Name(s)

Andrew R. Dutcher

Project Number

J1604

Project Title

Compressed for Time: Examining File Compression in Computers

Abstract

Objectives/Goals

My project's goal is to find out in which way files on computers can be most effectively compressed while retaining all the data inside them. I hypothesized that either the Bzip2 algorithm, an established entity in the world of compression, or an algorithm that began by rearranging the data of the file and ended with a function that took advantage of the orderliness of a file would be the most effective in compression.

Methods/Materials

To accomplish my objective, I built a computer program in C#, a programming language, that would be able to compress files in combinations of six different ways, and automatically cycle through all combinations of these methods for four different files and output the compressed file-sizes, in bytes. I then built a peripheral program that would process the raw numbers and return actual data.

Conclusions/Discussion

In the end, I found that the data support my hypothesis. Not only do many of the top algorithms closely resemble the Bzip2 algorithm, many other top algorithms also start with rearrangement functions and end with entropy functions.

Summary Statement

I am attempting to find the optimal form of losless file compression.

Help Received

Father helped with application coding.