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

What is important is how we will be implementing experiments for this project and how this will help us understand which algorithms are most efficient. The way that we will be doing this is by using implementing the lossless data compression algorithm Bit Code Complete Binary Tree (BCCBT) along with other compression techniques and testing multiple factors when files are compressed and decompressed using these. By being able to compare these factors that we will get by compressing and decompressing files we will be able to evaluate when certain algorithms are used for any given scenario. In order to be able to properly test and ensure that we have multiple different types of inputs to begin compressing and decompressing multiple different file types of sizes and content will be used in order to see if there is any effect on how certain files compare when using specific algorithms and if there is a case where one will be used over another this should be evident when comparing the different types of factors against each other for the different types of algorithms.