

ĐẠI HỌC THÁI NGUYÊN
KHOA CÔNG NGHỆ THÔNG TIN

VŨ TRÍ DŨNG

ỨNG DỤNG PHÉP DỊCH CHUYỂN
LƯỢC ĐỒ QUAN HỆ TRONG CƠ SỞ DỮ LIỆU

LUẬN VĂN THẠC SĨ CÔNG NGHỆ THÔNG TIN

Thái Nguyên - 2009

ĐẠI HỌC THÁI NGUYÊN
KHOA CÔNG NGHỆ THÔNG TIN

VŨ TRÍ DŨNG

ỨNG DỤNG PHÉP DỊCH CHUYỂN
LƯỢC ĐỒ QUAN HỆ TRONG CƠ SỞ DỮ LIỆU

CHUYÊN NGÀNH : KHOA HỌC MÁY TÍNH
MÃ SỐ : 60 48 35 01

LUẬN VĂN THẠC SĨ CÔNG NGHỆ THÔNG TIN

Người hướng dẫn khoa học
PGS. TSKH. NGUYỄN XUÂN HUY

Thái Nguyên - 2009

MỤC LỤC

LỜI NÓI ĐẦU	1
CHƯƠNG 1. TỔNG QUAN VỀ ĐỀ TÀI VÀ CÁC KHÁI NIỆM CƠ SỞ	4
1.1. TỔNG QUAN VỀ ĐỀ TÀI	4
1.1.1. Giới thiệu đề tài.	4
1.1.2. Nội dung của đề tài, các vấn đề cần giải quyết.	4
1.1.3. Phương pháp nghiên cứu.	5
1.1.4. Phạm vi ứng dụng.	5
1.1.5. Kết quả đạt được.	5
1.2. CÁC KHÁI NIỆM CƠ SỞ	6
1.2.1. Quan hệ, thuộc tính, bộ.	7
1.2.2. Đại số quan hệ.	10
1.2.3. Phụ thuộc hàm, Hệ tiên đề Armstrong, Lược đồ quan hệ.	13
1.2.4. Bao đóng của tập thuộc tính.	18
1.2.5. Phủ của tập phụ thuộc hàm	21
1.2.6. Khóa của lược đồ quan hệ.	27
1.2.7. Chuẩn hoá LĐQH trên cơ sở phụ thuộc hàm.	31
CHƯƠNG 2. PHÉP DỊCH CHUYỂN LƯỢC ĐỒ QUAN HỆ	36
2.1. Phép dịch chuyển LĐQH.	37
2.2. Thuật toán dịch chuyển LĐQH.	38
2.3. Định lý cơ bản của phép dịch chuyển LĐQH.	39
2.4. Dạng biểu diễn thứ nhất của khóa	43

2.5.	Dạng biểu diễn thứ hai của khóa	45
2.6.	Kết luận	50
CHƯƠNG 3. CÀI ĐẶT CHƯƠNG TRÌNH		51
3.1.	Giới thiệu.	51
3.2.	Các chức năng của chương trình.	51
3.3.	Một số giao diện của chương trình.	52
3.4.	Các thí dụ.	54
DANH MỤC BÀI BÁO, CÔNG TRÌNH NCKH		57
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN		58
TÀI LIỆU THAM KHẢO		60

DANH MỤC CÁC KÝ HIỆU, VIẾT TẮT

1NF	1 st normal form - dạng chuẩn 1
2NF	2 nd normal form - dạng chuẩn 2
3NF	3 rd normal form - dạng chuẩn 3
CSDL	Cơ sở dữ liệu
LĐQH	Lược đồ quan hệ
PTH	phụ thuộc hàm
FD	phụ thuộc hàm
\vDash	suy dẫn theo tiên đề (theo logic)
\vdash	suy dẫn theo quan hệ
\neq	khác
\forall	với mọi
\in	thuộc
\subseteq	là con
\supseteq	chứa
\cap	giao (của 2 tập thuộc tính)
\cup	hợp (của 2 tập thuộc tính)
X^+	bao đóng của tập thuộc tính X
\equiv	tương đương
$\not\equiv$	không tương đương
\setminus	phép trừ logic

LỜI NÓI ĐẦU

Trong quản lý các cơ sở dữ liệu (CSDL), phụ thuộc dữ liệu được hiểu là những mệnh đề mô tả các ràng buộc mà dữ liệu phải đáp ứng trong thực tế. Nhờ có những mô tả phụ thuộc này mà hệ quản trị cơ sở dữ liệu có thể quản lý tốt được chất lượng dữ liệu. Lý thuyết về các phụ thuộc dữ liệu đóng vai trò quan trọng trong việc mô tả thế giới thực, phản ánh ngữ nghĩa dữ liệu trong cơ sở dữ liệu. Phụ thuộc dữ liệu được Codd, tác giả của mô hình dữ liệu quan hệ đặt nền móng từ những năm 70 với khái niệm phụ thuộc hàm. Sau đó một loạt tác giả khác tiếp tục phát triển các dạng phụ thuộc bậc cao, phụ thuộc mờ cũng như xây dựng các hệ tiên đề cho các lớp phụ thuộc - tức là đặt cơ sở lý thuyết về phụ thuộc dữ liệu.

Một điều khá tự nhiên là ngay từ những ngày đầu phát triển lý thuyết thiết kế cơ sở dữ liệu, logic đã được chọn như một ngôn ngữ hữu hiệu để đặc tả phụ thuộc dữ liệu, do đó, trong số các loại hình phụ thuộc dữ liệu rất đa dạng được đề xuất và phát triển sau này, các phụ thuộc logic luôn luôn là trọng tâm chú ý của các nhóm nghiên cứu.

Đề tài này tập trung vào tìm hiểu và nghiên cứu khái niệm chuyên dịch lược đồ quan hệ, đưa chúng về dạng thu gọn và nhận được các biểu diễn quan trọng cho bao đóng, khóa và phân khóa. Các kết quả thu được sử dụng trong quá trình thiết kế các cơ sở dữ liệu.

Nội dung đề tài được cấu trúc như sau:

Chương 1 giới thiệu về đề tài và các *khái niệm chung về mô hình quan hệ* với trọng tâm là các khái niệm hình thức của mô hình quan hệ, trong đó vận dụng chủ yếu các cấu trúc rời rạc. *Phụ thuộc hàm* (PTH) là lớp phụ thuộc đầu tiên của *phụ thuộc logic* và đồng thời cũng là lớp phụ thuộc kinh điển theo nghĩa, được Codd, tác giả của mô hình dữ liệu quan hệ, đề xuất sớm nhất và được sử dụng như một công cụ thiết kế các cơ sở dữ liệu chuẩn hóa.

Chương 2 trình bày một kỹ thuật thu gọn lược đồ quan hệ (LĐQH) được gọi là *phép dịch chuyển lược đồ quan hệ*. Bản chất của kỹ thuật này là loại bỏ khỏi LĐQH ban đầu một số thuộc tính *không quan trọng* theo nghĩa chúng không làm ảnh hưởng đến kết quả tính toán các đối tượng đang được quan tâm như bao đóng, khóa, phản khóa... Mặc dù LĐQH thu được qua phép dịch chuyển không tương đương với LĐQH ban đầu, nhưng ta có thể thu được các đối tượng cần tìm bằng những phép toán đơn giản như loại bỏ hoặc thêm một số thuộc tính. Điều lý thú là sau khi loại bỏ một số thuộc tính thì một số PTH sẽ được loại bỏ theo vì chúng trở thành các PTH tầm thường (có vế trái chứa vế phải) hoặc mang thông tin tiền định (đó là các PTH dạng $\emptyset \rightarrow X$). Các phép dịch chuyển LĐQH được phát triển cho lớp các phụ thuộc logic đầu tiên là phụ thuộc hàm cho ta một số kết quả lý thú về biểu diễn bao đóng, khóa, phản khóa cùng một số dấu hiệu cần và đủ để nhận biết các đặc trưng tương quan giữa các đối tượng nói trên.

Chương 3 cài đặt chương trình mô phỏng *ứng dụng phép dịch chuyển lược đồ quan hệ* vào thiết kế cơ sở dữ liệu cùng với một số thí dụ.

Phần cuối của luận văn là kết luận và hướng phát triển và các tài liệu tham khảo.

Em xin bày tỏ lòng chân thành cảm ơn PGS TSKH Nguyễn Xuân Huy - người Thầy đã tận tình hướng dẫn, giúp đỡ em hoàn thành luận văn này.

Em xin chân thành cảm ơn Khoa Công nghệ thông tin - Đại học Thái Nguyên đã tạo điều kiện về tinh thần cũng như cơ sở vật chất để em được học tập, nâng cao kiến thức và thực hiện luận văn tốt nghiệp.

Em xin chân thành cảm ơn các Thầy, Cô giáo ở Viện Công nghệ thông tin - Viện Khoa học và Công nghệ Việt Nam, các Thầy, Cô giáo ở Khoa Công nghệ thông tin - Đại học Thái Nguyên đã nhiệt tình giảng dạy, hướng dẫn và cung cấp cho em những kiến thức vô cùng quý báu, để em có điều kiện nâng cao kiến thức và hiểu biết của mình trong lĩnh vực công nghệ thông tin.

Em cũng xin chân thành cảm ơn Ban lãnh đạo Liên đoàn Lao động tỉnh Hà Nam, Ban giám hiệu Trường trung cấp nghề Kinh tế - Kỹ thuật Hà Nam, gia đình, người thân và bạn bè đã tạo điều kiện thuận lợi, động viên và giúp đỡ em trong suốt thời gian học tập, nghiên cứu và làm luận văn tốt nghiệp.

Học viên

Vũ Trí Dũng

CHƯƠNG 1

TỔNG QUAN VỀ ĐỀ TÀI VÀ CÁC KHÁI NIỆM CƠ SỞ

1.1. TỔNG QUAN VỀ ĐỀ TÀI

1.1.1. Giới thiệu đề tài

Trong quản lý các cơ sở dữ liệu lớn và phức tạp đòi hỏi nhiều thuật toán hữu hiệu để tính toán các đối tượng như bao đóng, khóa, phủ... Một số thuật toán tốt theo nghĩa độ phức tạp tính toán giới hạn ở các hàm tuyến tính hoặc đa thức theo chiều dài dữ liệu vào đã được công bố như thuật toán tính bao đóng của tập thuộc tính, thuật toán tìm một khóa, thuật toán xác định thành viên hay thuật toán xác định phụ thuộc hàm suy dẫn, thuật toán tìm giao các khóa, thuật toán xác định một lược đồ quan hệ có một khóa duy nhất... [1, 2, 8].

Một nhận xét tự nhiên là nếu kích thước của lược đồ quan hệ càng nhỏ thì các thuật toán càng phát huy hiệu quả hơn. Một số hướng nghiên cứu tinh giản các lược đồ cơ sở dữ liệu được thực hiện thông qua các phép biến đổi tương đương, chẳng hạn đưa tập phụ thuộc hàm về dạng thu gọn hoặc thu gọn tự nhiên, dạng không dư, dạng tối ưu (chứa ít ký hiệu nhất)... đã được công bố [3, 5, 6, 7].

Trong *phép dịch chuyển lược đồ quan hệ*. Bản chất của kỹ thuật này là loại bỏ khỏi lược đồ quan hệ ban đầu một số thuộc tính *không quan trọng* theo nghĩa chúng không làm ảnh hưởng đến kết quả tính toán các đối tượng đang quan tâm như bao đóng, khóa,... Mặc dù lược đồ quan hệ thu được qua phép thu gọn không tương đương với lược đồ quan hệ ban đầu, nhưng ta có thể thu được các đối tượng cần tìm bằng những phép toán đơn giản như loại bỏ hoặc thêm một số thuộc tính. Điều lý thú là sau khi loại bỏ một số thuộc tính thì một số phụ thuộc hàm sẽ được loại bỏ theo, vì chúng trở thành các phụ thuộc hàm tầm thường (có vế trái chứa vế phải) hoặc mang thông tin tiên định (đó là các phụ thuộc hàm dạng $\emptyset \rightarrow X$).

1.1.2. Nội dung của đề tài, các vấn đề cần giải quyết

Luận văn tập trung tìm hiểu và cải tiến các kỹ thuật và thuật toán thu gọn lược đồ quan hệ p thông qua phép dịch chuyển lược đồ quan hệ theo một tập thuộc tính X . Khảo sát sự phụ thuộc của phép dịch chuyển thông qua các tính chất của tập thuộc tính X . Khảo sát hai dạng biểu diễn khóa của lược đồ quan hệ qua phép dịch chuyển. Xây dựng một hệ trình minh họa và đánh giá các kết quả lý thuyết.

1.1.3. Phương pháp nghiên cứu

1. Tiếp cận chủ yếu để giải quyết các vấn đề đặt ra trong phạm vi đề tài là tiên đề hóa. Các hệ tiên đề được xây dựng trên cơ sở một hệ suy dẫn hình thức với các tính chất cơ bản về các đối tượng cơ sở và các mối liên hệ giữa chúng. Cơ sở toán học của các hệ tiên đề là định lý về tính *xác đáng* và *đầy đủ* cùng với các định lý về *điều kiện cần và đủ* cho các *hệ tiên đề tương đương*.
2. Tiếp cận hình thức vận dụng chủ yếu các phương pháp và các cấu trúc của toán học rời rạc (bao gồm cả logic hình thức), kết hợp với các phương pháp đối sánh, mô hình hóa, tối ưu và quy hoạch rời rạc.
3. Kết hợp chặt chẽ giữa lý thuyết và thực hành, sử dụng và phát triển các phần mềm nói chung và các phần mềm toán học nói riêng để kiểm định và thể hiện các kết quả lý thuyết.

1.1.4. Phạm vi ứng dụng

Các kết quả thu được có thể vận dụng cho các quy trình thiết kế các cơ sở dữ liệu quan hệ dùng trong các hệ thống thông tin, cụ thể là:

- Tính bao đóng của các tập thuộc tính,
- Tìm khóa của các lược đồ quan hệ.
- Chuẩn hoá LĐQH