

Mô hình hóa dữ liệu ĐPT

Nguyễn Đình Hóa
dinhhoa@gmail.com 0942807711

Tóm tắt nội dung bài 5

- ▶ Kiến trúc hệ CSDL đa phương tiện
 - ▶ Hệ CSDL tập trung
 - ▶ Hệ CSDL phân tán
- ▶ Nguyên lý lưu trữ CSDL ĐPT: riêng biệt, đồng nhất, kết hợp
- ▶ Yêu cầu đối với máy chủ, siêu máy chủ
- ▶ Hệ thống khách-chủ (client-server)
- ▶ Hệ thống điểm-điểm (peer-to-peer)
- ▶ Hệ thống phân tán hỗn hợp

Vai trò mô hình hóa CSDL ĐPT

- ▶ Mô tả dữ liệu ĐPT với cấu trúc phức tạp một cách chính xác và đầy đủ
- ▶ Dữ liệu ĐPT rất phong phú, thay đổi liên tục
- ▶ Các loại dữ liệu đơn phương tiện thường có mối liên hệ mật thiết với nhau
- ▶ Mô hình hóa CSDL ĐPT giúp lưu trữ và tra cứu thông tin một cách hiệu quả

Mô hình dữ liệu quan hệ

- ▶ Dữ liệu được mô tả dưới dạng “phẳng” (các bảng CSDL), khó diễn đạt được các cấu trúc dữ liệu phức tạp
- ▶ Các mối liên kết dữ liệu giữa các thực thể thường ở dạng tĩnh,
- ▶ Các quan hệ giữa các bảng (các phần trong các bảng) được định nghĩa rõ ràng.

Mô hình dữ liệu quan hệ

- ▶ Các phép toán đại số quan hệ:
 - ▶ Phép chiếu
 - ▶ Phép chọn
 - ▶ Phép hợp
 - ▶ Phép trừ
 - ▶ Phép tích đề các
 - ▶ Các phép toán mở rộng: Phép giao, phép kết nối, phép kết nối tự nhiên, phép kết nối ngoài, phép bán kết nối, phép chia.
- ▶ Ngôn ngữ thao tác và truy vấn đại số quan hệ SQL

Mô hình dữ liệu quan hệ

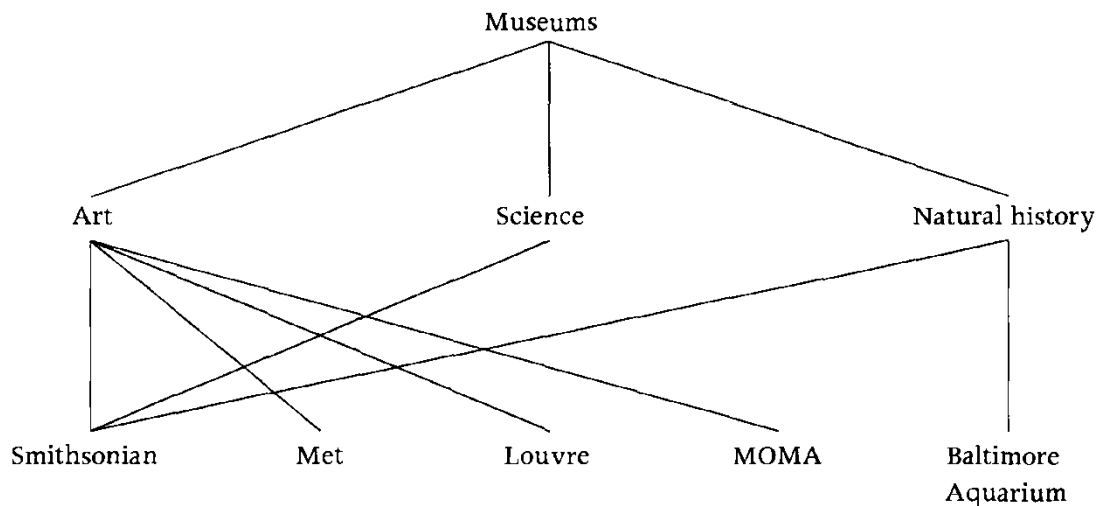
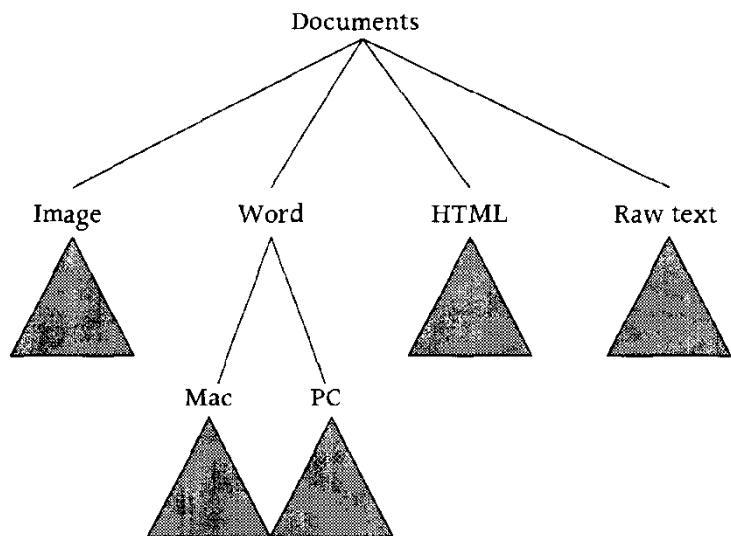
- ▶ Những bất cập trong lưu trữ dữ liệu đa phương tiện
 - ▶ Khó mô tả dữ liệu phức tạp
 - ▶ Khó thay đổi các cột, các hàng trong bảng dữ liệu
 - ▶ Quan hệ giữa các bảng dữ liệu tuân theo quy tắc vẹn toàn.

Mô hình dữ liệu hướng đối tượng (Object Oriented Database Systems)

- ▶ Dữ liệu được tổ chức theo 3 phần
 - ▶ Thực thể: tất cả dữ liệu được lưu trữ là các thành phần của thực thể
 - ▶ Lớp: là tập hợp các thực thể (một thực thể đơn lẻ cũng có thể coi là một lớp). Lớp cũng bao gồm các phương thức mô tả việc sử dụng các thực thể.
 - ▶ Phân cấp lớp (miền): bao gồm cấu trúc liên kết giữa các lớp.

Mô hình dữ liệu hướng đối tượng

- ▶ Mỗi văn bản thuộc một trong số các lớp ở trên nó (một số văn bản Word có thể xem được ở cả hai môi trường PC và MAC)
- ▶ Mỗi bảo tàng có thể coi là một thực thể. Các bảo tàng khác nhau có thể gom thành một nhóm (loại). Một bảo tàng có thể thuộc một hoặc nhiều nhóm.



Mô hình dữ liệu hướng đối tượng

- ▶ Thực thể: mỗi thực thể được mô tả bởi tập các thuộc tính
 - ▶ Văn bản: tác giả, ngày tạo lập, ngày chỉnh sửa, từ khóa...
 - ▶ Bảo tàng: địa chỉ, điện thoại, phí tham quan, giờ tham quan,...
- ▶ Mỗi thuộc tính có một miền giá trị xác định
- ▶ Một tập thực thể có thể coi là một miền không gian đa chiều trong đó mỗi thực thể là một tập giá trị cụ thể nào đó.

Mô hình dữ liệu hướng đối tượng

- ▶ Khái niệm tập hợp
 - ▶ Tập các thực thể O_{id}
 - ▶ Tập các lớp C_{id}
 - ▶ Tập các thuộc tính: Att

Mô hình dữ liệu hướng đối tượng

▶ Văn bản

- ▶ $O_{id} = \{b_1, b_2, \dots, b_6\}$
- ▶ $C_{id} = \{\#html, \#raw\ text, \#image, \#word\ mac, \#word\ pc\}$
- ▶ $Att = \{author, date_created, date_modified, key\ words, related_docs\}$

▶ Bảo tàng

- ▶ $O_{id} = \{b_{met}, b_{louvre}, \dots, b_{Smithsonian}\}$
- ▶ $C_{id} = \{\#art, \#history, \#nature\}$
- ▶ $Att = \{address, director, annual_budget, admission_fee, hours, special_exhibits\}$

Mô hình dữ liệu hướng đối tượng

- ▶ Một thực thể được xác định bởi cấu trúc các thuộc tính và bộ giá trị của các thuộc tính đó.
- ▶ Mô hình CSDL hướng đối tượng cho phép mô tả các thực thể theo một điểm chung (bằng cách gom nhóm các thực thể vào các lớp), nhưng cũng cho phép chúng có các thuộc tính riêng (hai thực thể riêng biệt thuộc cùng một lớp có thể có các tập thuộc tính hoàn toàn khác nhau)

Mô hình dữ liệu hướng đối tượng

► Khai báo các thực thể:

VD:

declare b_2

values [author = John Smith;

url = http://www.somewhere.com/index.html;

date-created= (15, Jan, 1996);

date-last-modified=(19, Nov, 1996)]

declare b_3

values [author = John Smith;

url = http://www.somewhereelse.com/index.html;

{ [link = b_1], [link = b_2] }]

Mô hình dữ liệu hướng đối tượng

```
declare bSmithsonian  
values [address = "The Mall, Washington DC";  
        director = "Jim Schwartz";  
        special-exhibits = {Caravaggio paintings,  
                           Persian manuscripts};  
        affiliated-museums = {bLouvre, bMOMA}]
```

MOMA: Museum of Modern Art, New York.

Mô hình dữ liệu hướng đối tượng

- ▶ Kiểu dữ liệu (định dạng) dữ liệu xác định miền giá trị cho từng thuộc tính của thực thể (Real, Boolean, Integer, Char, String ...)

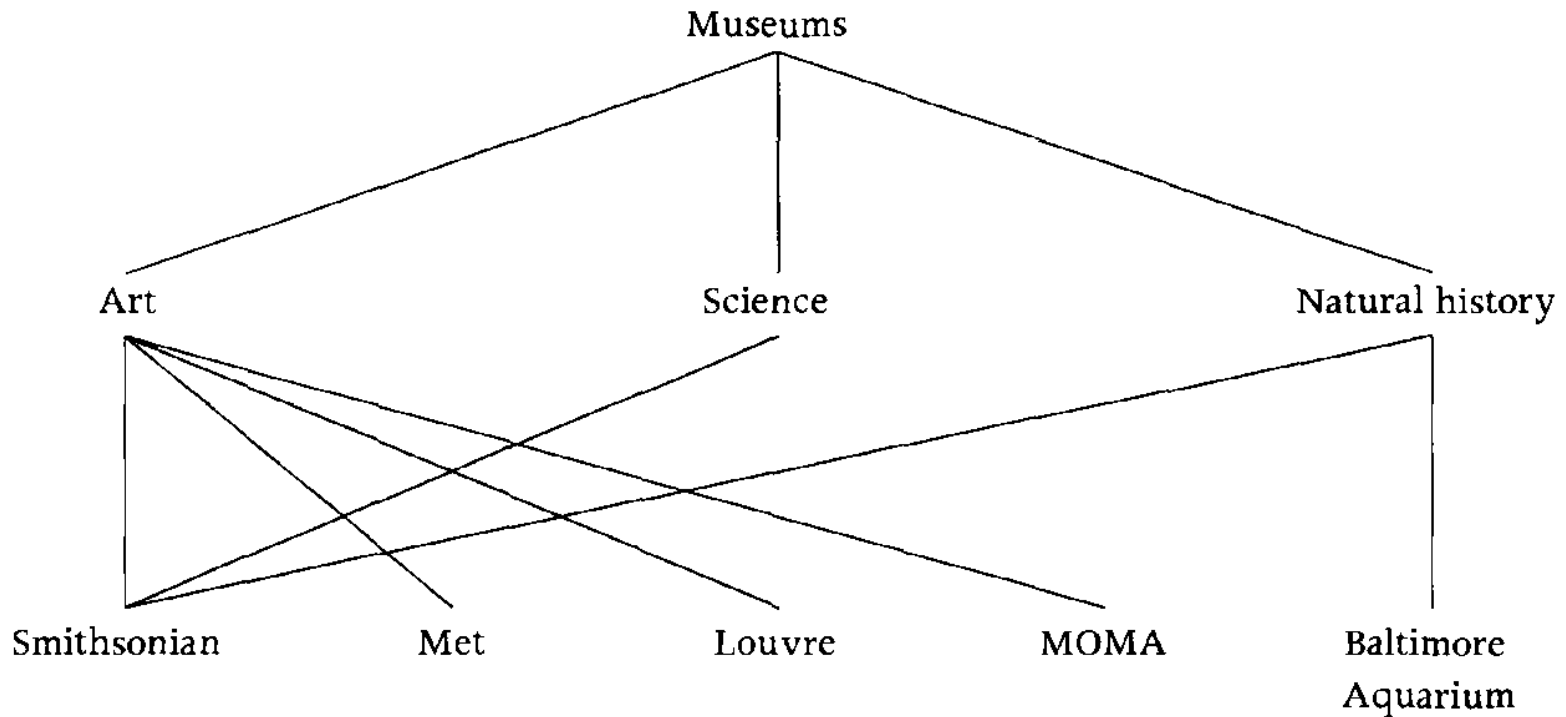
VD:

```
[ author: string;  
url: urltype;  
date-created: datatype;  
date-last-modified: datatype  
address: string;  
director: string;  
special-exhibits: { string }; // biểu diễn một tập các chuỗi  
affiliated-museums: { Oid_Set } ] // biểu diễn một tập các thực thể
```

Mô hình dữ liệu hướng đối tượng

► Phân cấp lớp (class hierarchy)

- Trong biểu diễn lớp phân cấp, để mô tả thực thể ở từng lớp, ta không đơn thuần liệt kê các kiểu dữ liệu kế thừa từ lớp cao xuống lớp thấp



Mô hình dữ liệu hướng đối tượng

Trong phân cấp lớp, các lớp con chứa toàn bộ các thuộc tính của lớp cha, cùng với các thuộc tính bổ sung

Museum	Art (museum)	Smithsonian
address: string; director: string; departments: { string }; budget: real	address: string; director: string; departments: { string }; budget: real ; old-master-collection: { string }; modern-art-collection: { string }; lithograph-collection: { string }; cartographic-collection: { string }	address: string; director: string; departments: { string }; budget: real; old-master-collection: { string }; modern-art-collection-. { string }; lithograph-collection: { string }; cartographic-collection: { string }; animal-collection: { string }; mineral-collection: { string }; aquatic-collection: { string }; airplane-collection: { string }; space-collection: { string }

Mô hình dữ liệu hướng đối tượng

- ▶ Trong biểu diễn phân cấp nói trên, Smithsonian bao gồm cả aquatic-collection và airplane-collection bởi vì nó là lớp con của cả Natural History Museum và Science Museum.

Mô hình dữ liệu hướng đối tượng

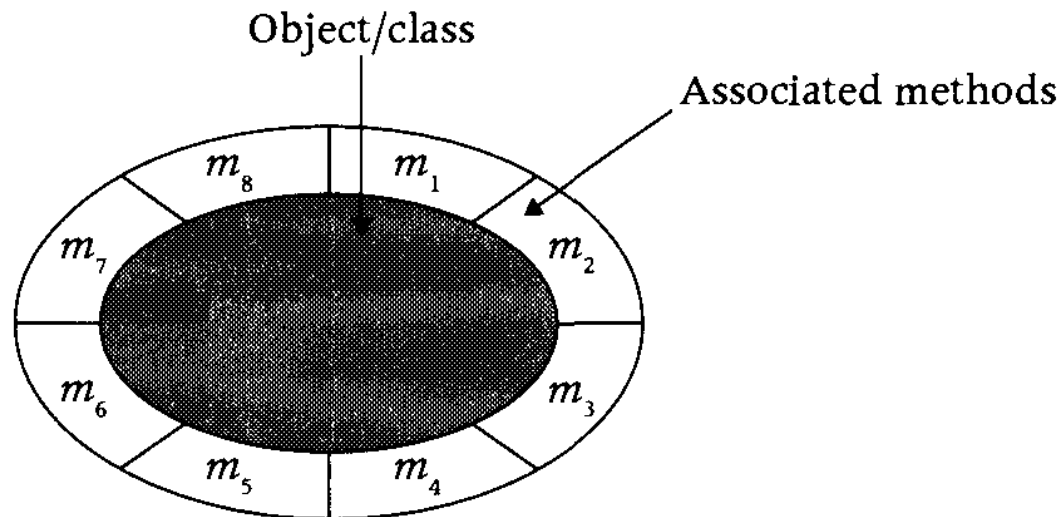
► Phân cấp lớp

- Các thành viên của một lớp là toàn bộ các lớp con được biểu diễn trong mô hình phân lớp.
- Khi biểu diễn các lớp con, ta không nhất thiết phải liệt kê hết các thuộc tính của lớp cha.
 - VD: lớp art có thể được biểu diễn bởi

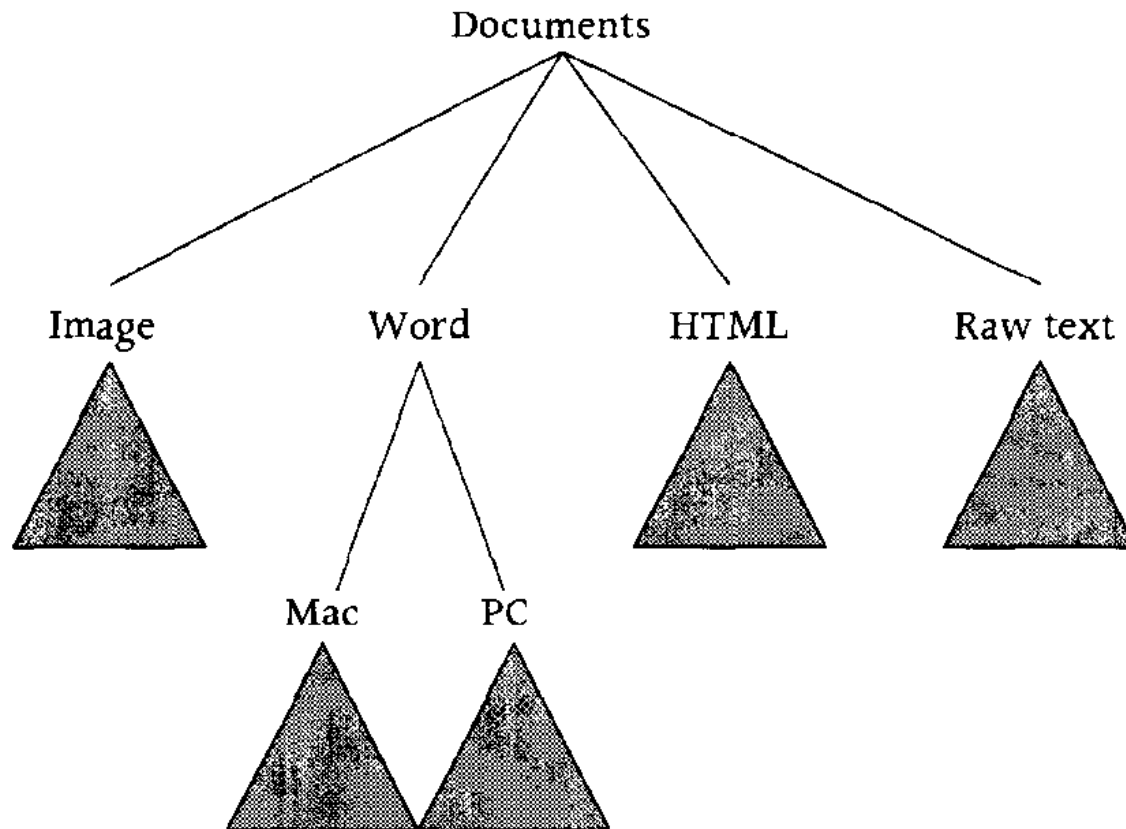
```
[ old-master-collection: { string };  
  modern-art-collection: { string };  
  lithograph-collection: { string };  
  cartographic-collection: { string }]
```

Mô hình dữ liệu hướng đối tượng

- ▶ Phương thức (Methods): là các thủ tục, thao tác thực hiện trên các lớp dữ liệu.
 - ▶ Mỗi lớp thực thể gắn liền với một (tập) phương thức nhất định
 - ▶ Các thực thể ở mỗi lớp chỉ có thể được thao tác (thực hiện) bởi các phương thức của lớp đó.



Mô hình dữ liệu hướng đối tượng



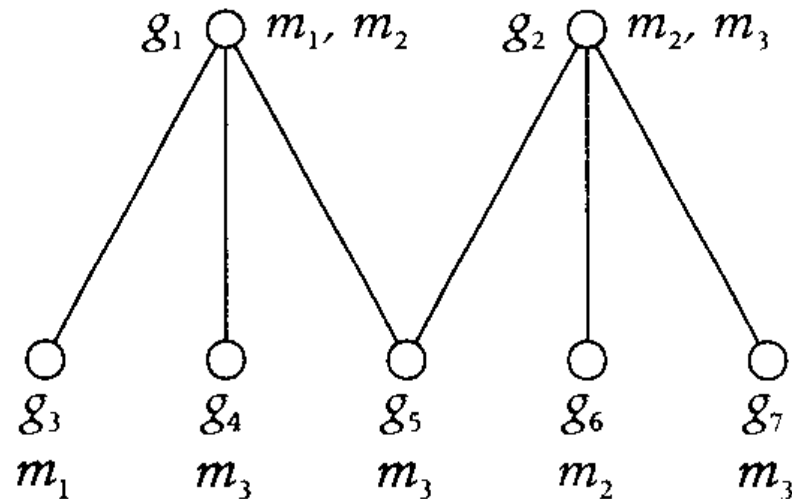
Mô hình dữ liệu hướng đối tượng

- ▶ Ví dụ: hàm FindDocs với đầu vào là tên của một người (kiểu String) để tìm tất cả các tài liệu liên quan đến người đó. Các phương thức tìm kiếm với các lớp khác nhau:
 - ▶ Lớp Raw_text có phương thức FindDocs_d với đầu ra là các văn bản có liên quan
 - ▶ Lớp Image có phương thức FindDocs_i với đầu ra là các hình ảnh có liên quan
 - ▶ Lớp Word có phương thức FindDocs_w với đầu ra là các văn bản Word có liên quan
 - ▶ Đối với lớp PC, không có hàm FindDocs tương ứng. Ta có thể sử dụng hàm tương ứng với lớp cha của nó (FindDocs_w)

Mô hình dữ liệu hướng đối tượng

- ▶ Các kiểu vào ra (I/O types)
 - ▶ Mỗi phương thức ứng với từng lớp thực thể có các kiểu dữ liệu vào/ra cụ thể
- ▶ Một phương thức có thể được dùng cho nhiều lớp thực thể. Để phân biệt phương thức m dùng cho lớp thực thể g , ta ký hiệu m_g
- ▶ Thừa kế: một phương thức của một lớp có thể được dùng cho các lớp con của lớp đấy.
 - ▶ Tương tự về cấu trúc (kiểu dữ liệu vào)
 - ▶ Giải quyết xung đột – conflict resolution policy (crp)

Mô hình dữ liệu hướng đối tượng



► Các giải pháp crp:

- Lựa chọn một lớp cha (theo thứ tự sắp xếp nhất định) từ tập nhiều lớp cha để kế thừa
- Lựa chọn phương thức mới nhất để kế thừa
- Lựa chọn một lớp cha theo độ ưu tiên (quan trọng) cao nhất.

Mô hình dữ liệu hướng đối tượng

- ▶ Ngôn ngữ định nghĩa đối tượng (Object Definition Language - ODL): không chỉ rõ các phương thức có thể thực hiện, nó chỉ nêu ra các kiểu dữ liệu vào/ra của các phương thức.
- ▶ Ngôn ngữ truy vấn đối tượng (Object Query Language - OQL): ngoài khả năng tìm kiếm dữ liệu theo yêu cầu, OQL còn cung cấp các phương tiện truy nhập vào các kiểu dữ liệu.