

Program 4

48 Points

Due: **October 6, 2008 at 8:00 a.m. {Modified September 29, 2008}**

A Cross Reference Generator in C++

This program focuses on using tree data structures to store and information. This assignment is to be done in C++ and the use of **binary trees** or another approved data structure is required for this assignment.

This assignment requires creating a case-insensitive cross reference generator. The program will take an ASCII text file as input (e.g. The Declaration of Independence). After reading in the file, the program prints out a **reverse** alphabetized list of all the words in the file, approximating the following format:

WORD	NUMBER OF OCCURENCES	LINES THAT THE WORD APPEARED ON
------	----------------------	---------------------------------

For example, given the poem “There are holes in The Sky” by Spike Milligan and excluding the title, the cross generator program would print out the tally below.

There are holes in The Sky

There are holes in the sky
 Where the rain gets in
 But they’re ever so small
 That’s why the rain is thin.

Word	Count	Lines
why	1	4
where	1	2
thin	1	4
they’re	1	3
there	1	1
the	3	1, 2, 4
that’s	1	4
so	1	3
small	1	3
sky	1	1
rain	2	2, 4
is	1	4
in	2	1, 2
holes	1	1
gets	1	2
ever	1	3
but	1	3
are	1	1
TOTAL WORDS	18	

Note – the words in the title of the document are excluded from the word count. To simplify your program skipping any lines at the beginning, these input lines will contain a '#' character in the first column of the line. Thus the first line of the actual text will be identified as line 1.

What to turn in for Program 4

Official test input files will be made available a few days before the due date. Turn in your assignment using the *turnin* program on the CCC machines. You should turn in a tarred file that includes your source code, a make file and a README file. The README file should provide any information to help the TA or SA test your program for grading purposes.