CPSC 530
Group 17 Project Proposal

Aiden Taylor - 30092686 - Computer Science
Noah Pinel - 30159409 - Computer Science
Ty Irving - 30105319 - Computer Science

Feb. 5th, 2023

Introduction to the Topic

The proposed project concerns the design and analysis of algorithms used in Data Compression. When storing, transferring, sending, or receiving large files it is important to have efficient data compression and decompression algorithms. However, this creates a problem, specifically, in the decision of which algorithm to use in different scenarios. In order to analyze if a choice of algorithm was the most efficient, one has to implement this algorithm then compare it against the other options, which is what we plan to do. In our case, we plan to implement the lossless data compression algorithm, Bit Code Complete Binary Tree (BCCBT), proposed in our first referenced paper. We will then compare this implementation against other sophisticated compression utilities, such as Huffman encoding, the Lempel-Ziv family of algorithms, Dynamic Markov Compression, and Arithmetic Encoding. We will mainly follow the first referenced paper, since we plan to implement the algorithm proposed in this paper, however, we plan to use the other two papers as inspiration for potential tests to analyze our implementation of BCCBT. In the end, the goals of this project are to gain insight into the different measures used to analyze compression algorithms, and to gain experience in implementing compression algorithms that are to be used in practical scenarios.

Papers and References

(1) Paper: A study in compression algorithms
Author: Mattias Håkansson Sjöstrand
Link to paper

(2) Paper: Comparison Study of Lossless Data Compression Algorithms for Text Data
Authors: A. Bhattacharjee, T. Bej, S. Agarwal
Link to paper

(3) Paper: Comparison of lossless data compression methods
Authors: D. Berz, M. Engstler, M. Heindl, F. Waibel
Link to paper

Outline of Proposed Work and Implementation

Proposed Work

The following is a basic outline of what needs to be done to complete this project:

(i) Analysis of BCCBT algorithm and its pseudocode.

(ii) Implement the BCCBT algorithm.

(iii) Compare our implementation of BCCBT against other compression utilities.

(iv) Compile the analysis and results of our comparisons into a 7-10 page technical paper.

(v) Design 3 quiz questions that are relevant to our paper.

(vi) Create a presentation that summarizes the main results of our paper.

Implementation

In our implementation we will be working on implementing the algorithm this will be done using a language of our choice. Once we are successfully able to implement the algorithm and have a way to compress and decompress files we will be working on analyzing multiple factors of the algorithm during run time and comparing these factors with other algorithms that are commonly used for compression the factors we will be using are as follows

1. Compression Time
2. Decompression Time
3. Saving Percentage = $\frac{Original\ File\ Size - Compressed\ File\ Size}{Original\ File\ Size}$
4. Compression Ratio = $\frac{Compressed\ File\ Size}{Original\ File\ Size}$

Using these factors we will successfully be able to see in what situations certain algorithms should and shouldn't be used. During our testing we plan on mainly using different size txt files in order to see if there are any changes in the factors we will be measuring however if time allows different types of files could be incorporated in order to test the compression algorithm on a wide variety of data