Algo Data Labb 3

1.

Output sample

```
editions  all of which are confirmed as Public Domain in the U S
unless a copyright notice is included   Thus  we do not necessarily
keep eBooks in compliance with any particular paper edition


Most people start at our Web site which has the main PG search facility

     http   www gutenberg org

This Web site includes information about Project Gutenberg tm
including how to make donations to the Project Gutenberg Literary
Archive Foundation  how to help produce our new eBooks  and how to
subscribe to our email newsletter to hear about new eBooks
non alphabetic: 61150
Charachters: 788059
```

Prints the filtered text, how many non-alphabetic characters and the total number of characters for the text.

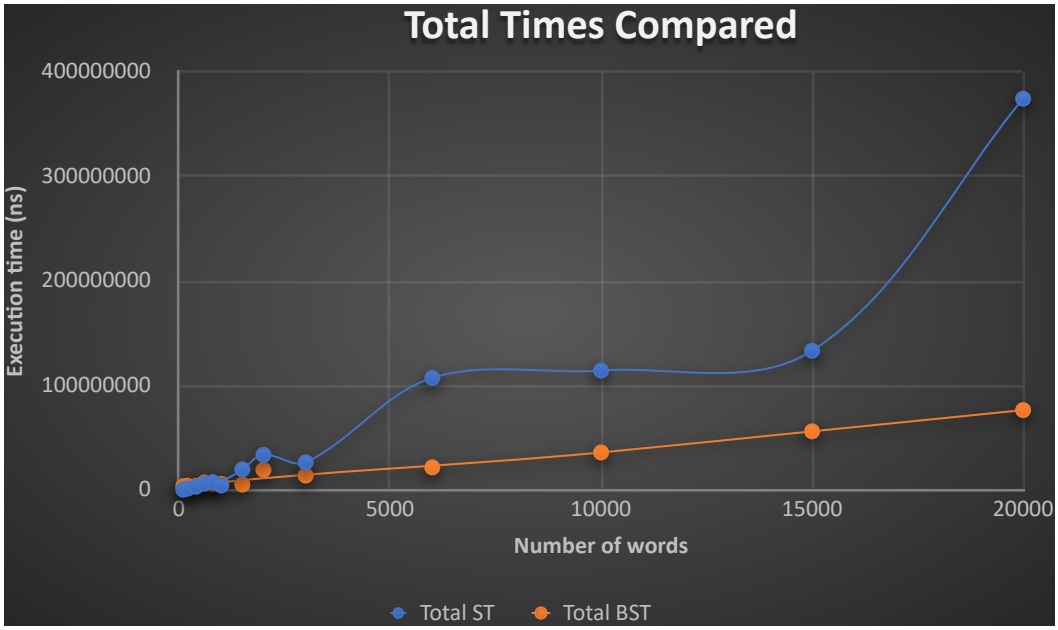
2.

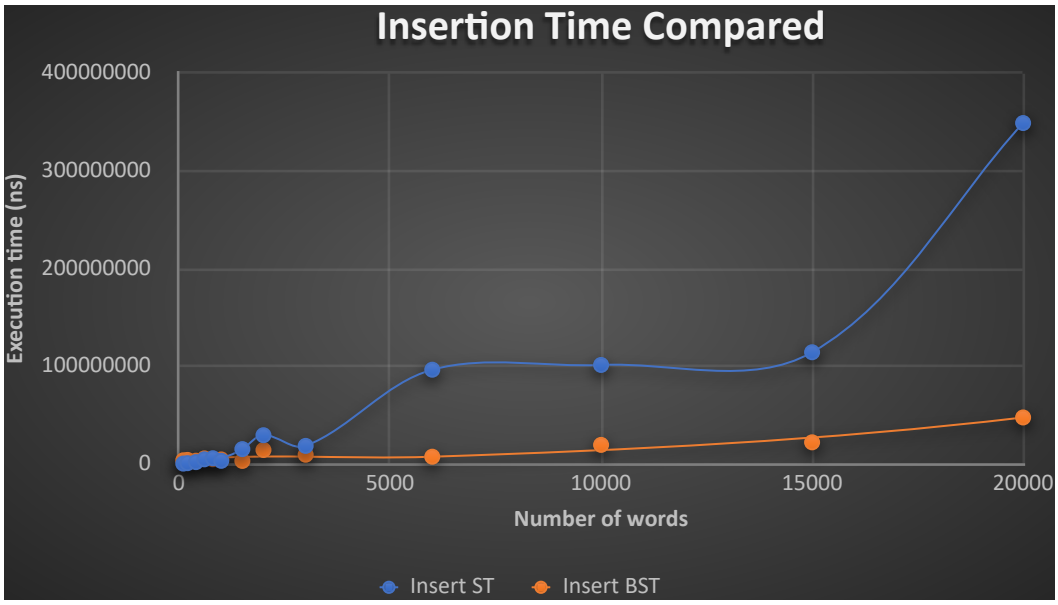
Output sample

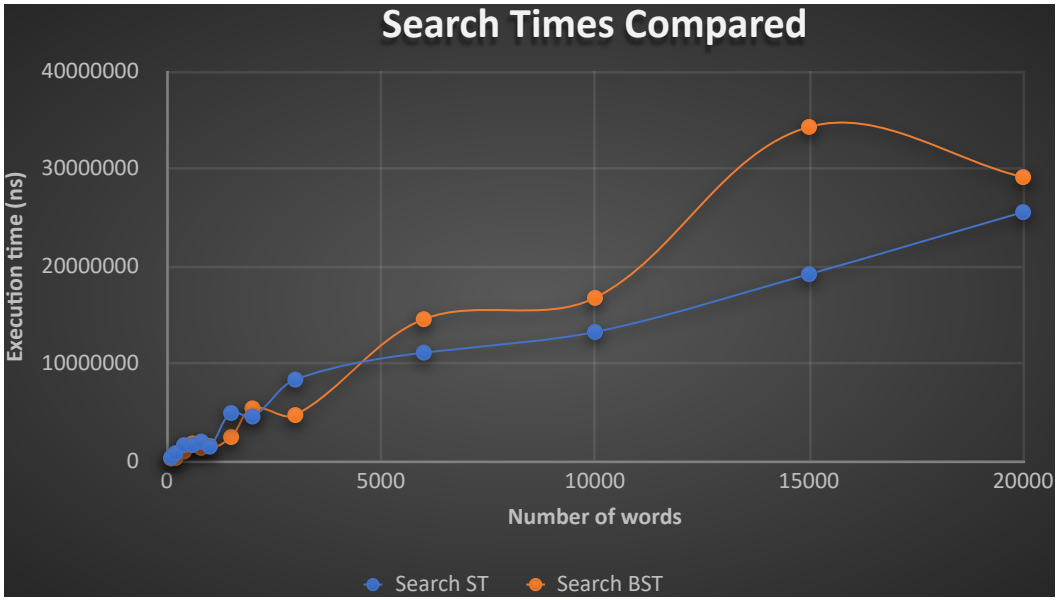
```
run:
Length: 1000
Ordered Array ST insert execution time: 2880098ns
Ordered Array ST search execution time: 988410ns
Most frequent word is "the" and it is used 78 times
Distinct words = 468
Total words    = 1000
MINLEN         = 1
Binary Search Tree insert execution time: 4323385ns
Binary Search Tree search execution time: 587106ns
Most frequent word is "the" and it is used 78 times
Distinct words = 468
Total words    = 1000
MINLEN         = 1
BUILD SUCCESSFUL (total time: 0 seconds)
```

Graphs

Number of Words (N)	Insert ST	Insert BST	Search ST	Search BST	Total ST	Total BST
100	347379	3703442	411004	305590	758383	4009032
200	775704	4114893	847800	353370	1623504	4468263
400	1861994	3320879	1682144	1035959	3544138	4356838
600	4714821	5665596	1686928	1853501	6401749	7519097
800	5836109	4844799	2044492	1393518	7880601	6238317
1000	3188335	4592097	1526991	1601979	4715326	6194076
1500	15175848	3088535	4982550	2513901	20158398	5602436
2000	29409732	14259452	4613103	5451583	34022835	19711035
3000	18491520	9453847	8412648	4769323	26904168	14223170
6000	96498014	7395653	11175641	14615467	107673655	22011120
10 000	101380293	19512072	13301959	16805024	114682252	36317096
15 000	114381849	22139289	19241105	34358332	133622954	56497621
20 000	349132514	47602053	25608255	29200854	374740769	76802907


Total Times Compared
400000000
300000000
200000000
100000000
0
Execution time (ns)
0
5000
10000
15000
20000
Number of words
Total ST
Total BST


Insertion Time Compared
400000000
300000000
200000000
100000000
0
Execution time (ns)
0
5000
10000
15000
20000
Number of words
Insert ST
Insert BST


Search Times Compared
40000000
30000000
20000000
10000000
0
Execution time (ns)
0
5000
10000
15000
20000
Number of words
Search ST
Search BST

Observations from graphs:

implementation	guarantee			average case			ordered ops?	key interface
	search	insert	delete	search hit	insert	delete		
sequential search (unordered list)	N	N	N	$\frac{1}{2} N$	N	$\frac{1}{2} N$		equals()
binary search (ordered array)	$\lg N$	N	N	$\lg N$	$\frac{1}{2} N$	$\frac{1}{2} N$	✓	compareTo()
BST	N	N	N	$1.39 \lg N$	$1.39 \lg N$	$\sqrt{N}$	✓	compareTo()
red-black BST	$2 \lg N$	$2 \lg N$	$2 \lg N$	$1.0 \lg N$	$1.0 \lg N$	$1.0 \lg N$	✓	compareTo()

Searching:
Binary search tree should be taking more time than the ordered symbol table, which is does for higher values.

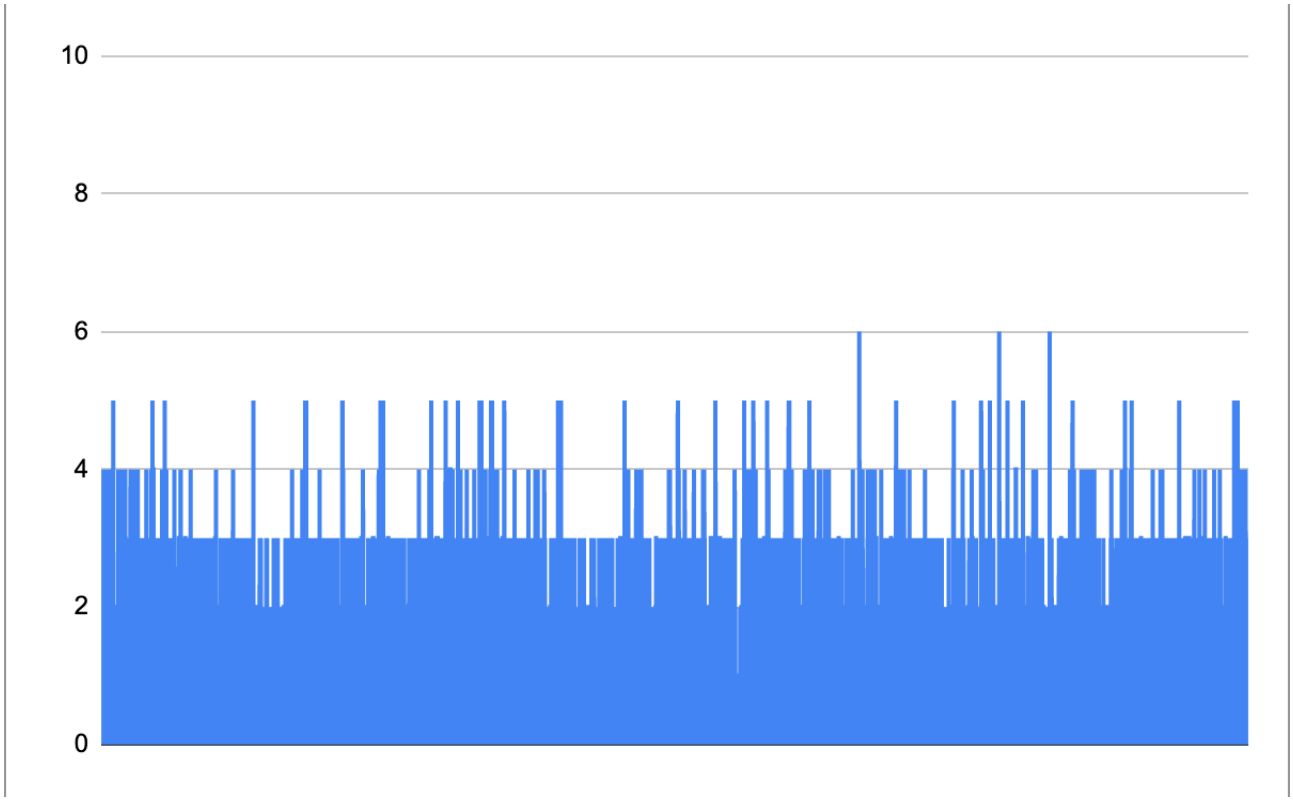
Insertion:
The ordered symbol table should be taking more time than binary search tree, which is pretty much always does except for some small values.

3.

Output sample

```
run:
In hash #1  there are  1000 nodes
In hash #2  there are  977 nodes
In hash #3  there are  939 nodes
In hash #4  there are  1026 nodes
In hash #5  there are  1007 nodes
In hash #6  there are  1011 nodes
In hash #7  there are  978 nodes
In hash #8  there are  964 nodes
In hash #9  there are  1009 nodes
In hash #10  there are  1033 nodes
BUILD SUCCESSFUL (total time: 4 seconds)
```

If I choose to divide the keys over 10 hashes we can see how evenly the keys are distributed over 10 hashes. To make a better observation I set the interval to 0-9943 (9944-1) since that is the number of distinct words, which is the same as the number of keys in the hash symbol table. I use those values in my graph.



4.

Output sample

```
run:
What word would you like to find?
tale
Word was found 7 times at positions: [34, 334, 614, 675, 720202, 769292, 769374]
BUILD SUCCESSFUL (total time: 8 seconds)
```

When I search for the word ”tale” I get a list of positions the word is found in the text.