$

1. Chứng minh rằng: Từ F , theo hệ tiên đề Armstrong ta có: $M \rightarrow R$
2. Các tập MQR, MQS có phải là khóa của α hay không? Vì sao?
3. Tìm toàn bộ các khóa của α ?
4. Có thể thêm hoặc bớt một phụ thuộc hàm trong F để lược đồ quan hệ α đã cho trên đây có đúng một khóa duy nhất hay không?
(giải thích rõ cách làm này)

Chương 5: PHỦ CỦA TẬP PHỤ THUỘC HÀM

MỞ ĐẦU

- Phủ của tập phụ thuộc hàm F là một tập phụ thuộc hàm G tương đương với F và G “đơn giản” hơn F . (đơn giản ở đây hiểu theo nghĩa: có số phụ thuộc hàm ít hơn và chiều dài vế phải/ vế trái của phụ thuộc hàm ngắn hơn)
- Việc nghiên cứu phủ của tập phụ thuộc hàm nhằm: Giúp tính toán/ giải quyết một cách nhanh chóng và chính xác bài toán tìm bao đóng của tập thuộc tính, tìm khoá của một lược đồ quan hệ cho trước, có tầm quan trọng trong việc phân tích và thiết kế các hệ cơ sở dữ liệu.
- Có nhiều loại phủ của tập phụ thuộc hàm nhưng trong tài liệu này chúng ta chỉ nghiên cứu 2 loại phủ cơ bản, đó là:
 - Phủ tự nhiên.
 - Phủ không dư.

§1 PHỦ CỦA TẬP PHỤ THUỘC HÀM

Cho 2 tập phụ thuộc hàm F và G trên U .

5.1.1. Định nghĩa

F tương đương với G (k/h: $F \sim G$) $\Leftrightarrow F^* = G^*$

5.1.2. Chú ý

Quan hệ \sim là một quan hệ tương đương, vì thỏa các tính chất sau:

- Tính chất phản xạ: $F^* = F^* \Rightarrow F \sim F$
- Tính chất đối xứng: $F \sim G \Rightarrow F^* = G^* \Rightarrow G^* = F^* \Rightarrow G \sim F$
- Tính chất bắc cầu: $\left\{ \begin{array}{l} F \sim G \Rightarrow F^* = G^* \\ G \sim H \Rightarrow G^* = H^* \end{array} \right\} \Rightarrow F^* = H^* \Rightarrow F \sim H$

5.1.3. Định nghĩa

G được gọi là phủ của F $\Leftrightarrow G \sim F$

5.1.4. Định nghĩa

G được gọi là phủ tự nhiên của F \Leftrightarrow thỏa các điều kiện sau đây:

- $G \sim F$ (G là phủ của F).
- $\forall X \rightarrow Y \in G: X \cap Y = \emptyset$
- $\forall X \rightarrow Y, Z \rightarrow W \in G: X \neq Z$

5.1.5. Định nghĩa

Phụ thuộc hàm f trong F được gọi là dư thừa $\Leftrightarrow F \setminus \{f\} \text{ — } f$

5.1.6. Nhận xét

Ta có: $F \setminus \{f\} \text{ — } f \Leftrightarrow f \in (F \setminus \{f\})^* \Leftrightarrow \text{right}(f) \subseteq (\text{left}(f))^+_{F \setminus \{f\}}$

5.1.7. Định nghĩa

F được gọi là tập không dư \Leftrightarrow Mọi phụ thuộc hàm f trong F đều là phụ thuộc hàm không dư thừa.

5.1.8. Định nghĩa

G được gọi là phủ không dư của F \Leftrightarrow thỏa mãn các điều kiện sau:

- G là phủ của F ($G \sim F$)
- G là tập không dư
- $G \subseteq F$

5.1.9. Nhận xét

- Phủ tự nhiên của F luôn tồn tại duy nhất.
- Phủ không dư của F luôn tồn tại và có thể không duy nhất.
- Thông thường để tìm bao đóng tập thuộc tính, khoá của một lược đồ quan hệ, chúng ta tiến hành tuần tự các phép biến đổi sau:
 - i) Tìm phủ tự nhiên G của F .
 - ii) Tìm phủ không dư H của G .

§2 CÁC KẾT QUẢ QUAN TRỌNG VỀ PHỦ CỦA TẬP PHỤ THUỘC HÀM

5.2.1. Định lý

$$F \sim G \Leftrightarrow X_F^+ = X_G^+, \forall X \subseteq U.$$

Chứng minh:

(\Rightarrow) Giả sử $F \sim G$. Ta cần chứng minh: $X_F^+ = X_G^+, \forall X \subseteq U$.

Ta c/m: $X_F^+ \subseteq X_G^+$

$$\left. \begin{array}{l} \forall A \in X_F^+ \Rightarrow X \rightarrow A \in F^* \\ F \sim G \Rightarrow F^* = G^* \end{array} \right\} \Rightarrow X \rightarrow A \in G^* \Rightarrow A \in X_G^+$$

Vậy: $X_F^+ \subseteq X_G^+ (1)$

Tương tự ta cũng chứng minh được: $X_G^+ \subseteq X_F^+ (2)$. Từ (1) và (2) ta suy ra: $X_F^+ = X_G^+$

(\Leftarrow) Giả sử ta có: $X_F^+ = X_G^+, \forall X \subseteq U$

Ta sẽ chứng minh: $F \sim G$ hay chứng minh: $F^* = G^*$

Ta c/m: $F^* \subseteq G^*$

Ta có:

$$\left. \begin{array}{l} \forall X \rightarrow Y \in F^* \Rightarrow Y \subseteq X_F^+ \\ X_F^+ = X_G^+ \end{array} \right\} \Rightarrow Y \subseteq X_G^+ \Rightarrow X \rightarrow Y \in G^* \text{ Vậy: } F^* \subseteq G^* (3)$$

Tương tự ta cũng chứng minh được: $G^* \subseteq F^* (4)$

Từ (3) và (4) ta có: $F^* = G^*$ Suy ra: $F \sim G$

5.2.2. Định lý

$$F \sim G \Rightarrow K_{F_0} = K_G$$

Ở đây: K_F là tập toàn thể các khoá của lược đồ quan hệ $\alpha = (U, F)$

K_G là tập toàn thể các khoá của lược đồ quan hệ $\beta = (U, G)$

§3 CÀI ĐẶT THUẬT TOÁN TÌM PHỦ CỦA TẬP PHỤ THUỘC HÀM

5.3.1. Thuật toán tìm phủ tự nhiên của F

Input: F

Output: Tìm phủ tự nhiên G của F?

Actions:

$G := \emptyset;$

For each $X \rightarrow Y \in F$ do

$Z := Y \setminus X;$

If $Z \neq \emptyset$ then

If $\exists X \rightarrow W \in G$ then

$G := G \setminus \{X \rightarrow W\} \cup \{X \rightarrow ZW\}$

Else

$G := G \cup \{X \rightarrow Z\};$

End if;

End if;

End for;

Return G;

End.

5.3.2. Ví dụ

Tìm phủ tự nhiên G của tập phụ thuộc hàm F sau đây:

$F = \{MN \rightarrow OP, O \rightarrow PO, N \rightarrow NO, M \rightarrow OM, MN \rightarrow NP, O \rightarrow ON, M \rightarrow MNO\}$

Đs: $G = \{MN \rightarrow OP, O \rightarrow NP, N \rightarrow O, M \rightarrow ON\}$

5.3.3. Thuật toán tìm phủ không dư của F

Input: F

Output: Tìm phủ không dư G của F?

Actions:

$G := F;$

For each $X \rightarrow Y \in G$ do

If $Y \subseteq X_{G \setminus \{X \rightarrow Y\}}^+$ then

$G := G \setminus \{X \rightarrow Y\};$

End if;

End for;

Return G;

End.

5.3.4. Ví dụ

Cho $F = \{O \rightarrow M, N \rightarrow P, ON \rightarrow M\}$

Kiểm tra xem phụ thuộc hàm: $ON \rightarrow M$ có dư thừa trong F ?

Đs: Ta có: $(ON)^+_{F \setminus \{ON \rightarrow M\}} = ONMP$

Mà $M \subseteq (ON)^+_{F \setminus \{ON \rightarrow M\}} \Rightarrow ON \rightarrow M$ là phụ thuộc hàm dư thừa trong F .

5.3.5. Ví dụ

Tìm phủ không dư H của tập phụ thuộc hàm F sau đây:

$F = \{AB \rightarrow CDE, B \rightarrow CE, C \rightarrow BD, AC \rightarrow BDE\}$

Đs: $H = \{B \rightarrow CE, C \rightarrow BD\}$

BÀI TẬP CHƯƠNG 5

Câu I.

1. Chứng minh định lý 2.
2. Cho một phản ví dụ để chứng tỏ rằng: chiều ngược lại của định lý 2 không xảy ra.
3. Cho $F = \{AC \rightarrow B\}$, $G = \{A \rightarrow C\}$. Hỏi: F có tương đương với G không? Tại sao?

Câu II.

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

$F = \{B \rightarrow OBE, OI \rightarrow BIC, C \rightarrow EG, G \rightarrow BI, OI \rightarrow I, C \rightarrow CE, C \rightarrow OC, B \rightarrow BI, BC \rightarrow BCIG\}$.

1. Chứng minh rằng: Từ F , theo hệ tiên đề Armstrong ta có: $B \rightarrow G$.
2. Tìm phủ tự nhiên F_1 của F ?
3. Tìm phủ không dư F_2 của F_1 ?

Chương 6: CHUẨN HÓA

MỞ ĐẦU

- Việc chuẩn hoá quan hệ/ lược đồ quan hệ trong cơ sở dữ liệu nhằm mục đích:
 - Đảm bảo lưu trữ dữ liệu tốn ít bộ nhớ.
 - Đảm bảo thời gian truy cập dữ liệu nhanh và chính xác.
 - Đảm bảo không nảy sinh những hiện tượng dị thường trong quá trình cập nhật dữ liệu.
- Năm 1972, E.F Codd đưa ra 3 dạng chuẩn của quan hệ/ lược đồ quan hệ như sau:
Dạng chuẩn 1 → Dạng chuẩn 2 → Dạng chuẩn 3
(1NF) (2NF) (3NF)
- Đến năm 1974, Boyce và Codd nhận thấy một quan hệ dù ở dạng chuẩn 3 nhưng vẫn còn xảy ra những dị thường trong khi cập nhật dữ liệu. Do đó cần phải xây dựng một dạng chuẩn mới mạnh hơn dạng chuẩn 3 nhằm tránh được những bất lợi nói trên và gọi tên là dạng chuẩn Boyce_Codd (BCNF)
- Một quan hệ ở dạng chưa chuẩn hoá \Leftrightarrow trong quan hệ tồn tại thuộc tính phức hợp. (Thuộc tính được gọi là phức hợp có nghĩa là thuộc tính này còn có thể phân chia nhỏ ra thành các thuộc tính con/ thuộc tính đơn)
- Một quan hệ ở dạng chưa chuẩn hoá sẽ tồn tại nhiều bất cập/ nhược điểm.
Vì vậy trong thực tế ứng dụng, để xây dựng và tổ chức một cơ sở dữ liệu quan hệ, thông thường người ta tiến hành chuẩn hoá một quan hệ cho trước ban đầu (quan hệ vũ trụ) ở dạng chưa chuẩn hoá thành các quan hệ con (quan hệ thành phần) đều ở dạng chuẩn 1 hoặc 2NF, 3NF, BCNF.

§1 TÁCH KHÔNG TỔN THẤT THÔNG TIN

6.1.1. Định nghĩa

Quan hệ $R(U)$ được gọi là tách không tổn thất thông tin thành các quan hệ con $R_i(U_i)$, với $i = 1..k \Leftrightarrow$ thỏa mãn các điều kiện sau đây:

$$i) \bigcup_{i=1}^k U_i = U$$

$$ii) R[U_i] = R_i(U_i), i = 1..k$$

$$iii) R(U) = \bigstar_{i=1}^k R_i(U_i)$$

6.1.2. Định lý Tách

Cho quan hệ $R(U)$.

Nếu quan hệ R thỏa phụ thuộc hàm $X \rightarrow Y$ (với $X, Y \subseteq U$ và $X \cap Y = \emptyset$) thì phép tách quan hệ $R(U)$ thành 2 quan hệ con: $R_1 = R[XY]$ và $R_2 = R[U \setminus Y]$ là không tổn thất thông tin.

§2 CÁC DẠNG CHUẨN CỦA QUAN HỆ

6.2.1. Dạng chuẩn 1 (1NF)

Quan hệ $R(U)$ được gọi là ở dạng chuẩn 1 \Leftrightarrow Mọi miền trị của thuộc tính trong R đều nguyên tố (tức: mọi thuộc tính trong R đều là những thuộc tính đơn)

6.2.2. Dạng chuẩn 2 (2NF)

6.2.2.1. Định nghĩa

Cho quan hệ $R(U)$ và $X, Y \subseteq U$. Ta nói:

Y phụ thuộc hàm toàn phần vào X (k/h: $X \twoheadrightarrow Y$) $\Leftrightarrow \begin{cases} X \rightarrow Y \\ \forall Z \subset X : Z \not\rightarrow Y \end{cases}$

6.2.2.2. Định nghĩa

Gọi K_α là tập toàn thể các khoá của lược đồ quan hệ α . Khi đó ta nói:

- Tập thuộc tính khóa của α (k/h: P) là: $P = \bigcup_{K \in K_\alpha} K$.

(Chú ý: P còn được gọi là tập thuộc tính nguyên thủy)

- Tập thuộc tính phi nguyên thủy của α (k/h: N) là: $N = U \setminus P$

6.2.2.3. Định nghĩa

Quan hệ R(U) ở dạng chuẩn 2 \Leftrightarrow Quan hệ này ở dạng chuẩn 1 và mọi thuộc tính phi nguyên thủy đều phụ thuộc hàm toàn phần vào mọi khoá.

$$\text{Nói gọn: } R - 2NF \Leftrightarrow \begin{cases} R - 1NF \\ \forall A \in N, \forall K \in K_{\alpha} : K \twoheadrightarrow A \end{cases}$$

6.2.3. Dạng chuẩn 3 (3NF)

6.2.3.1. Định nghĩa

Cho quan hệ R(U) và $X, Y \subseteq U$. Ta nói:

Y phụ thuộc hàm bậc cầu vào X (k/h: $X \rightsquigarrow Y$) $\Leftrightarrow \exists Z \subset U : X \rightleftarrows Z$
(với $Z \cap Y = \emptyset$).

\downarrow
Y

6.2.3.2. Định nghĩa

Quan hệ R(U) ở dạng chuẩn 3 \Leftrightarrow Quan hệ này ở dạng chuẩn 1 và mọi thuộc tính phi nguyên thủy đều không phụ thuộc hàm bậc cầu vào mọi khoá.

$$\text{Nói gọn: } R - 3NF \Leftrightarrow \begin{cases} R - 1NF \\ \forall A \in N, \forall K \in K_{\alpha} : K \not\rightsquigarrow A \end{cases}$$

6.2.4. Dạng Boyce_Codd (BCNF)

6.2.4.1. Định nghĩa:

Quan hệ R(U) ở dạng chuẩn Boyce_Codd \Leftrightarrow Quan hệ này ở dạng chuẩn 1 và mọi thuộc tính của U không phụ thuộc hàm bậc cầu vào mọi khoá.

$$\text{Nói gọn: } R - BCNF \Leftrightarrow \begin{cases} R - 1NF \\ \forall A \in U, \forall K \in K_{\alpha} : K \not\rightsquigarrow A \end{cases}$$

6.2.4.2. Định lý

$$R - BCNF \Rightarrow R - 3NF \Rightarrow R - 2NF \Rightarrow R - 1NF$$

6.2.4.3. Định lý

Quan hệ R(U) ở dạng chuẩn Boyce_Codd \Leftrightarrow Quan hệ này ở dạng chuẩn 1 và mọi phụ thuộc hàm không tầm thường $X \rightarrow Y$ đều cho X là một siêu khoá.

§3 CÁC THUẬT TOÁN CHUẨN HÓA

6.3.1. Thuật toán chuẩn hóa quan hệ ở dạng 1NF về dạng 2NF

Input: $R(U) - 1NF$ ($R(U)$ chưa ở dạng chuẩn 2)

Output: Tách không tổn thất thông tin quan hệ $R(U)$ ra thành các quan hệ con

$R_i(U_i)$ với $i = 1..k$, sao cho tất cả các quan hệ con này đều ở dạng chuẩn 2.

Actions:

B1. Tìm tập toàn thể các khoá của R (K_α)

B2. Tìm tập thuộc tính khoá P .

B3. Tìm tập thuộc tính phi nguyên thuỷ \mathcal{N} .

B4. Vì R chưa ở dạng chuẩn 2NF nên trong R : $\exists A \in \mathcal{N}, \exists K \in K_\alpha, \exists X \subset K$ sao cho: $X \rightarrow A$. Chọn tập X nhỏ nhất của K thỏa $X \rightarrow A$ rồi tách quan hệ R ra thành 2 quan hệ con R_1, R_2 với $R_1 = R[XA], R_2 = R[U \setminus A]$.

Rõ ràng việc tách này đảm bảo không tổn thất thông tin (theo định lý tách) và R_1 ở dạng chuẩn 2NF (vì $X \rightarrow A$ nên X là siêu khóa của R_1 , vì X được chọn là nhỏ nhất nên X chính là khóa của R_1 và cũng không tồn tại một bộ phận Y nào của X để $Y \rightarrow A$ nên $X \twoheadrightarrow A$).

Nếu R_2 cũng ở dạng chuẩn 2NF thì dừng thuật toán. Ngoài ra thì tiến hành thuật toán trên đối với R_2 .

Nhận xét: Sau mỗi lần tách, tập thuộc tính U sẽ giảm đi số thuộc tính. Vì U hữu hạn nên đến một số hữu hạn bước thực hiện nào đó thuật toán sẽ dừng.

6.3.2. Thuật toán chuẩn hóa quan hệ ở dạng 1NF về dạng 3NF

Input: $R(U)$ – 1NF ($R(U)$ chưa ở dạng chuẩn 3)

Output: Tách không tổn thất thông tin quan hệ $R(U)$ ra thành các quan hệ con $R_i(U_i)$ với $i = 1..k$, sao cho tất cả các quan hệ con này đều ở dạng chuẩn 3.

Actions:

B1. Tìm tập toàn thể các khóa của R (K_α)

B2. Tìm tập thuộc tính khóa P .

B3. Tìm tập thuộc tính phi nguyên thủy \mathcal{N} .

B4. Vì R chưa ở dạng chuẩn 3NF nên trong R : $\exists A \in \mathcal{N}, \exists K \in K_\alpha$ sao cho: $K \rightsquigarrow A$, hay tồn tại: $A \in \mathcal{N}, K \in K_\alpha, Z \subset U$ sao cho:

$$\begin{array}{ccc} K & \longleftrightarrow & Z \\ & \downarrow & \\ & A, A \notin Z & (*) \end{array}$$

Chọn tập Z nhỏ nhất của U thỏa mãn (*) rồi tách quan hệ R ra thành 2 quan hệ con R_1, R_2 với $R_1 = R[ZA], R_2 = R[U \setminus A]$.

Rõ ràng việc tách này đảm bảo không tổn thất thông tin (theo định lý tách) và R_1 ở dạng chuẩn 3NF (vì $Z \rightarrow A$ nên Z là siêu khóa của R_1 . Vì Z được chọn là nhỏ nhất của U nên Z chính là khóa duy nhất của R_1 và do đó ta có: $Z \not\rightsquigarrow A$).

Nếu R_2 cũng ở dạng chuẩn 3NF thì dừng thuật toán. Ngoài ra thì tiến hành thuật toán trên đối với R_2 .

Nhận xét: Sau mỗi lần tách, tập thuộc tính U sẽ giảm đi số thuộc tính. Vì U hữu hạn nên đến một số hữu hạn bước thực hiện nào đó thuật toán sẽ dừng.

BÀI TẬP CHƯƠNG 6

Câu I.

1. Chứng minh: định lý 1, định lý 2 và định lý tách.
2. Cho phản ví dụ chứng tỏ rằng: chiều ngược lại của định lý 1 không xảy ra.
3. Cho một ví dụ thực tế minh họa phép tách không tổn thất thông tin.
4. Cho ví dụ thực tế minh họa quan hệ ở dạng chuẩn 2, 3, BCNF.
5. Viết thuật toán chuẩn hóa quan hệ ở dạng 1NF về dạng BCNF.

Câu II.

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

và $F = \{ AB \rightarrow CDE, B \rightarrow BC, C \rightarrow CD, B \rightarrow E, AC \rightarrow BDE, C \rightarrow BC \}$

1. Xác định dạng chuẩn cao nhất của α ? (giải thích rõ)
2. Tách không tổn thất thông tin lược đồ quan hệ α đã cho thành những lược đồ quan hệ con đều ở dạng chuẩn 2/ 3/ Boyce_Codd.

Câu III.

Chuẩn hóa 3NF lược đồ quan hệ $\alpha = (U, F)$ sau đây:

$U = MLTGSDP$

$F = \{ M \rightarrow T, GP \rightarrow M, GT \rightarrow P, MS \rightarrow D, GS \rightarrow P \}$

Với ngữ nghĩa sau:

M: Môn học chuyên đề

L: Lớp học

T: Thầy - phụ trách chuyên đề

G: Giờ học chuyên đề

S: Sinh viên theo học chuyên đề

D: Số đăng ký của sinh viên trong chuyên đề đó

P: Phòng học dành cho chuyên đề

$M \rightarrow T$: Mỗi chuyên đề có một thầy phụ trách

$GP \rightarrow M$: Tại mỗi thời điểm, mỗi phòng học được dành cho không quá một môn

$GT \rightarrow P$: Tại mỗi thời điểm, mỗi thầy dạy trong không quá một phòng học

$MS \rightarrow D$: Mỗi sinh viên tham gia chuyên đề nào thì được cấp một mã số ghi danh theo chuyên đề đó

$GS \rightarrow P$: Tại mỗi thời điểm, mỗi sinh viên có mặt trong không quá một phòng học

TÀI LIỆU THAM KHẢO

- [1]. **Lê Tiến Vương**, *Nhập môn cơ sở dữ liệu quan hệ*, NXB Khoa học và Kỹ thuật, 2000.
- [2]. **Nguyễn Bá Tường**, *Cơ sở dữ liệu - lý thuyết và thực hành*, NXB Khoa học - Kỹ thuật Hà Nội, 2001.
- [3]. **Nguyễn Kim Anh**, *Nguyên lý của các hệ Cơ sở dữ liệu*, NXB Đại học Quốc gia Hà Nội, 2004.
- [4]. **David Maier**, *The theory of relational database*, Computer Science Press, 1983.
- [5]. **Jeffrey D. Ullman**, *The principles of database and knowledge base system* Vol. 1, 2, Computer Science Press, 1989.