

CMPSC 440
Operating Systems
Spring 2014

Laboratory Assignment Two: Implementing and Using a File System Traversal Tool

Introduction

The file system is one of the most heavily used components of an operating system. Figure 1-14 in your textbook shows a visualization of a file system that starts at the root directory and branches out to additional directories and files. Programming languages like Java and C provide libraries that support the traversal of a file system and the analysis of individual files and directories. In this laboratory assignment, you will implement your own file system traversal tool and use it to learn more about the characteristics of the file system used by the Linux operating system.

File System Traversal

For this laboratory assignment you should implement a file system traversal tool in the Java programming language (students who want to use a different programming language should first see the instructor before starting their implementation). Your program should use the JCommander command-line argument parsing tool, available for download from <http://jcommander.org/>, to accept as input all of the appropriate arguments. The tool should start at the directory specified by the user and then recursively traverse the file system until it either visits all of the files and directories or it visits up to a chosen maximum number of files and directories. Students should consider using the methods provided by the `java.io.File` class to perform the traversal.

Your file system traversal program should have features to calculate the minimum, maximum, and average size of the files that are accessible by recursive traversal from the specified root directory. The program should also be able to report the total number of files that it analyzed. In addition to determining the depth and breadth of the file system subject to analysis, the tool should output how long it took to perform the traversal. Students can earn extra credit if their traversal program produces a full-featured visualization of the file system using a program like Graphviz.

Experimental Study

Upon completing your file system traversal tool, you should use it to analyze at least five distinct regions of your file system. Your analysis should investigate all of the metrics that your tool is able to calculate (e.g., file size and the characteristics of the file system tree). Students should ensure that their analysis uses large representative file system regions so that the results are as generalizable as is possible. That is, for the purposes of this assignment, you need to run your analysis tool on directory structures that contain hundreds, or even thousands, of files. You should also use your tool to determine both the size of an empty file on the file system and the size of the Java program that you wrote to perform the traversal.

The results from your experimental study should be furnished in a written report that uses graphs and tables of data for summarization purposes. Whenever possible, your report should explain how your results might influence the design and implementation of future file systems.

Summary of the Required Deliverables

This assignment invites you to submit printed and signed versions of the following deliverables:

1. The source code of the file system traversal tool that you implemented in the Java language
2. The output of the file system traversal tool when run on several small file system regions
3. The report from an experimental study that characterizes the use of the Linux file system

Students are strongly encouraged to write their laboratory report in $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ and use tools such as the R language for statistical computation and Graphviz to better view and understand the empirical results. Please see the instructor if you have questions about these deliverables.

In adherence to the honor code, students should complete this assignment on an individual basis. While it is appropriate for students in this class to have high-level conversations about the assignment, it is necessary to distinguish carefully between the student who discusses the principles underlying a problem with others and the student who produces assignments that are identical to, or merely variations on, someone else's work. As such, deliverables that are nearly identical to the work of others will be taken as evidence of violating the Honor Code.