Name: Parth Motibhai Patel
NauId: pp594

Usage: ./executable <genome_file_path> <query_file_path> <(integer)hash_table_size> <search/hash>

Argument 1: genome file path.
Argument 2: query file path.
Argument 3: Hash table Size. → Integer number for the hash table size.
Argument 4: search/hash → whether to just create the hash table or search the whole genome file through it as well.

Example:

```
./executable /common/contrib/classroom/inf503/genomes/human.txt
/common/contrib/classroom/inf503/human_reads_2_trimmed.fa 1000000
```

Ans A.
1M

```
total collisions: 31668764
Time taken to create hash table: 30.6452 seconds
```

10M

```
total collisions: 29967957
Time taken to create hash table: 32.6999 seconds
```

30M

```
total collisions: 26815039
Time taken to create hash table: 32.9711 seconds
```

60M

```
total collisions: 23137590
Time taken to create hash table: 32.7122 seconds
```

Explanation behind the similar times for creating hash tables:

The relatively similar timings for populating different sizes of hash tables is because I am calculating the radix of all 16 characters of the query string and that will take O(n) (where n is 16) time and the number of queries doesn't change which makes this operation common to all, Apart from that for collision handling, I am adding the node to the head of a linked list which is O(1) time which means it doesn't matter how many times the collision occurs the node insertion time will always be O(1),hence the time taken to populate the hash table is majorly dependent on the number of queries and is independent of the Hash table size.

Ans B.

Time take to search all 16 char fragments : 4090.38 seconds

Total fragments found: 630073224

First 15 fragments found:

```
subsequence 0 : CTAACCCTAACCCTAA
subsequence 1 : TAACCCTAACCCTAAC
subsequence 2 : AACCCTAACCCTAACC
subsequence 3 : ACCCTAACCCTAACCC
subsequence 4 : CCCTAACCCTAACCCT
subsequence 5 : CCTAACCCTAACCCTA
subsequence 6 : CTAACCCTAACCCTAA
subsequence 7 : TAACCCTAACCCTAAC
subsequence 8 : AACCCTAACCCTAACC
subsequence 9 : ACCCTAACCCTAACCC
subsequence 10 : CCCTAACCCTAACCCT
subsequence 11 : CCTAACCCTAACCCTA
subsequence 12 : CTAACCCTAACCCTAA
subsequence 13 : TAACCCTAACCCTAAC
subsequence 14 : AACCCTAACCCTAACC
```

Raw Output for 60M search:

```
Sun Nov  5 16:40:25 MST 2023

./main /common/contrib/classroom/inf503/genomes/human.txt
/common/contrib/classroom/inf503/human_reads_2_trimmed.fa 60000000 search

Genome Length: 3057186663
created hash table
total queries: 31930886
total collisions: 23137590
Time taken to create hash table: 23.7766 seconds
searching queries:

subsequence 0 : CTAACCCTAACCCTAA found.
subsequence 1 : TAACCCTAACCCTAAC found.
subsequence 2 : AACCCTAACCCTAACC found.
subsequence 3 : ACCCTAACCCTAACCC found.
subsequence 4 : CCCTAACCCTAACCCT found.
subsequence 5 : CCTAACCCTAACCCTA found.
subsequence 6 : CTAACCCTAACCCTAA found.
subsequence 7 : TAACCCTAACCCTAAC found.
subsequence 8 : AACCCTAACCCTAACC found.
subsequence 9 : ACCCTAACCCTAACCC found.
subsequence 10 : CCCTAACCCTAACCCT found.
subsequence 11 : CCTAACCCTAACCCTA found.
subsequence 12 : CTAACCCTAACCCTAA found.
subsequence 13 : TAACCCTAACCCTAAC found.
subsequence 14 : AACCCTAACCCTAACC found.

threads created: 64
total fragments found: 630073224

Time taken to search 3057186648 fragments: 4090.38 seconds
Program ended
```