DATA STRUCTURES

ASSIGNMENT 1: HASH TABLES

SUGANDHI SAGGU

21355223

Hash Table is a data structure which stores data in an associative manner. In a hash table, data is stored in an array format, where each data value has its own unique index value. Access of data becomes very fast if we know the index of the desired data.

Thus, it becomes a data structure in which insertion and search operations are very fast irrespective of the size of the data. Hash Table uses an array as a storage medium and uses hash technique to generate an index where an element is to be inserted or is to be located from.

Task 1:

In this task we have to parse a .csv file to hash table in such a way that it stores the surnames of the people according to the provided hash function. We are using linear probing in this task. I created structures and several functions to make the code for efficient.

I took some help from the skeleton code provided as well. The general idea was to parse the file and keeping the count of unique surnames, frequency of repeated surnames and the number of collisions that occurred while storing the surnames into the hashtable. Practically, we intend to keep the size of array larger than what it is intend to use so that less collisions occurs.

Task 2:

In this task, we have to change the hash function in such a way that the number of collisions gets reduced.

Task 3:

The concept of double hashing is introduced to improve the code and to reduce the number of collisions. In this method we hashed the table once and if collision occurs the key is hashed again and then it stored into the table. Double hashing can be done using :

(hash1(key) + i * hash2(key)) % TABLE_SIZE

Where hash1 and hash2 are hash functions and table_size is the size of the hash table array and ‘i’ is used as an increment factor.

Task 4:

In this code, we parsed the different csv file in hash table such a way that the whenever we hashes the surname, we attach the linked list of the person’s entire available data and it is displayed when we enter the surname. The approach of double hashing is used in this code as well.