Project Proposal- Data Structures

CS-2001



**Group Members:**

Muhammad Bilal Khan 20k-1898

Talha Abdul 20k-1647

Taha khan 20k-0315

**Section: BCS-4A**

**Instructor: Anam Qureshi**

**Problem Statement:**

Data compression is one of the most important areas of computer science. It enables devices to transmit data with fewer bits i.e. saving both time and storage space .

In this project, a crucial lossless data compression algorithm (Huffman coding) will be implemented using the concepts of data structures.

In our Huffman coding project, our main concern is information redundancy, compression ratio and compression time based on different user defined bit lengths. Our main agenda was to proposed an optimum method which can potentially improve the compression ratio in order to perform lossless data compression.

. **Introduction and Solution:**

Huffman coding is a common lossless algorithm which is widely used in entropy encoding. Huffman Coding is generally useful to compress the data in which there are frequently occurring characters. The process of this particular algorithm is to assign variable-length code to characters (using different arrangement of 0 or 1 to represent characters). Length of the code depends on the frequency of the corresponding character, the most frequent character gets the smallest code and the least frequent character gets the longest code.

Points given below explain the procedure more clearly:

* Reading input data stream.
* Counting the frequency of every sigle input characters.
* Sorting input charaters and building Huffman tree.
* Encoding input characters.
* Recoding input data stream.

**Data structures that will be used:**

Trees,Linked List and priority queue.

**Real World Examples of the project:**

* Huffman encoding is widely used in compression formats like GZIP, PKZIP (WinZip) and BZIP2.
* Multimedia codecs like JPEG, PNG and MP3 uses Huffman encoding (to be more precise the prefix codes)