

UBND TỈNH THANH HÓA
TRƯỜNG CAO ĐẲNG NGHỀ CÔNG NGHIỆP THANH HÓA



GIÁO TRÌNH

CƠ SỞ DỮ LIỆU

NGÀNH, NGHỀ: CÔNG NGHỆ THÔNG TIN(U'DPM)

TRÌNH ĐỘ: CAO ĐẲNG



Năm 2018

TUYÊN BỐ BẢN QUYỀN

Tài liệu này thuộc loại sách giáo trình nên các nguồn thông tin có thể được phép dùng nguyên bản hoặc trích dùng cho các mục đích về đào tạo và tham khảo.

Mọi mục đích khác mang tính lệch lạc hoặc sử dụng với mục đích kinh doanh thiếu lành mạnh sẽ bị nghiêm cấm.

TaiLieu.vn

LỜI GIỚI THIỆU

Trong thời đại bùng nổ thông tin, vai trò của hệ thống thông tin ngày càng trở nên quan trọng đối với mọi lĩnh vực hoạt động của doanh nghiệp. Thuật ngữ Cơ sở dữ liệu giờ đây được áp dụng rộng rãi trong nhiều lĩnh vực ngoài Công nghệ thông tin, như là Tài chính - Ngân hàng, Kế toán - Kiểm toán, Thẩm định dự án, ... Trên thị trường hiện nay có khá nhiều tài liệu tham khảo về Cơ sở dữ liệu, nhưng hầu hết đều nhắm đến đối tượng người học là những sinh viên khối ngành Công nghệ thông tin hoặc có những kiến thức nhất định về công nghệ thông tin.

Giáo trình cơ sở dữ liệu này được biên soạn theo chương trình đào tạo chuyên ngành Công nghệ thông tin ở bậc cao đẳng nghề của Bộ Lao Động TBXH. Giáo trình trình bày những vấn đề cốt lõi nhất của môn cơ sở dữ liệu. Các bài học được trình bày ngắn gọn, có nhiều ví dụ minh họa. Cuối mỗi chương đều có bài tập để sinh viên luyện tập.

Giáo trình này có thể giúp các sinh viên trong việc học môn cơ sở dữ liệu ở bậc cao đẳng nghề cũng như trong các kỳ thi tốt nghiệp, trong các kỳ thi liên thông lên Đại Học. Tôi mong rằng các sinh viên tự tìm hiểu trước mỗi vấn đề và kết hợp với bài giảng trên lớp của giáo viên để việc học môn này đạt hiệu quả.

Mặc dù đã rất cố gắng trong quá trình biên soạn nhưng không thể tránh khỏi những sai sót, rất mong nhận được sự góp ý của người đọc và đồng nghiệp để giáo trình ngày càng được hoàn thiện hơn.

Xin chân thành cảm ơn!

Thanh Hóa, ngày 10 tháng 5 năm 2018

Tham gia biên soạn

1. Vũ Thị Tuyết
2. Lê Thị Bằng
3. Lê Ngọc Tâm

ĐỀ MỤC	MỤC LỤC	TRANG
TUYÊN BỐ BẢN QUYỀN		1
LỜI GIỚI THIỆU		3
1. NGUYÊN NHÂN RA ĐỜI CỦA MÔ HÌNH QUAN HỆ.....		8
2. HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU (Data Base Management System).....		9
3. MÔ HÌNH QUAN HỆ.....		10
3.1. Mô hình quan hệ là gì?		11
3.2. Thuộc Tính(attribute):.....		11
3.3. Lược đồ quan hệ (relation schema)		12
3.4. Quan Hệ (relation)		13
3.5. Bộ (Tuple).....		14
3. 6. Siêu Khoá – Khoá (<i>super key- key</i>)		14
4. MÔ HÌNH THỰC THỂ KẾT HỢP.....		15
4.1. Thực Thể (entity)		15
4.2. Thuộc tính (attribute).....		15
4.3. Loại thực thể (entity type)		16
4.4. Khoá (key):		16
4.5. Mối Kết Hợp (relationship)		17
5. NGÔN NGỮ ĐẠI SỐ QUAN HỆ		19
5.1. Phép hợp 2 quan hệ(Union).....		19
5.2. Phép Giao 2 quan hệ (Intersection)		20
5.3. Phép Trừ 2 quan hệ (Minus).....		21
5.4. Tích Decac của 2 quan hệ Cartesian Product).....		21
5.5. Phép chia 2 quan hệ:		22
5.6. Phép chiếu (projection).....		23
5.7. Phép chọn (Selection).....		24
5.8. Phép θ - Kết		24
5.9. Phép kết tự nhiên (natural join)		25
BÀI TẬP:		26
CHƯƠNG 2: NGÔN NGỮ TRUY VẤN SQL.....		29
1. TẠO BẢNG		29
2. XÓA BẢNG		30
3. KHỎI LỆNH SELECT.....		31
3.1. Cú pháp tổng quát.....		31

3.2. Các hàm thư viện của SQL	31
3.3. Tìm kiếm đơn giản.....	32
3.4. Tìm kiếm với câu hỏi phức tạp	36
BÀI TẬP:	41
CHƯƠNG 3: RÀNG BUỘC TOÀN VỆN	47
3.1. Khái niệm ràng buộc toàn vẹn.....	47
3.2 Các yếu tố của ràng buộc toàn vẹn	48
3.2.1. Điều kiện	48
3.2.2. Bối cảnh.....	48
3.2.3. Bảng tầm ảnh hưởng	48
3.2.4. Hành động cần phải có khi phát hiện có RBTV bị vi phạm:	49
3.3. Phân loại ràng buộc toàn vẹn.....	50
3.3.1. Ràng buộc toàn vẹn có bối cảnh là một quan hệ.....	51
3.3.2. Ràng buộc toàn vẹn có bối cảnh là nhiều quan hệ	52
BÀI TẬP:	53
CHƯƠNG 4: PHỤ THUỘC HÀM.....	56
1. ĐỊNH NGHĨA PHỤ THUỘC HÀM.....	56
2. HỆ LUẬT DẶN ARMSTRONG	60
3. BAO ĐÓNG CỦA TẬP THUỘC TÍNH X (closures of attribute sets).....	62
BÀI TẬP:	65
CHƯƠNG 5: PHỦ CỦA TẬP PHỤ THUỘC HÀM	67
1. ĐỊNH NGHĨA	67
2. PHỦ TỐI THIỂU CỦA MỘT TẬP PHỤ THUỘC HÀM.....	67
2.1. Phụ thuộc hàm có vế trái dư thừa	67
2.2. Tập phụ thuộc hàm có vế phải một thuộc tính.	68
2.3. Tập phụ thuộc hàm không dư thừa	68
2.4. Tập phụ thuộc hàm tối thiểu	68
3. KHÓA CỦA LƯỢC ĐỒ QUAN HỆ (Key).....	69
3.1. Định Nghĩa.....	69
3.2. Thuật toán tìm tất cả khóa	71
BÀI TẬP:	73
CHƯƠNG 6: CHUẨN HÓA CƠ SỞ DỮ LIỆU.....	75
1. CÁC DẠNG CHUẨN CỦA LƯỢC ĐỒ QUAN HỆ	75
1.1. Dạng chuẩn một (First Normal Form).....	76
1.2. Dạng chuẩn 2 (second normal form)	76

1.3. Dạng chuẩn 3 (third normal form).....	78
1.4. Dạng chuẩn BCNF (Boyce Codd Normal Form)	81
2. PHÉP TÁCH KẾT NỐI BẢO TOÀN.....	83
2.1. Phép tách kết nối bảo toàn thông tin.....	83
2.2. Phép tách kết nối bảo toàn phụ thuộc hàm	91
3. THIẾT KẾ CƠ SỞ DỮ LIỆU BẰNG CÁCH PHÂN RÃ.....	95
3.1. Phân rã thành dạng chuẩn BCNF (hay chuẩn 3) bảo toàn thông tin	95
3.2. Phân rã thành dạng chuẩn 3 vừa bảo toàn thông tin vừa bảo toàn phụ thuộc hàm.....	100
TÀI LIỆU THAM KHẢO.....	109

Tailieu.vn

GIÁO TRÌNH CƠ SỞ DỮ LIỆU

Tên môn học: Cơ sở dữ liệu

Mã môn học: MH12

I. Vị trí, ý nghĩa, vai trò môn học:

- Vị trí:

Cơ sở dữ liệu là môn học cơ sở nghề bắt buộc của chương trình đào tạo Cao đẳng nghề Công nghệ thông tin (ứng dụng phần mềm). Môn học này được học sau các môn học chung, tin học đại cương.

- Ý nghĩa:

Là môn học cơ sở trong chương trình đào tạo Cao đẳng nghề Công nghệ thông tin (Ứng dụng phần mềm).

- Vai trò:

Giúp cho sinh viên có thể ứng dụng các kiến thức về cơ sở dữ liệu vào thực tiễn, tiếp tục nghiên cứu sâu về lý thuyết cơ sở dữ liệu cũng như các môn tin học khác

Mục tiêu của môn học:

- Kiến thức:

- + Hiểu về các mô hình dữ liệu và các công cụ mô tả dữ liệu;
- + Hiểu về các khái niệm, tính năng và các phương thức xử lý dữ liệu của hệ quản trị cơ sở dữ liệu SQL;

- Kỹ năng:

- + Thiết kế được một số cơ sở dữ liệu quan hệ thông dụng: quản lý nhân sự, quản lý bán hàng,...;
- + Xây dựng được các phụ thuộc hàm, cách chuẩn hóa các cơ sở dữ liệu quan hệ cụ thể.

- Năng lực tự chủ và trách nhiệm:

- + Nghiêm túc, tỉ mỉ, sáng tạo trong quá trình tiếp thu kiến thức và vận dụng vào việc xây dựng các cơ sở dữ liệu cụ thể.
- + Chủ động, tích cực tìm hiểu các tài liệu và nguồn bài tập liên quan.

- Nội dung chính:

- + **Chương 1:** Mô hình quan hệ.
- + **Chương 2:** Ngôn ngữ truy vấn SQL.
- + **Chương 3:** Phụ thuộc hàm.
- + **Chương 4:** Phủ của tập phụ thuộc hàm.
- + **Chương 5:** Chuẩn hóa cơ sở dữ liệu.

CHƯƠNG 1: MÔ HÌNH QUAN HỆ

MÃ CHƯƠNG: MH 12/01

Giới thiệu:

Trong thời đại bùng nổ thông tin, vai trò của hệ thống thông tin ngày càng trở nên quan trọng đối với mọi lĩnh vực hoạt động của doanh nghiệp. Thuật ngữ Cơ sở dữ liệu giờ đây được áp dụng rộng rãi trong nhiều lĩnh vực ngoài Công nghệ thông tin, như là Tài chính - Ngân hàng, Kế toán - Kiểm toán, Thẩm định dự án, ... Trên thị trường hiện nay có khá nhiều tài liệu tham khảo về Cơ sở dữ liệu, nhưng hầu hết đều nhắm đến đối tượng người học là những sinh viên khối ngành Công nghệ thông tin hoặc có những kiến thức nhất định về công nghệ thông tin.

Mục tiêu:

Sau khi học xong bài này người học có khả năng:

- Hiểu được các khái niệm về cơ sở dữ liệu, hệ quản trị cơ sở dữ liệu và mô hình quan hệ.
- Biết cách chuyển đổi từ lược đồ cơ sở dữ liệu sang mô hình quan hệ dữ liệu.
- Áp dụng các phép toán đại số quan hệ để biểu diễn trên lược đồ quan hệ.
- Phát huy tính sáng tạo của sinh viên.
- Rèn luyện tính kiên trì cẩn thận.

Nội dung chính:

1. NGUYÊN NHÂN RA ĐỜI CỦA MÔ HÌNH QUAN HỆ

Trong nhiều năm, công nghệ tính toán và thông tin phát triển từ những hệ thống lớn, đắt tiền, độc quyền đến các hệ thống mở mạnh và không đắt tiền. Sự phát triển này mang lại lợi ích to lớn cho người dùng cuối bởi sự phát triển của các gói ứng dụng số như xử lý văn bản, bảng tính điện tử, văn phòng xuất bản, hệ quản lý cơ sở dữ liệu, máy tính trợ giúp công nghệ phần mềm...

Trước khi máy tính hóa cơ sở dữ liệu được giới thiệu, dữ liệu được lưu trữ theo kiểu điện tử thành nhiều tập tin riêng biệt sử dụng hệ tập tin (từ đây về sau ta gọi hệ tập tin theo lối cũ). Những tập tin này được xử lý bằng các ngôn ngữ thế hệ thứ ba như COBOL, FORTRAN, PASCAL và ngay cả BASIC để tạo ra các giải pháp cho các vấn đề của doanh nghiệp. Mỗi ứng dụng, chẳng hạn như hệ tính lương, hệ kho hay hệ thống kế toán sẽ có một tập các tập tin riêng chứa dữ liệu riêng. Các ứng dụng như vậy tạo ra ba vấn đề sau:

- Có sự liên kết chặt chẽ giữa cấu trúc luận lý và cấu trúc vật lý của các tập tin và chương trình ứng dụng khai thác chúng. Điều này khiến việc tạo nên các ứng dụng này rất khó khăn, tốn nhiều thời gian và do vậy mà tốn kém trong bảo trì hệ thống.

- Có sự dư thừa dữ liệu rất lớn qua việc trùng lặp các tập tin trong các ứng dụng khác nhau. Điều này tạo ra những vấn đề như: Dữ liệu thiếu nhất quán, không gian đĩa bị lãng phí, thời gian bảo trì và lưu phòng hồ các tập tin gia tăng, vấn đề về

quản trị như không chú trọng bảo mật và tổ chức dữ liệu thiếu thống nhất. Một ví dụ điển hình về sự trùng lặp dữ liệu là: Hệ quản lý nguồn nhân lực bao gồm ba hệ chính:

- + Hệ lương, hệ này duy trì ngày công và lương cho tất cả nhân viên.
- + Hệ nhân sự, hệ này duy trì lý lịch cá nhân, dữ liệu về tổ chức, công việc đào tạo và vị trí thăng tiến.
- + Hệ hưu, hệ này quản trị các qui tắc liên quan đến nghỉ hưu, loại nghỉ hưu. Chi tiết về hưu của từng nhân viên.

Vấn đề phức tạp là Hệ lương thông thường được quản lý bởi phòng tài chính, trong khi Hệ nhân sự và Hệ hưu được quản lý bởi phòng tổ chức. Rõ ràng, có nhiều dữ liệu về nhân viên là chung cho cả ba hệ. Thường những hệ này thực hiện và giữ gìn riêng biệt và chúng tạo ra trùng dữ liệu nhân viên mà chúng dùng.

- Người sử dụng có ít khả năng khai thác trực tiếp dữ liệu.

2. HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU (Data Base Management System)

Để giải quyết tốt những vấn đề mà cách tổ chức CSDL đặt ra như đã nói ở trên, cần thiết phải có những phần mềm chuyên dùng để khai thác chúng. Những phần mềm này được gọi là các hệ quản trị CSDL. Các hệ quản trị CSDL có nhiệm vụ hỗ trợ cho các nhà phân tích thiết kế CSDL cũng như những người khai thác CSDL. Hiện nay trên thị trường phần mềm đã có những hệ quản trị CSDL hỗ trợ được nhiều tiện ích như: MS Access, Visual Foxpro, SQL Server Oracle, ...

- Mỗi hệ quản trị CSDL đều được cài đặt dựa trên một mô hình dữ liệu cụ thể. Dù là dựa trên mô hình dữ liệu nào, một hệ quản trị CSDL cũng phải hội đủ các yếu tố sau:

- Ngôn ngữ giao tiếp giữa người sử dụng và CSDL, bao gồm :

+ **Ngôn ngữ mô tả dữ liệu:** Để cho phép khai báo cấu trúc của CSDL, khai báo các mối liên hệ của dữ liệu và các quy tắc quản lý áp đặt lên các dữ liệu đó.

+ **Ngôn ngữ thao tác dữ liệu:** Cho phép người sử dụng có thể cập nhật dữ liệu (thêm/sửa/xoá)

+ **Ngôn ngữ truy vấn dữ liệu:** Cho phép người khai thác sử dụng để truy vấn các thông tin cần thiết trong CSDL

+ **Ngôn ngữ quản lý dữ liệu:** Cho phép những người quản trị hệ thống thay đổi cấu trúc của các bảng dữ liệu, khai báo bảo mật thông tin và cấp quyền hạn khai thác CSDL cho người sử dụng,...

- Từ điển dữ liệu:

+ Dùng để mô tả các ánh xạ liên kết, ghi nhận các thành phần cấu trúc của CSDL, các chương trình ứng dụng, mật mã, quyền hạn sử dụng,...

- Cơ chế giải quyết vấn đề tranh chấp dữ liệu:

+ Mỗi hệ quản trị CSDL cũng có thể cài đặt một cơ chế riêng để giải quyết các vấn đề này. Một số biện pháp sau đây thường được sử dụng: thứ nhất: cấp quyền ưu

tiên cho từng người sử dụng; thứ hai: Đánh dấu yêu cầu truy xuất dữ liệu, phân chia thời gian, người nào có yêu cầu trước thì có quyền truy xuất dữ liệu trước,...

- Hệ quản trị CSDL cũng phải có cơ chế sao lưu (backup) và phục hồi (restore) dữ liệu khi có sự cố xảy ra.

+ Điều này có thể thực hiện sau một thời gian nhất định hệ quản trị CSDL sẽ tự động tạo ra một bản sao CSDL, cách này hơi tốn kém, nhất là đối với CSDL lớn.

- Hệ quản trị CSDL phải cung cấp một giao diện thân thiện, dễ sử dụng.

- HQTCSDDL biến đổi CSDL vật lý thành CSDL logic.

Dựa vào cách tổ chức dữ liệu, HQTCSDDL được chia thành năm loại:

+ Loại phân cấp như hệ IMS của IBM

+ Loại mạng như IDMS của Cullinet Software

+ Loại tập tin đảo như ADABAS của Software AG

+ Loại quan hệ như ORACLE của Oracle, DB2 của IBM, ACCESS của Microsoft Access

+ Loại đối tượng là một tiếp cận khá mới trong thiết kế HQTCSDDL và việc sử dụng loại này sớm trở nên phổ biến

Hiện tại, loại HQTCSDDL chính được sử dụng trong công nghệ là loại HQTCSDDL quan hệ (RDBMS). Loại này đã chiếm lĩnh trong công nghệ trên 10-15 năm cuối cùng khi đánh bật loại HQTCSDDL phân cấp và gần đây là HQTCSDDL mạng.

3. MÔ HÌNH QUAN HỆ

Mô hình dữ liệu là sự trừu tượng hoá môi trường thực. Mỗi loại mô hình dữ liệu đặc trưng cho một cách tiếp cận dữ liệu khác nhau của những nhà phân tích thiết kế CSDL. Mỗi loại mô hình dữ liệu đều có những ưu điểm và những mặt hạn chế của nó, nhưng vẫn có những mô hình dữ liệu nổi trội và được nhiều người quan tâm nghiên cứu.

Sau đây chúng ta sẽ đi qua lịch sử phát triển của các mô hình dữ liệu.

Vào những năm sáu mươi, thế hệ đầu tiên của CSDL ra đời dưới dạng mô hình thực thể kết hợp, mô hình mạng và mô hình phân cấp.

Vào những năm bảy mươi, thế hệ thứ hai của CSDL ra đời. Đó là mô hình dữ liệu quan hệ do EF. Codd phát minh. Mô hình này có cấu trúc logic chặt chẽ. Đây là mô hình đã và đang được sử dụng rộng khắp trong công tác quản lý trên phạm vi toàn cầu. Việc nghiên cứu mô hình dữ liệu quan hệ nhằm vào lý thuyết chuẩn hoá các quan hệ và là một công cụ quan trọng trong việc phân tích thiết kế các hệ CSDL hiện nay. Mục đích của nghiên cứu này nhằm bỏ đi các phần tử không bình thường của quan hệ khi thực hiện các phép cập nhật, loại bỏ các phần tử dư thừa.

Sang thập kỷ tám mươi, mô hình CSDL thứ ba ra đời, đó là mô hình cơ sở dữ liệu hướng đối tượng, mô hình cơ sở dữ liệu phân tán, mô hình cơ sở dữ liệu suy diễn,...

Trong phần tiếp theo sau đây, tôi sẽ trình bày về mô hình dữ liệu tiêu biểu nhất để thiết kế (*bước đầu*) một ứng dụng tin học đó là mô hình thực thể kết hợp. Trong các chương còn lại của giáo trình này tôi sẽ trình bày về mô hình dữ liệu quan hệ.

3.1. Mô hình quan hệ là gì?

Mô hình Cơ sở dữ liệu Quan hệ (*gọi tắt là Mô hình Quan hệ*) do E.F Codd đề xuất năm 1971. Mô hình này bao gồm:

- Một hệ thống các ký hiệu để mô tả dữ liệu dưới dạng dòng và cột như quan hệ, bộ, thuộc tính, khóa chính, khóa ngoại, ...
- Một tập hợp các phép toán thao tác trên dữ liệu như phép toán tập hợp, phép toán quan hệ.
- Ràng buộc toàn vẹn quan hệ.

Các hệ HQTCSDLQH ngày nay được xây dựng dựa vào lý thuyết của mô hình quan hệ.

Mục đích của môn học này giúp cho sinh viên nắm được kiến trúc tổng quát về mô hình quan hệ và áp dụng nó để lập mô hình dữ liệu quan hệ có hiệu quả trong lưu trữ và khai thác.

Chẳng hạn với bài toán quản lý điểm thi của sinh viên; nếu lưu trữ dữ liệu theo dạng bảng với các cột MASV, HOTEN, MONHOC, TENKHOA, DIEMTHI thì các giá trị của các cột HOTEN, MONHOC, TENKHOA sẽ bị trùng lặp. Sự trùng lặp này gây nên một số vấn đề:

- Ta không thể lưu trữ một sinh viên mới khi sinh viên này chưa có điểm thi
- Khi cần sửa đổi họ tên sinh viên thì ta phải sửa tất cả các dòng có liên quan đến sinh viên này. Điều này dễ gây ra tình trạng dữ liệu thiếu nhất quán.
- Khi có nhu cầu xóa điểm thi của một sinh viên kéo theo khả năng xóa luôn họ tên sinh viên đó.

Việc lưu trữ dữ liệu như trên không đúng với mô hình quan hệ. Để lưu trữ đúng với mô hình quan hệ ta phải thay MONHOC bằng MAMH, thay TENKHOA bằng MAKHOA, tách một bảng dữ liệu lớn đó ra thành nhiều bảng con.

3.2. Thuộc Tính (attribute):

Thuộc tính là các đặc điểm riêng của một đối tượng (đối tượng được hiểu như là một loại thực thể ở mô hình thực thể kết hợp), mỗi thuộc tính có một tên gọi và phải thuộc về một kiểu dữ liệu nhất định.

a. Kiểu dữ liệu (data type)

Các thuộc tính được phân biệt qua tên gọi và phải thuộc một kiểu dữ liệu nhất định (số, chuỗi, ngày tháng, logic, hình ảnh,...). Kiểu dữ liệu ở đây có thể là kiểu vô hướng hoặc là kiểu có cấu trúc. Nếu thuộc tính có kiểu dữ liệu là vô hướng thì nó được gọi là thuộc tính đơn hay thuộc tính nguyên tố, nếu thuộc tính có kiểu dữ liệu có cấu trúc thì ta nói rằng nó không phải là thuộc tính nguyên tố

Chẳng hạn với sinh viên Nguyễn Văn Thành thì các thuộc tính họ và tên, mã số sinh viên thuộc kiểu chuỗi, thuộc tính ngày sinh thuộc kiểu ngày tháng, hộ khẩu thường trú thuộc kiểu chuỗi, thuộc tính hình ảnh thuộc kiểu hình ảnh,...

b. Miền giá trị (domain of values)

Thông thường mỗi thuộc tính chỉ chọn lấy giá trị trong một tập con của kiểu dữ liệu và tập hợp con đó gọi là miền giá trị của thuộc tính đó. Chẳng hạn thuộc tính *Nữ* có miền giá trị là {nam,nữ}, thuộc tính *màu da* có miền giá trị là {da trắng, da vàng, da đen, da đỏ}, thuộc tính *điểm thi* là các số thuộc tập {0; 1;2;...,10}.

Lưu ý rằng nếu không lưu ý đến ngữ nghĩa thì tên của các thuộc tính thường được ký hiệu bằng các chữ cái in hoa đầu tiên trong bảng chữ cái la tinh: A,B,C,D,... Những chữ cái in hoa X,Y,Z,W,... thường dùng thay cho một nhóm nhiều thuộc tính. Đôi khi còn dùng các ký hiệu chữ cái với các chỉ số $A_1, A_2, \dots, A_n$ để chỉ các thuộc tính trong trường hợp tổng quát hay muốn đề cập đến số lượng các thuộc tính. Tên thuộc tính phải được đặt một cách gợi nhớ, không nên đặt tên thuộc tính quá dài (*vì như thế sẽ làm cho việc viết các câu lệnh truy vấn trở nên vất vả hơn*), nhưng cũng không nên đặt tên thuộc tính quá ngắn (*vì nó sẽ không cho thấy ngữ nghĩa của thuộc tính*), đặc biệt không đặt trùng tên hai thuộc tính mang ngữ nghĩa khác nhau thuộc hai đối tượng khác nhau.

Trong nhiều hệ quản trị cơ sở dữ liệu, người ta thường đưa thêm vào miền giá trị của các thuộc tính một giá trị đặc biệt gọi là giá trị rỗng (*NULL*). Tùy theo ngữ cảnh mà giá trị này có thể đặc trưng cho một giá trị không thể xác định được hoặc một giá trị chưa được xác định ở vào thời điểm nhập tin nhưng có thể được xác định vào một thời điểm khác.

3.3. Lược đồ quan hệ (relation schema)

Tập tất cả các thuộc tính cần quản lý của một đối tượng cùng với các mối liên hệ giữa chúng được gọi là lược đồ quan hệ. Lược đồ quan hệ *Q* với tập thuộc tính $\{A_1, A_2, \dots, A_n\}$ được viết là $Q(A_1, A_2, \dots, A_n)$, ký hiệu $Q^+ = \{A_1, A_2, \dots, A_n\}$.

Chẳng hạn lược đồ quan hệ *Sinhvien* với các thuộc tính như đã được liệt kê trong ví dụ 1.1 được viết như sau:

Sinhvien(MASV, HOTENSV, NU, NGAYSINH, NOISINH, TINH, MALOP)

Thường thì khi thành lập một lược đồ quan hệ, người thiết kế gán cho nó một ý nghĩa nhất định, gọi là *tân từ* của lược đồ quan hệ. chẳng hạn *tân từ* của lược đồ quan hệ *Sinhvien* là: ” Mỗi sinh viên có mỗi MASV duy nhất. Mỗi MASV xác định các thuộc tính còn lại của sinh viên đó như HOTENSV, NU, NGAYSINH, NOISINH, TINH, MALOP”

Khi phát biểu *tân từ* cho một lược đồ quan hệ, người thiết kế cần phải mô tả đầy đủ ý nghĩa để người khác tránh hiểu nhầm. Dựa vào *tân từ* này, người ta xác định được tập khoá, siêu khoá của lược đồ quan hệ (*sẽ được trình bày trong những mục kế tiếp*).

Nhiều lược đồ quan hệ cùng nằm trong một hệ thống thông tin được gọi là một lược đồ cơ sở dữ liệu.

Khái niệm lược đồ quan hệ ứng với khái niệm loại thực thể ở mô hình thực thể kết hợp.

Ví dụ lược đồ cơ sở dữ liệu để quản lý điểm sinh viên có thể gồm các lược đồ quan hệ sau:

Sv (MASV, HOSV, TENS, NU, NGÀY SINH, MALOP, TỈNH, HOCBONG)

Lop (MALOP, TENLOP, SISO, MAKHOA)

Kh (MAKHOA, TENKHOA, SOCBGD)

Mh (MAMH, TENMH, SOTIET)

Kq (MASV, MAMH, DIEMTHI)

Phân giải thích các thuộc tính:

MASV	Mã sinh viên
HOTEN	Tên sinh viên
NU	Nữ
NGAYSINH	Ngày sinh
LOP	Lớp
TENLOP	Tên lớp
SISO	Sĩ số lớp
MAKHOA	Mã khoa
HOCBONG	Học bổng
TINH	Tỉnh
TENKHOA	Tên khoa
SOCBGD	Số cán bộ giảng dạy
MAMH	Mã môn học
TENMH	Tên môn học
SOTIET	Số tiết
DIEMTHI	

3.4. Quan Hệ (relation)

Sự thể hiện của lược đồ quan hệ ở một thời điểm nào đó được gọi là quan hệ, rõ ràng là trên một lược đồ quan hệ có thể xác định nhiều quan hệ. Thường ta dùng các ký hiệu như R, S, Q để chỉ các lược đồ quan hệ, còn quan hệ thường được dùng bởi các ký hiệu là r, s, q,...

Về trực quan thì quan hệ (*hay bảng quan hệ*) như là một bảng hai chiều gồm các dòng và các cột.

Một quan hệ có n thuộc tính được gọi là quan hệ n ngôi.

Để chỉ quan hệ **r** xác định trên lược đồ quan hệ **Q** ta có thể viết **r(Q)**.

3.5. Bộ (Tuple)

Bộ là tập mỗi giá trị liên quan của tất cả các thuộc tính của một lược đồ quan hệ.

- Mỗi bộ là những thông tin về một đối tượng thuộc một quan hệ, bộ cũng còn được gọi là mẫu tin.

Chẳng hạn quan hệ sau có 2 bộ

MASV	HOTEN	NU	NGAYSINH	MALOP	TINH	HB
99001	TRAN DAN THU	TRU E	15-03-1977	CĐTH2B	TIEN GIANG	120000
99002	NGUYEN DA THAO	TRUE	25-04-1986	TCTH29C	TPHCM	120000

Thường người ta dùng các chữ cái thường (như t, p, q,...) để biểu diễn các bộ. Chẳng hạn để nói bộ t thuộc quan hệ r ta viết: t_r.

3. 6. Siêu Khoá – Khoá (*super key- key*)

S là siêu khoá (*super key*) của Q nếu với r là quan hệ bất kỳ trên Q, t₁, t₂ là hai bộ bất kỳ thuộc r thì t₁.S ≠ t₂.S.

Một lược đồ quan hệ có thể có một hoặc nhiều siêu khoá.

Chẳng hạn lược đồ quan hệ Sinhvien ở trên có các siêu khoá là: {MASV, HOTENSV}, {MASV, HOTENSV, NU}, {MASV, HOTENSV, NU, TINH },...

Siêu khoá không chứa một siêu khoá nào khác được gọi là **khoá chỉ định**, trong trường hợp lược đồ quan hệ có nhiều khoá chỉ định (*hay khoá nội*), thì khoá được chọn để cài đặt gọi là **khoá chính** (*trong các phần sau khoá chính được gọi tắt là khoá*). Chẳng hạn với lược đồ quan hệ Sinhvien trên có khoá là {MASV}. Thường các thuộc tính khoá được gạch dưới theo kiểu liền nét.

Một thuộc tính được gọi là thuộc tính khoá ngoại nếu nó không là thuộc tính khoá của một lược đồ quan hệ này nhưng lại là thuộc tính khoá của một lược đồ quan hệ khác, chẳng hạn như MALOP là khoá ngoại của lược đồ quan hệ Sinhvien. Thường các thuộc tính khoá ngoại được gạch dưới theo kiểu không liền nét.

Sinhvien (MASV, HOTENSV, NU, NGAYSINH, TINH, MALOP)

Lop (MALOP, TENLOP, MAKHOA)

Ý nghĩa thực tế của khoá là dùng để nhận diện một bộ trong một quan hệ, nghĩa là, khi cần tìm một bộ t nào đó, ta chỉ cần biết giá trị của thành phần khoá của t là đủ để dò tìm và hoàn toàn xác định được nó trong quan hệ.

Trong thực tế đối với các loại thực thể tồn tại khách quan (*ví dụ: Sinh viên, Giảng viên, Nhân viên, Hàng hoá, ...*) người thiết kế cơ sở dữ liệu thường gán thêm cho các lược đồ quan hệ này một thuộc tính giả gọi là mã số để làm khoá (*ví dụ: mã số sinh viên, mã số giảng viên, mã số nhân viên, mã số hàng hoá, ...*). Trong khi đó các lược đồ quan hệ biểu diễn cho sự trừu tượng hoá thường có khoá là một tổ hợp của hai hay nhiều thuộc tính của nó.

Một số hệ quản trị cơ sở dữ liệu hiện nay có tự động kiểm tra tính duy nhất trên khoá chính. Tức là nếu thêm một bộ mới q₂ có giá trị khoá chính trùng với giá

trị khoá chính của một bộ q1 nào đó đã có trong quan hệ thì hệ thống sẽ báo lỗi và yêu cầu nhập lại một giá trị khác.

- Các thuộc tính tham gia vào một khóa được gọi là thuộc tính khóa (*prime key*), ngược lại được gọi là thuộc tính không khóa (*non prime key*).

Ví dụ 1.1: Ta hãy xem lược đồ quan hệ sau:

Xe(SODANGBO, QUICACH, HINHDANG, MAUSAC, SOSUON, SOMAY, MAXE, QUOCGIA)

Siêu khóa: (SOSUON, QUICACH),...

Khóa chỉ định: (SODANGBO, QUOCGIA), (SOSUON), (SOMAY), (MAXE)

Khóa chính: MAXE

Thuộc tính khóa: SODANGBO, QUOCGIA, SOSUON, SOMAY, MAXE

Thuộc tính không khóa: QUICACH, HINHDANG, MAUSAC

- Khóa của Sv là (MASV), Khoá của Mh là (MAMH), khoá của Kh là (MAKHOA), khóa của Kq là (MASV, MAMH) khóa của Lop là MALOP, trong Lop thuộc tính MAKHOA là khóa ngoại

Người ta cũng quy ước rằng:

- Trong một bộ của quan hệ các thuộc tính khoá không chứa giá trị rỗng.

- Không được phép sửa đổi giá trị thuộc tính khoá của một bộ q. Nếu muốn sửa đổi giá trị thuộc tính khoá của một bộ q, người sử dụng phải huỷ bỏ bộ q và sau đó thêm một bộ q' với giá trị khoá đã được sửa đổi.

4. MÔ HÌNH THỰC THỂ KẾT HỢP

Hiện nay mô hình dữ liệu quan hệ thường được dùng trong các hệ quản trị CSDL, đây là mô hình dữ liệu ở mức vật lý. Để thành lập được mô hình này, thường là phải dùng mô hình dữ liệu ở mức quan niệm để đặc tả, một trong những mô hình ở dạng đó là mô hình thực thể kết hợp (*sau đó mới dùng một số quy tắc để chuyển hệ thống từ mô hình này về mô hình dữ liệu quan hệ – các quy tắc này sẽ được nói đến trong mục 2.2*).

Sau đây là các khái niệm của mô hình thực thể kết hợp.

4.1. Thực Thể (entity)

Thực thể là một sự vật tồn tại và phân biệt được, chẳng hạn sinh viên Nguyễn Văn Thành, lớp Cao Đẳng Tin Học 2A, môn học Cơ Sở Dữ Liệu, xe máy có biển số đăng ký 52-0549,... là các ví dụ về thực thể.

4.2. Thuộc tính (attribute)

Các đặc điểm riêng của thực thể gọi là các thuộc tính.

Chẳng hạn các thuộc tính của sinh viên Nguyễn Văn Thành là: mã số sinh viên, giới tính, ngày sinh, hộ khẩu thường trú, lớp đang theo học, ...

(Trong giáo trình này, tên thuộc tính được viết bằng chữ in hoa)

4.3. Loại thực thể (entity type)

Là tập hợp các thực thể có cùng thuộc tính. Mỗi loại thực thể đều phải được đặt tên sao cho có ý nghĩa. Một loại thực thể được biểu diễn bằng một hình chữ nhật.

- Ví dụ các sinh viên có mã sinh viên là “02CĐTH019”, “02CĐTH519”, “02TCTH465”,...nhóm lại thành một loại thực thể, được đặt tên là Sinhvien chẳng hạn.

Tương tự trong ứng dụng quản lý điểm của sinh viên (sẽ được trình bày ngay sau đây) ta có các loại thực thể như Monhoc, Lop, Khoa,...

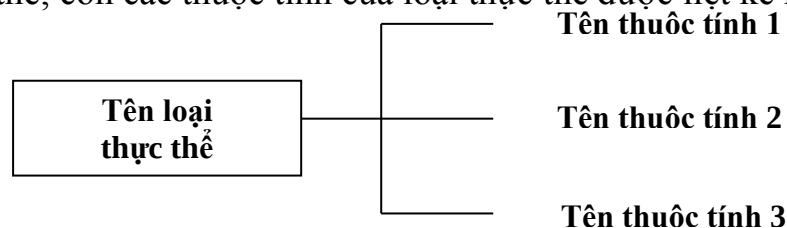
- (Trong giáo trình này, tên của loại thực thể được in hoa ký tự đầu tiên, các ký tự còn lại viết thường).

4.4. Khoá (key):

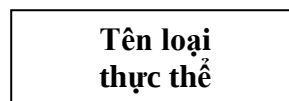
Khoá của loại thực thể E là một hay một tập các thuộc tính của E có thể dùng để phân biệt hai thực thể bất kỳ của E.

Ví dụ khoá của loại thực thể Sinhvien là MASV, của Lóp là MALOP, của Khoa là MAKHOA, của Monhoc là MAMH,...

Cần chú ý rằng khi biểu diễn một hệ thống bằng mô hình thực thể kết hợp thì tên của các loại thực thể phải khác nhau. Trong danh sách các thuộc tính của một loại thực thể thì tập thuộc tính khoá thường được gạch dưới liền nét. Nếu một hệ thống có nhiều loại thực thể, để đơn giản hoá mô hình, người ta có thể chỉ nêu tên các loại thực thể; còn các thuộc tính của loại thực thể được liệt kê riêng.



Hoặc ngắn gọn như sau



Hình 1-1. Mô tả tên loại thực thể

Ví dụ 1.2: Bài toán quản lý điểm của sinh viên được phát biểu sơ bộ như sau:

Mỗi sinh viên cần quản lý các thông tin như: họ và tên (HOTENSV), ngày tháng năm sinh(NGAYSINH), giới tính (NU), nơi sinh(NƠISINH), hộ khẩu thường trú (TINH). Mỗi sinh viên được cấp một mã số sinh viên duy nhất (MASV) để phân biệt với mọi sinh viên khác của trường, mỗi sinh viên chỉ thuộc về một lớp nào đó.

Mỗi lớp học có một mã số lớp (MALOP) duy nhất để phân biệt với tất cả các lớp học khác trong trường: có một tên gọi (TENLOP) của lớp, mỗi lớp chỉ thuộc về một khoa.

Mỗi khoa có một tên gọi (TENKHOA) và một mã số duy nhất (MAKHOA) để phân biệt với các khoa khác.

Mỗi môn học có một tên gọi (TENMH) cụ thể, được học trong một số đơn vị học trình (DONVIHT) và ứng với môn học là một mã số duy nhất (MAMH) để phân biệt với các môn học khác.

Mỗi giảng viên cần quản lý các thông tin: họ và tên (HOTENGV), cấp học vị (HOCVI), thuộc một chuyên ngành (CHUYENNGANH) và được gán cho một mã số duy nhất gọi là mã giảng viên (MAGV) để phân biệt với các giảng viên khác. Mỗi giảng viên có thể dạy nhiều môn ở nhiều khoa, nhưng chỉ thuộc về sự quản lý hành chính của một khoa.

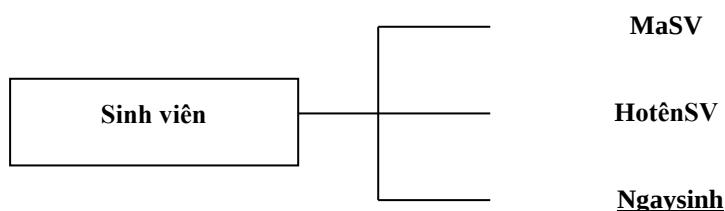
Mỗi sinh viên với một môn học được phép thi tối đa 3 lần, mỗi lần thi (LANTHI), điểm thi (DIEMTHI).

Mỗi môn học ở mỗi lớp học chỉ phân công cho một giảng viên dạy (tất nhiên là một giảng viên thì có thể dạy nhiều môn ở một lớp).

Với bài toán trên thì các loại thực thể cần quản lý như: Sinh viên, Môn học, Khoa, Lớp, Giảng viên.

Ví dụ với loại thực thể Sinh viên thì cần quản lý các thuộc tính như:

MASV, HOTENSV, NGAYSINH, ... và ta có thể biểu diễn như sau:



HÌNH 1-2. Loại thực thể sinh viên

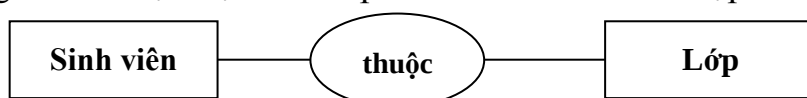
4.5. Môi Kết Hợp (relationship)

Môi kết hợp diễn tả sự liên hệ giữa các loại thực thể trong một ứng dụng tin học.

Ví dụ 1.3: Môi kết hợp giữa hai loại thực thể Sinh viên và Lop, môi kết hợp giữa Sinh viên với Môn học, ...

Môi kết hợp được biểu diễn bằng một hình elip và hai bên là hai nhánh gắn kết với các loại thực thể (hoặc môi kết hợp) liên quan, tên môi kết hợp thường là: thuộc, gồm, chứa, ...

- Chẳng hạn giữa hai loại thực thể Lớp và Khoa có môi kết hợp “thuộc” như sau:

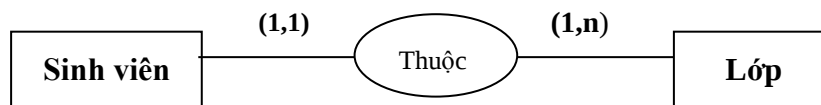


HÌNH 1-3. Mối quan hệ của loại thực thể sinh viên và Lớp

* Bản số của mỗi kết hợp:

- Bản số của một nhánh R trong mỗi kết hợp thể hiện số lượng các thực thể thuộc thực thể ở nhánh “bên kia” có liên hệ với một thực thể của nhánh R. Mỗi bản số là một cặp số (min,max), chỉ số lượng tối thiểu và số lượng tối đa của thực thể khi tham gia vào mỗi kết hợp đó.

Ví dụ 1.4:

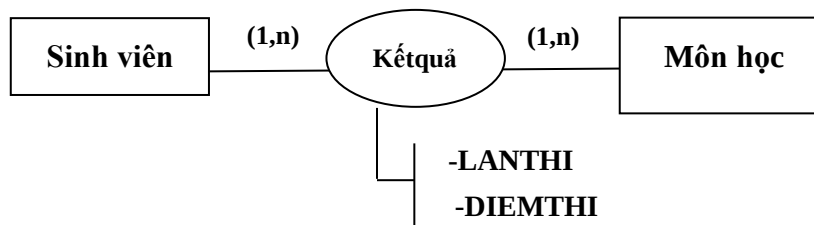


Hình 1- 4. Bản số của loại thực thể sinh viên và Lớp

Có nghĩa là: “mỗi sinh viên thuộc một và chỉ một lớp nên bản số bên nhánh Sinh viên là (1,1), mỗi lớp có 1 đến n sinh viên nên bản số bên nhánh Lớp là (1,n)”

Trong một số trường hợp đặc biệt, mỗi kết hợp có thể có các thuộc tính đi kèm và do đó chúng thường được đặt tên ý với nghĩa đầy đủ hơn.

- Ví dụ giữa hai loại thực thể Monhoc và Sinhvien có mỗi kết hợp ketqua với ý nghĩa: “mỗi sinh viên ứng với mỗi lần thi của mỗi môn học có một kết quả điểm thi duy nhất”.



Khoá của mỗi kết hợp: là hợp của các khoá của các loại thực thể liên quan. Chẳng hạn như thuộc tính MAGV là khoá của loại thực thể Giangvien, MALOP là thuộc tính khoá của loại thực thể Lop, MAMH là thuộc tính khoá của loại thực thể Monhoc, do đó mỗi kết hợp phancong (giữa các loại thực thể Giangvien, Lop, Monhoc) có khoá là {MAGV, MAMH, MALOP}

- phancong là mỗi kết hợp 3 ngôi. (Trong giáo trình này, tên của mỗi kết hợp được viết toàn bằng chữ thường).

- Việc thành lập mô hình thực thể kết hợp cho một ứng dụng tin học có thể tiến hành theo các bước sau:

B1. Xác định danh sách các loại thực thể

B2. Xác định các mối kết hợp giữa các loại thực thể để phác thảo mô hình.

B3. Lập bản số của các mối kết hợp.

5. NGÔN NGỮ ĐẠI SỐ QUAN HỆ

5.1. Phép hợp 2 quan hệ (Union)

Cho hai lược đồ quan hệ Q_1 và Q_2 có cùng tập thuộc tính $\{A_1, A_2, \dots, A_n\}$. r_1 và r_2 lần lượt là hai quan hệ trên Q_1 và Q_2 . Phép hợp của hai lược đồ quan hệ Q_1 và Q_2 sẽ tạo thành một lược đồ quan hệ Q_3 . Q_3 được xác định như sau:

$$Q_3^+ = \{A_1, A_2, \dots, A_n\}$$

$$r_3 = r_1 + r_2 = \{t / t \in r_1 \text{ hoặc } t \in r_2\}$$

Ví dụ 1.5: Cho hai quan hệ r_1 và r_2 như sau:

r_1

A	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

r_2

A	B	C	D
x ₁	y ₁	z ₁	v ₁
a ₂	b ₂	c ₂	d ₂
x ₃	y ₃	z ₃	v ₃
A	B	C	D

Khi đó nội dung của quan hệ $r_3 = r_1 + r_2$ là:

A	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
x ₁	y ₁	z ₁	v ₁
x ₃	y ₃	z ₃	v ₃

Do thứ tự trước/sau của các bộ trong các quan hệ là không quan trọng nên ta có:

$$\forall r_1, r_2 \text{ thì } r_1 + r_2 = r_2 + r_1$$

$$\forall r \text{ thì } r + r = r$$

Một cách tổng quát có thể lấy hợp của n quan hệ tương thích: cho n quan hệ tương thích $r_1, r_2, \dots, r_n$

Hợp của n quan hệ $r_1, r_2, \dots, r_n$ là một quan hệ $r_1 + r_2 + \dots + r_n$ gồm các phần tử thuộc r_1 hoặc thuộc $r_2 \dots$ hoặc thuộc r_n

Ví dụ 1.6: Cho hai quan hệ r_1 và r_2 như sau:

r_1

MASV	MAMH	DIEMTHI
99001	CSDL	5.0

99002	CTDL	2.0
99003	MANG	8.0

r_2

MASV	MAMH	DIEMTHI
99002	CTDL	2.0
99001	TTNT	5.0
99003	CSDL	6.0

$r_3 = r_1 + r_2$

MASV	MAMH	DIEMTHI
99001	CSDL	5.0
99002	CTDL	2.0
99003	MANG	8.0
99001	TTNT	5.0
99003	CSDL	6.0

5.2. Phép Giao 2 quan hệ (Intersection)

Cho hai lược đồ quan hệ Q_1 và Q_2 có cùng tập thuộc tính $\{A_1, A_2, \dots, A_n\}$. r_1 và r_2 lần lượt là hai quan hệ trên Q_1 và Q_2 . Phép giao của hai lược đồ quan hệ Q_1 và Q_2 sẽ tạo thành một lược đồ quan hệ Q_3 như sau:

$$Q_3^+ = \{A_1, A_2, \dots, A_n\}$$

$$r_3 = r_1 * r_2 = \{t / t \in r_1 \text{ hoặc } t \in r_2\}$$

Chẳng hạn với ví dụ 1.5 và ví dụ 1.6 ở trên thì $r_1 * r_2$ là:

A	B	C	D
a_2	b_2	c_2	d_2

r_1

MASV	MAMH	DIEMTHI
99001	CSDL	5.0
99002	CTDL	2.0
99003	MANG	8.0

r_2

MASV	MAMH	DIEMTHI
99002	CTDL	2.0
99001	TTNT	5.0
99003	CSDL	6.0

$r_3 = r_1 * r_2$

MASV	MAMH	DIEMTHI
99002	CTDL	2.0