

Chương 6 MÔ HÌNH HOÁ DỮ LIỆU

Mô hình quan niệm dữ liệu là một sự trình bày về dữ liệu được tổ chức. Mục đích của mô hình quan niệm dữ liệu là chỉ ra những qui tắc về ngữ nghĩa và mối quan hệ giữa các dữ liệu có thể được.

Chúng ta có thể làm công việc mô hình hoá dữ liệu song song với các công việc phân tích và cấu trúc hoá yêu cầu trong suốt quá trình phân tích hệ thống. Chúng ta thu thập các mô tả về nghiệp vụ cần thiết cho mô hình hoá dữ liệu từ các phương pháp thu thập thông tin như: phỏng vấn, bảng câu hỏi, quan sát,... Trong một môi trường phát triển hệ thống lớn, một nhóm có thể tập trung trong việc mô hình hoá dữ liệu trong khi nhóm khác tập trung làm những công việc khác như: mô hình hoá xử lý, luận lý,... Thông thường, chúng ta sẽ phải mô hình lại một mô hình quan niệm dữ liệu cho hệ thống hiện hành đang hoạt động và xây dựng một mô hình quan niệm phù hợp với phạm vi, yêu cầu cho hệ thống. Phần lớn trường hợp việc mô hình hoá mô hình dữ liệu quan niệm cho hệ thống hiện hành là nhằm để xây dựng mô hình quan niệm dữ liệu cho hệ thống mới bằng việc xem xét lại phạm vi, yêu cầu của hệ thống mới. Tuy nhiên, trong trường hợp một hệ thống là xây dựng mới hoàn toàn thì việc mô hình hoá dữ liệu sẽ phải bắt đầu mới hoàn toàn. Một số phương pháp khả sát hệ thống phải bao gồm luôn các câu hỏi để thu thập dữ liệu, sau đây là một bảng câu hỏi hướng dẫn:

1. Các chủ đề/đối tượng của nghiệp vụ là gì? Các loại người, vị trí, sự vật, nguyên liệu,... nào được sử dụng hoặc tương tác với nghiệp vụ sẽ được lưu trữ? Có bao nhiêu thể hiện cho mỗi đối tượng tồn tại? - Dữ liệu thực thể và mô tả
2. Đặc điểm phân biệt duy nhất cho một đối tượng từ những đối tượng khác? Đặc trưng này có thay đổi không hay cố định? - Khoá, định danh dữ liệu
3. Các đặc trưng khác của đối tượng là gì? để các đối tượng được tham chiếu, chọn, sắp xếp, phân loại,... - Thuộc tính và khác khoá ứng viên, hoặc định danh khác
4. Dữ liệu được sử dụng như thế nào? Ai được phép sử dụng dữ liệu này? Ai chịu trách nhiệm thiết lập giá trị hợp pháp cho dữ liệu này? - Các vấn đề về an toàn dữ liệu, quyền truy cập dữ liệu, và đối tượng chịu trách nhiệm thiết lập về dữ liệu.
5. Khoảng thời gian nào chúng ta sử dụng dữ liệu này? Dữ liệu này cần lưu vết quá trình của nó hay không? Nếu một đặc trưng của đối tượng thay đổi chúng ta có cần phải biết các giá trị cũ của nó hay không? - Bản số và chiều thời gian của dữ liệu.
6. Có phải tất cả thể hiện của mỗi đối tượng là tương tự nhau không? Đó là, có những loại đối tượng đặc biệt nào mà đơn vị cần mô tả và giải quyết khác nhau không? - Tổng quát hoá, chuyên biệt hoá
7. Các biến cố nào xảy ra ngụ ý các mối kết hợp giữa các đối tượng? – Các liên kết dữ liệu, bản số của liên kết
8. Xác lập về các luật toàn vẹn dữ liệu, ràng buộc về bản số tối thiểu, tối đa.

Sau đây chúng ta tìm hiểu về một dùng để biểu diễn mô hình dữ liệu ở mức quan niệm:

I. Mô hình thực thể - kết hợp (Entity Relationship Model)

Giáo trình này sẽ tập trung vào mô hình thực thể kết hợp (TT-KH) – mô hình được sử dụng nhiều nhất trong việc thiết kế quan niệm dữ liệu. Mô hình TT-KH được giới thiệu bởi CHEN vào năm 1976 và đã trở nên phổ biến ngày càng nhiều hơn. Đến năm 1988, ANSI chọn mô hình TT-KH là mô hình chuẩn cho Hệ thống tự diễn tài nguyên thông tin (IRDSS: Information Resource Dictionary System).

Đầu tiên, mô hình TT-KH chỉ bao gồm các khái niệm thực thể, kết hợp và thuộc tính. Về sau, một số khái niệm khác như là thuộc tính kết hợp, cấu trúc cây tổng quát hóa được bổ sung vào mở rộng hơn mô hình TT-KH. Trong phần tiếp theo chúng ta sẽ xem xét các khái niệm cơ bản, sau đó sẽ xét đến các khái niệm mở rộng.

Mô hình thực thể kết hợp là một sự trình bày chi tiết, luận lý về dữ liệu cho một đơn vị tổ chức hoặc một phạm vi nghiệp vụ xác định. Mô hình TT-KH được diễn đạt bằng các thuật ngữ thực thể trong môi trường nghiệp vụ, các mối kết hợp hoặc các liên kết giữa các thực thể này, và các thuộc tính hoặc các đặc trưng của tất cả thực thể và mối kết hợp. Một mô hình TT-KH thường được thể hiện dưới dạng sơ đồ và được gọi là sơ đồ TT-KH (ERD – Entity-Relationship Diagram). Cho đến bây giờ đã tồn tại nhiều khái niệm dùng để biểu diễn các thành phần của mô hình TT-KH, trong giáo trình này chúng ta sẽ tiếp cận một tập các khái niệm cơ bản, nếu chúng ta dùng một ký hiệu khác thì chúng ta cũng dễ dàng chuyển đổi giữa chúng.

I.1 Các yếu tố cơ bản của mô hình thực thể kết hợp

Mô hình TT-KH giới thiệu các khái niệm cơ bản là thực thể, mối kết hợp và thuộc tính. Nên lưu ý rằng chúng ta dùng thuật ngữ *thực thể* và *mối kết hợp* để biểu thị một lớp các đối tượng, Trong một số tài liệu khác, các tác giả có thể dùng thuật ngữ tương đương là *loại thực thể* và *loại mối kết hợp*.

I.1.1. Thực thể

Thực thể biểu diễn lớp các đối tượng của thế giới thực. Các đối tượng này có thể phân thành 2 loại:

- Đối tượng hữu hình: các đối tượng có thể quan sát một cách trực quan. Ví dụ: PHÒNG, TOÀ NHÀ, NHÂN VIÊN, SINH VIÊN,...
- Các đối tượng vô hình: là các đối tượng trừu tượng không cảm nhận được một cách trực quan. Ví dụ: DỰ ÁN, LỚP HỌC, PHÒNG BAN,...

Ký hiệu

Hình chữ nhật có tên thực thể nằm bên trong

Tên thực thể

Tên

Tên thực thể = Danh từ (cụm danh từ)

Ví dụ



1.1.2. *Mối kết hợp*

Mối kết hợp biểu diễn sự kết hợp giữa hai hay nhiều thực thể.

Ký hiệu

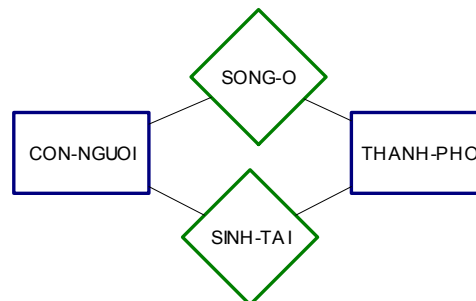


Tên

Tên MKH = Động từ (cụm danh từ)

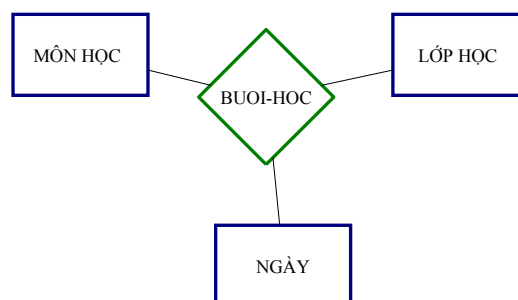
Ví dụ

Một ví dụ về mối kết hợp nhị phân SINH-TẠI là quan hệ giữa CON-NGƯỜI và THÀNH-PHỐ sinh ra. Một mối quan hệ nhị phân khác giữa hai thực thể trên là SỐNG-Ở cho biết thành phố mà con người đang sống. Mối kết hợp được biểu diễn đồ họa bởi hình thoi như trong hình



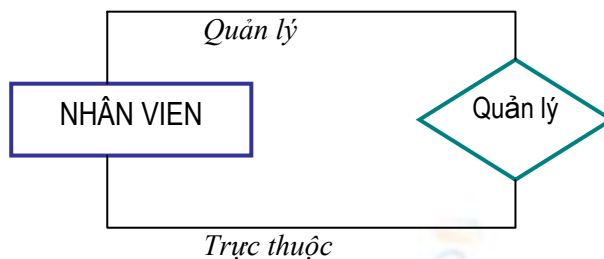
Một phần của lược đồ TT-KH biểu diễn các thực thể CON-NGƯỜI và THÀNH-PHỐ và các mối quan hệ SINH-TẠI và SỐNG-Ở.

Mối kết hợp đa phân nối kết nhiều hơn hai thực thể; ví dụ như quan hệ BUỔI-HỌC trong ví dụ hình



Hình 47. Mối quan hệ đa phân

Mối kết hợp phản thân là mối kết hợp giữa một thực thể với chính nó. Chúng còn được gọi là **mối kết hợp đệ quy**. Ví dụ như mối kết hợp QUẢN-LÝ trong hình 3.17 nối kết cán bộ quản lý và nhân viên trực thuộc, cả hai được biểu diễn bởi thực thể NHÂN-VIÊN. Để phân biệt hai *vai trò* của cùng một thực thể trong mối kết hợp, chúng ta kết thêm hai nhãn vào thực thể. Lược đồ dưới có hai vai trò QUẢN-LÝ và TRỰC THUỘC.



Hình 48. Mối kết hợp phản thân

Vai trò

Biểu diễn ngữ nghĩa của một thực thể tham gia vào mối kết hợp. Mỗi vai trò có một tên và một bản số

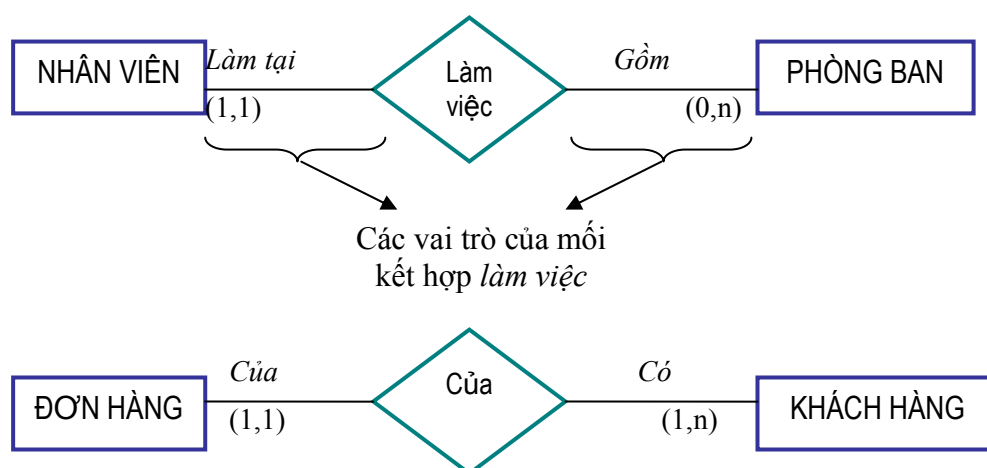
- Tên vai trò

Tên vai trò = động từ

- Bản số: bản số của một vai trò qui định ràng buộc về số lượng đối tượng của một thực thể có thể tham gia vào mối kết hợp. Một bản số được biểu diễn bởi một cặp giá trị là (mincard, maxcard) trong đó.
 - o Mincard: qui định giá trị tối thiểu của một đối tượng trong thực thể tham gia vào mối kết hợp. Giá trị của mincard từ $0 \rightarrow a$ (a là hàng số)
 - o Maxcard: qui định giá trị tối đa cho phép một đối tượng của thực thể có thể tham gia vào mối kết hợp. Một đối tượng tồn tại trong thực thể không thể có số lần tham gia vào mối kết hợp vượt con số này. Giá trị của maxcard từ $1 \rightarrow n$ (n có nghĩa là nhiều, không hạn chế)

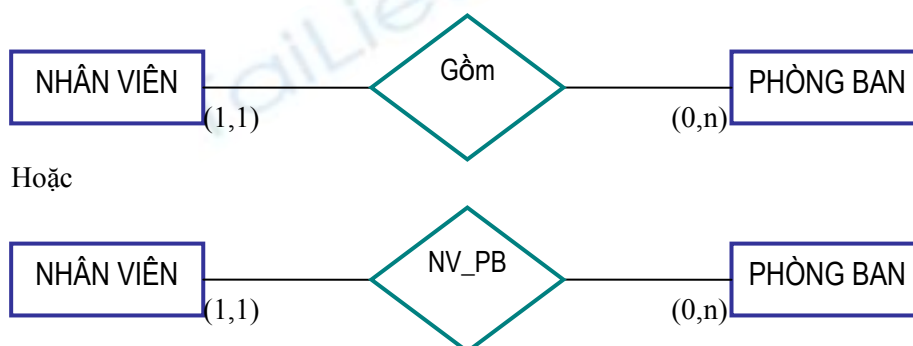
Các mẫu bản số cơ bản là : (0,1), (1,1), (2,5), (0,n), (1,n), (2,n)

Ví dụ:



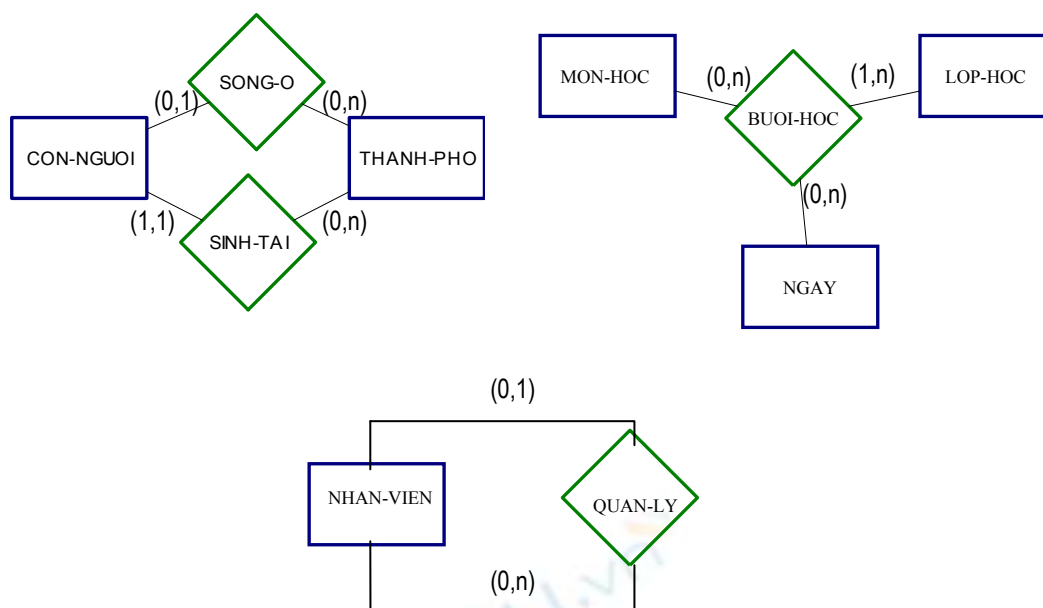
Hình 49. Mô tả vai trò của mỗi kết hợp

Thông thường tên của vai trò được che dấu và không được hiển thị trong sơ đồ, thay vào đó tên của mỗi kết hợp có thể sẽ lấy tên của một vai trò hoặc một tên ghép các tên thực thể

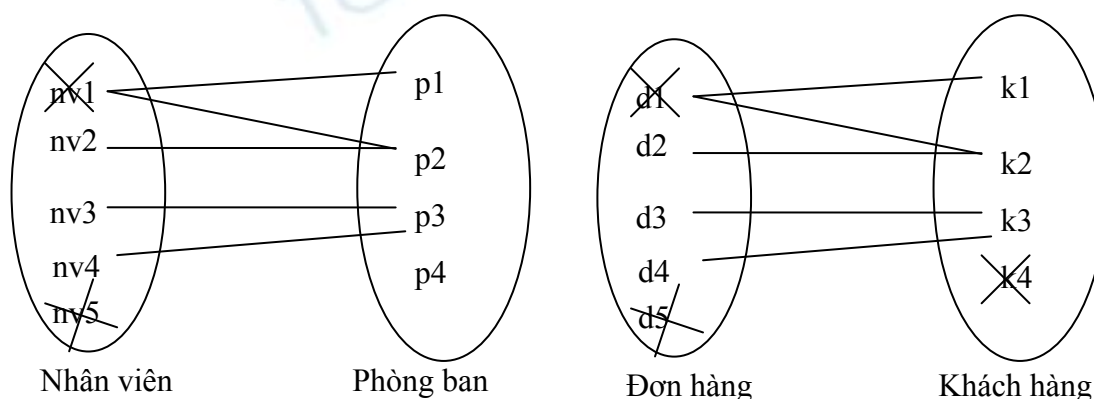


Mỗi kết hợp làm việc giữa hai thực thể Nhân viên và Phòng ban là mỗi kết hợp nhị phân bao gồm hai vai trò. Vai trò Làm tại mô tả nhân viên được phân công làm việc trong một phòng ban và bản số (1,1) của vai trò cho biết một thể hiện nhân viên của thực thể phải tham gia kết hợp với ít nhất một phòng ban và tối đa một phòng ban (duy nhất một phòng ban). Trong khi đó vai trò Gồm mô tả phòng ban gồm có những nhân viên trực thuộc và bản số (0,n) cho biết một thể hiện phòng ban có thể không hoặc chưa tham gia vào mỗi kết hợp Làm tại và có thể tham gia nhiều lần và mỗi kết hợp không hạn chế.

Trong các ví dụ trên chúng ta có bản số mô tả cho lần lượt các mối kết hợp SONG O, SINH TAI, BUOI HOC, QUAN LY như sau:



Một cách cụ thể hơn, chúng ta xem xét qua việc biểu diễn thể hiện sau:



Thể hiện nv1 không hợp lệ do vi phạm về bản số tối đa (tham gia kết hợp 2 lần); nv5 vi phạm bản số tối thiểu (không tham gia mỗi kết hợp). Tương tự, d1 của đơn hàng không hợp lệ do vi phạm bản số tối đa; d5, k4 vi phạm bản số tối thiểu.

Tổng quát, cho 2 thực thể E1, E2 và mỗi kết hợp R giữa chúng. Tùy theo giá trị bản số tối thiểu chúng ta có những trường hợp sau:

- Nếu mincard (E1,R) = 0 thì chúng ta nói rằng class E1 có sự tham gia tùy ý trong mỗi kết hợp bởi vì một đối tượng của class E1 có thể không tham gia kết hợp với đối tượng lớp E2 trong mỗi kết hợp R.
- Nếu mincard (E1,R) >0 thì chúng ta nói rằng class E1 có sự tham gia bắt buộc vào mỗi kết hợp bởi vì một đối tượng của class E1 phải bắt buộc tham gia kết hợp với ít nhất một phần tử của class E2 trong mỗi kết hợp R.

Tùy theo giá trị của bản số tối đa mà chúng ta có các trường hợp sau:

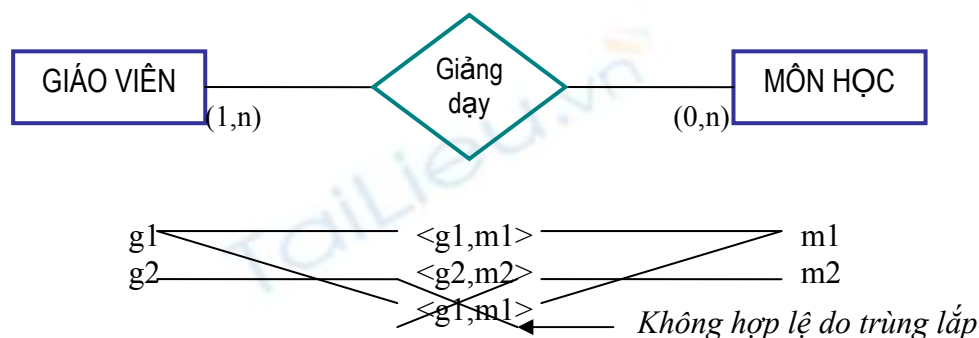
- Nếu maxcard(E1,R) = 1 và maxcard(E2,R) = 1 thì ta gọi là mỗi kết hợp một - một (one- to - one)

- Nếu $\text{maxcard}(E1,R) = 1$ và $\text{maxcard}(E2,R) = n$ thì ta gọi là mối kết hợp một - nhiều (one- to - many)
- Nếu $\text{maxcard}(E1,R) = n$ và $\text{maxcard}(E2,R) = 1$ thì ta gọi là mối kết hợp nhiều - một (many- to - one)
- Nếu $\text{maxcard}(E1,R) = n$ và $\text{maxcard}(E2,R) = n$ thì ta gọi là mối kết hợp nhiều - nhiều (many- to - many)

Thể hiện

Một thể hiện của mối kết hợp là một tổ hợp không trùng lặp của các thực thể tham gia vào mối kết hợp

Ví dụ:

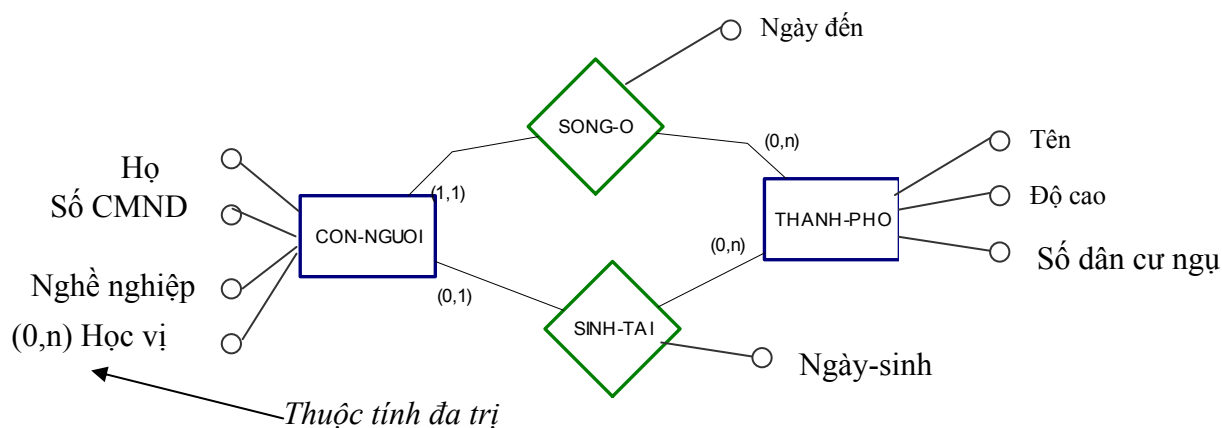
**1.1.3. Thuộc tính**

Thuộc tính: biểu diễn các đặc trưng cơ bản của thực thể hay mối kết hợp. Tất cả thông tin mở rộng được biểu diễn bởi thuộc tính.

Ký hiệu

—○ Tên thuộc tính

Hình mô tả thêm các thuộc tính cho ví dụ



Hình 50. Một lược đồ TT-KH với các thực thể, mối kết hợp và thuộc tính

Giống như mỗi kết hợp, thuộc tính cũng được đặc trưng bởi bản số tối thiểu và tối đa. Bản số tối thiểu (min-card) chỉ ra con số tối đa của giá trị thuộc tính liên kết với mỗi thể hiện của thực thể hay mỗi kết hợp.

Cho A là một thuộc tính của thực thể E:

- Nếu min-card (A,E) = 0 thì thuộc tính đó là tùy ý có thể có giá trị rỗng (Null) trong một số trường hợp.
- Nếu min-card (A,E) = 1 thì thuộc tính là bắt buộc và phải có ít nhất một giá trị thuộc tính được xác định cho tất cả các thể hiện thuộc tính.

Định nghĩa cho bản số của mỗi kết hợp cũng tương tự như bản số của thực thể.

Trong ví dụ của chúng ta, HỌ TÊN, SỐ CMND và NGHỀ NGHIỆP là thuộc tính bắt buộc nên chúng ta không thể chấp nhận thêm mới một con người vào CSDL nếu không đủ HỌ TÊN, SỐ CMND và NGHỀ NGHIỆP. HỌC-VỊ là tùy ý và chúng ta chấp nhận thêm mới một con người mà HỌC-VỊ là không xác định.

Bản số tối đa (max-card) chỉ ra con số tối đa các giá trị thuộc tính liên kết với mỗi thể hiện thực thể hay mỗi kết hợp.

Cho A là một thuộc tính của thực thể E:

- Nếu max-card (A,E) = 1 thì thuộc tính đó là thuộc tính đơn giá trị (single-valued).
- Nếu max-card (A,E) > 1 thì thuộc tính là đó là thuộc tính đa giá trị (multi-valued).

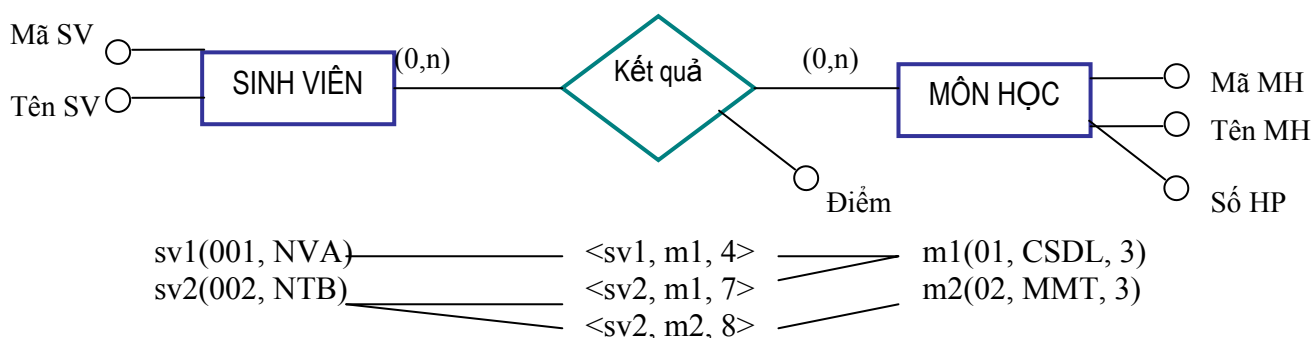
Định nghĩa cho bản số tối đa của mỗi kết hợp cũng tương tự như bản số của thực thể.

Trong ví dụ của chúng ta đang xét, HỌ TÊN, SỐ CMND và NGHỀ NGHIỆP là đơn giá trị. HỌC-VỊ là đa giá trị bởi vì mỗi người có thể có nhiều ngạch bậc ví dụ như tốt nghiệp phổ thông, đại học, cao học, tiến sĩ của các ngành khác nhau.

Chúng ta có thể tổng hợp bản số tối đa và tối thiểu thành một cặp (min-card, max-card) sẽ được trình bày trong lược đồ, bên cạnh tên thuộc tính. Giá trị thông thường là (1,1) được xem là giá trị ngầm định, do đó không cần thể hiện lên trên lược đồ.

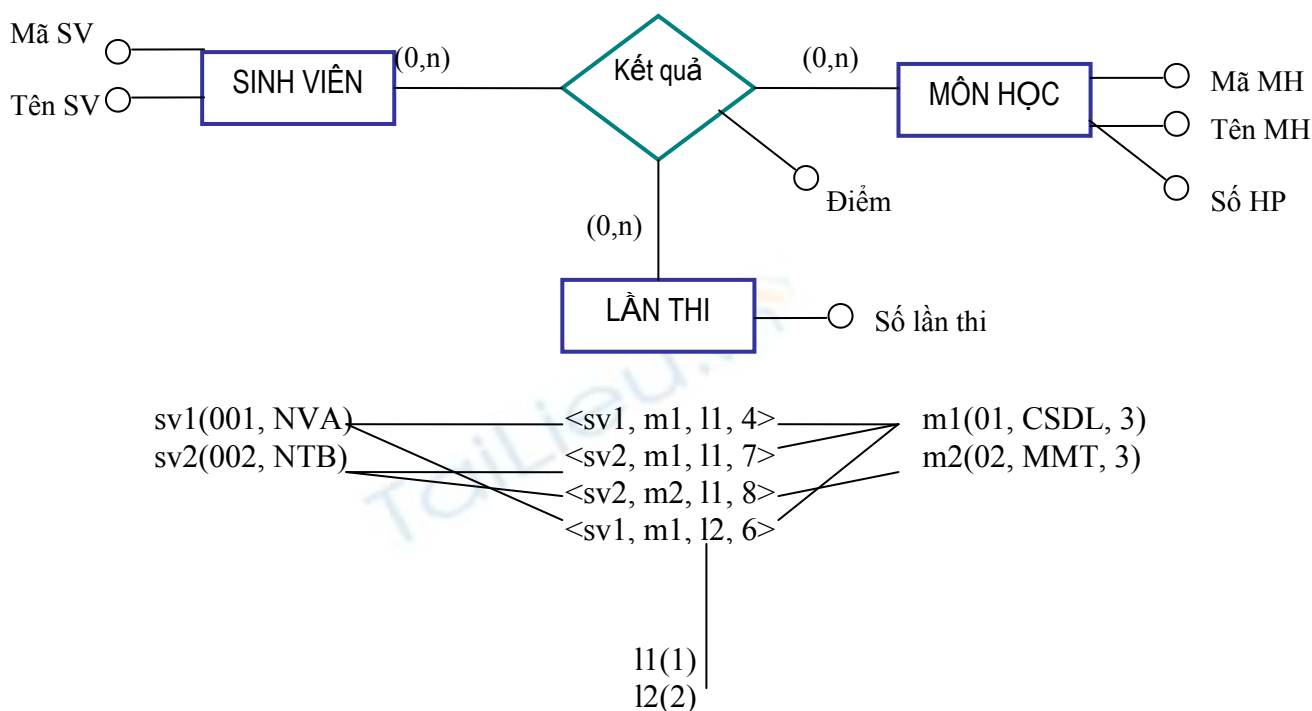
Mỗi thuộc tính sẽ liên quan đến một tập hợp các giá trị hợp lệ gọi là miền giá trị (domain) xác định. Khai báo miền giá trị tương ứng với khai báo kiểu trong các ngôn ngữ lập trình. Một thuộc tính đơn sẽ được định nghĩa trên một miền giá trị.

Một số ví dụ khác



⁸ Chú ý là do con người chúng ta quan tâm là người lớn: nhân viên hay sinh viên nên bắt buộc phải có số chứng minh nhân dân.

Mô hình trên cho phép chúng ta mô tả một sinh viên sẽ có một kết quả ghi nhận bằng điểm trên một môn học, và ứng với một môn học một sinh viên chỉ được phép ghi nhận một điểm số duy nhất. Nếu chúng ta muốn biểu diễn một sinh viên sẽ được ghi nhận trên nhiều điểm khác nhau trên một môn học ứng với nhiều lần thi khác nhau thì chúng ta phải chuẩn hoá các lần thi như là một tập hợp các đối tượng thuộc hệ thống quản lý và biểu diễn nó qua một thực thể Lần thi:



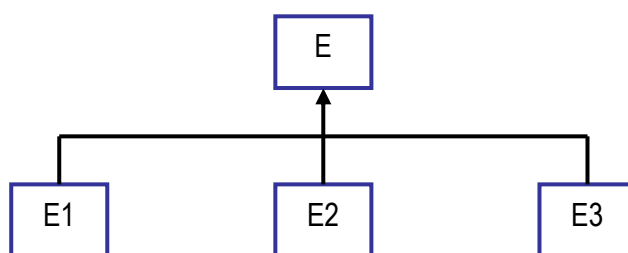
1.2 Mô hình thực thể kết hợp mở rộng

Các khái niệm mới được bổ sung vào mô hình thực thể kết hợp là cấu trúc cây tổng quát hóa, tập hợp con, thuộc tính kết hợp và định danh.

1.2.1. Cấu trúc cây phân cấp

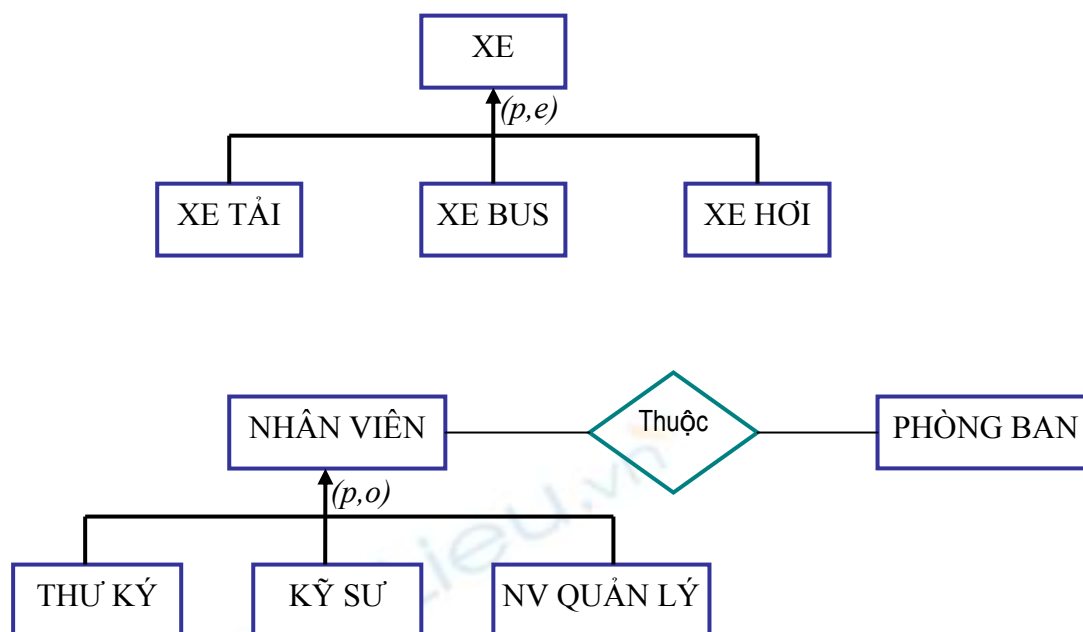
Trong mô hình TT-KH mở rộng, chúng ta có thể thiết lập cấu trúc cây phân cấp giữa các thực thể.

Một thực thể E là một tổng quát hóa của một nhóm các thực thể E1, E2,..., En nếu mỗi đối tượng của lớp E1, E2,..., En cũng là đối tượng của lớp E. Biểu diễn sơ đồ của tổng quát hóa được trình bày trong hình tiếp theo. Mũi tên của cung sẽ chỉ vào thực thể tổng quát hóa.

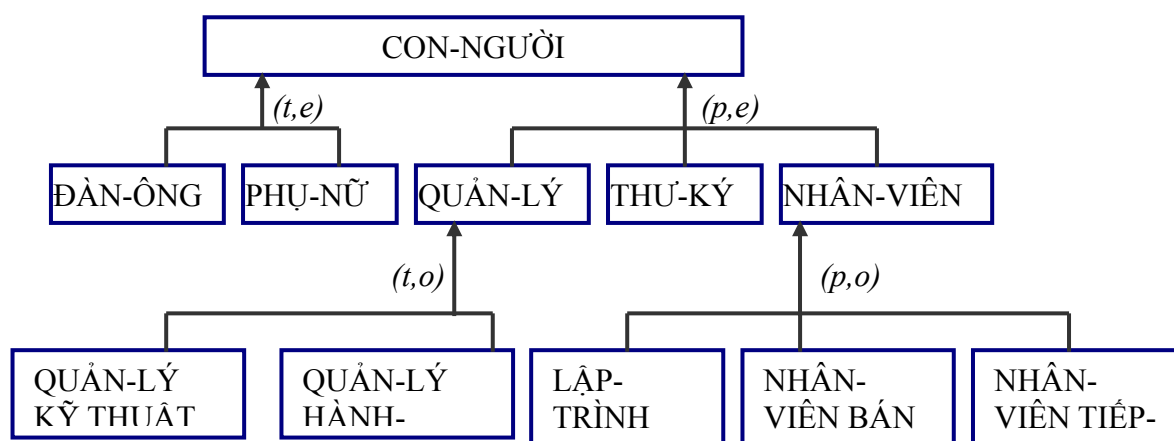


Hình 51. Biểu diễn tổng quát hóa trong mô hình TT-KH mở rộng

Ví dụ:



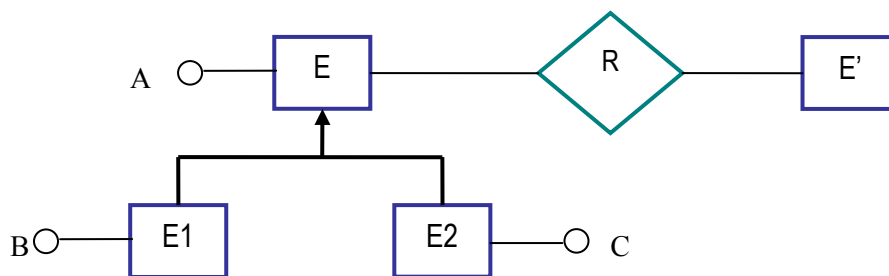
Mỗi thực thể có thể được bao gồm trong nhiều tổng quát hóa, có thể đồng thời đóng vai trò thực thể tổng quát của một quan hệ tổng quát hóa và là thực thể tập con của một quan hệ tổng quát hóa khác. Hình dưới giới thiệu một cấu trúc cây tổng quát hóa phức tạp cho thực thể CON-NGƯỜI. Đối lập với tổng quát hóa là chuyên biệt hóa.



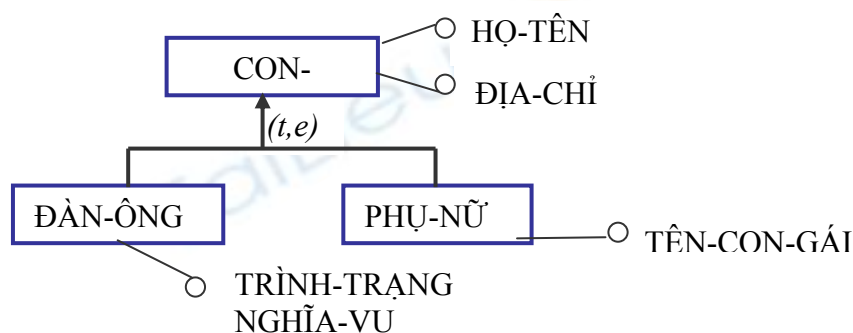
Hình 52. Cấu trúc cây tổng quát hóa của thực thể CON-NGƯỜI

Ý nghĩa của cấu trúc phân cấp là dùng để mô hình hoá việc phân loại một tập hợp các đối tượng thành những tập con chuyên biệt hoặc tổng hợp các thực thể chuyên biệt thành một thực thể tổng quát biểu diễn sự trừu tượng hoá ở mức cao hơn của những thực thể chuyên biệt này.

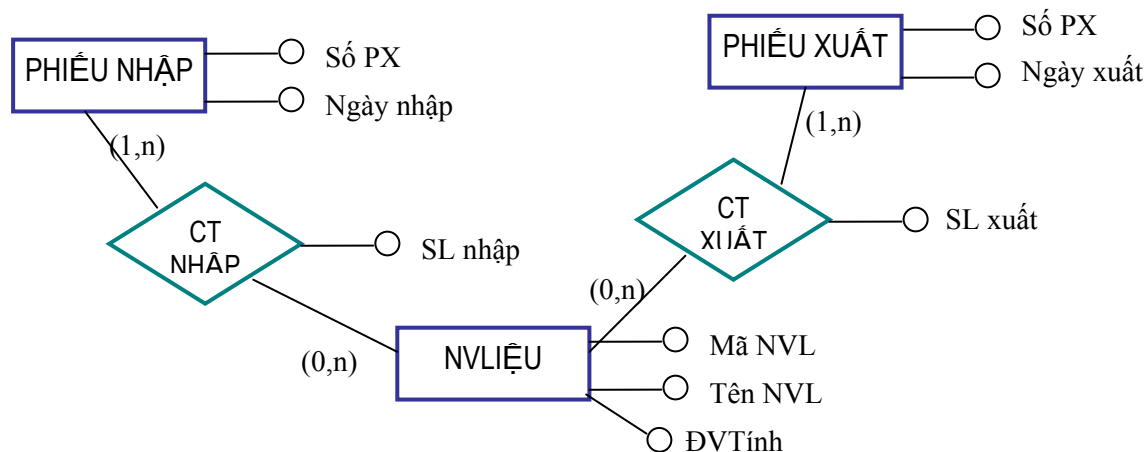
Tính chất quan trọng của cấu trúc phân cấp là tính thừa kế: tất cả các thực thể chuyên biệt sẽ thừa kế tất cả các đặc trưng của thực thể chuyên biệt (bao gồm cả thuộc tính và mối kết hợp). Sơ đồ dưới cho thấy thực thể E1 và E2 thừa hưởng từ E thuộc tính A và mối kết hợp R với thực thể E'.



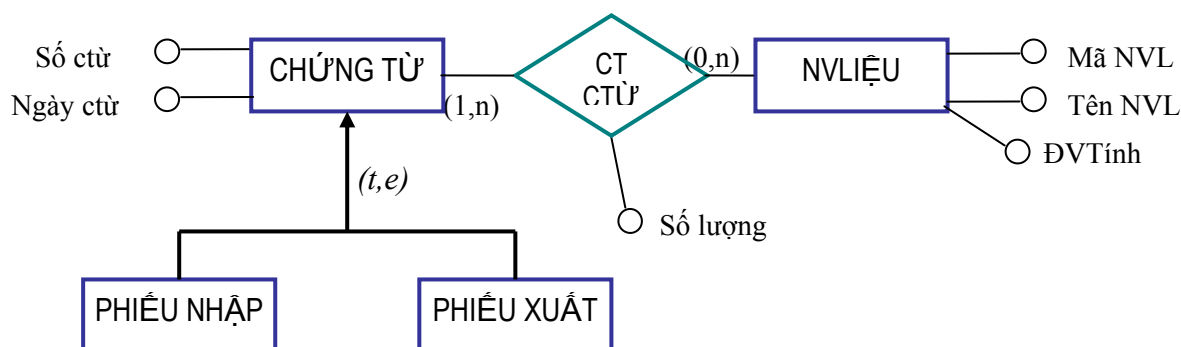
Ví dụ dưới đây, các thuộc tính HỌ-TÊN và ĐỊA-CHỈ của thực thể CON-NGƯỜI cũng là thuộc tính của thực thể ĐÀN-ÔNG và ĐÀN-BÀ. Ngoài ra, mỗi thực thể tập con ĐÀN-ÔNG và ĐÀN-BÀ cũng có thuộc tính riêng của mình là TRÌNH-TRẠNG NGHĨA-VỤ và TÊN-CON-GÁI



Ví dụ khác: giả sử chúng ta có một mô hình mô tả nhập xuất nguyên vật liệu:



Xây dựng một thực thể CHỨNG TỪ tổng quát cho hai thực thể PHIEU NHẬP và PHIEU XUẤT



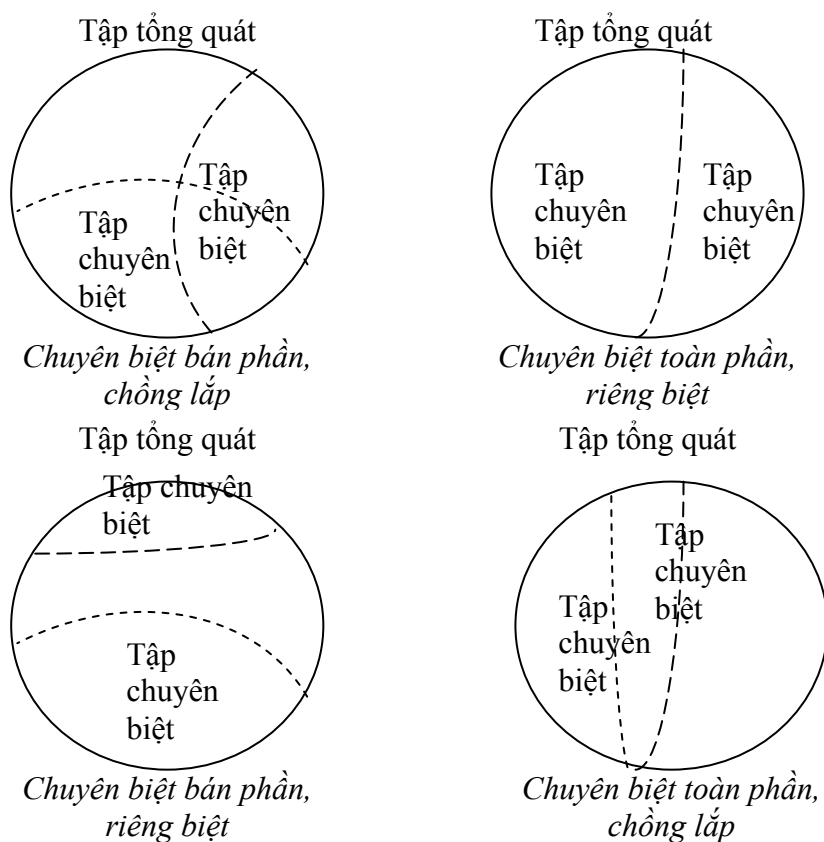
Sự tương quan giữa các đối tượng trong cấu trúc phân cấp

Sự tương quan giữa các thực thể chuyên biệt với thực thể tổng quát:

- Tập hợp các đối tượng của tất cả các thực thể chuyên biệt phủ toàn bộ tập đối tượng của thực thể tổng quát thì gọi là toàn phần (total), ký hiệu là t .
- Tập hợp các đối tượng của tất cả các thực thể chuyên biệt không phủ toàn bộ tập đối tượng của thực thể tổng quát thì gọi là bán phần (partial), ký hiệu là p .

Sự tương quan giữa các thực thể chuyên biệt:

- Không tồn tại một đối tượng của thực thể tổng quát thuộc hai thực thể chuyên biệt trở lên thì gọi là riêng biệt (exclusive), ký hiệu là e .
- Tồn tại một đối tượng của thực thể tổng quát thuộc hai thực thể chuyên biệt trở lên thì gọi là chồng lấp (overlapping), ký hiệu là o .



Hình 53. Sự tương quan giữa các class trong quan hệ tổng quát hoá

Như vậy, trong quan hệ tổng quát hoá, sự tương quan giữa các thực thể được biểu diễn quan bốn trường hợp (bán phần - chồng lấp (p,o), bán phần - riêng biệt (p,e), toàn phần - chồng lấp (t,o), toàn phần - riêng biệt (t,e)). Sự tương quan này phản ánh ràng buộc ngữ nghĩa trong tập hợp các đối tượng của quan hệ: một đối tượng của thực thể chuyên biệt này có thể là đối tượng trong thực thể chuyên biệt khác hay không? Và một đối tượng trong thực thể tổng quát có thể không thuộc một thực thể chuyên biệt nào hay không?.

Ví dụ, quan hệ tổng quát hoá giữa Xe – Xe tải, Xe bus, Xe hơi có sự tương quan là bán phần – riêng biệt (p,e). Quan hệ giữa Nhân viên – Thư ký, Kỹ sư, Nhân viên quản lý có sự tương quan là bán phần - chồng lấp (p,o).

Lưu ý: cặp giá trị thông thường là (t,e) được xem là ngầm định và có thể không cần biểu diễn trên sơ đồ.

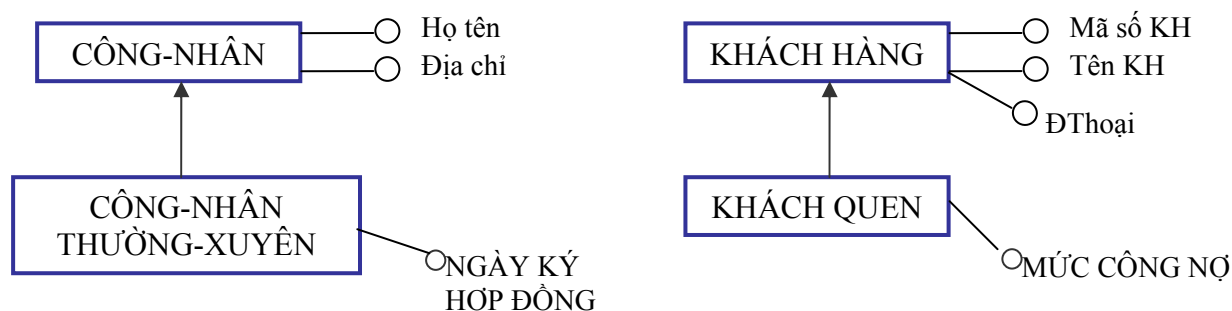
Trong hình 46, tính bao phủ được giải thích cụ thể như sau:

1. Sự tổng quát hóa dựa trên giới tính là toàn bộ và riêng biệt.
2. Giả sử trong lãnh vực ứng dụng mà chúng ta đang xét có thể bao gồm con người mà không phải là nhân viên, thư ký hay quản lý. Do đó tổng quát hóa dựa trên vai trò công việc này là bán phần và riêng biệt.
3. Giả sử mỗi người có thể đảm nhận nhiều loại công việc khác nhau do đó tổng quát hóa dựa vào loại công việc là bán phần và chồng chéo.
4. Giả sử mỗi nhân viên quản lý có thể nhận cả hai vai trò quản lý kỹ thuật và quản lý hành chính nên tổng quát hóa dựa trên vai trò quản lý sẽ là toàn bộ và chồng chéo.

Chúng ta cũng cần nhớ lại tính chất cơ bản của sự trừu tượng dạng tổng quát hóa: Tất cả các đặc trưng của thực thể tổng quát sẽ được kế thừa bởi các thực thể tập con. Trong mô hình TT-KH, điều này có nghĩa là mọi thuộc tính, mối kết hợp và tổng quát hóa định nghĩa cho thực thể tổng quát sẽ được tự động kế thừa bởi tất cả các thực thể tập con trong quan hệ tổng quát hóa. Tính chất này rất quan trọng bởi vì nó là cơ sở để xây dựng cấu trúc phân cấp tổng quát hóa.

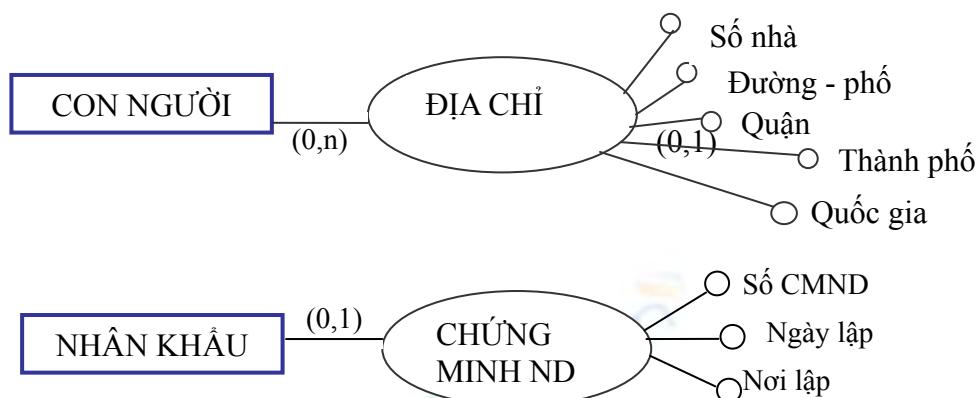
1.2.2. Tập con

Tập con là một trường hợp đặc biệt của tổng quát hóa chỉ có một thực thể tập con. Tính bao phủ của tập con rõ ràng là bán phần và riêng biệt. Chúng ta biểu diễn quan hệ tập con bởi một mũi tên chỉ vào thực thể tổng quát như trong hình sau. Tập con có thể có các thuộc tính riêng, ví dụ như là NGÀY-KÝ-HỢP ĐỒNG của CÔNG-NHÂN-THƯỜNG-XUYỀN.



1.2.3. Thuộc tính kết hợp

Thuộc tính kết hợp (subset) là một nhóm các thuộc tính mà có liên hệ trong ý nghĩa khi sử dụng. Ví dụ như thuộc tính địa chỉ biểu thị một nhóm các thuộc tính SỐ NHÀ, ĐƯỜNG PHỐ, QUẬN, THÀNH PHỐ VÀ QUỐC GIA. Chúng ta sẽ biểu diễn thuộc tính kết hợp bằng hình bầu dục như mô hình tiếp theo.



Bản số tối thiểu và tối đa áp dụng cho thuộc tính kết hợp cũng tương tự như định nghĩa cho thuộc tính cơ bản. Trong mô hình trên, một con người có thể có nhiều địa chỉ và mỗi địa chỉ bao gồm số nhà, đường phố, quận, thành phố và quốc gia. Trong đó thuộc tính quận là tùy ý, không bắt buộc.

1.2.4. Định danh

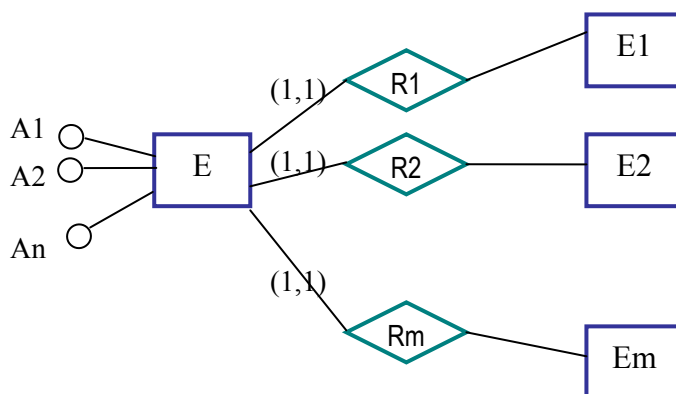
Định danh (Identifier) của một thực thể E là một tập hợp các thuộc tính hay của các thực thể liên quan đến E có tính chất đặc trưng duy nhất để có thể xác định phân biệt tất cả thể hiện của E. (Nhìn dưới quan điểm thuật ngữ, định danh thỉnh thoảng được xem là khóa (key) và khóa dự tuyển (candidate key))

Tiếp theo, chúng ta hãy xem xét định nghĩa hình thức của định danh như sau:

Cho E là một thực thể;

Cho A1, A2, A3,..., An là các đơn thuộc tính, bắt buộc của E;

Cho E1, E2, E3,..., Em là các thực thể khác được nối kết đến E bằng các mối kết hợp nhị phân, một-một, nhiều-một, bắt buộc (i.e min-card (E,Ri) = max-card (E,Ri) = 1)



Xem xét tập hợp có thể là định danh I như sau: $I = \{A_1, \dots, A_n, E_1, \dots, E_m\}$ với $n \geq 0, m \geq 0, n+m \geq 1$

Giá trị của định danh cho một thể hiện cụ thể của thực thể E là e được định nghĩa như là một tập hợp tất cả các giá trị tương ứng của thuộc tính $A_i \quad i = 1, 2, \dots, n$ và tất cả thể hiện của thực thể $E_j \quad j = 1, 2, \dots, m$ kết nối với E với $i \leq n$ và $j \leq m$.

Bởi vì giả thiết đã cho các thuộc tính là đơn trị và bắt buộc, các mối kết hợp là bắt buộc với bản số tối đa là 1, do đó mỗi thể hiện của E có thể ánh xạ tới hoặc một giá trị của thuộc tính A_i hoặc đến một thể hiện của thực thể E_j với $i \leq n$ và $j \leq m$.

I là định danh của E nếu các đặc trưng sau được đảm bảo:

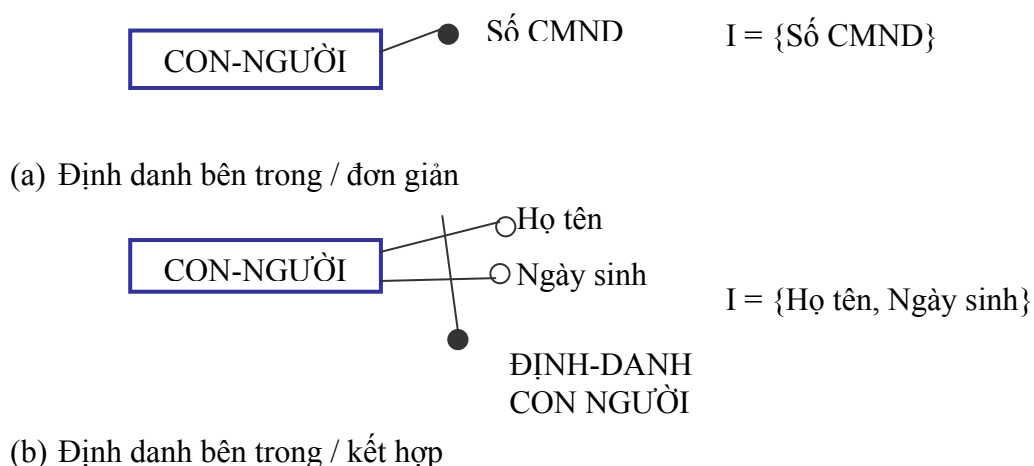
- *Đặc trưng duy nhất*: Không thể có hai thể hiện của E có chung một giá trị định danh chung.
- *Đặc trưng tối thiểu*: Nếu chúng ta bỏ một vài thuộc tính A_i hay thực thể E_j trong định danh, tính chất 1 không còn được đảm bảo.

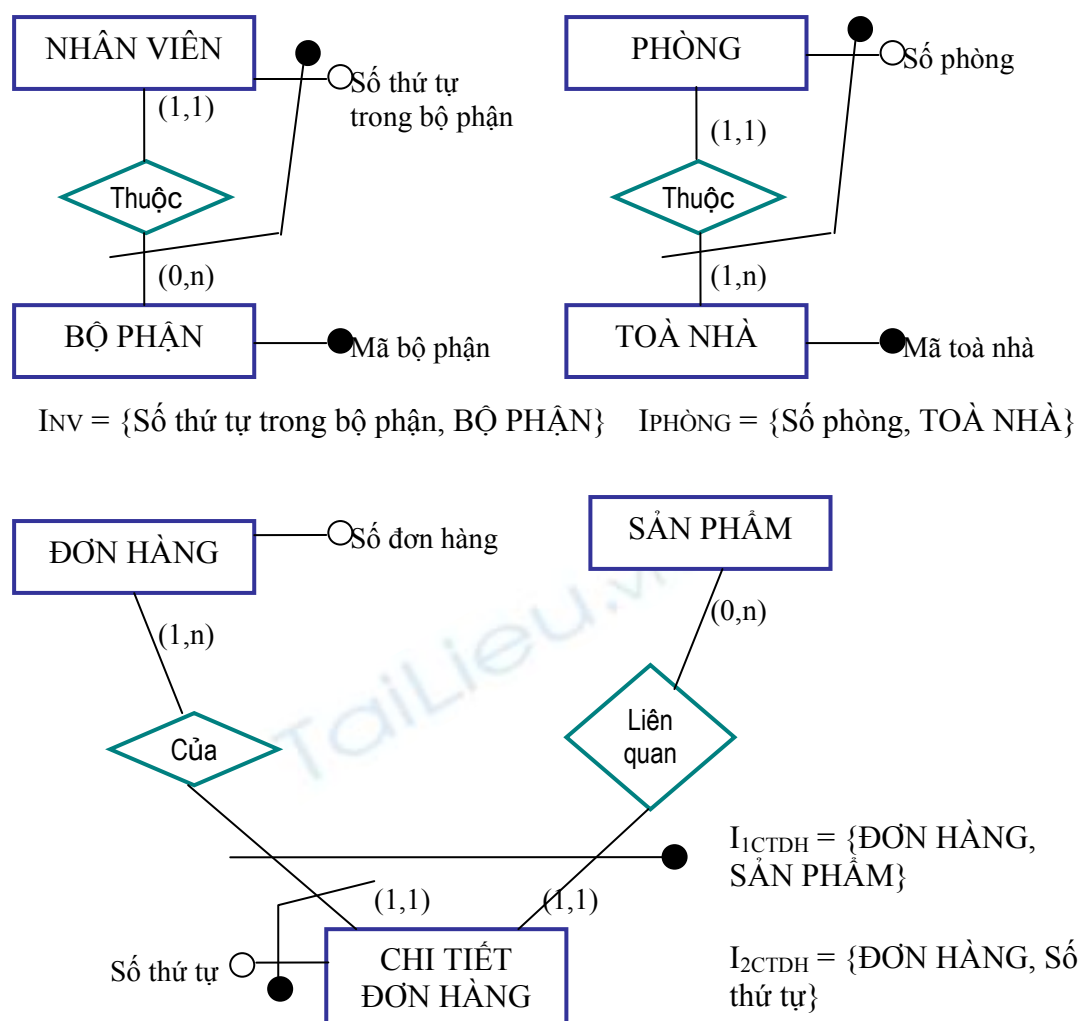
Mỗi thực thể có thể có nhiều định danh đồng thời. Chúng ta phân loại các định danh như sau:

1. Một định danh là **đơn giản** (simple) nếu $n+m = 1$; nó là **kết hợp** (composite) nếu $n+m > 1$.
2. Một định danh là **bên trong** (internal) nếu $m = 0$; nó là **bên ngoài** (external) nếu $n = 0$.
3. Một định danh là **phối hợp** (mixed) nếu $m > 0$ và $n > 0$.

Một điều quan trọng cần chú ý là nên tránh vòng lặp trong khi định nghĩa các định danh (ví dụ như sử dụng thực thể E_i khi định nghĩa định danh của E_j và thực thể E_j lại được định nghĩa bởi thực thể E_i). Với mục đích tránh các vòng lặp, một khi thực thể E_j được bao gồm trong định nghĩa của thực thể E , thì E_j phải có định danh riêng biệt không phụ thuộc vào E . Trong thực tế, việc xây dựng mô hình được thực hiện bằng cách bắt đầu quá trình với các thực thể có định danh bên trong. Các thực thể này trong một số trường hợp được gọi là thực thể mạnh (strong Entity). Sau đó, xác định định danh cho các thực thể có định danh bên ngoài. Các thực thể này còn được gọi là thực thể yếu (weak entities).

Hình tiếp theo sẽ giới thiệu các ví dụ về các định danh. Biểu diễn đồ họa của định danh là vòng tròn đen nhỏ.





(c) Định danh kết hợp – bên ngoài và phối hợp

Hình 54. Một số mô hình mô tả định danh

Trong hình (a), SỐ-CMND là định danh đơn-giản – bên trong của thực thể CON-NGƯỜI.

Hình (b), HỌ-TÊN và NGÀY-SINH cấu trúc nên định danh kết hợp – bên trong cho thực thể CON NGƯỜI. Định danh kết hợp được biểu diễn đồ họa bằng một đường phân đoạn nối kết hai hay nhiều thành phần của định danh. Một vòng tròn đen nhỏ đánh dấu các điểm nối giữa đường phân đoạn và các thành phần của định danh. Một đầu của đường phân đoạn sẽ có một vòng tròn đen và tên của định danh nếu có. Trong ví dụ này tên định danh là ĐỊNH-DANH-CON-NGƯỜI

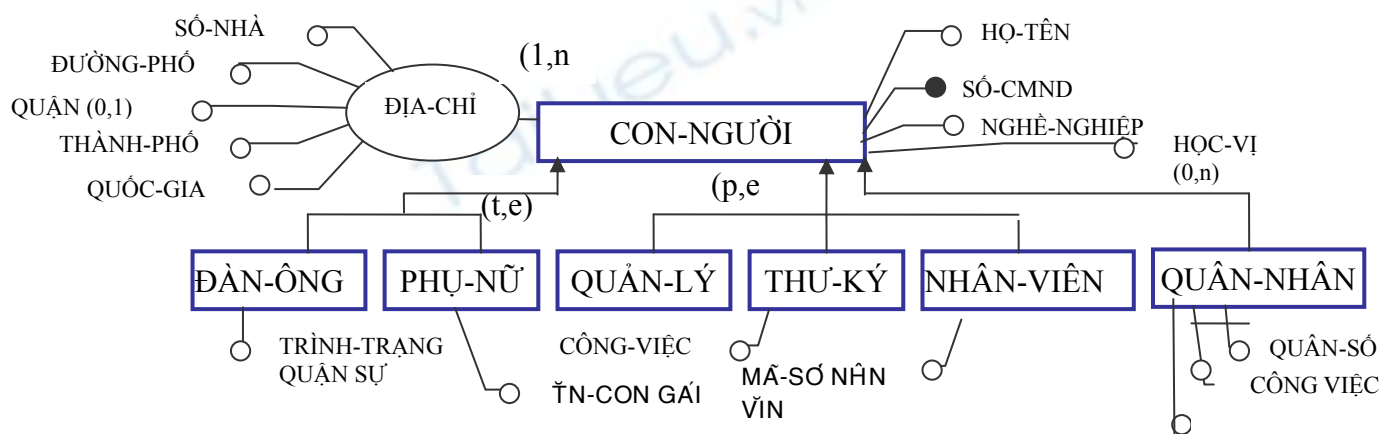
Chúng ta hãy xem xét thực thể NHÂN-VIÊN trong hình (c) kết nối đến thực thể BỘ-PHẬN bởi mỗi kết hợp THUỘC với bản số max-card (NHÂN-VIÊN, THUỘC) = 1. Do đó, thực thể BỘ-PHẬN và thuộc tính SỐ-THỨ-TỰ-TRONG-BỘ-PHẬN tạo nên một định danh bên ngoài, kết hợp và phối hợp cho thực thể NHÂN-VIÊN.

Bây giờ chúng ta tiếp tục xem xét thực thể ĐƠN-ĐẶT-HÀNG. Giả sử rằng mỗi CHI-TIẾT-ĐDH được kết nối đến thực thể SẢN-PHẨM bởi mỗi kết hợp LIÊN QUAN và kết nối đến thực thể ĐƠN-ĐẶT-HÀNG bằng mỗi kết hợp CỦA. Bản số của các mối kết hợp này là card (CHI-TIẾT-ĐDH, CỦA) = card (CHI-TIẾT-ĐDH, LIÊN QUAN) = (1,1). Cũng giả thiết rằng không có hai CHI-TIẾT-ĐDH cùng liên hệ đến cùng một sản phẩm

trong một đơn đặt hàng. Do đó, cặp các thực thể ĐƠN-ĐẶT-HÀNG (qua kết hợp CỬA) và SẢN-PHẨM (qua kết hợp LIÊN QUAN) tạo nên một định danh kết hợp – bên ngoài của thực thể CHI-TIẾT-ĐƠN với tên gọi là A. Mặt khác, giả sử một chi tiết đơn đặt hàng được đánh số thứ tự ghi nhận bởi thuộc tính SỐ-THỨ-TỰ. Cặp gồm thực thể ĐƠN-ĐẶT-HÀNG (qua mỗi kết hợp CỬA) và thuộc tính SỐ-THỨ-TỰ tạo nên một định danh khác tên gọi là B có tính chất bên ngoài, kết hợp và phối hợp.

Do định danh là một đặc trưng của thực thể nên nó cũng là một đặc trưng được kế thừa bởi các thực thể tập con tương ứng: một định danh của thực thể tổng quát cũng là định danh của các thực thể tập con. Ngoài ra, các thực thể tập con có thể có các định danh khác.

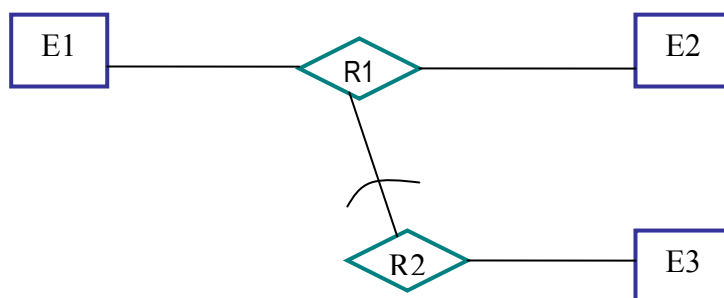
Trong hình tiếp theo, mỗi con người được định danh bởi số CMND; mỗi nhân viên được định danh bởi mã số nhân viên được gán trong công ty; và mỗi quân nhân được định danh bởi sư đoàn và số quân trong sư đoàn.



Hình 55. Cấu trúc cây tổng quát hóa, thuộc tính kết hợp và định danh trong mô hình TT-KH

1.2.5. Mỗi kết hợp mở rộng

Là mỗi kết hợp được định nghĩa trên ít nhất một mối kết hợp khác. Trong thực tế, một số tình huống của mỗi kết hợp được tạo ra không phải do ngữ nghĩa liên kết giữa hai thực thể mà dựa trên ít nhất một mối kết hợp đã tồn tại trước. Trường hợp này mô hình thực thể - kết hợp truyền thống không cho phép biểu diễn. trong mô hình thực thể kết hợp mở rộng, chúng ta biểu diễn điều này thông qua mỗi kết hợp mở rộng.

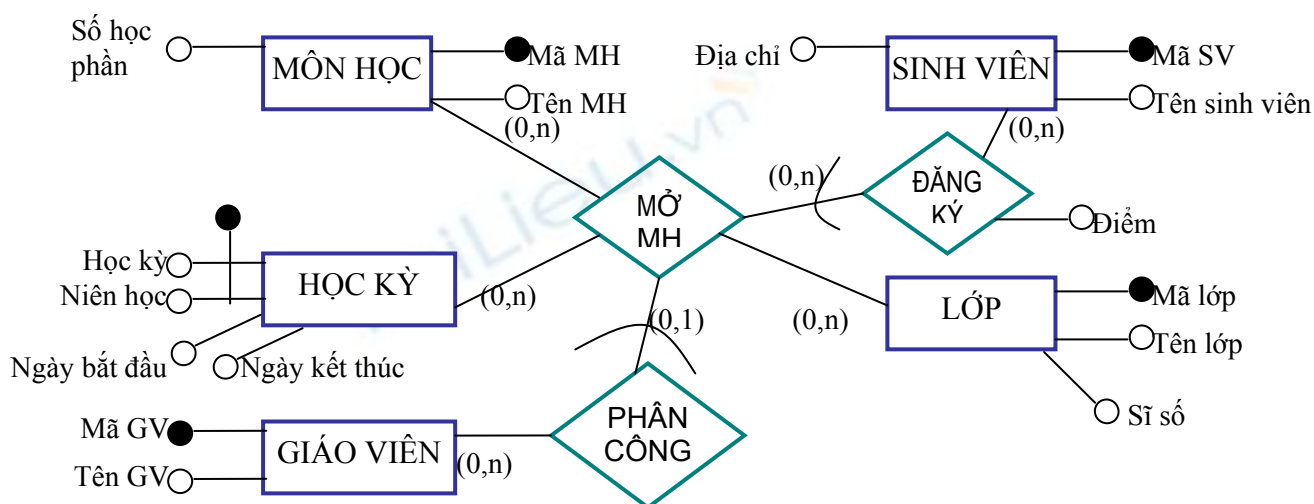


Trong mô hình, **R2** là mối kết hợp mở rộng bởi vì nó được xác định dựa trên mối kết hợp **R1** và **E3**. Do đó **R1** phải tồn tại trước **R2** và thể hiện của **R2** sẽ được xác định dựa trên thể hiện của **R1** và **E3**.

Ví dụ: xem mô tả của bài toán Quản lý giáo vụ của một trường học gồm các yêu cầu như sau:

- Lưu trữ các thông tin về sinh viên, môn học, giáo viên, lớp học và các học kỳ trong từng niên khóa.
- Lập danh sách mở các môn học cho một lớp trong một học kỳ.
- Phân công giảng dạy môn học được mở cho một giáo viên.
- Lưu thông tin đăng ký môn học của sinh viên trên môn học được mở
- Ghi nhận điểm kết quả học tập của sinh viên

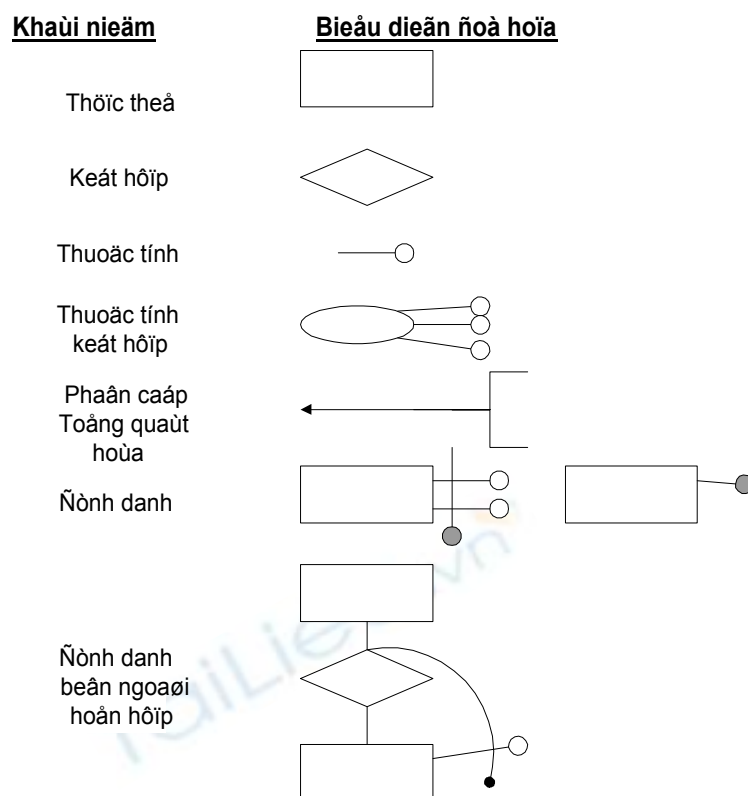
Kết quả mô hình hóa dữ liệu như sau:



Trong mô hình trên, ĐĂNG KÝ và PHÂN CÔNG là các mối kết hợp mở rộng bởi vì nó được định nghĩa dựa trên mối kết hợp MỞ MH. Dựa trên mối kết hợp MỞ MH hệ thống cho phép giáo vụ lưu trữ thông tin về môn học được mở ra cho một lớp trong một học kỳ. Mỗi kết hợp ĐĂNG KÝ cho phép lưu trữ thông tin đăng lý học phần của sinh viên trên môn học được mở. Mỗi kết hợp PHÂN CÔNG cho phép lưu trữ thông tin về việc phân công một giáo viên trên một môn học mở.

1.3 Sự trừu tượng hóa trong mô hình TT-KH

Để đánh giá sự trừu tượng hóa trong mô hình TT-KH, trước tiên chúng ta sẽ xem xét các khái niệm được biểu diễn trong mô hình TT-KH được nhắc lại trong hình 3.27 tiếp theo



Hình 56. các khái niệm trong mô hình TT-KH

Kế tiếp, chúng ta sẽ lần lượt xem xét từng loại trừ tượng hóa gồm trừ tượng hóa dạng phân loại, trừ tượng hóa dạng kết hợp và trừ tượng hóa dạng tổng quát hóa đã được sử dụng trong các khái niệm trong mô hình TT-KH biểu diễn như thế nào.

I.4 Sự trừ tượng hóa dạng phân loại

Trong mô hình TT-KH, các khái niệm cơ bản là thực thể, mối kết hợp và thuộc tính đều là sự trừ tượng hóa dạng phân loại:

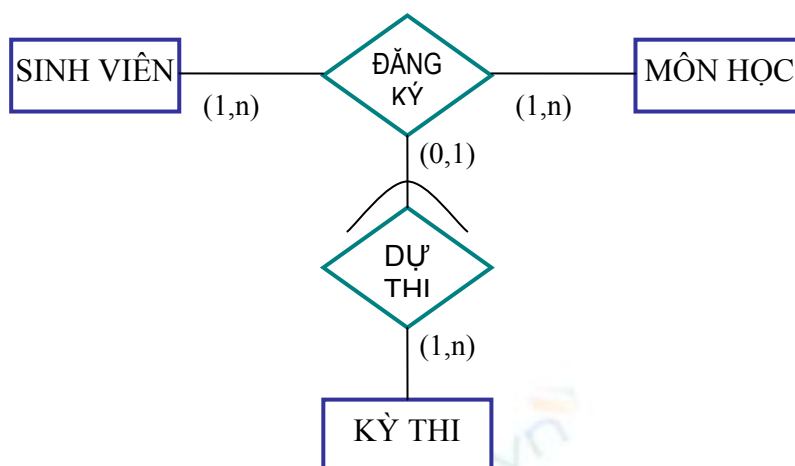
- *Thực thể*: là một lớp các đối tượng thể giới thực với các tính chất đặc trưng của chúng
- *Mối kết hợp*: là một lớp các sự kiện căn bản phản ánh sự liên hệ của hai hay nhiều thực thể
- *Thuộc tính*: là một lớp các giá trị biểu diễn các đặc trưng nguyên tố của thực thể hay mối kết hợp

I.5 Sự trừ tượng hóa dạng kết hợp

Tương tự như 3.4.3.2, các khái niệm cơ bản của mô hình TT-KH là thực thể, mối kết hợp và thuộc tính kết hợp đều là sự trừ tượng hóa dạng kết hợp:

- *Thực thể*: là sự kết hợp các thuộc tính
- *Mối kết hợp*: là sự kết hợp giữa các thực thể
- *Thuộc tính kết hợp*: là sự kết hợp các thuộc tính thành phần

Chú ý: trong một số trường hợp đặc biệt, một số mô hình TT-KH mở rộng cho phép chúng ta định nghĩa mối kết hợp giữa các thực thể và các mối kết hợp khác. Ví dụ như sau:



Hình 57. mối kết hợp trên mối kết hợp

Trong ví dụ trên, mối kết hợp DỰ-THI là mối kết hợp được định nghĩa trên mối kết hợp ĐĂNG-KÝ và thực thể KỶ-THI. Một cung tròn sẽ được vẽ gần mối kết hợp DỰ THI cho biết nó được định nghĩa trên mối kết hợp ĐĂNG KÝ.

Mối kết hợp ĐĂNG-KÝ được gọi là mối kết hợp một ngôi (nghĩa là định nghĩa giữa các thực thể), mối kết hợp DỰ-THI là mối kết hợp hai ngôi (nghĩa là định nghĩa giữa các thực thể và mối kết hợp một ngôi). Tương tự như trên, mối kết hợp ba ngôi sẽ là mối kết hợp định nghĩa giữa các thực thể, mối kết hợp một ngôi và mối kết hợp hai ngôi

1.6 Sự trừu tượng hóa dạng tổng quát hóa

Trong mô hình TT-KH, các khái niệm sau diễn tả sự trừu tượng hóa dạng tổng quát hóa:

- Phân cấp tổng quát hóa
- Thực thể tập con

Chú ý: khái niệm tổng quát hóa và tập con thường áp dụng cho các thực thể, tuy nhiên trong một số mô hình TT-KH mở rộng, tổng quát hóa có thể áp dụng cho cả mối kết hợp. Ví dụ như sau: