DATA COMPRESSION

Abstract

From the start we always dream for more and more storage space. This has led to rapid improvements and innovations in disks and data storing techniques. Space is still at a premium primarily because as the machines have gotten bigger and faster, so have the problems we want to solve. In 2013, **5 Exabytes** of data was created every day. We are facing a severe crisis of data storage space.

Data compression can solve the problem partially by reducing the space required for each file. Many algorithms such as Huffman coding,

Shannon- Fano and variable length encoder are used currently. Huffman coding is a way to reduce the space required.

It does it by prioritizing the patterns that repeat and replaces it with a simple binary digit. This can sometimes compress the file up to half of its size. In this I have tried to compress the text data using python language for its better data structures specifications and usability.

Aim

In this we want to reduce the space required for storing text data. Using python language, we are implementing Huffman algorithm.

Method

Language: Python

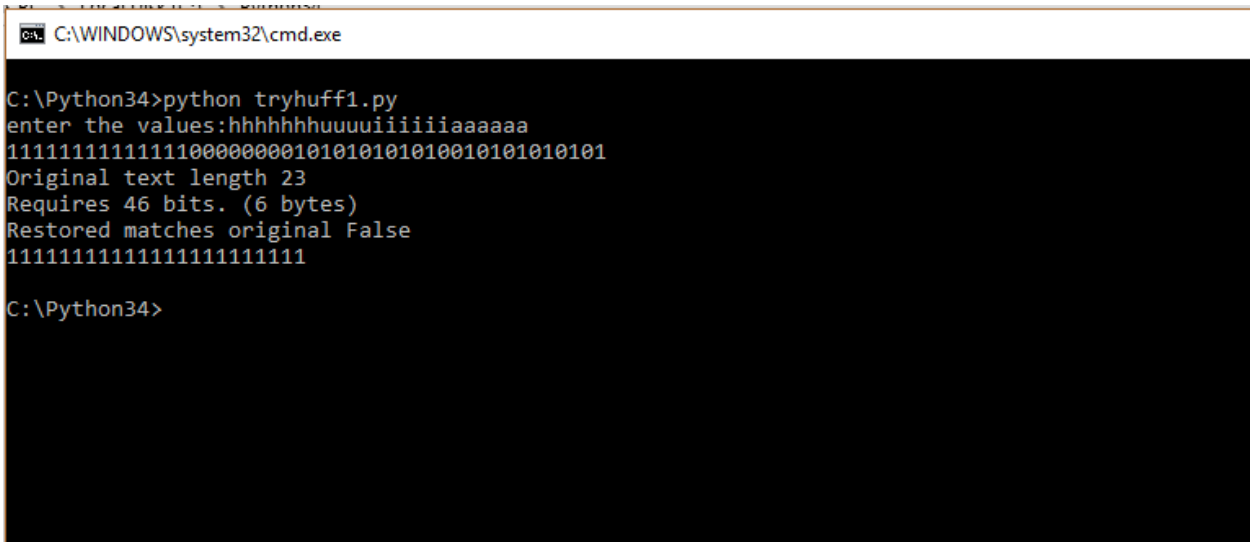
Algorithm: Huffman

Technique

We implement Huffman algorithm in the following steps:

- Encounter the text data and store it in a string.
- Store it in a dictionary in a non-increasing order of frequencies.
- Make a tree by giving priority to the character with highest frequency till each and every character is added in the tree.
- Traverse the tree by naming each right branch as 1 and left branch as 0 till we reach each and every node once.
- Store the binary code for each and every character by reaching it through the root.
- Replace the character by its binary code and display the final output.

Results



Here it can be seen that to store the entered text 23 bytes were needed. But after encoding using Huffman coding only 46 bits i.e. 6 bytes are required. This has helped to compress that data up to **70%**. This if used while storing text data in appliances can lead to more free space.

Future Modifications

It can also be used to compress image and audio files.

This can lead to less internet usage for downloading and accessing media