

Lab 06: Hash Table

Thời gian dự kiến: 01 tuần

1 Hash Table

1.1 Mô tả dữ liệu

Dữ liệu trong bài là mã số thuế của 1250 công ty Việt Nam, được lưu trong file *MST.txt*. Các dòng đầu của file có nội dung như hình bên dưới:

```
1 Ten cong ty|MST|Dia chi
2 CONG TY TNHH BEE VIET NAM|0108927262|So 8 - K8, Khu nha o lien ke trung tam 75, Tong cuc II, Bo Quoc Phong, thon Lai Xa, Xa Kim Chung, Huyen
  Hoai Duc, Thanh pho Ha Noi
3 CONG TY CO PHAN THUONG MAI CHAU DUC PHAT|3502406778|So 266 Ap Phuoc Trung, Xa Tam Phuoc, Huyen Long Dien, Tinh Ba Ria - Vung Tau
4 CONG TY CO PHAN XAY DUNG DAU TU PHAT TRIEN DI SAN SAO VIET|0315938079|30/18 Truong Sa, Phuong 17, Quan Binh Thanh, Thanh pho Ho Chi Minh
5 CONG TY TNHH MTV THAN TAN HOANG LONG|0315938103|2/47 Duong Thanh Loc 31, Khu Pho 3C, Phuong Thanh Loc, Quan 12, Thanh pho Ho Chi Minh
6 CONG TY TNHH NONG NGHIEP CONG NGHE CAO MIEN DONG VIET|3401194911|Thon 5, Xa Tan Phuoc, Huyen Ham Tan, Tinh Binh Thuan
7 CONG TY TNHH XAY DUNG VINH GIA PHAT|3603671412|So 171, Xom 4, Khu 2, Ap Bau Ca, Xa Trung Hoa, Huyen Trang Bom, Tinh Dong Nai
8 CONG TY TNHH THUONG MAI DICH VU PHU LONG RIVERSIDE|3401194929|243 Huynh Thuc Khang, KP1, Phuong Mui Ne, Thanh pho Phan Thiet, Tinh Binh Thuan
9 CONG TY TNHH HANH THANH PHAT|3603671155|To 5, Ap Thanh Binh, Xa Loc An, Huyen Long Thanh, Tinh Dong Nai
10 CONG TY TNHH THUONG MAI DICH VU THIET BI M.K.K|0315932380|154/1/34 Cong Lo, Phuong 15, Quan Tan Binh, Thanh pho Ho Chi Minh
11 CONG TY TNHH TM - DV - NHA HANG HAI SAN KY QUANG|0315933352|So 526 Duong Pham Van Dong, Phuong 13, Quan Binh Thanh, Thanh pho Ho Chi Minh
12 CONG TY TNHH MTV KIM LONG|1201613551|So 170 Nguyen Minh Duong, Ap 1, Xa Dao Thanh, Thanh pho My Tho, Tinh Tien Giang
13 CONG TY TNHH SONA AGENCY VIET NAM|0108926660|So 333 Bach Mai, Phuong Bach Mai, Quan Hai Ba Trung, Thanh pho Ha Noi
```

Trong đó:

- Dòng đầu là tên các trường dữ liệu.
- Các dòng tiếp theo, mỗi dòng là thông tin của một công ty. Các trường dữ liệu được ngăn cách nhau bởi dấu gạch đứng "|".

Trong bài thực hành này, sinh viên cần đọc thông tin các công ty trong file *MST.txt* vào một struct tên **Company**, sau đó lưu các struct này bằng cấu trúc dữ liệu hash table.

1.2 Bài tập

Cho định nghĩa struct **Company** như sau:

```
1 struct Company {
2     std::string name;
3     std::string tax_code;
4     std::string address;
5 };
```

Thực hiện các yêu cầu sau:

1. Đọc thông tin một công ty từ một file

```
std::vector<Company> readCompanyList(std::string fileName);
```

- Input: `fileName` – tên file dữ liệu, trong lab này là *MST.txt*
- Output: vector lưu các công ty đọc từ `fileName`

2. Hash một chuỗi tên công ty

```
long long hashString(std::string companyName);
```

- Input: chuỗi `companyName` – tên của công ty.
- Output: một số nguyên dương kiểu `long long` – kết quả hash `companyName` bằng hàm hash định nghĩa bên dưới.
- Sử dụng hàm hash sau:

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

trong đó

- s : 20 ký tự cuối của `companyName`. Nếu `companyName` không vượt quá 20 ký tự thì sử dụng toàn bộ chuỗi.
- $s[i]$: giá trị ASCII của ký tự ở vị trí i trong chuỗi s .
- $p = 31$
- $m = 10^9 + 9$

3. Tạo một hash table có 2000 phần tử từ danh sách các công ty. Nếu có đụng độ, sử dụng cơ chế Linear Probing (dò tuyến tính).

```
Company* createHashTable(std::vector<Company> listCompany);
```

- Input: `listCompany` – danh sách các công ty đã đọc ở câu 1.
- Output: hash table tạo từ danh sách trên.

4. Thêm một công ty vào hash table cho trước

```
void Insert(Company* hashTable, Company company);
```

- `hashTable`: hash table đã tạo trước.
- `company`: struct công ty cần thêm vào hash table (ta sẽ hash `company.name`)

5. Tìm một công ty theo tên

```
Company* Search(Company* hashTable, std::string companyName);
```

- Input:
 - `hashTable`: hash table cho trước.
 - `companyName`: tên công ty cần tìm.
- Output: Thông tin của một công ty, lưu trong struct `Company`. Nếu không tìm thấy, trả về con trỏ NULL