

BÀI TẬP TUẦN 6

Môn học: Cấu Trúc Dữ Liệu và Giải Thuật.

GV hướng dẫn thực hành: Nguyễn Khánh Toàn (ktoan271199@gmail.com).

Nội dung chính:

- Hàng đợi ưu tiên.
- Bảng băm.

Thời gian thực hiện: 1.5 tuần.

Bài tập được biên soạn lại và tham khảo từ tài liệu thực hành môn Cấu Trúc Dữ Liệu và Giải Thuật của quý Thầy Cô Trợ Giảng khóa trước của bộ môn Khoa Học Máy Tính và quý Thầy Cô của bộ môn Công Nghệ Tri Thức. Trân trọng cảm ơn quý Thầy Cô của các bộ môn.

1 Cài đặt hàng đợi ưu tiên

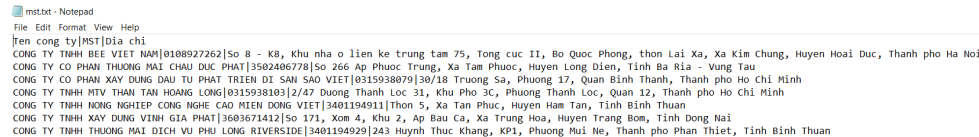
Các bạn sử dụng cấu trúc dữ liệu **min-heap** để cài đặt hàng đợi ưu tiên. Lưu ý cần thực hiện các thao tác sau:

- **getMin:** Trả về phần tử nhỏ nhất (nút gốc) của hàng đợi ưu tiên.
- **extractMin:** Trả về phần tử nhỏ nhất (nút gốc) của hàng đợi ưu tiên và thực hiện xóa nút gốc. Độ phức tạp của hàm này là $O(\log(n))$, các bạn sẽ cần phải hoán đổi nút gốc với nút cuối cùng của heap để thực hiện xóa và thực hiện heapify lại nút gốc.
- **insert:** Thêm phần tử mới vào hàng đợi ưu tiên. Các bạn sẽ thêm phần tử đó vào cuối của hàng đợi và thực hiện *heapify ngược* với nút đó để đưa nó về vị trí đúng, độ phức tạp của thao tác này là $O(\log(n))$.

2 Cài đặt bảng băm

2.1 Mô tả dữ liệu

Dữ liệu bao gồm mã số thuế và địa chỉ của 1250 công ty ở Việt Nam. Nội dung của dữ liệu được minh họa như hình dưới đây:



```
mst.txt - Notepad
File Edit Format View Help
[ten cong ty][MST][Địa chỉ]
CÔNG TY TNHH BEE VIỆT NAM[0108927262][Số 8 - K8, Khu nhà ở liên kết trung tâm 75, Tổng cục II, Bồ Quốc Phong, thôn Lai Xá, Xã Kim Chung, Huyện Hoài Đức, Thành phố Hà Nội]
CÔNG TY CỔ PHẦN THƯƠNG MẠI CHÂU DỤC PHÁT[3502406778][Số 266 Ấp Phước Trung, Xã Tân Phước, Huyện Long Điền, Tỉnh Bà Rịa - Vũng Tàu]
CÔNG TY CỔ PHẦN XÂY DỰNG ĐẦU TƯ PHÁT TRIỂN DI SẢN SÁO VIỆT[0315938079][30/18 Trương Sa, Phường 17, Quận Bình Thạnh, Thành phố Hồ Chí Minh]
CÔNG TY TNHH MTV THẠM TÀI HOÀNG LONG[0315938103][2/47 Đường Thanh 31, Khu Phố 3C, Phường Thanh Lộc, Quận 12, Thành phố Hồ Chí Minh]
CÔNG TY TNHH NÔNG NGHIỆP CÔNG NGHỆ CAO MIỀN ĐÔNG VIỆT[3401194911][Thôn 5, Xã Tân Phúc, Huyện Ham Tân, Tỉnh Bình Thuận]
CÔNG TY TNHH XÂY DỰNG VINH GIA PHÁT[3603671412][Số 171, Xóm 4, Khu 2, Ấp Bàu Ca, Xã Trung Hòa, Huyện Trang Bom, Tỉnh Đồng Nai]
CÔNG TY TNHH THƯƠNG MẠI DỊCH VỤ PHỤ LONG RIVERSIDE[3401194929][243 Huỳnh Thúc Kháng, KP1, Phường Mũi Né, Thành phố Phan Thiết, Tỉnh Bình Thuận]
```

Hình 1: Demo dữ liệu của các công ty.

Trong đó:

- Dòng đầu tiên mô tả thông tin của các trường dữ liệu (tên công ty, mã số thuế, địa chỉ).
- Những dòng tiếp theo, là nội dung dữ liệu của một công ty, cách nhau bởi dấu |.

Các bạn download dữ liệu ở đây: [Dữ liệu](#)

Trong bài tập này, các bạn được yêu cầu lưu trữ thông tin của toàn bộ các công ty trong một cấu trúc dữ liệu, nhằm phục vụ mục đích tìm kiếm nhanh thông tin của một công ty, cụ thể là lưu trữ với bảng băm.

2.2 Cài đặt các hàm chức năng

Mỗi cấu trúc dữ liệu công ty được định nghĩa ở hình 2.

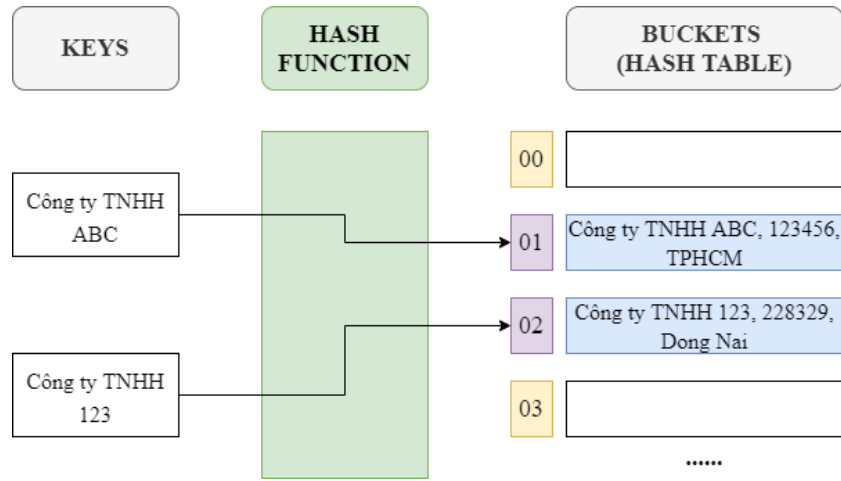
```
struct CongTy {
    string tenCongTy;
    string maSoThue;
    string diaChi;
};
```

Hình 2: Khai báo cấu trúc công ty

Các bạn có thể sử dụng *Class* nếu các bạn đã học các kiến thức về lập trình hướng đối tượng (OOP).

Tổ chức công ty bằng cấu trúc dữ liệu bảng băm

Các bạn tham khảo hình vẽ sau (hình 3):



Hình 3: Khai báo cấu trúc công ty

Ở đây ý tưởng là các bạn sẽ sử dụng một hàm hash để biến đổi từ tên công ty ra một giá trị index trong bảng băm và thực hiện lưu trữ dữ liệu của công ty tại đó.

Hàm chức năng 1 (Hàm băm).

Các bạn sẽ xây dựng hàm băm (hash function) nhận đầu vào là tên công ty và trả ra giá trị băm.

Chữ ký gợi ý: **long long** HashString(const **string** &tenCongTy)

Công thức của hàm băm:

$$hash(s) = \left(\sum_{i=0}^{n-1} (s[i] \cdot p^i) \right) \bmod m \quad (1)$$

trong đó:

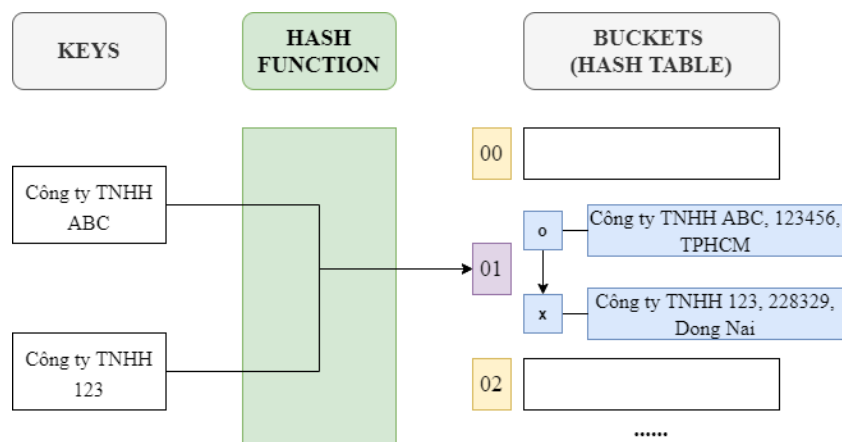
- **s**: 18 ký tự cuối cùng của tên công ty, nếu kích thước không đủ 18 thì sẽ lấy toàn bộ tên công ty làm giá trị đầu vào.
- **s[i]**: Giá trị ASCII của ký tự thứ *i* của chuỗi *s*.
- $p = 31$.
- $m = 100019$.

Hàm chức năng 2 (Xử lý đụng độ).

Lưu ý rằng, để kiểm soát đụng độ, các bạn sẽ chọn 1 trong 2 cách giải quyết sau:

Kết nối (chaining): Ứng với mỗi địa chỉ (chỉ số) trong bảng băm sẽ là một danh sách liên kết, mảng động hoặc có thể là một cấu trúc dữ liệu nào đó để tiện cho việc tìm kiếm hơn (mảng động đã được sắp xếp, cây nhị phân tìm

kiểm, v.v.). Chúng sẽ thực hiện lưu trữ toàn bộ các giá trị bị đụng độ tại mỗi vị trí trong bảng băm. Tham khảo hình vẽ sau (hình 4).



Hình 4: Xử lý bằng phương pháp kết nối trong trường hợp xảy ra đụng độ

Trong hình vẽ này, khi thêm công ty TNHH 123 vào bảng băm, sẽ xảy ra đụng độ tại vị trí bảng băm thứ 01, do đó thuật toán sẽ tiến hành thêm công ty TNHH 123 này vào vị trí thứ 2 của danh sách liên kết hoặc mảng động được tổ chức tại vị trí này của bảng băm. Lưu ý rằng danh sách liên kết tại vị trí 01 này trước đó đã chứa giá trị của công ty TNHH ABC.

Địa chỉ mở (open addressing): Khi xảy ra đụng độ, các bạn sẽ dò tìm vị trí thích hợp trong bảng băm để chứa phần tử đang cần thêm vào. Có 3 cách dò thường xuyên được sử dụng: dò tuyến tính (linear probing), dò bậc 2 (quadratic probing), băm kép (double hashing). Các bạn sẽ chọn 1 trong 3 cách để thực hiện nếu các bạn chọn giải quyết đụng độ bằng phương pháp địa chỉ mở.

Hình vẽ minh họa sử dụng dò tuyến tính (hình 5):

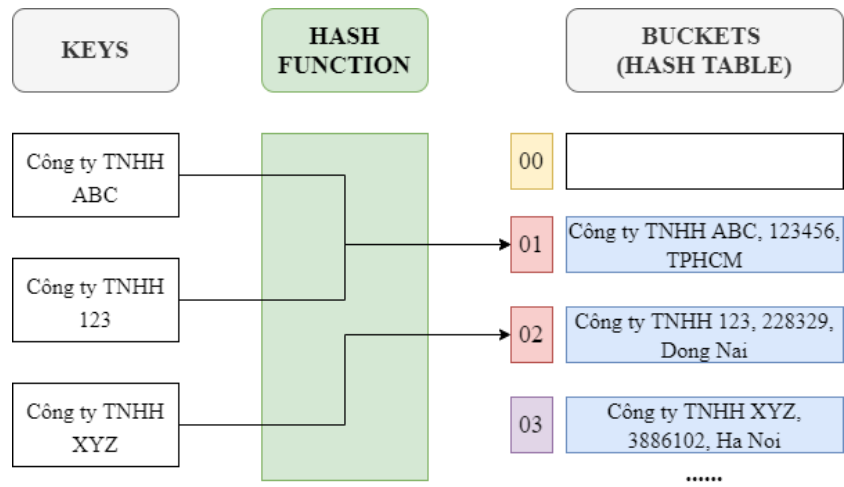
Trong hình vẽ này, khi thêm công ty TNHH 123 vào bảng băm, sẽ xảy ra đụng độ tại vị trí bảng băm thứ 01, do đó thuật toán dò tiếp tục ở các vị trí bảng băm tiếp theo. Thuật toán phát hiện thấy vị trí 02 trống, tiến hành thêm công ty TNHH 123 vào bảng băm vị trí thứ 02. Tương tự khi thêm công ty TNHH XYZ bị đụng độ tại bảng băm vị trí thứ 02, thuật toán tiến hành tìm vị trí thích hợp tiếp theo là 03 để thêm công ty này vào.

Hàm chức năng 3 (Xây dựng bảng băm).

Các bạn xây dựng hàm đọc thông tin toàn bộ công ty từ file, hàm này sau đó sẽ **thêm mỗi công ty đọc được vào bảng băm** đã định nghĩa. Hay nói cách khác các bạn cần cài đặt hàm thêm một phần tử vào bảng băm, lưu ý xử lý trường hợp đụng độ.

Hàm chức năng 4 (Tìm kiếm phần tử trong bảng băm).

Các bạn xây dựng hàm tìm kiếm thông tin của một công ty. Với giá trị cần tìm kiếm là tên của công ty, hàm này sẽ trả ra các thông tin của công ty đó nếu được lưu trong bảng băm, nếu không tồn tại sẽ trả ra giá trị -1.



Hình 5: Xử lý bằng phương pháp dò tuyến tính trong trường hợp xảy ra đụng độ

Hết