CS 101 Spring 2018 Project 3

Due date: April 3,2018

Description

This project will be about building a hash table. You will read words, separated by whitespace, from a given file into your program. As you read the words, you should insert each word into a hash table. The word itself is the key value, and each entry in the hash table also needs to store a list of the line numbers where the word appears. When inserting a word that already appears in the hash table, simply add the current line number to the list for the word. You will also count the number of collisions that occur while inserting words into the hash table. Every probe to a cell that contains a word other than the one being inserted or searched should count as an additional collision, so that inserting a word may have multiple collisions. Probes to empty cells or cells already containing the word do not count as collisions.

Once you have read all of the words, you should output the total number of words read, the number of distinct words, and the total number of collisions that occurred in the hash table.

Finally, you will read words from the query file and report the line numbers where the word occurs, and the number of collisions while searching for the word in the table.

Requirements

Please carefully read the following requirements:

- You must supply a makefile that builds an executable named project3
- You must use C++ streams for all I/O
- You must format your output as shown in the example below.
- You must do your own work, you must not share code.
- You must submit your project in a zipfile as specified under submission
- There will be four command line arguments to your program.
    1. The name of the input file
    2. The name of the query file
    3. The size of the hash table
    4. The collision resolution strategy, lp for linear probing, qp for quadratic probing, and dh for double hashing. If the strategy is double hashing, the double hash function will be of the form h2(x) = a – (x % a), and the integer parameter a will be the fifth command line argument.
- You should use the “djb2” hash function. (http://www.cse.yorku.ca/~oz/hash.html)

Example

```
brandondixon:~/workspace $ more hashtest1.txt
This is a sample input file
It is ten words
brandondixon:~/workspace $ more hashq1.txt
sample
is
ten
this
a
void
five
brandondixon:~/workspace $
brandondixon:~/workspace $ make
g++ hash.cpp -o project3

brandondixon:~/workspace $ ./project3 hashtest1.txt hashq1.txt 13 qp

The number of words found in the file was 10
The number of unique words found in the file was 9
The number of collisions was 5

sample appears on lines [1]
The search had 0 collisions

is appears on lines [1,2]
The search had 0 collisions

ten appears on lines [2]
The search had 0 collisions

this appears on lines []
The search had 1 collisions

a appears on lines [1]
The search had 0 collisions

void appears on lines []
The search had 1 collisions

five appears on lines []
The search had 6 collisions


brandondixon:~/workspace $ ./project3 hashtest1.txt hashq1.txt 13 lp
The number of words found in the file was 10
The number of unique words found in the file was 9
The number of collisions was 3

sample appears on lines [1]
The search had 0 collisions

is appears on lines [1,2]
The search had 0 collisions
```

ten appears on lines [2]
The search had 0 collisions

this appears on lines []
The search had 1 collisions

a appears on lines [1]
The search had 0 collisions

void appears on lines []
The search had 1 collisions

five appears on lines []
The search had 3 collisions

brandondixon:~/workspace $ ./project3 hashtest1.txt hashq1.txt 13 dh 3

The number of words found in the file was 10
The number of unique words found in the file was 9
The number of collisions was 6

…

brandondixon:~/workspace $ ./project3 big.txt hashq2.txt 91111 qp

The number of words found in the file was 1095695
The number of unique words found in the file was 81397
The number of collisions was 370273

twelve appears on lines
[1759,1829,3152,3807,7594,7615,9913,10786,13261,13412,18823,24515,25454,25454,30604,30655,35927,36156,36607,39977,39978,41131,43717,56494,59389,68293,72830,74120,86640,94115,117167,128073]
The search had 0 collisions

Euclidean appears on lines []
The search had 63 collisions

Earth appears on lines [7371]
The search had 0 collisions

brandondixon:~/workspace $ ./project3 big.txt hashq2.txt 113813 dh 74327

The number of words found in the file was 1095695
The number of unique words found in the file was 81397
The number of collisions was 288122

twelve appears on lines
[1759,1829,3152,3807,7594,7615,9913,10786,13261,13412,18823,24515,25454,25454,30604,30655,35927,36156,36607,39977,39978,41131,43717,56494,59389,68293,72830,74120,86640,94115,117167,128073]
The search had 0 collisions

Euclidean appears on lines []
The search had 5 collisions

Earth appears on lines [7371]
The search had 2 collisions

brandondixon:~/workspace $