CMP-210: Data Structures and Algorithms
Instructor: Waheed Iqbal

Assignment 03: Search Engine Using Binary Search Tree

Submission Deadline: Sunday 10:00pm, 3rd January, 2017

Search Engine

Suppose you have developed a small search engine that can help users to search over the Internet. Your search engine provides set of URLs in response to specific search queries issued by the users. We have provided you two files:

a. **`search-results.txt:`** each line in provides search query and set of URLs containing the result of the search query

```
aalii: URL-24403, URL-18952, URL-26651, URL-17969, URL-14486, URL-7807, URL-20901, URL-25351, URL-16591, URL-13707, URL-1693, URL-1536, URL-29845, URL-8190, URL-20835, URL-20087, URL-29873, URL-1266, URL-10556, URL-26041, URL-0
Aani: URL-29108, URL-18702, URL-0
aardvark: URL-4789, URL-24573, URL-9520, URL-7938, URL-14038, URL-19686, URL-10922, URL-13944, URL-1310, URL-15810, URL-345, URL-12337, URL-12694, URL-28583, URL-4741, URL-8001, URL-21842, URL-0
aardwolf: URL-5869, URL-2406, URL-500, URL-26090, URL-19842, URL-20036, URL-7401, URL-15287, URL-9882, URL-239, URL-0
……..
```

b. **`search-history.txt`**: Each line in search-history contains the search queries issued by the users.

```
nonevolutionist
lacunosity
unjustifiably
dawny
sunspot
autointoxication
aboriginary
bekinkinite
rachilla
katacrotism
……….
```

You need to build a binary search tree based on **`search-results.txt`** by using the following guidelines:

1. You need to identify top 10,000 search queries based on probability from the **`search-history.txt`**
2. Use search query as a key in the binary search tree.
3. The comparison between keys should follow lexicographic (dictionary) order.
4. Each node contains a set of URLs against the key.
5. For top 1,000 nodes you should have the set of URLs in memory but for the rest of the nodes, each node points to a file containing set of URLs against the search query.

Once you build the tree based on the above guidelines, your program should read an input from the user as a search query and print appropriate set of URLs by querying the binary search tree. The output should also mention that URLs are read from the memory or from the file. For example:

```
Please enter search query: Aaru
Output: URL-2661, URL-21975, URL-519, URL-17318, URL-20031, URL-8652, URL-22149, URL-0
URLs Read From: memory
```