

CS2420 Assignment12 Analysis

Daxton Wilson

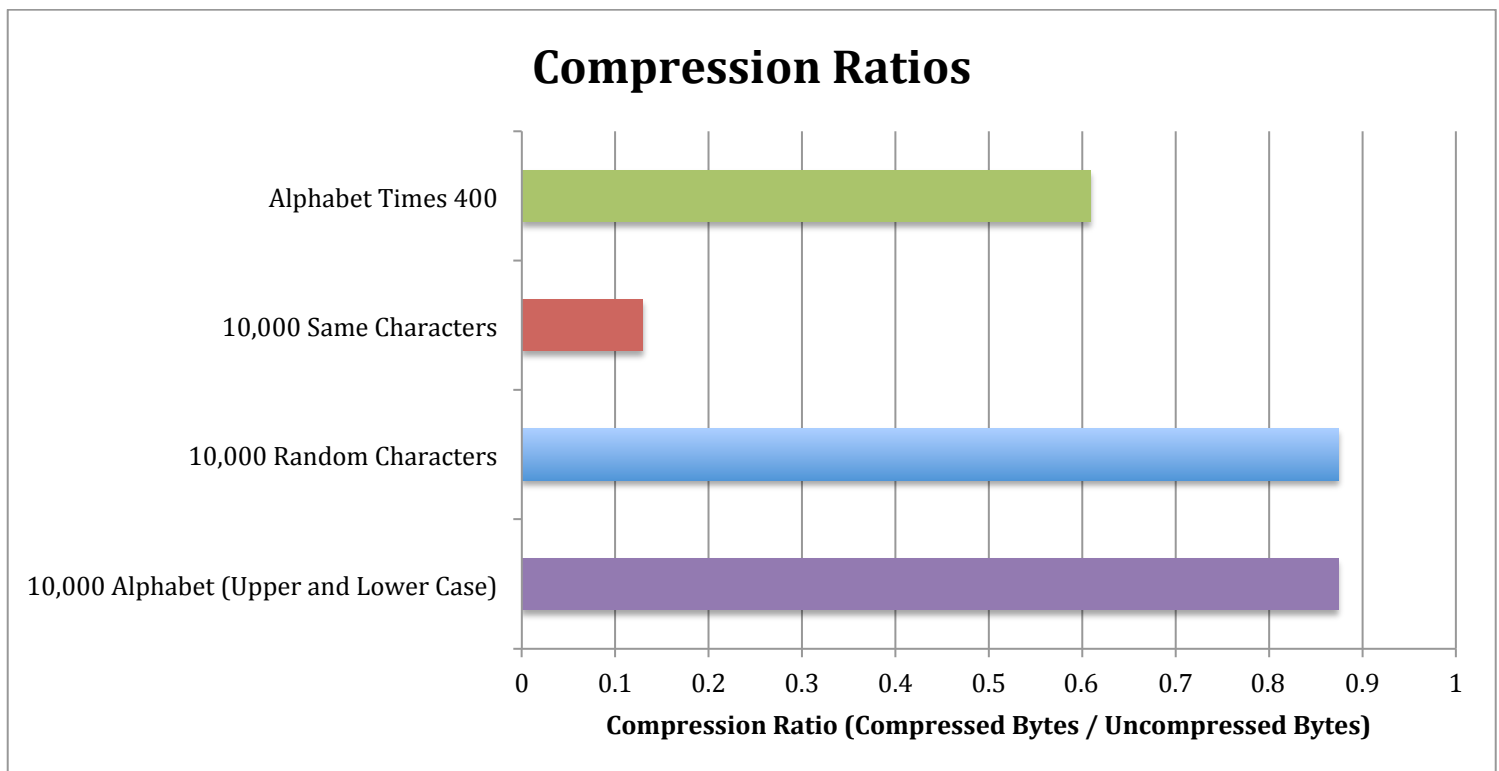
U0264580

1. Experiment Steps:

- 1 - Create a 4 .txt files with 10,000 characters in each: all lower-case and upper-case letters in random amounts, all 95 unique characters on my Mac keyboard in random amounts, one with the lower-case alphabet in equal amounts and one with 10,000 of the same character.
- 2 - Plot the file (x-axis) vs. the compression ratio (y-axis) for each of the .txt files.
- 3 - Analyze the compressed ratio for each of the four different files.

Interpretation:

The input files that most significantly reduced the number of bits in the compressed file were the Alphabet Times 400 and 10,000 of the same character. This is because characters that are simply repeated can be compressed down to just a few bits that contain that information, which is to be repeated however many times it occurs in decompression. The other two have such random sets and ordering of characters that their information cannot be compressed effectively.



2. The input files that will result in the fewest bits are those that have much less information (ie the number of different characters present) and have repeating characters (ie how many times in a row a character may occur). Files that will receive little or no savings are completely randomly ordered and have the widest variety of possible values.
3. If Huffman's algorithm were to repeatedly merge the two largest-weight trees during compression, the most frequently occurring values would have the most bits and the least occurring values would have the fewest bits. Huffman's merges the smallest-weight trees first so that the most frequently occurring elements will result in the smallest number of bits possible.
4. Huffman's algorithm employs lossless data compression because the .txt files are the exact same originally as after they are decompressed. This is common for .txt files because all the information present is important and the size is not dramatically affected by more or less text. Lossy data compression is used for video and pictures where some of the data is sacrificed to maximize memory and run-time efficiency.
5. I spent 6 hours on this assignment.