CSCE 221 Cover Page – Programming Assignment 4

First Name: Maria Last Name: Dmitrievskaia UIN: 927009911

User Name: E-mail address: dmimar382@tamu.edu

Please list all sources in the table below including web pages which you used to solve or implement the current homework. If you fail to cite sources you can get a lower number of points or even zero, read more on Aggie Honor System Office website: http://aggiehonor.tamu.edu/

Type of sources				
People				
Web pages (provide URL)	http://www.cplusplus.com/reference/string/stoi/			
Printed material				
Other Sources				

I certify that I have listed all the sources that I used to develop the solutions/codes to the submitted work. On my honor as an Aggie, I have neither given nor received any unauthorized help on this academic work.

Your Name: Maria Dmitrievskaia Date: 04/11/20

Description of the assignment objective:

The purpose of the assignment is to write a program that would help record every student's grades in a spreadsheet roster. This program can assist graders in gathering data from online platforms and keep track of grades and completed assignments for all the students.

Description of data structures and algorithms used by my program:

The primary data structure used in my program is a hash table. The hash table data structure is used for storing data in an organized manner. Here, I create a hash table and insert a students UIN and grade into the table. In order to decide where to place the data, a hash function. This hash function calculates the remainder of the UIN divided by the size of the hash table, thus returning the index at which to insert the data. If something already exists at that particular index, then the data is added to the "chain" of values at that index (linked list at that index). While searching for the UIN, all that we need to do is go to the index that the data was inserted at, and traverse through the linked list to find the object or to verify that that object is not in the list or hash table. I also implemented functions to calculate the minimum, maximum, and average chain lengths. In these calculations, I am taking empty lists into account. So, if my vector consists of 3 linked lists, one of which is empty, the second of which has 2 members and the third of which has 5 members, min will return 0, max will return 5, and average will return (0+2+5)/3 = 2.3.

Description of input and output data. List all restrictions and assumptions that you have imposed on your input data and program.

The input data consists of the input and roster files. The input file contains data about each student who has completed the assignment: their name, e-mail, UIN, and grade for an assignment. The roster file contains information about each student in the class: their name, e-mail, and UIN. In order to create the output file, the program goes through the roster list and searches for the students UIN in the created hash table. If the grade exists for that student, then the grade, along with the student's name, e-mail and UIN are printed to the output file. If there is no grade for the student, then only their name, e-mail and UIN are printed to the output file. An assumption I have placed on the input data is that the input data exists in the file. The number of students in the program is designated as an integer m. If that integer is 0, then the program returns.

How have you tested your program for corrections?

To test my program for correctness, I verified that the insert function was working properly with different sizes of the hash table. I did this by creating output statements in my minimum and maximum chain length functions that would show how many elements are chained together at a certain vector position. Through doing this, I could also see what the minimum, maximum, and average chain lengths should be. The search function was tested through the write to file function, where the UINs from the roster were looked up in the hash table to see if the student had a grade for the assignment. The correct output verified that the search function was working correctly.

Which C++ features or standard library classes have you used in your program?

Throughout this programming assignment, I have used the lists, vectors, pairs, regex, and strings standard library classes.

Provide the statistics about the hash table. Are the computational results about the hashing consistent with the expected running time for the hashing algorithm? Justify your answer.

The statistics about the hash table are provided through the min chain length, max chain length, and average chain length functions. For the provided inputs, the minimum chain length is 0, the maximum chain length is 1, and the average chain length is 0.17. These results are expected, as the unique UINs will never cause collisions, so that the maximum chain length is always 1. Because the hash table has zero entries, the minimum is 0. Because the hash table has a size of 100 and only 17 entries not equal to 0, the average size is 0.17. These results are consistent with the expected running time of the hashing algorithm of O (1). Traversing through the whole hash table is not necessary to find or insert a value, because of the know index at which we need to insert a value, or to look for it. Thus, the running time is O (1) and is consistent with the expected running time.

Conclusion:

Through this assignment, I have learned about the advantages of using a hash table and the ease of accessing information thanks to hash tables. I am able to access information in a faster manner, as compared to arrays, linked lists, and binary trees, because of knowing where the item should be located in the hash table. I am also now familiar with the regex standard library class.