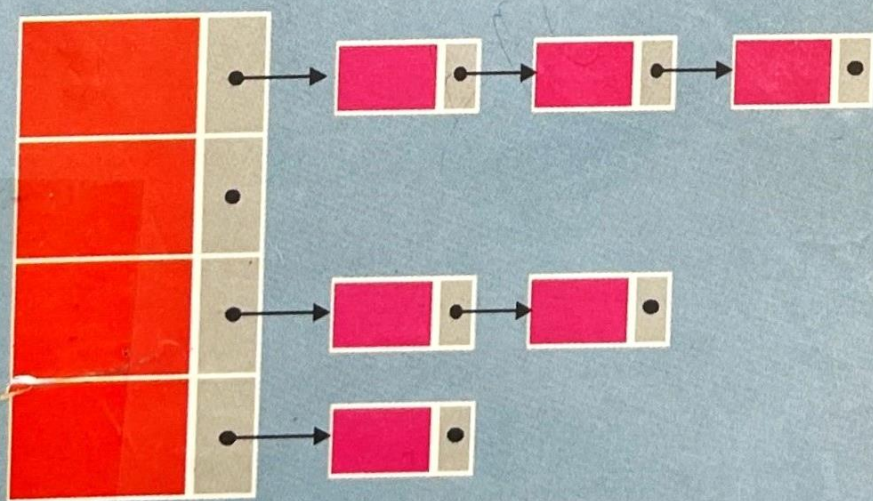


ĐINH MẠNH TƯỜNG

CẤU TRÚC DỮ LIỆU & THUẬT TOÁN



NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT

LỜI NÓI ĐẦU

Cấu trúc dữ liệu (*CTDL*) và thuật toán là môn học đóng vai trò quan trọng trong quá trình đào tạo cử nhân các ngành khoa học máy tính và công nghệ thông tin. Cuốn sách này được hình thành trên cơ sở giáo trình *CTDL* và thuật toán mà chúng tôi đã giảng dạy trong nhiều năm tại khoa Công nghệ thông tin, Đại học *KHTN*, Đại học quốc gia Hà Nội.

Trong quá trình phát triển chương trình, nhất là khi phát triển các hệ mềm lớn, chúng ta cần đến hai dạng biểu diễn dữ liệu: biểu diễn trừu tượng và biểu diễn cụ thể. Trong giai đoạn thiết kế chương trình ta cần sử dụng biểu diễn trừu tượng của dữ liệu, tới giai đoạn cài đặt chương trình ta cần sử dụng biểu diễn cụ thể của dữ liệu. Biểu diễn trừu tượng của dữ liệu được xác định bởi mô hình dữ liệu (*MHDL*), đó là mô hình toán học của các đối tượng dữ liệu cùng với các phép toán thực hiện trên các đối tượng đó. Chẳng hạn, các *MHDL* quan trọng là danh sách, cây, tập hợp,.... Khi chúng ta sử dụng *MHDL* với một số xác định các phép toán nào đó, chúng ta sẽ có một kiểu dữ liệu trừu tượng (Abstract Data Type). Chẳng hạn, một danh sách với hai phép toán xen vào và loại bỏ chỉ được phép thực hiện ở một đầu của danh sách lập thành kiểu dữ liệu trừu tượng (*KDLTT*) ngăn xếp (Stack).

Biểu diễn cụ thể của dữ liệu là biểu diễn xác định cách lưu giữ vật lý của dữ liệu trong bộ nhớ của máy tính. Biểu diễn cụ thể của dữ liệu được xác định bởi các *CTDL*. Các *CTDL* được mô tả trong ngôn ngữ lập trình mà ta sử dụng. Từ biểu diễn trừu tượng, ta có thể chuyển dịch thành các biểu diễn cụ thể khác nhau. Hay nói cách khác, từ các *MHDL*, hoặc từ các *KDLTT*, ta có thể chuyển dịch thành các *CTDL* khác nhau. Chẳng hạn, ta có thể cài đặt danh sách bởi *CTDL* mảng hoặc bởi *CTDL* danh sách liên kết. Khi cài đặt *MHDL* bởi *CTDL* nào đó, thì các phép toán trên mô hình được thực hiện bởi các thao tác cần thiết trên *CTDL* đó.

Vì vậy, trong phần nói về *CTDL*, phương pháp trình bày của chúng tôi như sau: Đầu tiên chúng tôi đặc tả các *MHDL* (hoặc *KDLTT*), sau đó đưa ra các *CTDL* biểu diễn chúng. Với mỗi *CTDL* được sử dụng để biểu diễn *MHDL* (hoặc *KDLTT*), chúng ta đưa ra các thủ tục hoặc hàm thực hiện các phép toán cần thiết. Chúng ta mô tả các *CTDL* và các thủ tục (hoặc hàm) trong ngôn ngữ Pascal. Tuy nhiên, bạn đọc sẽ không khó khăn gì khi chuyển dịch các mô tả đó sang các ngôn ngữ lập trình khác, chẳng hạn C.

Trong phần thuật toán, chúng tôi tập trung vào các kỹ thuật thiết kế thuật toán, chẳng hạn các chiến lược chia để trị, tham ăn, quay lui,...., bởi vì có nắm được các kỹ thuật này chúng ta mới có thể tìm ra thuật toán cho các vấn đề được đặt ra. Các lớp thuật toán giải quyết các vấn đề trong các lĩnh vực khác nhau: khoa học, kỹ thuật, công nghệ, kinh tế,.... là hết sức phong phú. Chúng ta

chỉ trình bày hai lớp thuật toán: các thuật toán sắp xếp và các thuật toán đồ thị. Sở dĩ như vậy, vì đó là các thuật toán được sử dụng rộng rãi nhất, và hơn nữa, các thuật toán này chứa đựng nhiều kỹ thuật quan trọng mà chúng ta cần biết để sử dụng.

Chúng tôi không chú ý nhiều đến phân tích thuật toán. Song trong phần lớn các thuật toán được trình bày chúng tôi cũng đưa ra đánh giá cận trên của thời gian chạy.

Nội dung cuốn sách gồm mười chương:

- Chương 1 trình bày khái niệm thuật toán, đưa ra ký hiệu ω lớn và các phương pháp đánh giá thời gian thực hiện thuật toán bởi ký hiệu ω lớn.
- Chương 2 trình bày các khái niệm cơ bản: kiểu dữ liệu, *CTDL*, *MHDL*, *KDLTT*. Chương này cũng trình bày hệ kiểu của ngôn ngữ Pascal để qua đó chúng ta thấy được phương pháp kiến tạo nên các kiểu dữ liệu của người sử dụng trong các ngôn ngữ lập trình bậc cao.
- Chương 3 trình bày *MHDL* danh sách, các *CTDL* cài đặt danh sách, thủ tục thực hiện các phép toán trên danh sách trong mỗi cách cài đặt. Hai *KDLTT* quan trọng là hàng đợi và ngăn xếp sẽ được nghiên cứu.
- Chương 4 trình bày các *CTDL* biểu diễn cây, và nghiên cứu một lớp cây đặc biệt là cây nhị phân.
- Chương 5 trình bày *MHDL* tập hợp và hai *KDLTT* là từ điển và hàng ưu tiên. Chương này chú trọng đến phương pháp băm, một kỹ thuật quan trọng để cài đặt tập hợp.
- Chương 6 nghiên cứu *MHDL* bảng và các kỹ thuật cài đặt bảng chữ nhật.
- Chương 7 trình bày các *CTDL* ở bộ nhớ ngoài: file băm, file có chỉ số, B-cây.
- Chương 8 trình bày các kỹ thuật thiết kế thuật toán.
- Chương 9 nghiên cứu các thuật toán sắp xếp.
- Chương 10 nghiên cứu các thuật toán đồ thị.

Để đọc cuốn sách này độc giả chỉ cần biết lập trình trên ngôn ngữ Pascal, ngoài ra không đòi hỏi ở độc giả kiến thức chuyên môn nào khác. Ngoài đối tượng chính là sinh viên các ngành Khoa học máy tính và Công nghệ thông tin, cuốn sách còn cần cho tất cả mọi người (kể cả học sinh các lớp chuyên chọn về Tin học) muốn nâng cao kiến thức cơ bản về Tin học.

Cuốn sách chắc chắn không tránh khỏi những thiếu sót, chúng tôi chân thành mong nhận được ý kiến phê bình của độc giả. Thư góp ý xin gửi về Nhà xuất bản Khoa học và Kỹ thuật - 70 Trần Hưng Đạo Hà Nội.

Tác giả

Chương 1

THUẬT TOÁN VÀ PHÂN TÍCH THUẬT TOÁN

1.1. THUẬT TOÁN

1.1.1. Khái niệm thuật toán

Thuật toán (algorithm) là một trong những khái niệm quan trọng nhất trong tin học. Thuật ngữ thuật toán xuất phát từ nhà toán học Ả-rập Abu Ja'far Mohammed ibn Musa al Khowarizmi (khoảng năm 825). Tuy nhiên lúc bấy giờ và trong nhiều thế kỷ sau, nó không mang nội dung như ngày nay chúng ta quan niệm. Thuật toán nổi tiếng nhất, có từ thời cổ Hy Lạp là thuật toán Euclid, thuật toán tìm ước chung lớn nhất của hai số nguyên. Có thể mô tả thuật toán này như sau :

Thuật toán Euclid

Input : m, n nguyên dương

Output : g , ước chung lớn nhất của m và n .

Phương pháp :

Bước 1 : Tìm r , phần dư của phép chia m cho n

Bước 2 : Nếu $r = 0$, thì $g \leftarrow n$ (gán giá trị của n cho g) và dừng lại. Trong trường hợp ngược lại ($r \neq 0$), thì $m \leftarrow n, n \leftarrow r$ và quay lại bước 1.

Chúng ta có thể quan niệm các bước cần thực hiện để làm một món ăn, được mô tả trong các sách dạy chế biến món ăn, là một thuật toán. Cũng có thể xem các bước cần tiến hành để gấp đồ chơi bằng giấy, được trình bày trong sách dạy gấp đồ chơi bằng giấy, là thuật toán. Phương pháp thực hiện phép cộng, nhân các số nguyên, chúng ta đã học ở cấp I cũng là các thuật toán.

MỤC LỤC

Lời nói đầu	3
Chương 1. Thuật toán và phân tích thuật toán	5
1.1. Thuật toán	5
1.2. Phân tích thuật toán	10
Chương 2. Kiểu dữ liệu, cấu trúc dữ liệu và mô hình dữ liệu	23
2.1. Biểu diễn dữ liệu	23
2.2. Kiểu dữ liệu và cấu trúc dữ liệu	24
2.3. Hệ kiểu của ngôn ngữ Pascal	27
2.4. Mô hình dữ liệu và kiểu dữ liệu trừu tượng	31
Chương 3. Danh sách	36
3.1. Danh sách	36
3.2. Cài đặt danh sách bởi mảng	38
3.3. Tìm kiếm trên danh sách	42
3.4. Cấu trúc dữ liệu danh sách liên kết	46
3.5. Các dạng danh sách liên kết khác	53
3.6. Ứng dụng danh sách: Các phép tính số học trên đa thức	58
3.7. Ngăn xếp	63
3.8. Giá trị của một biểu thức	69
3.9. Hàng	74
Chương 4. Cây	83
4.1. Cây và các khái niệm về cây	83
4.2. Các phép toán trên cây	87
4.3. Cài đặt cây	93
4.4. Cây nhị phân	103
4.5. Cây tìm kiếm nhị phân	108
4.6. Thời gian thực hiện các phép toán trên cây tìm kiếm nhị phân	117
4.7. Cây cân bằng	119
4.8. Thời gian thực hiện các phép toán trên cây cân bằng	131
Chương 5. Tập hợp	134
5.1. Tập hợp và các phép toán trên tập hợp	134
5.2. Cài đặt tập hợp	138
5.3. Từ điển	144
5.4. Cấu trúc dữ liệu bảng băm. Cài đặt từ điển bởi bảng băm	146

5.5. Phân tích và đánh giá các phương pháp băm	156
5.6. Hàng ưu tiên	160
5.7. Cây thứ tự bộ phận và cài đặt ưu tiên bởi cây thứ tự bộ phận	161
Chương 6. Bảng	170
6.1. Kiểu dữ liệu trừu tượng bảng	170
6.2. Cài đặt bảng	171
6.3. Bảng chữ nhật	174
6.4. Trò chơi đời sống	180
Chương 7. Các cấu trúc dữ liệu ở bộ nhớ ngoài	191
7.1. Mô hình tổ chức dữ liệu ở bộ nhớ ngoài	191
7.2. File băm	194
7.3. File có chỉ số (Indexed File)	196
7.4. B-Cây	199
Chương 8. Các chiến lược thiết kế thuật toán	207
8.1. Chia - để - trị	207
8.2. Phương pháp tham ăn	214
8.3. Quay lui	221
8.4. Nhánh và cận	227
8.5. Quy hoạch động	232
Chương 9. Sắp xếp	241
9.1. Vấn đề sắp xếp	241
9.2. Các phương pháp sắp xếp đơn giản	242
9.3. Sắp xếp nhanh	247
9.4. Sắp xếp sử dụng cây thứ tự bộ phận	253
Chương 10. Các thuật toán trên đồ thị	256
10.1. Một số khái niệm cơ bản	256
10.2. Biểu diễn đồ thị	258
10.3. Đi qua đồ thị	261
10.4. Thành phần liên thông và thành phần liên thông mạnh	269
10.5. Đồ thị định hướng không có chu trình. Sắp xếp topo	273
10.6. Đường đi ngắn nhất	275
10.7. Cây bao trùm ngắn nhất	285
Tài liệu tham khảo	295