TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÀI TẬP LỚN/ĐỒ ÁN CUỐI KÌ MÔN**

**Phân Tích Và Thiết Kế Giải Thuật**

**Final Report**

*Người hướng dẫn*: **TS NGUYỄN CHÍ THIỆN**

*Người thực hiện*: **ĐINH HỒNG HÀ – 518H0171**

**PHÙNG QUỐC ĐẠT – 518H0482**

Lớp **: 18H50208**

Khoá  **: 22**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2020**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÀI TẬP LỚN/ĐỒ ÁN CUỐI KÌ MÔN**

**Phân Tích Và Thiết Kế Giải Thuật**

**Final Report**

*Người hướng dẫn*: **TS NGUYỄN CHÍ THIỆN**

*Người thực hiện*: **ĐINH HỒNG HÀ – 518H0171**

**PHÙNG QUỐC ĐẠT – 518H0482**

Lớp **: 18H50208**

Khoá  **: 22**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2020**

LỜI CẢM ƠN

Em cảm ơn thầy rất nhiều vì đã rất nhiệt tình giúp đỡ chỉ bảo với bọn em trong quá trình học tập. Những bài học thầy giảng cho bọn em thật bổ ích , thầy đã tận tình chỉ những lỗi cho bọn em biết để sửa chửa cảm ơn thầy vì điều đó. Em xin chúc thầy luôn thành công trong công việc và cuộc sống và thực hiện tâm huyết của mình với công việc.

**ĐỒ ÁN ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là sản phẩm đồ án của riêng tôi / chúng tôi và được sự hướng dẫn của TS Nguyễn Chí Thiện;. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình.** Trường đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày 11 tháng 10 năm 2020*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

*Đinh Hồng Hà*

*Phùng Quốc Đạt*

PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN

**Phần xác nhận của GV hướng dẫn**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

**Phần đánh giá của GV chấm bài**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

TÓM TẮT

* Chúng ta sẽ đi nghiên cứu trong 2 phần :
  + Phần 1: Giới thiệu về phân tích và thiết kế giải thuật
  + Phần 2 : Chúng ta sẽ làm về phần đã học , được học và tự học
* Về phần đã học và được học , chúng ta sẽ nghiên cứu và giải quyết về các kỹ thuật:
  + Brute-force
  + Divide-and-conquer
  + Greedy Algorithms
  + Dynamic Programming.
* Về phần tự học , chúng ta sẽ giới thiệu và hiểu về kỹ thuật Backtracking.

MỤC LỤC

[LỜI CẢM ƠN i](#_Toc54214577)

[PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN iii](#_Toc54214578)

[TÓM TẮT iv](#_Toc54214579)

[MỤC LỤC 1](#_Toc54214580)

[CHƯƠNG 1 – GIỚI THIỆU 4](#_Toc54214581)

[1.1 Thuật toán là gì 4](#_Toc54214582)

[1.2 Các nguyên tắc cơ bản giải quyết vấn đề về thuật toán 4](#_Toc54214583)

[1.2.1 Hiểu vấn đề 4](#_Toc54214584)

[1.2.2 Xác minh khả năng của thiết bị tính toán 4](#_Toc54214585)

[1.2.3 Lựa chọn giữa giải quyết vấn đề chính xác và gần đúng 5](#_Toc54214586)

[1.2.4 Kỹ thuật thiết kế thuật toán 5](#_Toc54214587)

[1.2.5 Thiết kế thuật toán và cấu trúc dữ liệu 5](#_Toc54214588)

[1.2.6 Phương pháp chỉ định một thuật toán 6](#_Toc54214589)

[1.2.7 Chứng minh tính đúng đắn của thuật toán 6](#_Toc54214590)

[1.2.8 Phân tích một thuật toán 7](#_Toc54214591)

[1.3 Important Problem Types 7](#_Toc54214592)

[1.3.1 Sorting 8](#_Toc54214593)

[1.3.2 Searching 8](#_Toc54214594)

[1.3.3 String Processing 8](#_Toc54214595)

[1.3.4 Graph Problems 9](#_Toc54214596)

[1.3.5 Geometric Problems 9](#_Toc54214597)

[1.3.6 Numerical Problems 9](#_Toc54214598)

[1.4 Cấu trúc dữ liệu cơ bản 10](#_Toc54214599)

[1.4.1 Cấu trúc dữ liệu tuyến tính 10](#_Toc54214600)

[1.4.2 Đồ thị 11](#_Toc54214601)

[1.4.3 Trees 11](#_Toc54214602)

[CHƯƠNG 2 – PHẦN ĐÃ HỌC 13](#_Toc54214603)

[2.1 Brute Force 13](#_Toc54214604)

[2.1.1 Khái niệm: 13](#_Toc54214605)

[2.1.2 Selection Sort và Bubble Sort: 13](#_Toc54214606)

[2.1.2.1 Selection Sort: 13](#_Toc54214607)

[2.1.2.2 Bubble Sort: 14](#_Toc54214608)

[2.1.3 Search: 15](#_Toc54214609)

[2.1.3.1 Sequential Search. 15](#_Toc54214610)

[2.1.4 Kết luận: 16](#_Toc54214611)

[2.2 Divide and Conquer 18](#_Toc54214612)

[2.2.1 Merge Sort 18](#_Toc54214613)

[2.2.2 Quick Sort 20](#_Toc54214614)

[2.2.3 Binary Tree Traversals 23](#_Toc54214615)

[2.2.4 Kết luận: 24](#_Toc54214616)

[2.3 Greedy Algorithms 24](#_Toc54214617)

[2.3.1 Thuật toán Prim 25](#_Toc54214618)

[2.3.2 Kruskal’s Algorithm 26](#_Toc54214619)

[2.3.3 Dijkstra’s Algorithm 27](#_Toc54214620)

[2.3.4 Huffman Trees and Codes 28](#_Toc54214621)

[2.3.5 Kết luận 29](#_Toc54214622)

[2.4 Dynamic Programming 30](#_Toc54214623)

[2.4.1 Three Basic Examples 30](#_Toc54214624)

[2.4.1.1 Coin Row Problem 30](#_Toc54214625)

[2.4.2 Change Making Problem: 32](#_Toc54214626)

[2.4.3 Coin Collecting Problem: 33](#_Toc54214627)

[2.4.2 Kết Luận 35](#_Toc54214628)

[CHƯƠNG 3 – PHẦN TỰ HỌC 37](#_Toc54214629)

[3.1 Backtraking 37](#_Toc54214630)

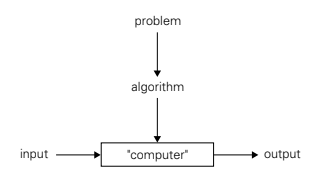
CHƯƠNG 1 – GIỚI THIỆU

Tại sao bạn cần nghiên cứu các thuật toán? Nếu bạn định trở thành một chuyên gia máy tính, có cả lý do thực tế và lý thuyết để nghiên cứu nhịp điệu thuật toán. Từ quan điểm thực tế, bạn phải biết một bộ tiêu chuẩn các thuật toán quan trọng từ các lĩnh vực máy tính khác nhau; Ngoài ra, bạn sẽ có thể thiết kế các thuật toán mới và phân tích hiệu quả của chúng. Từ lập trường lý thuyết- quan điểm, việc nghiên cứu các thuật toán, đôi khi được gọi là thuật toán, đã được công nhận là nền tảng của khoa học máy tính.

* 1. Thuật toán là gì

Thuật toán là một chuỗi các hướng dẫn rõ ràng để giải quyết một vấn đề, tức là để có được đầu ra bắt buộc cho bất kỳ đầu vào hợp pháp nào trong một khoảng thời gian hữu hạn.

Định nghĩa này có thể được minh họa bằng một sơ đồ đơn giản :



1.2 Các nguyên tắc cơ bản giải quyết vấn đề về thuật toán

1.2.1 Hiểu vấn đề

Từ góc độ thực tế, điều đầu tiên bạn cần làm trước khi thiết kế một thuật toán là hiểu hoàn toàn vấn đề được đưa ra. Đọc kỹ mô tả vấn đề và đặt câu hỏi nếu bạn có bất kỳ nghi ngờ nào về vấn đề, làm một vài ví dụ nhỏ bằng tay, suy nghĩ về những trường hợp đặc biệt, và đặt câu hỏi nếu cần.

1.2.2 Xác minh khả năng của thiết bị tính toán

Khi bạn hoàn toàn hiểu một vấn đề, bạn cần xác định khả năng của thiết bị tính toán mà thuật toán hướng tới. Phần lớn các thuật toán được sử dụng ngày nay vẫn được lập trình chặt chẽ cho máy tính .

1.2.3 Lựa chọn giữa giải quyết vấn đề chính xác và gần đúng

Quyết định chính tiếp theo là lựa chọn giữa giải quyết vấn đề chính xác hoặc giải quyết nó một cách gần đúng. Trong trường hợp trước đây, một thuật toán được gọi là một thuật toán chính xác; trong trường hợp thứ hai, một thuật toán được gọi là thuật toán xấp xỉ. Tại sao một người sẽ chọn một thuật toán xấp xỉ? Đầu tiên, có những vấn đề quan trọng không thể được giải quyết chính xác đối với hầu hết các trường hợp của chúng; các ví dụ bao gồm trích xuất căn bậc hai, giải câu lệnh phi tuyến và đánh giá định nghĩa

tích phân inite. Thứ hai, các thuật toán có sẵn để giải quyết chính xác một vấn đề có thể chậm đến mức không thể chấp nhận được vì độ phức tạp nội tại của vấn đề. Điều này đặc biệt xảy ra đối với nhiều vấn đề liên quan đến một số lượng rất lớn các lựa chọn; bạn sẽ xem ví dụ về những vấn đề khó như vậy trong Chương 3, 11 và 12. Thứ ba, thuật toán ap-proximation có thể là một phần của thuật toán phức tạp hơn giúp giải quyết chính xác vấn đề.

1.2.4 Kỹ thuật thiết kế thuật toán

Kỹ thuật thiết kế thuật toán (hay “chiến lược” hoặc “mô hình”) là một cách tiếp cận chung để giải quyết các vấn đề theo thuật toán có thể áp dụng cho nhiều vấn đề từ các lĩnh vực máy tính khác nhau.

1.2.5 Thiết kế thuật toán và cấu trúc dữ liệu

Mặc dù các kỹ thuật thiết kế thuật toán cung cấp một tập hợp các ứng dụng chung mạnh mẽ hướng dẫn giải quyết vấn đề theo thuật toán, thiết kế một thuật toán cho một vấn đề cụ thể vẫn có thể là một nhiệm vụ đầy thách thức. Một số kỹ thuật thiết kế có thể đơn giản là không thể áp dụng cho vấn đề được đề cập. Đôi khi, một số kỹ thuật cần được kết hợp với nhau và có những thuật toán khó xác định là ứng dụng của các kỹ thuật thiết kế đã biết. Ngay cả khi một kỹ thuật thiết kế cụ thể là phù hợp, việc tạo ra một thuật toán thường đòi hỏi sự khéo léo không hề nhỏ của người thiết kế thuật toán. Với thực hành, cả hai nhiệm vụ lựa chọn trong số các kỹ thuật chung và áp dụng chúng — trở nên dễ dàng hơn, nhưng chúng hiếm khi dễ dàng.

1.2.6 Phương pháp chỉ định một thuật toán

Sử dụng một ngôn ngữ tự nhiên có một sức hấp dẫn rõ ràng; tuy nhiên, đặc tính xung quanh vốn có của bất kỳ ngôn ngữ tự nhiên nào khiến cho việc mô tả các thuật toán ngắn gọn và rõ ràng trở nên khó khăn một cách đáng ngạc nhiên. Tuy nhiên, có thể làm được điều này là một kỹ năng quan trọng mà bạn nên cố gắng phát triển trong quá trình học thuật toán.

Mã giả là một hỗn hợp của ngôn ngữ tự nhiên và các cấu trúc giống như ngôn ngữ lập trình. Mã giả thường chính xác hơn ngôn ngữ tự nhiên và việc sử dụng nó thường mang lại các mô tả thuật toán ngắn gọn hơn.

Hiện đại của máy tính vẫn chưa đạt đến mức mô tả thuật toán — có thể là bằng ngôn ngữ tự nhiên hoặc mã giả — có thể được đưa trực tiếp vào máy tính điện tử. Thay vào đó, nó cần được chuyển đổi thành chương trình máy tính được viết bằng một ngôn ngữ máy tính cụ thể. Chúng ta có thể xem một chương trình như vậy là một cách khác để chỉ định thuật toán, Mặc dù nên coi nó như là cách triển khai của thuật toán.

1.2.7 Chứng minh tính đúng đắn của thuật toán

Khi một thuật toán đã được chỉ định, bạn phải chứng minh tính đúng đắn của nó. Đó là, bạn phải chứng minh rằng thuật toán mang lại kết quả bắt buộc cho mọi đầu vào hợp pháp trong một khoảng thời gian hữu hạn.

Đối với một số thuật toán, một bằng chứng về tính đúng đắn là khá dễ dàng; đối với những người khác, nó có thể khá phức tạp. Một kỹ thuật phổ biến để chứng minh tính đúng đắn là sử dụng quy nạp toán học vì các phép lặp của thuật toán cung cấp một chuỗi các bước tự nhiên cần thiết cho các chứng minh đó. Điều đáng nói là mặc dù việc theo dõi hiệu suất của thuật toán cho một số đầu vào cụ thể có thể là một hoạt động rất đáng giá, nhưng nó không thể chứng minh tính đúng đắn của thuật toán một cách rõ ràng. Nhưng để chỉ ra rằng một thuật toán không chính xác, bạn chỉ cần một trường hợp đầu vào của nó mà thuật toán không thành công.

Khái niệm về tính đúng cho các thuật toán xấp xỉ ít đơn giản hơn so với các thuật toán chính xác. Đối với một thuật toán gần đúng, chúng tôi thường muốn có thể chỉ ra rằng lỗi do thuật toán tạo ra không vượt quá một giới hạn được xác định trước.

1.2.8 Phân tích một thuật toán

Chúng tôi thường muốn các thuật toán của mình có một số phẩm chất. Sau sự đúng đắn, cho đến nay, điều quan trọng nhất là hiệu quả. Trên thực tế, có hai loại hiệu quả của thuật toán: hiệu quả về thời gian, cho biết tốc độ chạy của thuật toán và hiệu quả về không gian, cho biết nó sử dụng thêm bao nhiêu bộ nhớ.

Một đặc tính mong muốn khác của một thuật toán là tính đơn giản. Không giống như hiệu quả, có thể được xác định chính xác và nghiên cứu bằng sự nghiêm ngặt của toán học, tính đơn giản, giống như vẻ đẹp, ở một mức độ đáng kể trong mắt người xem.

Tuy nhiên, một đặc tính mong muốn khác của một thuật toán là tính tổng quát. Trên thực tế, có hai vấn đề ở đây: tính tổng quát của vấn đề mà thuật toán giải quyết và tập hợp các đầu vào mà nó chấp nhận. Về vấn đề đầu tiên, hãy lưu ý rằng đôi khi việc thiết kế một thuật toán cho một vấn đề được đặt ra theo các thuật ngữ chung hơn lại dễ dàng hơn.

1.3 Important Problem Types

Trong vô số các vấn đề mà người ta gặp phải trong máy tính, có một số lĩnh vực thu hút sự chú ý đặc biệt của các nhà nghiên cứu. Nhìn chung, sự quan tâm của họ đã được thúc đẩy bởi tầm quan trọng thực tế của vấn đề hoặc bởi một số đặc điểm cụ thể khiến vấn đề trở thành một chủ đề nghiên cứu thú vị; may mắn thay, hai động lực thúc đẩy này củng cố lẫn nhau trong hầu hết các trường hợp.

Trong phần này, chúng tôi sẽ giới thiệu các dạng vấn đề quan trọng nhất:

+ Sorting

+ Searching

+ String processing

+ Graph problems

+ Combinatorial problems

+ Geometric problems

+ Numerical problems

1.3.1 Sorting

Vấn đề sắp xếp là sắp xếp lại các mục của một danh sách nhất định theo thứ tự không giảm. Tất nhiên, để vấn đề này có ý nghĩa, bản chất của các mục danh sách phải cho phép một thứ tự như vậy. (Các nhà toán học sẽ nói rằng phải tồn tại một quan hệ của tổng thứ tự.) Như một vấn đề thực tế, chúng ta thường cần sắp xếp danh sách các số, các ký tự từ bảng chữ cái, các chuỗi ký tự và, quan trọng nhất,hồ sơ tương tự như hồ sơ do trường học duy trì về học sinh của họ, thư viện về tài sản của họ và công ty về nhân viên của họ. Trong trường hợp hồ sơ, chúng ta cần chọn một phần thông tin để hướng dẫn sắp xếp.

1.3.2 Searching

Bài toán tìm kiếm đề cập đến việc tìm một giá trị nhất định, được gọi là khóa tìm kiếm, trong một tập hợp nhất định (hoặc tập hợp nhiều, cho phép một số phần tử có cùng giá trị). Có rất nhiều thuật toán tìm kiếm để lựa chọn. Chúng bao gồm từ tìm kiếm tuần tự đơn giản đến tìm kiếm nhị phân hiệu quả ngoạn mục nhưng có giới hạn và các thuật toán dựa trên việc biểu diễn tập hợp cơ bản ở một dạng khác có lợi hơn cho việc tìm kiếm. Các thuật toán thứ hai có tầm quan trọng đặc biệt đối với các ứng dụng trong thế giới thực vì chúng không thể thiếu để lưu trữ và truy xuất thông tin từ cơ sở dữ liệu lớn.

1.3.3 String Processing

Trong những thập kỷ gần đây, sự gia tăng nhanh chóng của các ứng dụng xử lý dữ liệu phi số đã tăng cường sự quan tâm của các nhà nghiên cứu và các nhà thực hành máy tính trong các thuật toán xử lý chuỗi. Chuỗi là một chuỗi các ký tự từ một bảng chữ cái.Các chuỗi được quan tâm đặc biệt là các chuỗi văn bản, bao gồm các chữ cái, số và các ký tự đặc biệt; chuỗi bit, bao gồm số không và số một; và trình tự gen, có thể được mô hình hóa bởi các chuỗi ký tự từ bảng chữ cái bốn ký tự {A, C, G, T}. Tuy nhiên, cần phải chỉ ra rằng các thuật toán xử lý chuỗi có quan trọng đối với khoa học máy tính trong một thời gian dài cùng với máy tính ngôn ngữ và các vấn đề biên dịch.

1.3.4 Graph Problems

Một trong những lĩnh vực lâu đời nhất và thú vị nhất trong thuật toán là thuật toán đồ thị. Một cách không chính thức, đồ thị có thể được coi như một tập hợp các điểm được gọi là đỉnh, một số điểm được nối với nhau bằng các đoạn thẳng được gọi là các cạnh. Đồ thị là một đối tượng thú vị để nghiên cứu, vì lý do cả lý thuyết và thực tiễn. Đồ thị có thể được sử dụng để lập mô hình cho nhiều ứng dụng, bao gồm giao thông vận tải, truyền thông, mạng xã hội và kinh tế, lập kế hoạch dự án và trò chơi. Nghiên cứu các khía cạnh kỹ thuật và xã hội khác nhau của Internet nói riêng là một trong những lĩnh vực nghiên cứu tích cực hiện nay liên quan đến các nhà khoa học máy tính, nhà kinh tế học và nhà khoa học xã hội (xem, ví dụ: [Eas10]).

Các thuật toán đồ thị cơ bản bao gồm Graph Travelsal Algorithms , Shortest Path Algorithms , Topological Sorting For Graphs With Directed Edges.

1.3.5 Geometric Problems

Các thuật toán hình học xử lý các đối tượng hình học như điểm, đường thẳng và đa giác. Các thuật toán chỉ cho hai vấn đề kinh điển của tính toán hình học: bài toán cặp gần nhất và bài toán lồi cầu.

1.3.6 Numerical Problems

Các vấn đề số, một lĩnh vực ứng dụng đặc biệt lớn khác, là những vấn đề liên quan đến các đối tượng toán học có tính chất liên tục: giải phương trình và hệ thống phương trình, tính toán tích phân xác định, đánh giá các hàm, v.v. Phần lớn các vấn đề toán học như vậy chỉ có thể được giải quyết một cách gần đúng. Một khó khăn chính khác bắt nguồn từ thực tế là những vấn đề như vậy thường yêu cầu thao tác các số thực, chỉ có thể được biểu diễn trong máy tính xấp xỉ. Hơn nữa, một số lượng lớn các phép toán số học được thực hiện trên các con số được đại diện gần đúng có thể dẫn đến sự tích lũy của sai số làm tròn đến một điểm mà nó có thể làm sai lệch đáng kể đầu ra được tạo ra bởi một thuật toán âm thanh.

1.4 Cấu trúc dữ liệu cơ bản

Vì phần lớn các thuật toán quan tâm hoạt động dựa trên dữ liệu, các cách tổ chức dữ liệu cụ thể đóng một vai trò quan trọng trong việc thiết kế và phân tích các thuật toán. Cấu trúc dữ liệu có thể được định nghĩa là một sơ đồ cụ thể để tổ chức các mục dữ liệu liên quan. Bản chất của các mục dữ liệu được quyết định bởi vấn đề hiện tại; họ có thể phạm vi từ các kiểu dữ liệu cơ bản đến cấu trúc dữ liệu. Có một số cấu trúc dữ liệu đã được chứng minh là đặc biệt quan trọng đối với các thuật toán máy tính.

1.4.1 Cấu trúc dữ liệu tuyến tính

Hai cấu trúc dữ liệu cơ bản quan trọng nhất là mảng và danh sách liên kết. Mảng (một chiều) là một chuỗi gồm n mục có cùng kiểu dữ liệu được lưu trữ liền kề trong bộ nhớ máy tính và có thể truy cập được bằng cách chỉ định một giá trị của chỉ mục của mảng

Mảng được sử dụng để triển khai nhiều cấu trúc dữ liệu khác nhau. Nổi bật trong số đó là chuỗi, một chuỗi các ký tự từ bảng chữ cái được kết thúc bằng một ký tự đặc biệt cho biết kết thúc của chuỗi. Các chuỗi bao gồm các số không và một được gọi là chuỗi nhị phân hoặc chuỗi bit. Chuỗi là không thể thiếu để xử lý dữ liệu dạng văn bản, xác định ngôn ngữ máy tính và biên dịch các chương trình được viết trong đó và nghiên cứu các mô hình tính toán trừu tượng. Các hoạt động chúng ta thường thực hiện trên chuỗi khác với các hoạt động chúng ta thường thực hiện trên các mảng khác (ví dụ, mảng số). Chúng bao gồm tính toán độ dài chuỗi, so sánh hai chuỗi để xác định chuỗi nào đứng trước chuỗi kia theo thứ tự từ vựng (tức là theo thứ tự bảng chữ cái) và nối hai chuỗi (tạo thành một chuỗi từ hai chuỗi đã cho bằng cách nối chuỗi thứ hai vào cuối chuỗi thứ nhất).

Danh sách liên kết là một chuỗi gồm không hoặc nhiều phần tử được gọi là các nút, mỗi phần tử chứa hai loại thông tin: một số dữ liệu và một hoặc nhiều liên kết được gọi là con trỏ đến các nút khác của danh sách liên kết. (Một con trỏ đặc biệt được gọi là "null" được sử dụng để chỉ ra sự vắng mặt của nút kế nhiệm.) Trong danh sách được liên kết đơn lẻ, mỗi nút ngoại trừ nút cuối cùng chứa một con trỏ đến phần tử tiếp theo

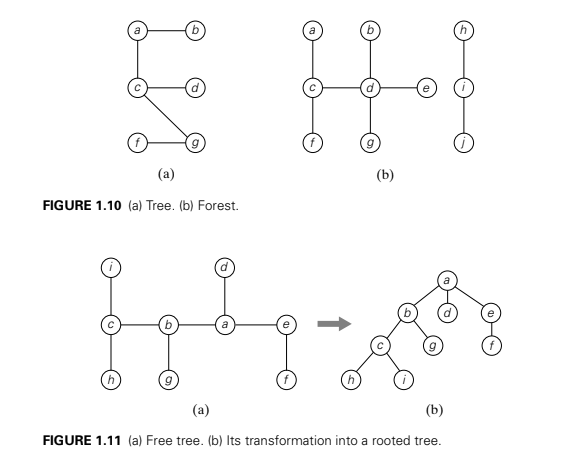
1.4.2 Đồ thị

Như chúng ta đã đề cập trong phần trước, một biểu đồ được coi là một tập hợp các điểm trong mặt phẳng được gọi là “đỉnh” hoặc “nút”, một số trong số chúng được nối với nhau bằng các đoạn thẳng được gọi là “cạnh” hoặc “cung”. Về mặt hình thức, một đồ thị G = V, E được xác định bởi một cặp hai tập hợp: tập hợp hữu hạn vô hạn V gồm các phần được gọi là đỉnh và tập E gồm các cặp phần này được gọi là các cạnh. Nếu các cặp đỉnh này không có thứ tự, tức là một cặp đỉnh (u, v) giống với cặp (v, u), chúng ta nói rằng các đỉnh u và v kề nhau và chúng được nối với nhau bởi cạnh vô hướng (u, v). Chúng ta gọi các đỉnh u và v là điểm cuối của cạnh (u, v) và nói rằng u và v là sự cố đối với cạnh này; chúng tôi cũng nói rằng cạnh (u, v) là sự cố tới các điểm cuối u và v của nó. Một đồ thị G được gọi là vô hướng nếu mọi cạnh trong nó là vô hướng.

1.4.3 Trees

Cây (chính xác hơn, cây tự do) là một đồ thị mạch hở liên thông (Hình 1.10a).

Một đồ thị không có chu trình nhưng không nhất thiết phải liên thông được gọi là rừng: mỗi thành phần liên thông của nó là một cây (Hình 1.10b).



CHƯƠNG 2 – PHẦN ĐÃ HỌC

*Tổng quan*: Những khái niệm, thuật toán đã được thầy chỉ dạy, hướng dẫn tại lớp;

2.1 Brute Force

2.1.1 Khái niệm:

- Brute Force với ý tưởng sử dụng “Force” với cách tiếp cận vấn để theo kiểu “Brute”.

- Những thuật toán dạng Brute Force tuy không thực sự “Thông minh” và hiệu quả về mặt hiệu suất tuy nhiên chúng có thể áp dụng được cho rất nhiều vấn đề.

- Đối với một số vấn đề như sort, matching .., Brute Force lại là cách tiếp cận hợp lý.

- Đối với các vấn đề nhỏ, Brute Force vẫn thực sự hữu ích để giải quyết.

2.1.2 Selection Sort và Bubble Sort:

- Nếu hỏi một sinh viên về thuật toán sắp xếp nào hiệu quả nhất thì có thể họ sẽ không có câu trả lời, tuy nhiên nếu hỏi sinh viên sử dụng thuật toán nào thì chắc chắn sẽ là 2 thuật toán dưới đây.

- Trong phần này, chúng ta sẽ thử tiếp cần Brute Force thông qua 2 thuật toán sắp xếp Select Sort và Bubble sort.

2.1.2.1 Selection Sort:

- Thuật toán với ý tưởng sẽ tìm phần tử nhỏ nhất từ phần tử thứ 0 tới n-1 và hoán đổi với phần tử thứ 0, sau đó tiếp tục tìm phần tử nhỏ nhất cho vị trí tiếp theo( vị trí thứ 1) từ vị trí đó tới phần tử cuối cùng và lặp lại cho tới khi tìm được phần tử nhỏ nhất cho vị trí thứ n-2.

- Mã giả:

1. selectionSort(A[0,…n-1]):
2. // Purpose: Sắp xếp mảng theo thứ tự tăng dần.
3. // Input: Một mảng có kích thước n phần tử
4. //Output: Mảng đầu vào đã được sắp xếp theo thứ tự tăng dần
5. For I = 0 to n-2:
6. Min =i
7. For j = i+1 to n-1:
8. If ( A[j]<A[min]): min = j
9. Swap (A[min],A[i])
10. Endfor
11. Endfor
12. Phép toán cơ bản nhất:
    * Phép so sánh tại vị trí 8 A[j]<A[min]
13. Độ phức tạp thuật toán:
    * T(n)=1.(n-1 + n-2 + n-3 + …. + 1) =
    * T(n) O()

2.1.2.2 Bubble Sort:

- Khác với thuật toán Selection Sort, Bubble Sort được xây dựng bằng cách thực hiện “chuyền” lần lượt phần tử có giá trị lớn nhất từ vị trí đầu tiên cho tới ví trí cuối cùng của lần lặp. Thuật toán sẽ thực hiện so sánh lần lượt 2 cặp số liên tiếp j và j+1.

- Mã giả:

1. selectionSort(A[0,…n-1]):
2. // Purpose: Sắp xếp mảng theo thứ tự tăng dần.
3. // Input: Một mảng có kích thước n phần tử
4. //Output: Mảng đầu vào đã được sắp xếp theo thứ tự tăng dần
5. For i= 0 to n-2 :
6. For j=0 to n-2-i:
7. If( A[j+1] <A[j]) :
8. Swap(A[j+1],A[j])
9. Endfor
10. Endfor
11. Phép toán cơ bản nhất:
    * Phép so sánh tại dòng 7 A[j+1]<A[j]
12. Độ phức tạp thuật toán:
    * T(n)=1.(n-1 + n-2 + n-3 + …. + 1) =
    * T(n) O()

2.1.3 Search:

- Phần trước chúng ta đã phân tích một số thuật toán Sort theo chiến lược Brute Force, tiếp theo chúng ta sẽ tìm hiểu một thuật toán tìm kiếm cũng rất phổ biến đó là Sequential Search.

2.1.3.1 Sequential Search.

- Thuật toán tìm kiếm tuần tự với ý tưởng như tên gọi của nó, tìm kiếm lần lượt từng phần tử để tìm ra phần tử key. Tuy nhiên, hôm nay chúng ta sẽ thực hiện một bản nâng cấp của thuật toán tìm kiếm tuần tự.

- Thuật toán tìm kiếm tuần tự 2 sẽ thêm khóa key vào cuối mảng cần tìm, điều này giúp chúng ta không cần phải kiểm tra khi nào duyệt hết phần tử.

- Mã giả:

1. sequentialSearch(A[0,…n],k):
2. // Purpose:Tìm kiếm key k trong mảng một cách tuần tự.
3. // Input: Một mảng cần duyệt có kích thước n phần tử A[0,n-1] và giá trị cần tìm k
4. //Output: index của phần tử cần tìm trong A hoặc -1 nếu không tìm thấy.
5. A[n]=k
6. i=0
7. While A[i]≠k do:
8. i=i+1
9. endwhile
10. If i<n return i
11. Return -1
12. Phép toán cơ bản nhất:
    * Phép so sánh tại dòng 7 A[i]≠k
13. Độ phức tạp thuật toán:
    * T(n)=1.n = n
    * T(n) O(n)

2.1.4 Kết luận:

- Sau khi tìm hiểu qua 3 đại diện tiêu biểu về Brute Force. Selection Sort, Bubble Sort và Sequential Search. Chúng ta có thể rút ra một số nhận định sau:

* Brute Force là một cách đơn giản để tiếp cận vấn đề. “Nghĩ sao làm vậy, không chú trọng hiệu năng, chỉ quan tâm kết quả”.
* Điểm mạnh của phương pháp này đó là có thể áp dụng được vào hầu hết các vấn đề, tuy nhiên điều này có thể dẫn nhiều vấn đề trong kĩ thuật lập trình.
  + Ví dụ, Một bài toán với đầu vào là một mảng đã được sắp xếp, không nên sử dụng thuật toán tìm kiếm tuần tự cho điều này. Hay thuật toán Bubble Sort và Selection Sort với độ phức tạp O() do đó hiện nay không ai sử dụng chúng vào các dự án phức tạp.
* Tuy nhiên, nói như vậy không có nghĩa không nên sử dụng Brute Force. Ở một vài khía cạch, chúng ta buộc phải sử dụng tới chiến lược này, ví dụ như thuật toán tìm kiếm đối với môt mảng đầu vào chưa được sắp xếp, chúng ta buộc phải sử dụng thuật toán tìm kiếm tuần tự thay vì tìm kiếm nhị phân, hay trong các vấn đề về bảo mật, các hacker thường xuyên thực hiện chiến lược này để tìm thông tin người dùng, do đó chúng ta cần phải hiểu thật rõ chiến lược này.

2.2 Divide and Conquer

- Trong phần này chúng ta sẽ tìm hiểu về chiến lược “Chia để trị”, cái tên mà học sinh phổ thông nào cũng được nghe tới thông qua môn lịch sử.

- Divde: chia nhỏ vấn đề và Conquer: giải quyết các vấn đề nhỏ đó.

- Các thuật toán Divide and Conquer hoạt động theo cách sau:

* Một vấn đề được chia làm nhiều bài toán con cùng loại. (Thường là cùng kích thước).
* Thường được thực hiện bởi các thuật toán đệ quy.
* Các giải pháp giải quyết các vấn đề con được kết hợp để giải quyết vấn đề ban đầu.

2.2.1 Merge Sort

- Merge Sort là một ví dụ hoàn hảo về việc áp dụng thành công kĩ thuật Divide and Conquer để thực hiện sắp xếp.

- Merge Sort thực hiện trên ý tưởng chia mảng hiện tại thành 2 nửa A[0,…n/2 -1] và A[n/2,…n-1] cho đến khi mảng còn 1 phâng tử, và tiến hành sắp xếp chúng theo cách đệ quy. Và cuối cùng hợp nhất mảng con thành một mảng được sắp xếp.

- Mã giả:

1. Mergesort(A[0,…n-1]):
2. // sắp xếp mảng bởi đệ quy merge sort.
3. //input: 1 mảng A có kích thước n phần tử.
4. //output: Mảng được sắp xếp tăng dần.
5. If(n >1):
6. Copy A[0,..n/2-1] to B[0,…n/2-1]
7. Copy A[n/2, …n-1] to C[0,…n/2-1]
8. Mergesort(B)
9. Mergesort(C)
10. Merge(B,C,A) \*(1)\*
11. Endif
12. Enfalgorithm

* (1) Giai đoạn merge 2 mảng con thành một mảng sorted A.
* Hàm Merge sẽ thực hiện so sanh lần lượt 2 cặp số từ 2 mảng đầu vào, giá trị nhỏ hơn sẽ được add vào phần tử đầu tiên của mảng A. Quá trinh này sẽ được lặp đi lặp lại cho tới khi duyệt hết 1 trong 2 mảng đầu vào. Và tiếp tục copy các giá trị còn lại của mảng chưa được duyệt hết vào mảng A.
* Mã giả:

1. Merge(B,C,A):
2. // purpose : Hợp 2 mảng B C thành mảng A theo thứ tự tăng dần.
3. // Input: mảng B có kích thước p, mảng C có kích thước q, mảng A có kích thước p+q
4. // Ouput: Sắp xếp mảng A từ mảng B và C
5. i, j, k=0
6. While i <p and j <q do:
7. If B[i]≤C[j]:
8. A[k]=B[i]; i+=1
9. Else :
10. A[k]=C[j]; j+=1
11. Endif
12. k+=1
13. Endwhile
14. If i =p:
15. Copy C[j,q-1] to A[k,…p+1-1]
16. Else:
17. Copy B[i,p-1] to A[k,…p+1-1]
18. Endif
19. Endalgorithm.

* Độ phức tạp của thuật toán:
* T(n) = 2T(n/2) + fmerge (n)
* fmerge (n), chúng ta sẽ xem xét số lần thực hiện phép so sánh tại dòng 7 B[i]≤C[j].
* Phép so sánh sẽ được thực hiện cho tới khi duyệt hết một trong 2 mảng B và C, trong trường hợp tệ nhất, ví dụ mảng B[1,3,5] và mảng C[2,4,6], tổng số lần thực hiện phép so sánh sẽ là 5. Vậy trong trường hợp tệ nhất fmerge (n)= n -1.
* Vậy T(n) = 2T(n/2)+n-1.
* Áp dụng định lý Master Theorem. a = 2, b=2, k=1
  + => T(n) (nlogn)

2.2.2 Quick Sort

- Tiếp theo chúng ta sẽ xem xét một thuật toán Sort khác, Quick Sort.

- Cũng giống như Merge Sort, thuật toán Quick Sort cũng là một thuật toán chia để trị.

- Ý tưởng của thuật toán Quick Sort là lựa chọn một điểm đánh dấu pivot, thuật toán sẽ thực hiện chia mảng thành các mảng con dựa vào pivot đã chọn. Di chuyển tất cả các giá trị nhỏ hơn pivot sang bên trái và các giá trị lớn hơn qua bên phải so với pivot.

- Mã giả:

1. Quicksort(A,l,r)
2. // mục đích: sắp xếp mảng theo thuật toán quicksort
3. // input: một mảng A và l là index đầu và r là index của phần tử cuối trong mảng.
4. // output: mảng A được sắp xếp tăng dần.
5. If l < r
6. s = partition(A,l,r) \*(1)\*
7. Quicksort(A,l,s-1)
8. Quicksort(A,s+1,r)
9. Endif
10. Endalgorithm

* (1) Hàm partition sẽ thực hiện trả về vị trí của pivot.
* Mã giả:

1. partition(A,l,r):
2. // mục đích, đẩy các giả trị nhỏ hơn pivot sang bên trái và giá trị lớn hơn sang bên phải
3. // input Mảng A và index l của phần tử đầu tiền và index r của phần tử cuối cùng trong Mảng A hiện tại
4. // output Trả về index của pivot hiện tại trong mảng A.
5. p = A[l]
6. i = l
7. j = r
8. while i < j do:
9. while A[i]<p do :
10. i++
11. endwhile
12. While A[j]>p do:
13. J --
14. Endwhile
15. Swap(A[i],A[j])
16. Endwhile
17. Swap(A[i],A[j])
18. Swap(A[l],A[j]) // đổi vị trí pivot ban đầu so với vị trí pivot mới tìm
19. Return j

* Độ phức tạp của thuật toán:
  + T(n) = 2T(n/2)+ fparition(n)
  + fparition(n) = n
  + Trong trường hợp tệ nhất pivot được chọn là giá trị nhỏ nhất hoặc lớn nhất trong mảng A, thì biểu thức T(n) sẽ trở thành T(n)=T(n-1) + n => T(n) O()
  + Trường hợp tốt nhất pivot được chọn là giá trị trung bình (số phần tử bên trái = số phần tử bên phải) thì biểu thức T(n) sẽ trở thành
    - T(n) = 2T(n/2)+n
    - Áp dụng Master Theorem: a= 2, b= 2 k =1
    - => T(n) (nlogn)

2.2.3 Binary Tree Traversals

- Trong phần này chúng ta sẽ xem chiến lược Divide and Conquer đối với cây nhị phân.

- Cây nhị phân được biết đến là cấu trúc dữ liệu dạng cây và có mỗi nút nhiều nhất gồm 2 node con, node trái và phải.

- Chiều cao của cây nhị phân được tính từ node root đến node con xa nhất.

- Thuật toán đệ quy tính chiều cao cây nhị phân T áp dụng Divide and Conquer.

- Mã giả:

1. Height(T):
2. // Tính chiều cao của cây nhị phân
3. // input: Cây nhị phân T
4. // output chiều cao cây T
5. // if T = : return -1
6. // return max( Height(T­trái,Tphải))+1
7. Endalgorithm

* Thao tác cơ bản nhất:
  + So sánh tại dòng 5 T =
* Hệ thức truy hồi:
  + A(Tn)= A(Ttrái)+A(Tphải) + 1 ; n là tổng số node
* Độ phức tạp của thuật toán:
  + A(Tn) = n

2.2.4 Kết luận:

- Sau khi đi qua 3 thuật toán Merge Sort, Quick Sort và Binary Tree Traversals chúng ta có thể thấy rằng :

* Divide and Conquer là một kỹ thuật thiết kế thuật toán chung bằng cách chia nhỏ bài toán thành các bài toán nhỏ hơn cùng loại, và giải quyết từng bài toán nhỏ bằng đệ quy và kết hợp chúng để giải quyết vấn đề ban đầu.
* Thời gian chạy của các thuật toán chia để trị hoàn toàn tối ưu hơn so với kĩ thuật Brute Force, cụ thể với quick sort hay merge sort có độ phức tạp nhỏ hơn nhiều so với 2 thuật toán sort sử dụng kĩ thuật brute force. Tuy nhiên, với thuật toán merge sort lại đến từ việc hao tốn bộ nhớ khi thực hiện khởi tạo quá nhiều mảng con.
* Do vậy, không phải mọi thuật toán Divide and Conquer đều hiệu quả hơn Brute Force, trên thực tế cách tiếp cận chia để trị mang lại một số thuật toán quan trọng và hiệu quả nhất trong khoa học máy tính, trong đó khi các vấn đề con được giải quyết đồng thời.

2.3 Greedy Algorithms

- Thuật toán Greedy là một thuật toán đơn giản, trực quan được sử dụng trong các bài toán tối ưu hóa. Thuật toán đưa ra lựa chọn tối ưu ở mỗi bước khi nó cố gắng tìm ra cách tối ưu tổng thể để giải quyết toàn bộ vấn đề. Các thuật toán Greedy khá thành công trong một số vấn đề, chẳng hạn như mã hóa Huffman được sử dụng để nén dữ liệu, hoặc thuật toán Dijkstra, được sử dụng để tìm đường đi ngắn nhất qua biểu đồ.

- Thuật toán Greedy lấy tất cả dữ liệu trong một vấn đề cụ thể, sau đó đặt quy tắc cho các yếu tố nào cần thêm vào giải pháp ở mỗi bước của thuật toán.

- Nếu cả hai thuộc tính dưới đây đều đúng,thuật toán Greedy có thể được sử dụng để giải quyết vấn đề :

+ Thuộc tính lựa chọn Greedy : Có thể đạt được giải pháp tối ưu toàn cầu (tổng thể) bằng cách chọn lựa chọn tối ưu ở mỗi bước.

+ Cấu trúc con tối ưu: Một bài toán có cấu trúc con tối ưu nếu một giải pháp tối ưu cho toàn bộ bài toán chứa các giải pháp tối ưu cho các bài toán con.

- Nói cách khác, các thuật toán Greedy hoạt động trên các bài toán mà đúng là ở mỗi bước, có một lựa chọn tối ưu cho bài toán cho đến bước đó và sau bước cuối cùng, thuật toán tạo ra giải pháp tối ưu hoàn chỉnh vấn đề.

- Để thực hiện một thuật toán Greedy, hãy xác định một cấu trúc con hoặc bài toán con tối ưu trong bài toán. Sau đó, xác định giải pháp sẽ bao gồm những gì (ví dụ: tổng lớn nhất, đường đi ngắn nhất, v.v.). Tạo một số cách lặp đi lặp lại để đi qua tất cả các bài toán con và xây dựng giải pháp.

2.3.1 Thuật toán Prim

- Định nghĩa : Cây bao trùm của đồ thị được kết nối vô hướng là đồ thị con xoay chiều được kết nối của nó (tức là cây) chứa tất cả các đỉnh của đồ thị. Nếu một đồ thị như vậy có các trọng số được gán cho các cạnh của nó, thì cây khung tối thiểu là cây khung có trọng số nhỏ nhất của nó, trong đó trọng số của cây được định nghĩa là tổng các trọng số trên tất cả các cạnh của nó. Bài toán cây bao trùm tối thiểu là bài toán tìm cây bao trùm tối thiểu cho một đồ thị liên thông có trọng số cho trước.

- Nếu chúng ta cố gắng xây dựng một cây bao trùm tối thiểu bằng cách tìm kiếm đầy đủ, chúng ta sẽ gặp phải hai trở ngại nghiêm trọng. Đầu tiên, số lượng cây bao trùm phát triển theo cấp số nhân với kích thước đồ thị (ít nhất là đối với đồ thị dày đặc). Thứ hai, tạo ra tất cả các cây bao trùm cho một đồ thị đã cho là không dễ dàng; trên thực tế, khó hơn việc tìm một cây bao trùm tối thiểu cho một đồ thị có trọng số bằng cách sử dụng một trong một số thuật toán hiệu quả có sẵn cho vấn đề này. Trong phần này, chúng tôi trình bày sơ lược về thuật toán Prim.

- Mã giả :

1. Thuật toán Prim tạo cây bao trùm tối thiểu

2. Input : Đồ thị liên thông có trọng số G = V,E

3. Output : ET  tập hợp các cạnh tạo nên một cây khung tối thiểu của G

VT 🡨 {v0} // tập hợp các đỉnh cây có thể được khởi tạo với bất kì đỉnh nào

ET 🡨 ∅

4. for i 🡨 1 to |V| - 1 do

5. Tìm cạnh tối thiểu e\* = (v\*,u\*) trong tất cả các cạnh (v,u) sao cho v trong VT và u trong V – VT

VT 🡨 VT ∪ {u\*}

ET 🡨 ET ∪ {e\*}

6. return ET

2.3.2 Kruskal’s Algorithm

Trong phần trước, chúng ta đã xem xét thuật toán tham lam “phát triển” một cây bao trùm tối thiểu thông qua việc đưa vào tham lam của đỉnh gần nhất với các đỉnh đã có trong cây. Đáng chú ý, có một thuật toán tham lam khác cho bài toán cây bao trùm tối thiểu cũng luôn đưa ra giải pháp tối ưu. Nó được đặt tên là thuật toán Kruskal theo tên của Joseph Kruskal, người đã khám phá ra thuật toán này khi anh ấy là sinh viên năm thứ hai tốt nghiệp [Kru56]. Thuật toán của Kruskal xem xét cây bao trùm tối thiểu của một đồ thị liên thông có trọng số G = V, E dưới dạng một đồ thị con xoay chiều với | V | - 1 cạnh mà tổng trọng số các cạnh là nhỏ nhất. (Không khó để chứng minh rằng một đồ thị con như vậy phải là một cái cây.) Do đó,thuật toán xây dựng một cây bao trùm tối thiểu dưới dạng một chuỗi mở rộng của các đồ thị con luôn có tính tuần hoàn nhưng không nhất thiết phải được kết nối trên các giai đoạn trung gian của thuật toán.

Thuật toán bắt đầu bằng cách sắp xếp các cạnh của biểu đồ theo thứ tự không giảm về trọng số của chúng. Sau đó, bắt đầu với đồ thị con trống, nó sẽ quét danh sách đã sắp xếp này, thêm cạnh tiếp theo trên danh sách vào đồ thị con hiện tại nếu việc bao gồm như vậy không không tạo một chu trình và chỉ cần bỏ qua cạnh khác.

- Mã giả :

1. //Thuật toán của Kruskal để xây dựng cây bao trùm tối thiểu

2. //Input : Đồ thị được kết nối có trọng số G = {V,E}

3.//Output: ET, tập hợp các cạnh tạo cây khung tối thiểu G sắp xếp E theo thứ tự không giảm của trọng số cạnh w(ei1)≤… ≤w(ei|E|)

4.while ecounter < |V| - 1 do

5.k <- k + 1

6. if ET ∪ {eik} là mạch vòng

7. ET <- ET ∪ {eik}; ecounter <- ecounter + 1

8. return ET

2.3.3 Dijkstra’s Algorithm

Thuật toán Dijkstra là một thuật toán để tìm các đường đi ngắn nhất giữa các nút trong biểu đồ, có thể biểu diễn. Nó được hình thành bởi nhà khoa học máy tính Edsger W. Dijkstra vào năm 1956 và được xuất bản ba năm sau đó.

Thuật toán tồn tại trong nhiều biến thể. Thuật toán ban đầu của Dijkstra đã tìm ra đường đi ngắn nhất giữa hai nút nhất định, nhưng một biến thể phổ biến hơn sửa một nút duy nhất làm nút "nguồn" và tìm đường đi ngắn nhất từ ​​nguồn đến tất cả các nút khác trong biểu đồ, tạo ra một đường ngắn nhất cây.

Thuật toán Dijkstra sử dụng các nhãn là số nguyên dương hoặc số thực, được sắp xếp hoàn toàn. Nó có thể được tổng quát để sử dụng bất kỳ nhãn nào được sắp xếp một phần, miễn là các nhãn tiếp theo (nhãn tiếp theo được tạo ra khi đi qua một cạnh) là đơn nguyên không giảm. Tổng quát hóa này được gọi là thuật toán đường đi ngắn nhất Dijkstra chung.

- Mã giả :

1. // Thuật toán Dijkstra cho các đường đi ngắn nhất nguồn đơn

2. // Đầu vào: Đồ thị liên thông có trọng số G = V, E với các trọng số không âm   
và đỉnh của nó là s

3. // Đầu ra: Độ dài dv của đường đi ngắn nhất từ ​​s đến v và đỉnh áp chót của nó pv cho mọi đỉnh v trong V. Khởi tạo (Q) khởi tạo hàng đợi ưu tiên để trống

4. for every vertex v in V

5. dv <- ∞; pv <- null

6. Insert (Q,v,dv) // khởi tạo ưu tiên đỉnh trong hàng đợi ưu tiên ds <- 0;

Giảm (Q,s,ds) // cập nhật mức độ ưu tiên của s với ds VT <- ∅

7.for i <- 0 to |V| - 1 do

8. u\* <- DeleteMin(Q) //xóa phần tử ưu tiên tối thiểu

9. VT <- VT ∪ {u∗}

10. for every vertex u in V – VT that is adjacent to u\* do

11.if du\* + w (u\* , u ) < du

12.du <- du\* + w (u\* , u ) ; Pu <- u\*

13.Giảm xuống(Q,u,du)

2.3.4 Huffman Trees and Codes

Mã hóa Huffman gán mã cho các ký tự sao cho độ dài của mã phụ thuộc vào tần suất tương đối hoặc trọng lượng của ký tự tương ứng. Mã Huffman có độ dài thay đổi và không có tiền tố (không có mã nào là tiền tố của bất kỳ mã nào khác). Bất kỳ mã nhị phân không có tiền tố nào cũng có thể được hình dung như một cây nhị phân với các ký tự mã hóa được lưu trữ ở các lá.

Huffman coding tree hay H uffman tree là một cây nhị phân đầy đủ, trong đó mỗi lá của cây tương ứng với một chữ cái trong bảng chữ cái nhất định.

Xác định chiều dài đường đi có trọng số của một chiếc lá bằng trọng lượng của nó nhân với chiều sâu. Huffman Tree là cây nhị phân có trọng số đường dẫn bên ngoài tối thiểu, tức là cây có tổng chiều dài đường dẫn có trọng số tối thiểu cho tập hợp các lá đã cho. Vì vậy, mục tiêu là xây dựng một cây có trọng lượng đường đi bên ngoài tối thiểu.

- Mã giả :

1. Khởi tạo n cây một nút và gắn nhãn chúng bằng các ký hiệu của bảng chữ cái đã cho. Ghi lại tần suất của từng ký hiệu trong gốc cây của nó để cho biết trọng lượng của cây. (Nói chung, trọng lượng của cây sẽ bằng tổng các tần số trên lá của cây.)

2. Lặp lại thao tác sau cho đến khi thu được một cây. Tìm thấy hai cây có khối lượng nhỏ nhất (các dây buộc có thể bị bẻ gãy tùy ý, nhưng hãy xem Vấn đề 2 trong các bài tập của phần này). Đặt chúng thành cây con bên trái và bên phải của một cây mới và ghi lại tổng trọng lượng của chúng ở gốc của cây mới làm trọng lượng của nó.

2.3.5 Kết luận

- Kỹ thuật Greedy đề xuất xây dựng một giải pháp cho một vấn đề tối ưu hóa thông qua một chuỗi các bước, mỗi bước mở rộng một phần giải pháp thu được cho đến khi đạt được giải pháp hoàn chỉnh cho vấn đề. Trên mỗi bước, lựa chọn được đưa ra phải khả thi, tối ưu cục bộ và không thể thu hồi.

- Thuật toán của Prim là một thuật toán Greedy để xây dựng một khoảng mở rộng tối thiểu cây của một đồ thị liên thông có trọng số. Nó hoạt động bằng cách gắn vào một cây con đã xây dựng một đỉnh gần nhất với các đỉnh đã có trong cây.

- Thuật toán của Kruskal là một thuật toán Greedy khác để kéo dài tối thiểu vấn đề cây. Nó xây dựng một cây bao trùm tối thiểu bằng cách chọn các cạnh theo thứ tự không giảm về trọng lượng của chúng với điều kiện là sự bao gồm không tạo ra một chu kỳ. Kiểm tra điều kiện sau một cách hiệu quả yêu cầu một ứng dụng của một trong những thuật toán được gọi là liên hợp tìm.

- Thuật toán của Dijkstra giải quyết vấn đề tìm đường ngắn nhất nguồn đơn

đường đi ngắn nhất từ ​​một đỉnh đã cho (nguồn) đến tất cả các đỉnh khác của đồ thị hoặc đồ thị có trọng số. Nó hoạt động như thuật toán của Prim nhưng so sánh đường dẫn độ dài chứ không phải độ dài cạnh. Thuật toán của Dijkstra luôn mang lại kết quả chính xác giải pháp cho một đồ thị có trọng số không âm.

- Huffman Tree là cây nhị phân giảm thiểu độ dài đường dẫn có trọng số từ

gốc đến lá của trọng lượng xác định trước. Ứng dụng quan trọng nhất của Huffman Tree là mã Huffman.

- Mã Huffman là một lược đồ mã hóa độ dài thay đổi không có tiền tố tối ưu gán các chuỗi bit cho các ký hiệu dựa trên tần số của chúng trong một văn bản nhất định. Điều này được thực hiện bằng cách xây dựng một cây nhị phân có lá đại diện cho các ký hiệu trong bảng chữ cái và có các cạnh được gắn nhãn 0 và 1.

2.4 Dynamic Programming

- Sau khi đã tìm hiểu xong kĩ thuật Divide and Conquer – Giải quyết vấn đề bằng cách chia nhỏ vấn đề ban đầu thành vấn đề nhỏ hơn và sử dụng thuật toán đệ quy để giảivấn đề kích thước lớn tới nhỏ (trên xuống dưới). Tuy nhiên nhược điểm chính của kĩ thuật này là tiêu tốn bộ nhớ sau mỗi lần gọi đệ quy.

- Do vậy, Dynamic Programming hay còn gọi là lập trình động được đề xuất để tránh sử dụng thêm dung lượng. Dynamic Programming cũng gần tương tự kĩ thuật Divide and Conquer với ý tưởng chia nhỏ bài toán thành các vấn đề nhỏ hơn, tuy nhiên DP sử dụng vòng lặp để giải quyết các vấn đề từ nhỏ tới lớn ( dưới lên trên ) khác với DAC.

2.4.1 Three Basic Examples

- Sau đây chúng ta sẽ tìm hiểu cơ bản về kĩ thuật Dynamic Programming thông qua 3 ví dụ cơ bản và phổ biến.

2.4.1.1 Coin Row Problem

- Giả sử có một dãy đồng xu có giá trị lần lượt là fi, giá trị mỗi đồng có thể bằng nhau.

* Mục tiêu thiết kế giải thuật để có thể lấy các đồng xu, sao cho tổng giá trị lấy được là lớn nhất, và mỗi đồng xu được lấy không được nằm kế tiếp nhau trên dãy.
* Để giải quyết được vấn đề này, chúng ta cùng đi phân tích hướng giải quyết.
  + Giả sử tổng số đồng xu trên dãy là n
  + Chúng ta sẽ có 2 cách để lấy các đồng xu: lấy các đồng xu có vị trí lẻ hoặc lấy các đồng xu có vị trí chẵn.
    - Cách thứ nhất ( tổng số đồng xu ở vị trí chẵn), Achẵn = fn + A(n-2), với Ak là tổng giá trị các đồng xu lấy được tại vị trí k.
    - Cách thứ hai ( tổng số đồng xu ở vị trí lẻ), Alẻ = A(n-1­).
  + Vậy công thức tính sẽ là Amax = max(Achẵn , Alẻ ) (\*)

* Mã giả:

1. CoinRow(C[1,..n]):
2. // Mục tiêu: áp dụng công thức \* để giải quyết bài toán từ dưới lên.
3. // Input: Một mảng C các đồng xu với các giá trị C[i] là dương.
4. // Ouput: Tổng giá trị nhiều nhất các đồng xu có thể lấy được.
5. F=[] // khởi tạo mảng F có n phần tử.
6. F[0]=0, F[1]=C[1]
7. For i = 2 to n-1 do :
8. F[i] = max(C[i] + F[i-2], F[i-1])
9. Endfor
10. Return F[n-1]

* Độ phức tạp thuật toán:
  + T(n) (n)
  + Nếu giải bài toán bằng phép đệ quy kĩ thuật Divide and Conquer thì độ phức tạp thuật toán sẽ vẫn là O(n). Tuy nhiên, nó sẽ tiêu tốn thêm bộ nhớ tại mỗi lần gọi đệ quy.­

2.4.2 Change Making Problem:

- Bạn nhận được yêu cầu thiết kế một thuật toán để đổi tiền lẻ.

- Giả sử các đồng xu có giá trị d1, d2, d3, …dm, với d1 = 1. Và số lượng các đồng xu là không giới hạn.

- Thiết kế thuật toán sao cho số lượng tổng các loại đồng xu sau khi đổi là nhỏ nhất.

- Ý tưởng thực hiện :

* Gọi F(n) là tổng số đồng xu it nhất để quy đổi cho n. => F(0) = 0.
* Chúng ta biết rằng, ở mỗi loại đồng xu, để đổi được thì n≥ dj.
* Và sau mỗi lần đổi F(n) cộng thêm 1, và giá trị tiền hiện tại lúc này là n-d­j.
* Do đó, ta có công thức tổng quát: F(n) = min(F(n-dj)) +1.
* Và ta tìm min của F(n-dj) bằng cách thử tất cả trường hợp đổi với từng loại giá trị đồng xu và chọn cách đổi được ít xu nhất.

- Mã Giả:

1. ChangeMaking(D[1,…m],n):
2. // Mục đích: Áp dụng Dynamic Programming để tìm số đồng xu tối thiểu để quy đổi cho n.
3. // Input: Mảng D gồm các giá trị của các loại đồng xu có kích thước m, và giá trị tiền cần đổi n. Và D[1] = 1.
4. // Output: Tổng số đồng xu tối thiểu.
5. F[0]=0
6. For i= 1 to n-1 do:
7. temp= ∞; j=1;
8. While j ≤ m and i ≥ D[j] do:
9. temp = min(F[i-D[j]],temp)
10. j=j+1
11. Endwhile
12. F[i]= temp +1
13. Endfor
14. Return F[n-1]
15. EndAlgorithm

- Độ phức tạp thuật toán:

* + T(n,m) (nm)
* Không gian của thuật toán:

2.4.3 Coin Collecting Problem:

- Có một bàn cờ có n hàng và m cột. Ở mỗi ô đặt một đồng xu, hãy tính số xu nhiều nhất có thể nhặt được khi đi lần lượt từ ô hàng 1 cột 1 đến ô hàng n cột m. Với điều kiện mỗi lần di chuyển chỉ có thể đi được một ô và từ trên xuống dưới hoặc trái sang phải.

- Mục tiêu của bài toán là tìm max tại F[n][m], với F là tổng số đồng xu nhặt được nhiều nhất tại hàng n cột m.

- Vì chỉ có thể đi được từ trên xuống hoặc trái sang phải nên tại F[n][m] bằng giá trị tại ô cạch trái hoặc ô phía trên và + 1 nếu tại vị trí n, m tồn tại đồng xu hoặc không.

- Do đó ta có công thức sau: F[n][m] = max( F[n-1][m] , F[n][m-1] ) + C[n][m]; với C[n][m] là có hay không đồng xu ở vị trí n, m.

* Mã giả:

1. RobotCoinCollection(C[1,..n],[1,…m]):
2. // Mục tiêu: Áp dụng kĩ thuật Dynamic Programming để tìm số lượng đồng xu nhiều nhất nhặt được tại vị trí n m khi đi từ vị trí (1,1). Và chỉ được di chuyển từ trên xuống hoặc trái qua phải.
3. // Input: Một mảng C 2 chiều có kích thước n x m, nếu tại vị trí (i,j) có đồng xu thì C[i][j] = 1, ngược lại = 0.
4. // Output: Số lượng đồng xu nhiều nhất có thể nhặt được.
5. F[n][m] khởi tạo một mảng 2 chiều F có kích thước n x m
6. F[0][0] = C[0][0]
7. For j = 1 to m -1 do :
8. F[0][j] = F[0][j-1] + C[0][j]
9. Endfor
10. For i = 1 to n-1 do:
11. F[i,0]= F[i-1,0]+ C[I,0]
12. For j = 1 to m-1 do:
13. F[i][j] = max( F[i-1][j] , F[i][j-1] ) + C[i][j]
14. Endfor
15. Endfor
16. Return F[n-1][m-1]
17. Endalgorithm.

- Độ phức tạp thuật toán:

* + T(n,m) = m + n\*m
  + => T(n,m) (nm)
* Không gian của thuật toán:

2.4.2 Kết Luận

- Lập trình động là một kỹ thuật để giải quyết các vấn đề chồng chéo các bài toán con.

- Thông thường, các vấn đề con này phát sinh từ sự lặp lại liên quan đến giải pháp cho một vấn đề nhất định với các giải pháp cho các vấn đề con của nó.

- Dynamic Programming giải quyết bài toán bằng các ghi lại kết quả của mỗi vấn đề con vào một bảng và từ đó đưa ra giải pháp cho vấn đề ban đầu.

- Dynamic Programming hoạt động theo một nguyên tắc: giải pháp để giải quyết bài toán phải tối ưu và được tạo thành từ cách giải pháp tối ưu cho từng vấn đề con.

CHƯƠNG 3 – PHẦN TỰ HỌC

*Phần sinh viên phải tự tìm hiểu và hoàn thành*;

3.1 Backtraking

**TÀI LIỆU THAM KHẢO**

**Tiếng Việt**

1. Quách Ngọc Ân (1992), “Nhìn lại hai năm phát triển lúa lai”, *Di tuyền học ứng dụng*, 98(1), tr. 10-16.
2. Bộ nông nghiệp & PTNT (1996), *Báo cáo tổng kết 5 năm (1992-1996) phát triển lúa lai,* Hà Nội.
3. Nguyễn Hữu Đống, Đào Thanh Bằng, Lâm Quang Dụ, Phan Đức Trực (1997), *Đột biến –* *Cơ sở lý luận và ứng dụng,* Nhà xuất bản nông nghiệp, Viện khoa học kỹ thuật nông nghiệp Việt Nam, Hà Nội.
4. Nguyễn Thị Gấm (1996), *Phát hiện và đánh giá một số dòng bất dục đực cảm ứng nhiệt* *độ,* Luận văn thạc sĩ khoa học nông nghiệp, Viện khoa học kỹ thuật nông nghiệp Việt Nam, Hà Nội.

……….

1. Võ Thị Kim Huệ (2000), *Nghiên cứu chẩn đoán và điều trị bệnh…,* Luận án Tiến sĩ y khoa, Trường đại học y Hà Nội, Hà Nội.

**Tiếng Anh**

1. Anderson J.E. (1985), The Relative Inefficiency of Quota, The Cheese Case, *American* *Economic Review*, 75(1), pp. 178-90.
2. Borkakati R. P.,Virmani S. S. (1997), Genetics of thermosensitive genic male sterility in Rice, *Euphytica* 88, pp. 1-7.
3. Boulding K.E. (1955), *Economics Analysis*, Hamish Hamilton, London.
4. Burton G. W. (1988), “Cytoplasmic male-sterility in pearl millet (penni-setum glaucum L.)”, *Agronomic Journal* 50, pp. 230-231.
5. Central Statistical Oraganisation (1995), *Statistical Year Book*, Beijing.
6. FAO (1971), *Agricultural Commodity Projections (1970-1980)*, Vol. II. Rome.
7. Institute of Economics (1988), *Analysis of Expenditure Pattern of Urban Households in* *Vietnam,* Departement pf Economics, Economic Research Report, Hanoi.

**PHỤ LỤC**

Phần này bao gồm những nội dung cần thiết nhằm minh họa hoặc hỗ trợ cho nội dung luận văn như số liệu, biểu mẫu, tranh ảnh. . . . nếu sử dụng những câu trả lời cho một *bảng câu hỏi thì bảng câu hỏi mẫu này phải được đưa vào phần Phụ lục ở dạng nguyên bản* đã dùng để điều tra, thăm dò ý kiến; **không được tóm tắt hoặc sửa đổi**. Các tính toán mẫu trình bày tóm tắt trong các biểu mẫu cũng cần nêu trong Phụ lục của luận văn. Phụ lục không được dày hơn phần chính của luận văn