# Huffman Coding

Morse code appears to be a binary code, but it isn't. For example, "dot" "dash" could be either A or ET. A *prefix-free* (or simply "prefix") code with two symbols is binary. In such a code, no codon is a prefix of any other codon, unlike Morse code, where the codon for E is a prefix of the codon for A.

Huffman's algorithm finds an optimal prefix-free binary code. Given an alphabet and a frequency of each letter, Huffman's algorithm finds a code such that the expected number of bits in a message is minimized.

In the example below, the frequencies of the symbols of an alphabet Σ of order six are given in the table. Huffman's algorithm creates the binary tree, and the binary code for a symbol is read from the tree by following the path from the root down to the leaf labeled by the symbol.

Given the frequencies shown in the table, Huffman's algorithm creates the binary tree shown.

A	6	101
B	4	001
C	5	100
D	8	01
E	14	11
F	3	000
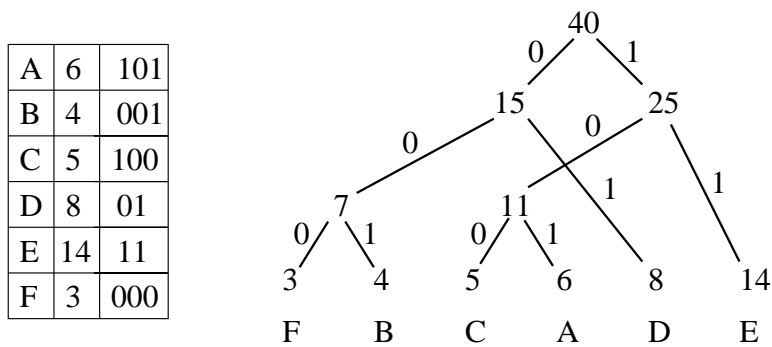
Each node of the tree shows the combined frequency of the symbols in its subtree. The tree is constructed bottom-up by combining the two smallest frequencies at each step.

The prefix code enables unambiguous decoding of a binary string. For example, what string of symbols of the alphabet Σ is encoded by the binary string "10010100011"?