DSII Project Specification

Team Members:

Anosha Fatima Fazli and Muminah Khurram

Idea:

For our project, we are using Pagoda which is an efficient implementation of priority queue. It is a type of binary tree-based data structure that is usually utilized for effective representation of priority queues. It handles an arbitrary sequence of n primitive operations chosen from MIN, INSERT, UNION, EXTRACT and EXTRACTMIN in time o(n log n). The constant factors affecting these asymptotic run time are small enough to make the pagoda competitive with any other priority queue. (Francon, Viennot and Vuillemin 1978)

We are using this data structure to implement the Huffman coding. Huffman Coding is a technique of compressing data to reduce its size without losing any of the details. By keeping a priority queue of Huffman tree nodes, the pagoda data structure can be used for Huffman coding and can help in data compression. There are various functionalities of using this technique like file compression, images compression, and encryption however we will be focusing on file compression as well as its visualization with GUI.

Input and Output:

Since uncompressed data can take up a lot of space, which is inconvenient for device storage and file sharing. The input for the project will be a file where each character in the text file will be assigned variable length binary codes. This algorithm will receive input as array of characters in the text file. The output of the Huffman coding algorithm would be a compressed version of the input file that uses fewer bits to represent the same information.

References:

<https://ieeexplore.ieee.org/abstract/document/4567956>