LSDS-HW3

Student Name: Nitin Prakash
Student Email Id: np679@nau.edu
Student NAU ID: 6255364

A. **(30 pts) Assess the impact of the hash table size.** Set the hash table to a fixed value (**m**, see below). Set the size of your hash table (***m***) to 1 million, 10 million, 30 million, and 60 million elements. Populate the hash table with the sequence fragments from the *query dataset*.

• For each of your 4 hash table sizes, how many collisions did you observe while populating the hash?

M Size	collisions
1M	27510406
10M	25869924
30M	22972624
40M	19946973

• For each of your 4 hash table sizes, how long did it take you to populate the hash table? Do the timing results make sense? Explain.

M Size	collisions	Time Taken
1M	27510406	46.39765
10M	25869924	45.00872
30M	22972624	42.66553
40M	19946973	39.71903

When we are increasing the hash table size than number of collisions will be reduced due to this we can insert data faster in Hash table which has bigger size. This also depends on number of that which we are inserting into hash table.

B. **(30 pts) Searching speed**: Set the hash table size to 60 million. Populate the hash table with the sequence fragments from the *query dataset*. Read in the entire *subject dataset* into a single, concatenated character array (same way you did it in HW#1). Implement a search function which would search for 16-character fragments of the subject sequence within the Queries_HT object. Iterate through all 16-character long fragments of the *subject dataset*, searching for each one in the *query dataset*.

• How long did it take to search for every possible 16-character long fragment of the *subject dataset* within the *query dataset*?

.Our Progrom took 4705.65 seconds to search for every possible 16-character long fragment of the subject dataset within the query dataset for m value 60 Million into hash table.

• How many such fragments did you find?

fragment found: 542456454

• Print the first 10 fragments of the *subject dataset* that you found within the Query_HT.

Found :CTAACCCTAACCCTAA

Found :TAACCCTAACCCTAAC

Found :AACCCTAACCCTAACC

Found :ACCCTAACCCTAACCC

Found :CCCTAACCCTAACCCT

Found :CCTAACCCTAACCCTA

Found :CTAACCCTAACCCTAA

Found :TAACCCTAACCCTAAC

Found :AACCCTAACCCTAACC

Found :ACCCTAACCCTAACCC