Sam Feig

CSCI 2275

November 15, 2018

Final Project Proposal

My idea for my final project is to implement and compare different data compression algorithms in C++ since we do not learn them in this class. I would implement Huffman, LZW, and Run Length compression algorithms then compare their compression rations and runtimes for both a small (less than 100 character) text file and for large (closer to 3000 character) text file. This would work with the tree required for Huffman compression and the code table for LZW compression. I could also compare my implementations to a third-party library that does standard ZIP compression that modern programs use. I find data compression really interesting and want to learn more about it, so this project would give me an excuse to research it more as I develop the code and learn how to implement these compression algorithms and read binary data in C++. I can see issues both attempting to read binary data as opposed to standard text like I know how to. I will need to research and figure out how to do that. I also might have trouble implementing the tree for Huffman compression and making sure it is structured properly. The two different size file comparisons help to show that one algorithm might be good for small data but terrible for larger data sets when a different algorithm would be greatly more efficient on large data.

Explain difference between entropy type and dictionary type.