In computer science and information theory, Huffman coding is an encoding algorithm used for lossless data compression. The term refers to the use of a variable-length code table for encoding a source symbol (such as a character in a file) where the variable-length code table has been derived in a particular way based on the estimated probability of occurrence for each possible value of the source symbol. It was developed by David A. Huffman while he was a Ph.D. student at Massachusetts Institute of Technology, and published in the 1952 paper "A Method for the Construction of Minimum-Redundancy Codes".

Huffman coding uses a specific method for choosing the representation for each symbol, resulting in a prefix code (that is, the bit string representing some particular symbol is never a prefix of the bit string representing any other symbol) that expresses the most common characters using shorter strings of bits than are used for less common source symbols. 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 produce a smaller average output size when the actual symbol frequencies agree with those used to create the code. A method was later found to do this in linear time if input probabilities (also known as weights) are sorted.

The algorithm basically encodes a string of symbols (example: strings of characters or bytes in a file) into a prefix code (an optimal Huffman code) of codewords (symbols used in the encoding of the symbols) (made up of symbols from an alphabet) with the minimum expected codeword length (and consequently encoded string length) using a tree constructed from the weights of the symbols, to encode and to decode.

For a set of symbols with a uniform probability distribution and a number of members which is a power of two, Huffman coding is equivalent to simple binary block encoding, such as ASCII coding. Huffman coding is such a widespread method for creating prefix codes that the term "Huffman code" is widely used as a synonym for "prefix code" even when such a code is not produced by Huffman's algorithm.

В информатике и теории информации кодирование Хаффмана - это алгоритм энтропийного кодирования, используемый для сжатия данных без потерь. Термин относится к использованию кодовой таблицы переменной длины для кодирования исходного символа (например, символа в файле), где кодовая таблица переменной длины была получена определенным образом на основе предполагаемой вероятности появления для каждого возможное значение исходного символа. Он был разработан Дэвидом А. Хаффманом, когда он был доктором философии. студент Массачусетского технологического института и опубликованный в статье 1952 года «Метод построения кодов с минимальной избыточностью».

Кодирование Хаффмана использует определенный метод для выбора представления для каждого символа, в результате чего получается префиксный код (то есть битовая строка, представляющая некоторый конкретный символ, никогда не является префиксом битовой строки, представляющей любой другой символ), который выражает наиболее распространенные символы с использованием более короткие строки битов, чем используются для менее распространенных исходных символов. Хаффману удалось разработать наиболее эффективный метод сжатия этого типа: никакое другое отображение отдельных исходных символов в уникальные строки битов не приведет к меньшему среднему выходному размеру, когда фактические частоты символов совпадают с частотами, использованными для создания кода. Позже был найден метод, позволяющий делать это за линейное время, если входные вероятности (также известные как веса) отсортированы.

Алгоритм в основном кодирует строку символов (пример: строки символов или байтов в файле) в префиксный код (оптимальный код Хаффмана) кодовых слов (символы, используемые при кодировании символов) (состоящий из символов алфавита). ) с минимальной ожидаемой длиной кодового слова (и, следовательно, длиной закодированной строки) с использованием дерева (дерево - это ациклический связанный граф точек), построенного из весов символов, для кодирования и декодирования.

Для набора символов с равномерным распределением вероятностей и числом членов, равным степени двойки, кодирование Хаффмана эквивалентно простому двоичному блочному кодированию, например, кодированию ASCII. Кодирование Хаффмана является настолько широко распространенным методом создания префиксных кодов, что термин «код Хаффмана» широко используется как синоним «префиксного кода», даже если такой код не создается алгоритмом Хаффмана.