HUFFMAN CODING

INTRODUCTION

Huffman coding is a lossless data compression algorithm. Used to assign variable-length codes to input characters. Lengths of the assigned codes are based on the frequencies of corresponding characters. The most frequent character gets smallest code and least frequent character gets the largest code.

Huffman coding is widely used for creating prefix codes that the term "Huffman code" is sometimes used as a synonym for "prefix code" even when such a code is not produced by Huffman's algorithm.

Huffman was able to design the most efficient compression method of this type: no other mapping of individual source symbols to unique strings of bits will require lesser space for storing a piece of text when the actual symbol frequencies agree with those used to create the code.

There are mainly two parts.

1. To create Huffman tree
2. To traverse the tree to find codes.

ADVANTAGES

This encoding scheme results in saving lot of storage space since the binary codes generated are variable in length.

It generates shorter binary codes for encoding symbols/characters that appear more frequently in the input string

The binary codes generated are prefix-free

DRAWBACKS

This compression algorithm is mainly efficient in compressing text or program files

Static Huffman is not suitable for image compression

For small file overhead is significant and hence less compression ratio

Not suitable for live data compression as prior data analysis is not available

Scan input data twice

APPLICATIONS

Huffman compression is used in programs like pkZIP, gz, zoo, and arj.

It is also used within JPEG and MPEG compression