HƯỚNG DẪN THỰC HÀNH

CÁC THUẬT TOÁN TÌM KIẾM

I. Mục tiêu

Sinh viên cài đặt các thuật toán tìm kiếm và ứng dụng nó để giải quyết các bài toán đặt ra.

II. Qui định nộp

- Sinh viên nộp một tập tin nén, có tên là ***<MSSV>.zip*** hoặc ***<MSSV>.rar*** chứa source code và báo cáo của chương trình.
- Sinh viên nộp kèm một file báo cáo ghi mức độ hoàn thành công việc của mình, các bộ dữ liệu mà mình test ở mỗi bài.
- Tất cả các bài tập sẽ lập trình theo command line (tham số dòng lệnh).
- Kiến trúc thư mục nộp:
    - MSSV
        - Document: báo cáo mức độ hoàn thành của mỗi bài tập, những vấn đề gặp phải dẫn đến không hoàn thành được các phần của bài tập.
        - MSSV_BT1:
            - Source: chứa các file .cpp
            - Header: chứa các file .h
            - Test case: chứa các file input để test thử (ít nhất 3 bộ test có đặc trưng khác nhau)
            - …
        - MSSV_BT2: ….
- Môi trường làm việc: Visual Studio 2015 hoặc các môi trường tương đương. Không sử dụng các hàm bị lỗ hổng bảo mật như gets, …
- Hạn nộp: xem link trên Moodle.
- Bài giống nhau hay nộp file rác sẽ 0 điểm MÔN HỌC.

III. Nội dung

BT1. Cài đặt các thuật toán tìm kiếm để xác định tất cả vị trí của phần tử cần tìm trong mảng một chiều các số nguyên:

- (1) Tìm kiếm tuần tự
- (2) Tìm kiếm tuần tự có lính canh
- (3) Tìm kiếm nhị phân

Command line: MSSV_BTx.exe *ThuatToan x* Input.txt Output.txt

Ví dụ: 1512345_BT1.exe 1 3 test1.txt out1.txt

Nghĩa là chạy thuật toán tìm kiếm tuần tự để tìm phần tử x=3 chứa trong mảng một chiều trong file test1.txt và xuất kết quả ra file out1.txt

Định dạng input:

- Dòng đầu tiên chứa số phần tử của mảng
- Dòng tiếp theo chứa các phần tử, mỗi phần tử cách nhau 1 khoảng trắng.
- Ví dụ:

5
2 8 3 9 1 3

Định dạng output:

- Tất cả các vị trí xuất hiện với ví trí được đếm từ 0. Mỗi vị trí cách nhau khoảng trắng. Nếu không có phần tử thì xuất ra giá trị -1.
- Lưu ý không có khoảng trắng ở cuối dòng.
- Ví dụ cho ví dụ ở input:

2 5

Lưu ý: đối với thuật toán tìm kiếm nhị phân, mảng phải có thứ tự tăng dần.

BT2. Tìm cách tăng kích thước mảng dần lên từ 1000, 10000, 1000000, … phần tử. Đo thời gian thực thi của mỗi thuật toán ở BT1. Mảng và giá trị cần tìm được tạo ngẫu nhiên. Đối với thuật toán tìm kiếm nhị phân, mảng phát sinh phải có thứ tự tăng dần.

Command line: MSSV_BTx.exe *ThuatToan* KichThuoc KetQua

Ví dụ: 1512345_BT1.exe 1 1000 out1000.txt

Nghĩa là chạy thuật toán tìm kiếm tuần tự để tìm kiếm trên mảng ngẫu nhiên có 1000 phần tử và xuất ra thời gian mili giây với độ chính xác làm tròn đến 3 chữ số.

BT3. Thực hiện thuật toán tìm kiếm sử dụng hàm băm lấy dư (k mod N) với các cách xử lý đụng độ khác nhau:

- (1) Xử lý bằng nối kết (cho phép sử dụng thư viện danh sách liên kết)
- (2) Xử lý bằng dò tuyến tính
- (3) Xử lý bằng dò bậc 2
- (4) Xử lý bằng hàm băm kép.

Giả sử dữ liệu đưa vào không có giá trị trùng nhau.

Command line: MSSV_BTx.exe XuLy *x* Input.txt Output.txt

Ví dụ: 1512345_BT1.exe 1 7 test1.txt out1.txt

Nghĩa là chạy thuật toán tìm kiếm sử dụng hàm băm với phương pháp xử lý đụng độ bằng nối kết để chứa và tìm phần tử x = 7 trong mảng một chiều. Nhận dữ liệu từ test1.txt và xuất kết quả ra file out1.txt

Định dạng input:

- Dòng đầu tiên chứa kích thước không gian.
- Dòng đầu tiên chứa số phần tử của mảng cần lưu trữ
- Dòng tiếp theo chứa các phần tử, mỗi phần tử cách nhau 1 khoảng trắng.

- Ví dụ:

100
5
2 8 3 9 1 4 6

Định dạng output:

- Không gian chứa phần tử đã thêm vào. Mỗi ô trong không gian sẽ tương ứng với một dòng. Dòng nào không có dữ liệu thì để trống. Đối với nối kết thì các phần tử trong danh sách liên kết sẽ ghi cùng một dòng.
- Dòng cuối cùng là vị trí của phần tử cần tìm với phần tử đầu tiên đếm là 0. Nếu không có phần tử thì xuất ra giá trị -1.
- Lưu ý không có khoảng trắng ở cuối dòng.
- Ví dụ:

3
8 4 6
…
2

1
-1

Giải thích: các dòng trước số -1 là mỗi ô trong không gian tìm kiếm. Ô đầu chứa giá trị 3. Ô thứ hai, chứa giá trị 8 nhưng do đụng độ nên các giá trị sau là 4 6 được xử lý bằng phương pháp nối kết. Lưu ý, chỉ có phương pháp xử lý bằng nối kết mới có một dòng có nhiều giá trị. Dòng cuối cùng in ra -1 vì không tìm thấy giá trị cần tìm trong không gian chứa bằng phương pháp tìm kiếm theo bảng băm.

Lưu ý: các test case phải đa dạng có chứa đụng độ xảy ra.

BT4. (Nâng cao - Không bắt buộc)

Viết chương trình kiểm tra một đoạn văn bản tiếng Anh có từ nào viết sai lỗi chính tả không. Giả sử, đoạn văn bản chỉ chứa các từ ở thể nguyên mẫu (không chia, không số ít/số nhiều) và không xuống dòng. Bạn sẽ được cho trước một từ điển, nếu từ nào không xuất hiện trong từ điển nghĩa là bị lỗi chính tả.

Cho trước: từ điển tiếng Anh (https://gist.github.com/deekayen/4148741)

Command line: MSSV_BTx.exe Input.txt Output.txt

- Với input.txt chứa đoạn văn bản cần kiểm tra.

  Ví dụ:

  I luve you very mach.
- Output.txt liệt kê các từ bị lỗi có trong doạn văn và vị trí của nó. Mỗi từ một dòng. Giả sử từ đầu tiên trong văn bản được đếm là 1. Nếu không lỗi thì xuất ra một file rỗng.

  Ví dụ:

  luve 2

mach 5