

Patrick Berry & Zach Hegemann

CDA5125

March 5th, 2022

Term Project Proposal

Our term project for the semester will be a software development project, developing and optimizing the Huffman Coding algorithm. The Huffman Coding algorithm is considered the most foundational algorithm technique for data compression and is still commonly used in today's software utilities. GZIP is a common, easy to use compression and decompression utility used in Unix systems, producing a file with a gzip extension. This utilities' software is derived from the Huffman Coding algorithm. Additionally, there are other software compression utilities that have the Huffman algorithm implemented.

To better appreciate the algorithm, we will develop the sequential code from scratch. Furthermore, we will measure how much each operation takes regarding the overall time. For example, we would measure how much time encoding a file takes compared to the population of a data compression dictionary. This will help us gauge the different possible ways to reduce the runtime most effectively. We will also survey how to strategically utilize threads while still obtaining the correct data for compression. After producing the sequential code, we will utilize the APIs learned in this course, such as OpenMP and SIMD intrinsics, to optimize our code. We would then compare runtimes between the two coding methods and observe the performance gained. We will also provide a final report that will detail some of the key attributes in designing the Huffman algorithm, the runtimes of the sequential code, the processes used to reduce runtime, and a justification of the improved runtimes.