Problem Set 6: Searching

Please send back via NYU Classes

- A zip archive named as
  `PS06_<your name as FirstLast>.zip`
  containing the C code files that implements all aspects of all problems.

Total points: 100

In this problem set, you will write a program that searches a phone book containing 500 names and returns the information associated with a last name entered at the terminal. Your program will implement three search methods:

- Linear Search
- Binary Search
- Hash Search

This scenario is somewhat of a "toy" problem, in that a 500-name phone book is small, and the user must enter the exact last name string. However, it is sufficient to illustrate how the three searching methods work.

Your program will be structured into 4 files:

- **`main.c`** (portions provided by instructor)
- a framework that can be used for all searching methods
- **`main.h`** (provided by instructor – no need to add anything) parameters and data structures that are used by all files in the program
- **`search.c`** (portions provided by instructor) will be the code that performs the actual sorting
- **`search.h`** (provided by instructor – no need to add anything) the search function prototypes that will be called in **`main.c`**

This program is structured as multiple files, and we can compile and link it using the following command line:

```
gcc –o search main.c search.c
```

We can put this in a bash shell script **`build.sh`** to make life even easier.

If you are using CLion, you can drag the four files to the project directory and the IDE will take care of configuring the build.

Problem 1 (20 points): Parsing command line and reading data

Your program must have its command-line arguments as follows:

```
./search linear|binary|hash phone_book_file.csv
```

The | character indicates a choice amongst possible arguments.

`linear` specifies linear search
`binary` specifies binary search
`hash` specifies hash search
`phone_book_file.csv` is a file supplied by instructor containing an ASCII text, in the comma separated values format.

a. Parse the command line. Write code in `main.c` that parses the command line.

- If your program is executed without any command-line arguments, or with an incorrect number of command-line arguments, it should print a "usage" message and then exit.
- Determine if linear, binary, or hash is present and set a mode variable accordingly.
- Open the input file. Print an error with a diagnostic message if the file cannot be opened.

b. Initialize data structure. Your main() contains a data structure declaration:
`struct PhoneBook phone_book[PB_LEN]`

Initialize the data structure so that all strings are empty and the chain field that will be used in the hashed search has the value -1 (which is not a valid chain value)

```
    for (i=0; i<PB_LEN; i++) {
        phone_book[i].first[0] = 0;
        phone_book[i].last[0] = 0;
        phone_book[i].phone[0] = 0;
        phone_book[i].chain = -1;

}
```

c. Read phone book file. Read each line from the input file and enter the data into the phone book, one line per struct array element. After all data has been read, close the input file and optionally print the phone book, depending on the value of `db_flag`.

Using `strncpy`() is recommended for filling in the string fields. Adding a NULL character at the end of the character array will prevent array overflow when copying a string longer than the size of the string field, and will also ensure that the string is always NULL-terminated.

Problem 2 (15 points): Linear Search

Implement linear search, by filling in code for the function in **search.c**:

```
int linear_search(char *target, struct PhoneBook *phone_book, int num_entries)
```

The first argument is the last name to search, provided by the user in the terminal, and the second and third arguments specifies the phone book array and its length.

When an entry is found, return the index of the entry. In addition, be sure to treat the case in which the entry is not found, in which case the function should return a value of -1.

Problem 3 (20 points): Binary Search

Implement binary search, by filling in code for the function in search.c:

```
int binary_search(char *target, struct PhoneBook *phone_book, int num_entries)
```

This function uses the same argument, but leverages the fact that the phone book is ordered by the last name. The sorting is already implemented in the provided **`main.c`** using **`qsort`** and **`strncmp()`**.

Similarly, when an entry is found, return the index of the entry. In addition, be sure to treat the case in which the entry is not found, in which case the function should return a value of -1.

Problem 4 (35 points): Hash Search

Implement hash search. In **search.c**, a simple hash function is provided:

```
int hash_func(char *s)
```

This function returns an integer hash, that tries to produce drastically different values for even slightly different strings. For example, the hash value of "aa" is 6208, and of "aaa" is 14369. In the hash map, these numbers will be used as indices of an array, to quickly access the information corresponding to a key string.

This problem consists of two subproblems; the first one is to build a hash map in order to use it later for searching repeatedly, and the second one is to actually perform searching.

a. Create hash map. Complete the function:

```
void create_hash_map(struct PhoneBook *phone_book, int num_entries)
```

The data structure is set up to use chaining to resolve hash collisions, with field **`chain`** of **`struct PhoneBook`**. Your implementation should:

- For each last name in the phone book, call **`hash_func()`** using the last name string. That is, for each **`i`**, use **`phone_book[i].last`** as the argument to **`hash_func().`**

Use the value **`j`** returned from **`hash_func()`** as the index in the global array **`hash_map[]`**. The array value for that index **`j`** is the phone book index **`i`** for the last name **`phone_book[i].last`** that is being hashed.

- Resolve hash collisions. For a given hash value **`j`**, if the value in **`hash_map[j]`** is already "taken" (unused entries were initialized to -1), then use the chain feature to resolve the hash collision conflict:
    - Given **`j = hash_func(phone_book[i].last)`**
    - If (**`hash_map[j] == -1)`** then set **`hash_map[j] = i`** and you are done.
    - Otherwise, **`k = hash_map[j],`** and this is a hash collision.
    - Inspect **`phone_book[k].chain.`** If this is -1, set **`phone_book[k].chain = i`** and you are done.
    - Otherwise, follow the chain until you fine a chain value = -1, and set it equal to **`i`**.

The code provided by the instructor prints statistics of the hash map. It should look like:

```
Hash table statistics:
2.99% map occupancy; top map entry 16290 out of 16384 entries
Collisions: 10 ( 0.06%)
Average chain length 1.00
```

b. Implement hash search. Complete the function:

`int hash_search(char *target, struct PhoneBook *phone_book)`

This should:

- Calculate the hash value for the target last name, that was entered by the user. Use your function hash_func() for this.
- Get the entry in hash_map[] using the hash value as the index. This is the index into the
- phone book data structure. Check the value of chain at that index. If it is -1 (no collision)
- then you are done, just return the index.
- If not -1 (that is, if there was a collision for this hash value):
    - Check to see if the current last name is the one that you are looking for. If so, then return the index.
    - Otherwise, follow the chain index to the indicated phone book structure, and check to see if the current last name is the one that you are looking for. Repeat until the desired entry is found.
    - If you should get to a chain value of -1 without finding the desired entry, it means there are no such last name in the phone book. Return -1.

Problem 5 (10 points): Time Complexity

What are the Big-O time complexities of the three search algorithms? What are the tradeoffs of binary search and hash search, for being faster than linear search?

Put your answers in file TimeComplexity.txt and submit this file along with your C source code files.