CENG 301
Spring 2015-2016

Assignment #3
Counting Words

In this assignment your goal is to count the words in a text file. Your program take a text file as input and you will print out the number of times each word occurs in the text file. For simplicity no punctuation marks will be present in the input text file. The counting will not be case sensitive in other words "THIS", "this", "This", "ThIS" are all the same word. Below is an example text file:

"This is an example In this example some words occur more than once"

Your output should look like:

this 2
is 1
an 1
example 2
in 1
some 1
words 1
occur 1
more 1
than 1
once 1

You are going to implement two different strategies for counting and compare their timing for the provided example input files. In the first strategy you are going to maintain a list of words and their respective counts as a linear list (such as linked lists or arrays). As new words appear you will append them to the list and if existing words appear you are going to update their count accordingly. In the second strategy, you will use a binary search tree to store the encountered words. Each node in the tree is going to store the word as the key and also it will store a count for that word. Again you will add new nodes in the tree for new words and update the counts for existing words. Compare the running times of both strategies for the 6 example input files provided. Analyse and reason about your results. Which one is faster, why?

Submission and Deliverables:

Submit your source code named <student_id>.c and a short report that shows the running time results via ODTU-Class before the deadline. Late submissions are allowed with 20 pts penalty per day.