Huffman coding

The Huffman coding can be used to encode characters (from an alphabet) using variable length codes.

- In order to reduce the total number of bits needed to encode a message, characters that appear more frequently have shorter codes.
- Since we use variable length code for each character, no code can be the prefix of any other code (if we encode letter E with 01 and letter X with 010011, during decoding, when we find a 01, we will not know whether it is E or the beginning of X).

Huffman coding

- When building the Huffman encoding for a message, we first have to compute the frequency of every character from the message, because we are going to define the codes based on the frequencies.

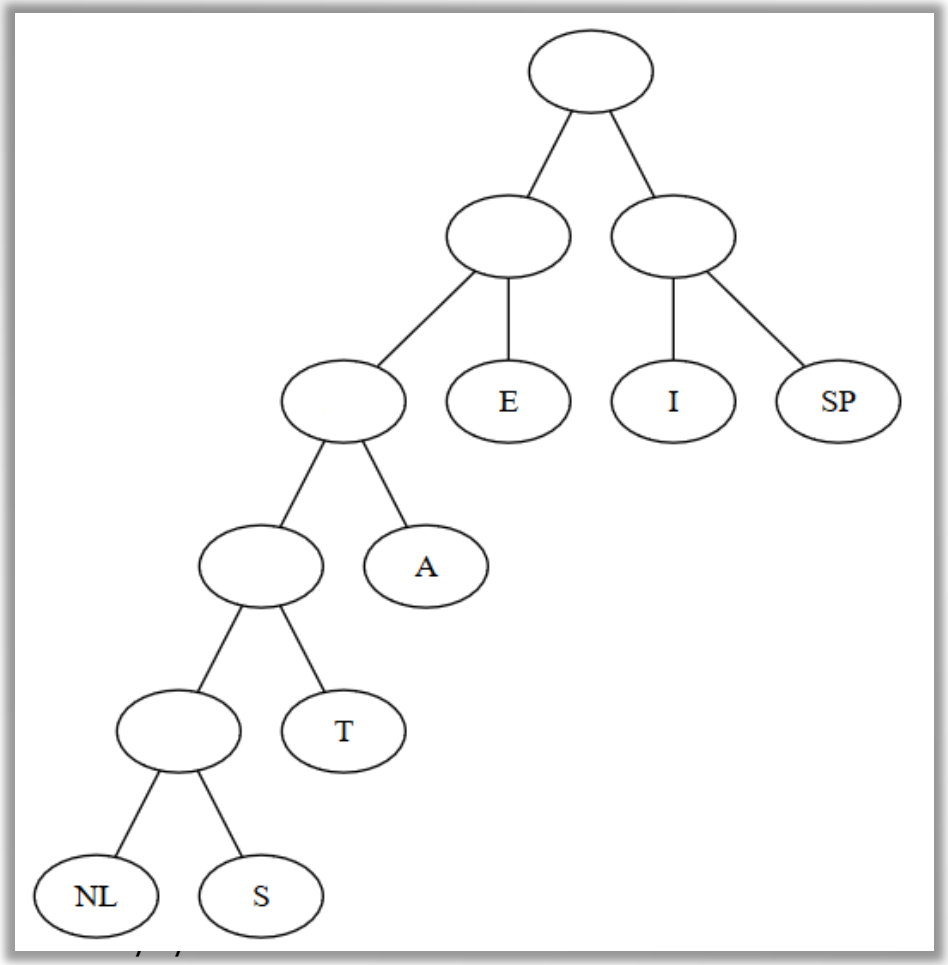
For defining the Huffman code a binary tree is build in the following way:

- Start with trees containing only a root node, one for every character. Each tree has a weight, which is frequency of the character.
- Get the two trees with the least weight (if there is a tie, choose randomly), combine them into one tree which has as weight the sum of the two weights.
- Repeat until we get only one tree

Code for each character can be read from the tree in the following way: start from the root and go towards the corresponding leaf node. Every time we go left add the bit 0 to encoding and when we go right add bit 1.

Huffman coding

Character	a	e	i	s	t	space	newline
Frequency	10	15	12	3	4	13	1



Code for the characters:

- NL - 00000
- S - 00001
- T - 0001
- A - 001
- E - 01
- I - 10
- SP - 11

Decode the message:
01101100010001001110010000

Huffman coding

Assume we have the following code and we want to decode it:

01101100010001001110010000 0

We do not know where the code of each character ends, but we can use the previously built tree to decode it.

Start parsing the code and iterate through the tree in the following way:

- Start from the root
- If the current bit from the code is 0 go to the left child, otherwise go to the right child
- If we are at a leaf node we have decoded a character and have to start over from the root