VỤ GIÁO DỤC CHUYÊN NGHIỆP

GIÁO TRÌNH
CƠ SỞ DỮ LIỆU

SÁCH DÙNG CHO CÁC TRƯỜNG ĐÀO TẠO HỆ TRUNG HỌC CHUYÊN NGHIỆP

NHÀ XUẤT BẢN GIÁO DỤC

TÔ VĂN NAM

GIÁO TRÌNH
CƠ SỞ DỮ LIỆU

(Sách dùng cho các trường đào tạo hệ Trung học chuyên nghiệp)

(Tái bản lần thứ nhất)

NHÀ XUẤT BẢN GIÁO DỤC

Lời giới thiệu

Năm 2002, Vụ Giáo dục Chuyên nghiệp – Bộ Giáo dục và Đào tạo đã phối hợp với Nhà xuất bản Giáo dục xuất bản 21 giáo trình phục vụ cho đào tạo hệ THCN. Các giáo trình trên đã được nhiều trường sử dụng và hoan nghênh. Để tiếp tục bổ sung nguồn giáo trình đang còn thiếu, Vụ Giáo dục Chuyên nghiệp phối hợp cùng Nhà xuất bản Giáo dục tiếp tục biên soạn một số giáo trình, sách tham khảo phục vụ cho đào tạo ở các ngành : Điện – Điện tử, Tin học, Khai thác cơ khí. Những giáo trình này trước khi biên soạn, Vụ Giáo dục Chuyên nghiệp đã gửi đề cương về trên 20 trường và tổ chức hội thảo, lấy ý kiến đóng góp về nội dung đề cương các giáo trình nói trên. Trên cơ sở nghiên cứu ý kiến đóng góp của các trường, nhóm tác giả đã điều chỉnh nội dung các giáo trình cho phù hợp với yêu cầu thực tiễn hơn.

Với kinh nghiệm giảng dạy, kiến thức tích luỹ qua nhiều năm, các tác giả đã cố gắng để những nội dung được trình bày là những kiến thức cơ bản nhất nhưng vẫn cập nhật được với những tiến bộ của khoa học kỹ thuật, với thực tế sản xuất. Nội dung của giáo trình còn tạo sự liên thông từ Dạy nghề lên THCN.

Các giáo trình được biên soạn theo hướng mở, kiến thức rộng và cố gắng chỉ ra tính ứng dụng của nội dung được trình bày. Trên cơ sở đó tạo điều kiện để các trường sử dụng một cách phù hợp với điều kiện cơ sở vật chất phục vụ thực hành, thực tập và đặc điểm của các ngành, chuyên ngành đào tạo.

Để việc đổi mới phương pháp dạy và học theo chỉ đạo của Bộ Giáo dục và Đào tạo nhằm nâng cao chất lượng dạy và học, các trường cần trang bị đủ sách cho thư viện và tạo điều kiện để giáo viên và học sinh có đủ sách theo ngành đào tạo. Những giáo trình này cũng là tài liệu tham khảo tốt cho học sinh đã tốt nghiệp cần đào tạo lại, nhân viên kỹ thuật đang trực tiếp sản xuất.

Các giáo trình đã xuất bản không thể tránh khỏi những sai sót. Rất mong các thầy, cô giáo, bạn đọc góp ý để lần xuất bản sau được tốt hơn. Mọi góp ý xin gửi về : Công ty Cổ phần sách Đại học – Dạy nghề, 25 Hàn Thuyên – Hà Nội.

VỤ GIÁO DỤC CHUYÊN NGHIỆP - NXB GIÁO DỤC

Mở đầu

***Cơ sở dữ liệu** (CSDL) là một trong những bộ môn không thể thiếu được trong chương trình của bất cứ môi trường đào tạo chuyên nghiệp nào, bởi vì có tới hơn 80% các ứng dụng của tin học là phục vụ công tác quản lý, mà **quản lý** thực chất là **quản lý thông tin**. Song, mọi thông tin cần quản lý trên máy tính cũng đều phải được thể hiện bằng các **dữ liệu**, do vậy khi nói đến **quản lý thông tin** tức là nói đến **quản lý dữ liệu**.*

*Cuốn "**Cơ sở dữ liệu**" bao gồm hai phần:*

Phần I : dành phần lớn nội dung giới thiệu các cấu trúc và khái niệm cơ bản về Cơ sở dữ liệu.

Phần II : cung cấp một số hướng dẫn đầy đủ cho bạn đọc về Hệ Quản trị Cơ sở dữ liệu SQL Server 2000.

Cuốn sách được biên soạn nhằm phục vụ công tác đào tạo và học tập trong các trường Trung học chuyên nghiệp và Cao đẳng, nhưng cũng là một tư liệu tham khảo tốt cho tất cả những ai quan tâm đến CSDL.

Tác giả rất hoan nghênh và cảm ơn mọi ý kiến đóng góp của các bạn đọc ở khắp mọi nơi. Thư góp ý xin gửi về theo địa chỉ:

TÔ VĂN NAM, Bộ môn Hệ thống thông tin, Khoa Công nghệ Thông tin - Trường Đại Học Bách Khoa, Hà Nội.

Email: namtv@it-hut.edu.vn

TÁC GIẢ

Phần 1. CẤU TRÚC VÀ KHÁI NIỆM CƠ BẢN VỀ CƠ SỞ DỮ LIỆU

Chương 1
CÁC KHÁI NIỆM CƠ BẢN

I. CÁC KHÁI NIỆM VỀ QUAN HỆ

1.1. Tập hợp

1.1.1. Các khái niệm về tập hợp

- Tập hợp là khái niệm đầu tiên của toán học, không định nghĩa. Ví dụ: Tập hợp các số nguyên tự nhiên; tập hợp các sinh viên trong một lớp; tập hợp các nghiệm của phương trình, ...
- Một tập hợp được tạo thành từ những phần tử của nó. Phần tử cũng là khái niệm đầu tiên của toán học, không định nghĩa.
- Người ta thường dùng các chữ cái viết hoa để đặt tên cho một tập hợp, chữ cái viết thường để chỉ một phần tử của tập hợp.

Ví dụ: N là tập hợp các số nguyên tự nhiên; X là tập hợp các sinh viên lớp 97 ĐIA, ...

- x góp phần tạo nên tập hợp A, ta nói x là một phần tử của A và ký hiệu là $x \in A$, đọc là: x thuộc A; y không là phần tử của A, ta ký hiệu $y \notin A$, đọc là: y không thuộc A.
- Có hai cách để biểu diễn một tập hợp:
  - Cách liệt kê: liệt kê tất cả các phần tử của tập hợp trong hai dấu ngoặc {}, hai phần tử cách nhau bởi một dấu phẩy ",".

  Ví dụ: X = {cam, quýt, ổi, xoài, me}; A = {1, 2, 3, 4}; ...
  - Cách đặc trưng: dùng một tính chất đặc trưng P để mô tả tập hợp: A={x|x thoả mãn P}, nghĩa là: mọi phần tử của A phải thoả mãn tính chất P và mọi phần tử thoả mãn tính chất P phải thuộc A.

  Ví dụ: A = {x| x là sinh viên khoá 41 trường ĐHBK}

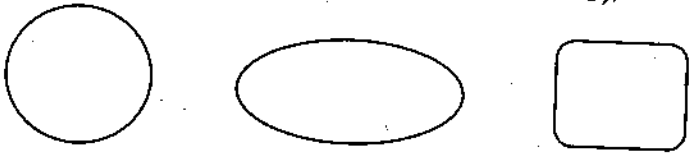
  $B = \{m/n \mid m, n \in Z \text{ và } n \neq 0\}$.

- Tập hợp rỗng: một tập hợp không có phần tử nào được gọi là tập hợp rỗng, ký hiệu là ϕ.

Ví dụ: $A = \{x \in R \mid x^2 + 2x + 3 < 0\} = \phi$

$B = \{x \mid x$ là sinh viên khoá 38 trường ĐHBK và $x > 60$ tuổi$\} = \phi$

- Giản đồ Venn: Để biểu diễn trực quan trên giấy, người ta dùng phần mặt phẳng giới hạn bởi một đường cong kín để mô tả một tập hợp, bên cạnh ghi tên tập hợp ấy (Ví dụ: A, B, C trên hình 1.1).



Hình 1.1. Mô tả tập hợp

- Tập hợp con: Cho hai tập hợp A và B, ta nói B là tập con của A nếu và chỉ nếu mọi phần tử của B cũng là phần tử của A, ký hiệu: $B \subset A$

$$B \subset A \Leftrightarrow \forall x \in B \Rightarrow x \in A$$

- Hai tập hợp bằng nhau:

$$A = B \Leftrightarrow A \subset B \text{ và } B \subset A$$

1.1.2. Các phép toán về tập hợp

Cho hai tập hợp A và B, ta có các phép toán trên hai tập hợp như sau:

a) Phép giao: Giao của hai tập hợp A và B là một tập hợp, ký hiệu $A \cap B$ gồm những phần tử vừa thuộc A, vừa thuộc B.

$$A \cap B = \{x \mid x \in A \text{ và } x \in B\}$$

b) Phép hội: Hội của hai tập hợp A và B là một tập hợp, ký hiệu $A \cup B$, gồm những phần thuộc A hay thuộc B.

$$A \cup B = \{x \mid x \in A \text{ hay } x \in B\}$$

c) Phép hiệu: Hiệu của hai tập hợp A và B là một tập hợp, ký hiệu A\B, gồm những phần tử thuộc A và không thuộc B.

$$A \setminus B = \{x \mid x \in A \text{ và } x \notin B\}$$

Ví dụ: $A = \{1, 2, 3, 4, 5\}$; $B = \{2, 4, 6, 8\}$

$\Rightarrow A \cap B = \{2, 4\}$; $A \cup B = \{1, 2, 3, 4, 5, 6, 8\}$; $A \setminus B = \{1, 3, 5\}$

1.1.3. Tích Đề-các của các tập hợp

Cho hai tập hợp A và B, tích Đề-các của hai tập A và B là một tập hợp, ký hiệu $A \times B$, được định nghĩa:

$$A \times B = \{(x, y) \mid x \in A \text{ và } y \in B\}$$

Ví dụ: A = {1, 2, 3}; B = {a, b}

$\Rightarrow$ A×B = {(1, a), (1, b), (2, a), (2, b), (3, a), (3, b)}

Cho n tập hợp $A_1, A_2, ..., A_n$, tích Đề-các của n tập hợp $A_1, A_2, ..., A_n$ là một tập hợp được ký hiệu và định nghĩa như sau:

$$\prod_{i=1}^{n} A_i = A_1 \times A_2 \times ... \times A_n = \left\{ (x_1, x_2, ... x_n) \middle| x_i \in A_i, i = \overline{1..n} \right\}$$

Ví dụ: A = {1, 2, 3}; B = {a, b}; C = {α, β}, lúc đó:

A×B×C = {(1, a, α), (1, b, α), (2, a, α), (2, b, α), (3, a, α), (3, b, α),
(1, a, β), (1, b, β), (2, a, β), (2, b, β), (3, a, β), (3, b, β)}

1.2. Quan hệ

1.2.1. Định nghĩa 1

Một quan hệ n ngôi trên một tập hợp $A \neq \phi$ là một tập con của A^n.

Ví dụ: Cho A = {1, 2, 3, 4}

Q_1 = {(1, 1), (1, 2), (2, 1), (2, 2)} là một quan hệ hai ngôi trên A.

Q_1 = {(1, 1, 1), (1, 2, 1), (2, 1, 2)} là một quan hệ ba ngôi trên A.

1.2.2. Định nghĩa 2

Một quan hệ ngôi trên n tập hợp $A_1, A_2, ..., A_n \neq \phi$ là một tập con của tích Đề-các m tập hợp $A_{j1} \times A_{j2} \times ... \times A_{jm}$ với $A_{ji} \in \{A_1, A_2, ..., A_n\}$.

Ví dụ: Cho A = {1, 2, 3, 4}, B = {a, b, c}, C = {α, β, γ, δ}.

Lúc đó: Q = {(1, a, 2, α), (2, b, 3, β), (4, a, 1, γ), (1, b, 4, β)} là một quan hệ bốn ngôi trên A, B, C.

1.3. Ánh xạ

1.3.1. Định nghĩa 1 (ánh xạ)

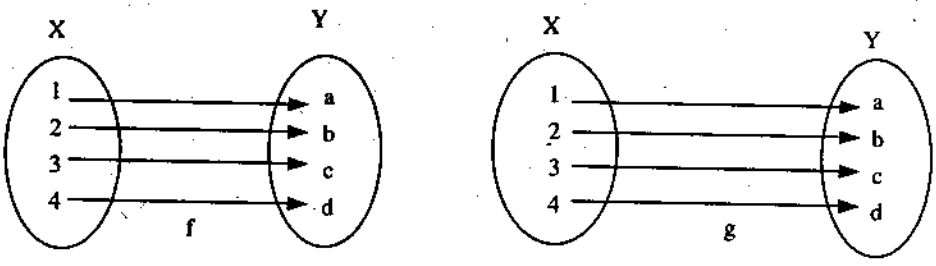
Cho tập hợp X, $Y \neq \phi$. Một ánh xạ f từ X vào Y là một quy luật tương ứng mỗi phần tử $x \in X$ với một phần tử $y \in Y$.

Ký hiệu $f: X \rightarrow Y$

$$x \rightarrow y = f(x)$$

y = f (x) gọi là ảnh của x qua ánh xạ f.

Ví dụ 1: Cho X = {1, 2, 3, 4}, Y = {a, b, c, d} và các phép ứng f, g, h, k sau:



Hình 1.2. Ví dụ về ánh xạ của một tập hợp X đến tập hợp Y

Ta có: f, k là hai ánh xạ từ X vào Y, còn g, h không phải là hai ánh xạ từ X vào Y.

Ví dụ 2: Cho R là tập hợp các số thực và hai phép ứng f và g từ R vào R như sau: $f(x) = x^2 - 3x + 4$ và $g = 1/x \quad \forall x \in R$

Thì f là một ánh xạ, g không là một ánh xạ từ R vào R.

1.3.2. Định nghĩa 2 (ánh xạ hợp)

Cho ba tập hợp X, Y, $Z \neq \phi$ và hai ánh xạ $f: X \to Y$ và $g: Y \to Z$ ta định nghĩa một ánh xạ từ X vào Z như sau:

$gof: X \to Z$

$x \longmapsto z = gof(x) = g[f(x)]$

gof(x) gọi là ánh xạ hợp của f và g, đọc là g tròn f.

Ví dụ: Cho f, g là hai ánh xạ: $f(x) = \sin(x)$, $g(x) = e^x$ thì $gof = e^{\sin(x)}$, $fog(x) = \sin(e^x)$

1.3.3. Ánh xạ thu hẹp

Cho hai tập hợp X, $Y \neq \phi$, f là một ánh xạ X vào Y, $A \subset X$. Ta định nghĩa một ánh xạ: $f_A: A \to Y$ cho bởi $f_A(x) = f(x) \ \forall x \in A$, gọi là ánh xạ thu hẹp của f lên A.

Khi f_A là ánh xạ thu hẹp của X lên A thì ta cũng nói f là ánh xạ mở rộng của f_A lên X.

Ví dụ: Cho $f: R \to R$ như sau: $f(x) = x^2 + 1 \ \forall x \in R$

và $g: N \to R$ như sau: $x(x) = x^2 + 1 \ \forall x \in N$ thì $g = f_N$.

1.3.4. Định nghĩa 4 (ánh xạ chiếu)

Cho hai tập hợp A, $B \neq \phi$, $X = A \times B$, khi đó ánh xạ $\Pi_A: X \to A$ được định nghĩa: $\Pi_A(a, b) = a \ \forall (a, b) \in X$, được gọi là ánh xạ chiếu của X lên A.

Cho $X = \prod_{i=1}^{n} A_{ri}$, $K = \prod_{j=1}^{m} A_{ij}$ với $A_{ij} \in \{A_1, A_2, \ldots, A_n\}$, X, ta định nghĩa ánh xạ chiếu của X lên K như sau: Π_K: $X \to K$ với $\forall t = (a_1, a_2, \ldots, a_n) \in X$, ta có:

$\Pi_K(t) = t(K) = (a_{i1}, a_{i2}, \ldots, a_{im})$ với $a_{ij} \in A_{ij} \subset K$.

Ví dụ: A = {1, 2, 3}, B = {a, b, c}, C = {x, y}; X = A×B×C, K = B×C, lúc ấy :

$\Pi_K(1, a, x) = (a, x)$; $\Pi_K(1, a, y) = (a, y)$, ...

II. CƠ SỞ DỮ LIỆU, HỆ QUẢN TRỊ CƠ SỞ DỮ LIỆU

2.1. Một số ví dụ về cơ sở dữ liệu

Cơ sở dữ liệu là một khái niệm đầu tiên và chủ yếu liên quan tới bất kỳ hệ thống thông tin nào, nhất là đối với một hệ thống thông tin quản lý. Để bắt đầu, trước hết ta xét một vài ví dụ.

Ví dụ 1: Hệ thống bán vé máy bay của một hãng hàng không.

Để lựa chọn chuyến bay, khách hàng có thể tham khảo lịch bay của hãng, thông tin về các chuyến bay được lập theo bảng:

Mã chuyến bay	Loại máy bay	Sân bay đi	Sân bay đến	Ngày bay	Giờ bay	Giờ đến
VN272	ART72	SGN	NHA	30/06/02	7:20	08:05p
VN372	ART72	NHA	SGN	30/06/02	8:45	09:40p
VNA32	A320A	SGN	HAN	30/06/02	12:05	13:35p
VN472	ART72	SGN	DAK	30/06/02	12:25	13:25p
VNB77	BOE77	SGN	HAN	30/06/02	14:05	15:45p
VNB67	BOE67	HAN	SGN	30/06/02	14:05	15:45p
VNT06	TU106	HAN	DAN	30/06/02	16:25	17:35p
...	....	....	....	....	...	...

Mỗi chuyến bay là một mối quan hệ giữa các thuộc tính: Mã chuyến bay; Loại máy bay; Sân bay đi; Sân bay đến; Ngày bay; Giờ bay; giờ đến.

Bảng trên được gọi là một bảng dữ liệu, tuỳ từng đối tượng mà khai thác dữ liệu trong bảng trên theo mục đích của mình, chẳng hạn:

- Một khách hàng muốn bay từ Hà Nội vào Nha Trang, lúc đó anh ta chỉ quan tâm tới các dòng chứa thông tin về các chuyến bay từ Hà Nội tới Nha Trang.
- Trong khi đó đối với một nhà quản lý, ví dụ như Tổng Giám đốc hãng Hàng không Việt Nam, có thể ông ta chỉ cần biết tới số lượng chuyến bay thực hiện trong ngày bằng cách đếm số dòng trên bảng.
- ...

Điều này có nghĩa là các dữ liệu của bảng trên độc lập với các xử lý (về sau này là các chương trình khai thác nó) tác động lên chúng.

Dữ liệu trong bảng trên được tổ chức thành một bảng gồm các cột và các hàng. Các cột được gọi là lược đồ (SCHEME) hay bộ khung của bảng dữ liệu, mỗi cột được gọi là một thuộc tính (attribute) hay một trường, mỗi hàng (dòng) được gọi là một thể hiện (instance), một bộ (tuple) hay một bản ghi (record) của bảng dữ liệu.

Các thao tác thường áp dụng lên bảng dữ liệu là:

- Thêm một bản ghi.
- Xoá một bản ghi.
- Sửa một bản ghi.
- ...

Chú ý: Thao tác sửa một bản ghi thực chất là thực hiện hai thao tác: xóa bản ghi cũ rồi sửa vào bản ghi mới.

Ví dụ 2: Hệ thông tin thương mại

Trong một hệ thống thương mại, thông tin về các công ty thường được lưu trữ dưới dạng như bảng sau:

S#	Tên NCC	Trạng thái	Địa chỉ
S1	Anh Tú	25	HAN
S2	Thái Bình	10	SGN
S3	Tự Cường	30	SGN

Trong khi đó tình hình kinh doanh lại có thể mô tả nhờ vào bảng sau:

S#	P#	Số lượng
S1	P1	300
S1	P2	200
S1	P3	400
S2	P1	300
S2	P2	400
S3	P2	200