Huffman coding

- Definition: a prefix code that is used to compressed data base on the frequency of occurrence.

- Purpose:

Normally, every character has the same weight (8 bits).

HC try to create a tree that total weight of all characters is the smallest.

- Algorithm

1/ Scan the text to find the frequency of occurrence of each character.

2/ Sort the frequencies in-order level.

3/ Build Huffman tree based on the list of sorting

- Create nodes contain frequency for every character.

- Drawing from the leaf nodes up to the root.

- Choose 2 minimum nodes.

- Create a new parent node contains the key value equals to the sum of 2 children’s frequency.

Repeat this step until the remaining node is root node.

4/ Create tranversal for the tree to illustrate code words. (for eg: write 0 if it travel to the left, write 1 if it travel to the right).

5/ Scan the text again to find Huffman codes.