Shweta Jones
CSE 13S - Spring 2021
Prof Darrell Long

Assignment 7: Write-Up

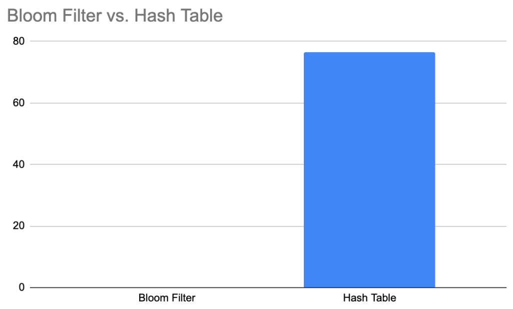
Description:

This program uses several different ADTs including nodes, linked lists, hash tables, bit vectors and bloom filters. The program keeps track of badspeak and oldspeak words to decide whether or not the user gets accused of a thoughtcrime. When a basspeak doesn't have a translation, the user is accused of thoughtcrime and sent to joycamp, and otherwise they are sent to counsel. There are a few command line options: h, t, f, m, s. The '-h' option points out the help information which provides program usage information. The '-t' option specifies the size of the hash table, if this is not specified, the default size is set to 10000. If the user writes '-f', this specifies the size of the bloom filter (the default is 2^20). The '-m' option enables mtf, or move-to-front rule. If this is not selected, this rule is not applied. This '-s' option allows for the user to see the stats of the program which are: total number of seeks, average seek length, hash table load, and the bloom filter load. This program is run on a terminal and requires linux/unix. THe program also requires several c files: node.c, ll.c, ht.c, bf.c, bv.c, parser.c, and speck.c, along with their corresponding h files. The program also uses badspeak.txt and newspeak.txt. And finally the main file we run is banhammer.c.

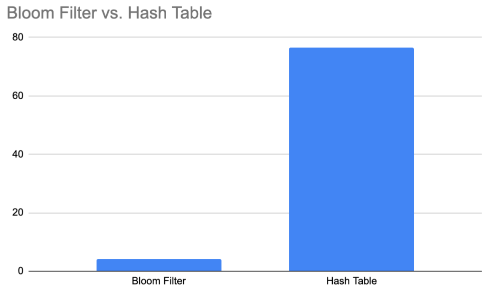
Graphs:

Comparing Hash Table and Bloom Filter load values with Default:

Mine:



Prof. Long's:



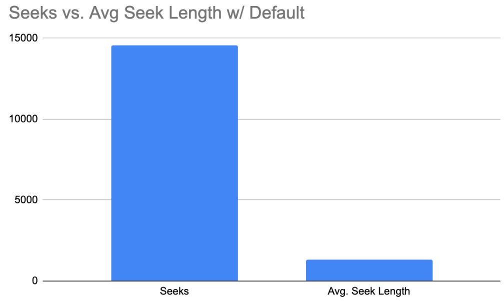
Graph Analysis:

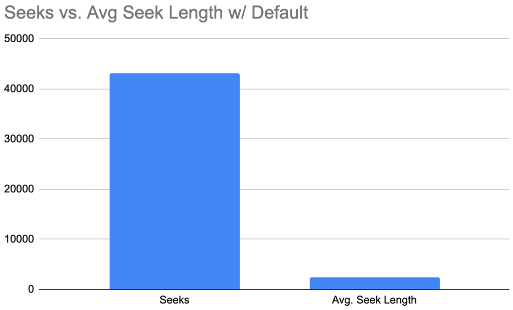
Based on the graphs you can observe significant differences between the expected vs. present values. Based on the values that were determined by my program, the percentage determined for bloom filters is an extremely miniscule amount for the default values, however, the hash table value was the exact same. <u>The linked lists do not change in size.</u>

Comparing Seeks and Avg. Seek Length with Default:

Mine:

Prof. Long's:





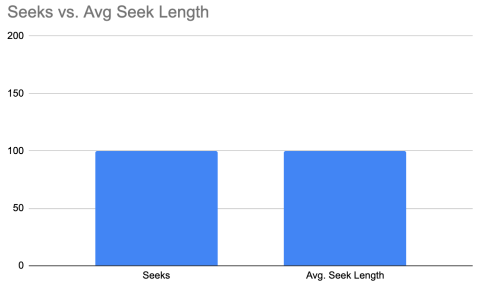
Graph Analysis:

While at first glance, these graphs look similar you can notice that these have different values, however, the difference is comparable. The similarity could be a result of incorrect calculations within the ll.c file where the numbers of seeks and links are calculated.

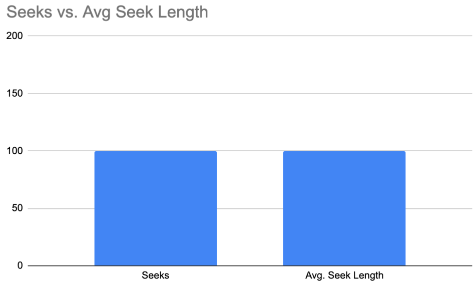
Seeks vs. Avg. Seek Length

Equivalent Hash Table and Bloom Filter:

With MTF:



No MTF:

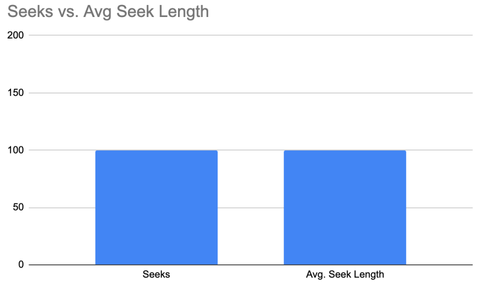


Graph Analysis:

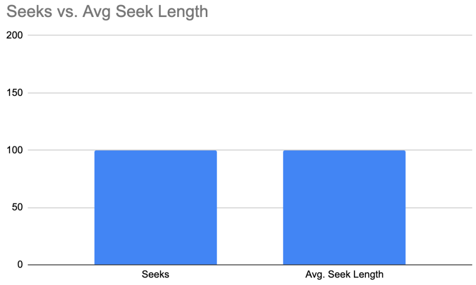
If the size of the bloom filter and hash table are the same, the load values are also the same. The second graph shows when the mtf is enabled, and you can observe that the value for seeks and average seek length is the same.

Larger Hash Table:

With MTF:



No MTF:

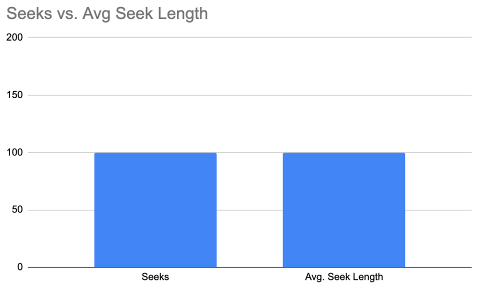


Graph Analysis:

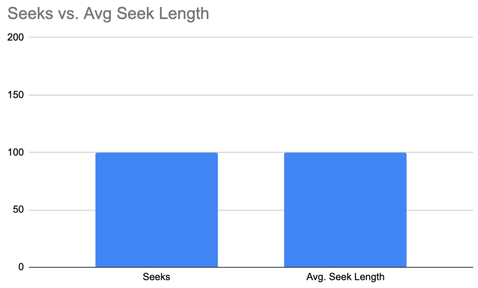
If the size of the hash table is smaller, the load values are the same. The second graph shows when the mtf is enabled, and you can observe that the value for seeks and average seek length is the same.

Larger Bloom Filter:

With MTF:



No MTF:

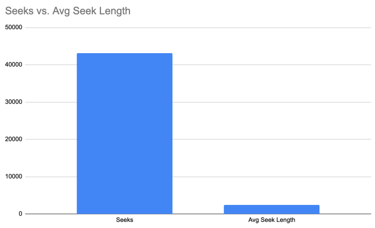


Graph Analysis:

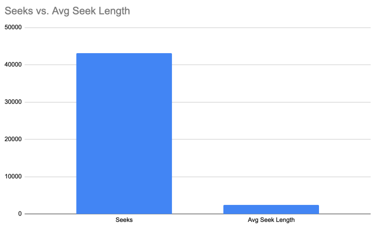
If the size of the bloom filter is larger than the hash table, the load values are the same. The second graph shows when the mtf is enabled, and you can observe that the value for seeks and average seek length is the same.

The Effect of Bloom Filter on Number of Lookups:

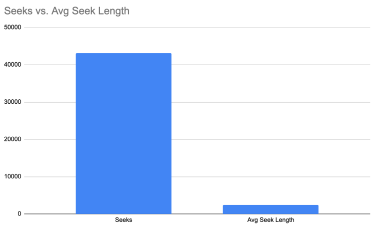
Bloom Filter Size 10:



Bloom Filter Size 50:



Bloom Filter Size 100:



Graph Analysis:

You can see that the size of the bloom filter does not affect the number of lookups, or the links or seeks count. So changing the bloom filter size does not affect the number of lookups performed in the hash table.