Lab Assignment 2: CS2233

August 28, 2024

Question:

Consider the following dataset click here. The dataset consists of a key parameter - `NHS No.`, and three value parameters, `first name, email and gender`, respectively. The number of elements in the datasets is $n = 1000$.

Recall the static `dictionary problem` in which, after creating the dictionary, no further insertion/deletion is allowed. Create a static dictionary data structure using `perfect hashing` algorithm. The space required to store the elements should be at most cn, for some constant c.

`Hint:` First, you need to create a family of universal hash functions from $\mathcal{H} : \mathcal{U} \mapsto \{0, 1, \ldots, m-1\}$, where $|\mathcal{U}| = 10^{10}$, and m denotes the number of slots in the hash table. You can use the dot-product hash function for the same (covered in the class). You need to create two levels of hashing using a universal hash function. For the first level of hashing pick $h_1 \in_R \mathcal{H}$ by setting $m = 5n$. If $\sum_{i=0}^{m-1} l_i^2 > 5n$, then try with another hash function h_1, where l_i denotes the number of keys mapped in the i-th slot. For each of the keys mapped in i-th slot, pick another hash function $h_{2,i} \in_R \mathcal{H}$ by setting $m = 10l_i^2$. If there are collisions in the second level repick $h_{2,i}$ for $i \in \{0, \ldots, m-1\}$.

Test criteria: Once you have created the static dictionary, to evaluate the code, write a search function in which you pass the key "`search(NHS No)`", and it should return the corresponding parameters `first name, email and gender` if the key is present, otherwise it should report `key not found`.

20 Marks