Homework 4

Alexander Starr

22C:016:A01

00567613

March 31, 2013

Problem 1

File	Total number of words	Percentage of repeat words
hyde.txt	13,525	72.44
treasure.txt	35,174	84.20
war.txt	313,490	94.59

Yes, I think that this test justifies the proposed changes to speed up the `computeFrequencies` function. For each novel, around three-quarters or more of the words are repeats. Every instance of a repeat word is a point at which time can be saved with a more efficient search algorithm. So it would seem that the number of costly inserts is outweighed by the number of searches from which time can be gained.

Problem 4

Version	Time Taken (seconds)
Original	29.73
Modified	4.71

The plan for speeding up the `computeFrequencies` algorithm appears to have worked. The program showed a time reduction of about 84% for the long, repeat-ridden novel "War and Peace". This time reduction is dramatic, and would be very useful if we planned on using our program for many novels. It would be less useful the less books examined, the shorter the book, and the less repeated words there are per book.