

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

LÊ MẠNH HÀ

**NGHIÊN CỨU PHỤ THUỘC MẠNH TRONG
CƠ SỞ DỮ LIỆU**

Chuyên ngành: Khoa học máy tính

Mã số: 60 48 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC

GS.TS. VŨ ĐỨC THI

Thái Nguyên - 2014

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn là công trình nghiên cứu của riêng cá nhân tôi, không sao chép của ai do tôi tự nghiên cứu, đọc, dịch tài liệu, tổng hợp và thực hiện. Nội dung lý thuyết trong luận văn tôi có sử dụng một số tài liệu tham khảo như đã trình bày trong phần tài liệu tham khảo. Chương trình phần mềm và những kết quả trong luận văn là trung thực và chưa được công bố trong bất kỳ một công trình nào khác.

Thái Nguyên, ngày 7 tháng 6 năm 2014

Học viên thực hiện

Lê Mạnh Hà

LỜI CẢM ƠN

Lời đầu tiên, em xin gửi lời biết ơn sâu sắc đến GS.TS Vũ Đức Thi người đã tận tình hướng dẫn, chỉ bảo, giúp đỡ em trong suốt quá trình làm luận văn.

Em cũng xin gửi lời cảm ơn đến các giảng viên trường Đại học Công nghệ thông tin và Truyền thông - Đại học Thái Nguyên, các thầy Viện Công nghệ thông tin đã truyền đạt những kiến thức và giúp đỡ em trong suốt quá trình học của mình.

Tôi cũng xin gửi lời cảm ơn tới các đồng nghiệp, gia đình và bạn bè những người đã ủng hộ, động viên tạo mọi điều kiện giúp đỡ để tôi có thể hoàn thành tốt luận văn.

Tôi cũng xin gửi lời cảm ơn tới Ban giám hiệu Trường Cao đẳng Sư phạm Quảng Ninh đã tạo kiện thuận lợi cho tôi tham gia khóa học và trong suốt quá trình hoàn thành luận văn.

Một lần nữa, xin chân thành cảm ơn.

Thái Nguyên, ngày 7 tháng 6 năm 2014

Học viên thực hiện

Lê Mạnh Hà

MỤC LỤC

Lời cam đoan	
Lời cảm ơn	
Mục lục	
Phần mở đầu	
Bảng các kí hiệu	
Bảng các hình vẽ	
Chương 1: Phụ thuộc hàm	
1.1. Định nghĩa.....	
1.2. Hệ tiên đề ArmStrong	
1.3. Bao đóng của tập phụ thuộc hàm và tập thuộc tính.....	
1.4. Khoá tối thiểu của sơ đồ quan hệ và quan hệ	
1.5. Các dạng chuẩn.....	
1.6. Hệ Sperner	
1.7. Các dạng tương đương của họ phụ thuộc hàm	
1.8. Kết luận.....	
Chương 2: Phụ thuộc mạnh và một số tính chất đặc trưng của phụ thuộc mạnh	
2.1. Định nghĩa.....	
2.2. Hệ tiên đề phụ thuộc mạnh	
2.3. Bao đóng của tập phụ thuộc mạnh và tập thuộc tính.....	
2.4. Khoá tối thiểu của sơ đồ mạnh và quan hệ.....	
2.5. Các dạng tương đương của họ phụ thuộc mạnh	
2.6. Một số tính chất cơ bản của bao đóng của tập thuộc tính	
2.7. Thuật toán tính bao đóng của tập thuộc tính trên quan hệ.....	
2.8. Họ các tập tối thiểu	
2.9. Quan hệ ArmStrong của phụ thuộc mạnh	

2.9.1. Sự tồn tại của quan hệ ArmStrong	
2.92. Các thuật toán	
2.9.3. Một số bài toán quan trọng	
2.10. Kết luận	
Chương 3: Cài đặt một số thuật toán về phụ thuộc mạnh trong CSDL	
3.1. Lựa chọn bài toán.	
3.2 Thuật toán sử dụng trong chương trình.....	
3.3.Cài đặt chương trình.....	
3.4.Một số mã lệnh.....	
3.5. Hướng dẫn sử dụng chương trình.	
3.6Chương trình minh họa.....	
3.7 Đánh giá kết quả thực nghiệm	
Kết luận	
Tài liệu tham khảo	

MỞ ĐẦU

Cơ sở dữ liệu (CSDL) là một trong những lĩnh vực được tập trung nghiên cứu và phát triển của công nghệ thông tin, nhằm giải quyết các bài toán quản lý, tìm kiếm thông tin trong những hệ thống lớn đa dạng, phức tạp cho nhiều người sử dụng trên máy tính điện tử. Có thể nói E. F. Codd là người đầu tiên đề xuất mô hình dữ liệu quan hệ cho CSDL với công trình [6] mà ngày nay đã trở thành kinh điển. Đây là mô hình được xây dựng trên cơ sở lý thuyết toán học về các quan hệ, bao gồm các thực thể (đối tượng) và các mối quan hệ qua lại giữa chúng. Chỉ điều này đã tạo cơ sở toán học với cấu trúc hoàn chỉnh làm nền tảng cho các vấn đề nghiên cứu lý thuyết về CSDL.

Người ta xem CSDL quan hệ như là một tập hợp hữu hạn các quan hệ. Trong đó mỗi quan hệ có thể được hình dung một cách trực quan như là một bảng chữ nhật gồm có các hàng và các cột. Mỗi hàng là một bản ghi (record) lưu trữ các dữ liệu. Mỗi cột là một thuộc tính.

Trong lý thuyết thiết kế CSDL quan hệ, ràng buộc dữ liệu hay còn gọi là phụ thuộc dữ liệu có một ý nghĩa quan trọng trong việc đảm bảo tính nhất quán của dữ liệu. Nghiên cứu ràng buộc dữ liệu là một vấn đề cần thiết. Ý nghĩa của việc nêu ra khái niệm ràng buộc dữ liệu là nhằm đảm bảo cho dữ liệu trong CSDL không mâu thuẫn, phản ánh đúng thế giới hiện thực. Các nhà nghiên cứu đã đưa ra nhiều loại ràng buộc dữ liệu khác nhau để đáp ứng phù hợp với thực tế rất phong phú và đa dạng. Loại ràng buộc dữ liệu đầu tiên là phụ thuộc hàm được giới thiệu bởi E. F. Codd [6] vào những năm 1970. Ba loại ràng buộc dữ liệu khác sau đó cũng được xem xét đến là phụ thuộc đối ngẫu, phụ thuộc mạnh và phụ thuộc yếu bởi Czedli [7, 8] (1980). Tiếp sau đó J. Demetrovics và G.Gyepesi [11] (1983), và những người khác [1, 15, 26]

cũng tiếp tục nghiên cứu các ràng buộc dữ liệu này. Với ba loại ràng buộc dữ liệu này, người sử dụng đôi khi có thể lấy được các thông tin thực mong muốn, ngay cả khi không tồn tại một phụ thuộc hàm nào giữa các tập thuộc tính và chỉ cần biết ít nhất một giá trị của các thuộc tính chứ không phải tập toàn bộ các giá trị của các thuộc tính của vế trái. Hơn thế nữa, đôi khi việc xử lý và tìm kiếm thông tin được tiến hành nhanh chóng hơn vì chỉ cần phải tìm kiếm trên một phần của quan hệ mà thôi.

Mục tiêu của luận văn là tiếp tục nghiên cứu phụ thuộc hàm và phụ thuộc mạnh. Luận văn bao gồm: Phần mở đầu, 3 chương và phần kết luận.

Chương 1: Nhắc lại một số khái niệm cơ bản về phụ thuộc hàm và quan hệ Armstrong.

Chương 2: Mục đích của chương này là trình bày nghiên cứu quan hệ Armstrong đối với phụ thuộc mạnh. Có thể nói, trong nghiên cứu về các ràng buộc dữ liệu nói chung và phụ thuộc mạnh nói riêng, khái niệm bao đóng của tập thuộc tính thật sự đóng một vai trò quan trọng. Kết quả chính là trình bày một số nghiên cứu về quan hệ Armstrong. Đầu tiên, khái niệm họ các tập tối tiểu của thuộc tính của một sơ đồ mạnh được đề xuất. Đây là khái niệm đóng vai trò quan trọng trong việc xây dựng quan hệ Armstrong của sơ đồ mạnh

Cuối cùng, luận văn đề cập đến bốn bài toán quan trọng đối với việc nghiên cứu cấu trúc và logic của họ phụ thuộc mạnh: bài toán xây dựng quan hệ Armstrong của một sơ đồ mạnh cho trước, bài toán xây dựng sơ đồ mạnh đúng trên một quan hệ cho trước, bài toán kéo theo phụ thuộc mạnh-quan hệ và bài toán tương đương phụ thuộc mạnh-quan hệ. Tất cả các bài toán này được chứng tỏ có thể được giải quyết bằng các thuật toán thời gian đa thức.

Chương 3: Cài đặt chương trình để minh họa phụ lý thuyết phụ thuộc mạnh.

BẢNG CÁC KÍ HIỆU

Trong luận văn, có sử dụng các quy ước về kí hiệu và chữ viết tắt sau:

$S=(U,F)$: Sơ đồ quan hệ, với U là tập các thuộc tính và F là tập các phụ thuộc hàm trên U

$G=(U,S)$: Sơ đồ mạnh, với U là tập thuộc tính và S là tập các phụ thuộc mạnh trên U .

R : quan hệ trên tập thuộc tính U .

F_R : (tương ứng S_R) tập tất cả các phụ thuộc hàm (tương ứng phụ thuộc mạnh) đúng trên quan hệ R .

F^+ : (tương ứng S^+) bao đóng của tập phụ thuộc hàm (tương ứng phụ thuộc mạnh).

K_S : (tương ứng K_G, K_R) tập tất cả các khóa tối thiểu của sơ đồ quan hệ s (tương ứng sơ đồ mạnh, quan hệ R).

K_S^{-1} : (tương ứng K_R^{-1}) tập phản khóa của K_S (tương ứng K_R).

K_a : họ các tập tối thiểu của thuộc tính a .

F_n : (tương ứng D_n) tập tất cả các thuộc tính không cơ bản (tương ứng thuộc tính phụ thuộc).

\mathcal{E}_R : (tương ứng N_R) hệ bằng nhau (tương ứng không bằng nhau) quan hệ của R .

I_S : họ các tập độc lập tối thiểu của sơ đồ quan hệ s .

L_S : (tương ứng L_R) họ tất cả các bao đóng của tập thuộc tính của sơ đồ quan hệ s (tương ứng quan hệ R).

CSDL: cơ sở dữ liệu.

BẢNG CÁC HÌNH VẼ

Hình 1.1: Phụ thuộc hàm.

Hình 1.2 : Tương quan giữa lớp quan hệ với lớp phụ thuộc hàm.

Hình 1.3 : Tương quan giữa lớp phụ thuộc hàm với lớp hàm đóng.

Hình 1.4 : Lớp các dạng chuẩn.

Hình 1.5 : Tương quan giữa họ phụ thuộc hàm với hệ Sperner.

Hình 1.6 : Tương quan lớp các hàm đóng với hàm chọn đặc biệt.

Hình 1.7 : Tương quan giữa lớp các hàm đóng và nửa dàn giao.

Hình 1.8 : Tương quan giữa lớp các nửa dàn giao và tập không giao.

Chương 1: PHỤ THUỘC HÀM

Khái niệm phụ thuộc hàm được đề xuất bởi E.F.Codd và sau đó lần đầu tiên được tiên đề hoá bởi W.W.Armstrong [4](1974). Đây là một loại ràng buộc dữ liệu xảy ra tự nhiên nhất giữa các tập thuộc tính, về cơ bản các hệ quản trị cơ sở dữ liệu lớn đều sử dụng phụ thuộc này.

1.1 Định nghĩa

Cho U là tập hữu hạn khác rỗng các thuộc tính. Một *phụ thuộc hàm* là một mệnh đề có dạng $X \rightarrow Y$, trong đó $X, Y \subseteq U$. Ta nói phụ thuộc hàm $X \rightarrow Y$ *đúng* trong quan hệ $R = \{h_1, \dots, h_m\}$ trên U nếu:

$$(\forall h_i, h_j \in R)((\forall a \in X)(h_i(a) = h_j(a)) \implies (\forall b \in Y)(h_i(b) = h_j(b)))$$

Ta cũng nói rằng R *thỏa* phụ thuộc hàm $X \rightarrow Y$.

Gọi F_R là họ tất cả các phụ thuộc hàm đúng trên quan hệ R .

Ta có thể nhận thấy rằng X mà phụ thuộc hàm vào Y , nếu hai dòng bất kỳ mà các giá trị của tập thuộc tính X mà bằng nhau từng cặp một, thì kéo theo các giá trị trên tập thuộc tính Y cũng phải bằng nhau từng cặp một.

Với định nghĩa này dễ thấy rằng trong các file dữ liệu cột, mã số hoặc thứ tự không thể bằng nhau.