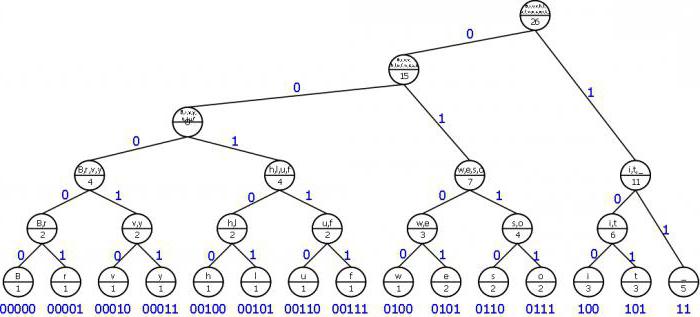
**Коды Хаффмана**

На данный момент мало кто задумывается над тем, как же работает сжатие файлов. Практически каждый человек, работающий с файловой системой, пользуется архивами. Но мало кто задумывается над тем, как они работают и по какому принципу происходит сжатие файлов. Самым первым вариантом этого процесса стали коды Хаффмана, и их используют по сей день в различных популярных архиваторах. Многие пользователи даже не задумываются, насколько просто происходит сжатие файла и по какой схеме это работает.

История алгоритма

Самым первым алгоритмом проведения эффективного кодирования электронной информации стал код, предложенный Хаффманом еще в середине двадцатого века, а именно в 1952 году. Именно он на данный момент является основным базовым элементом большинства программ, созданных для сжатия информации.

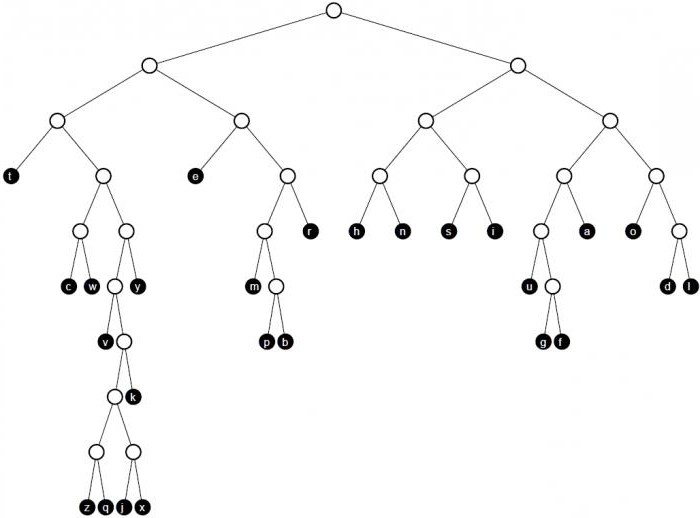


Код Хаффмана

Также данный алгоритм Хаффмана применяется для сжатия JPEG-изображений и других графических объектов. Ну и все современные факсы также используют кодирование, изобретенное в 1952 году.

Принцип эффективного кодирования.

В основу алгоритма по Хаффману входит схема, позволяющая заменить самые вероятные, чаще всего встречающиеся символы кодами двоичной системы. А те, которые встречаются реже, заменяются более длинными кодами. Переход на длинные коды Хаффмана происходит только после того, как система использует все минимальные значения. Такая методика позволяет минимизировать длину кода на каждый символ исходного сообщения в целом.

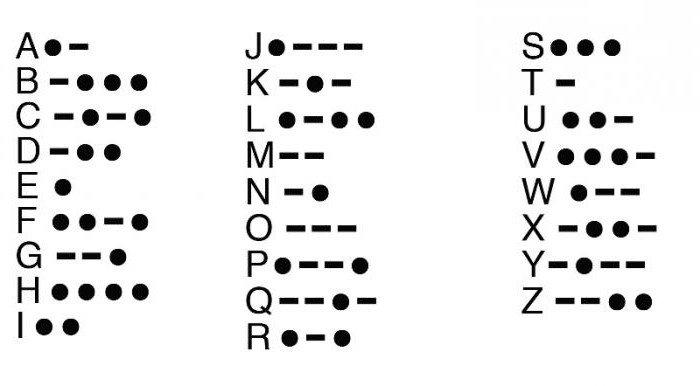


алгоритм хаффмана

Важным моментом является то, что в начале кодирования вероятности появления букв должны быть уже известны. Именно из них и будет составляться конечное сообщение. Исходя из этих данных, осуществляется построение кодового дерева Хаффмана, на основе которого и будет проводиться процесс кодирования букв в архиве.

Код Хаффмана (пример)

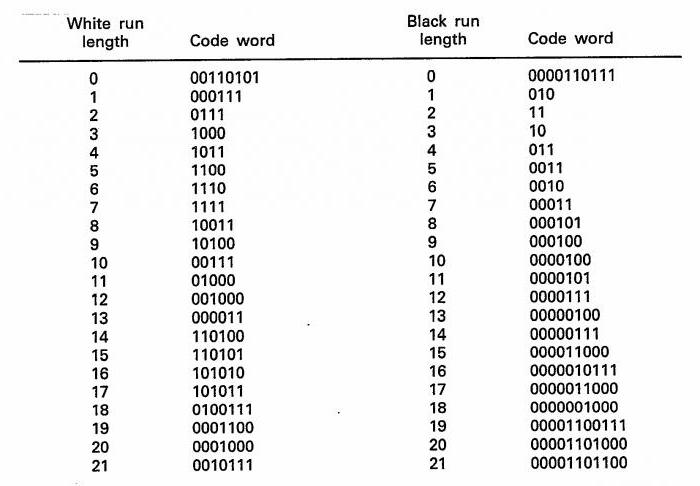
Чтобы проиллюстрировать алгоритм, возьмем графический вариант построения кодового дерева. Чтобы использование этого способа было эффективным, стоит уточнить определение некоторых значений, необходимых для понятия данного способа. Совокупность множества дуг и узлов, которые направлены от узла к узлу, принято называть графом. Само дерево является графом с набором определенных свойств: в каждый узел может входить не больше одной из дуг; один из узлов должен быть корнем дерева, то есть в него не должны входить дуги вообще; если от корня начать перемещение по дугам, этот процесс должен позволять попасть совершенно в любой из узлов.

код хаффмана пример

Существует также такое понятие, входящее в коды Хаффмана, как лист дерева. Он представляет собой узел, из которого не должно выходить ни одной дуги. Если два узла соединены дугой, то один из них является родителем, другой ребенком, в зависимости от того, из какого узла дуга выходит, и в какой входит. Если два узла имеют один и тот же родительский узел, их принято называть братскими узлами. Если же, кроме листьев, у узлов выходит по несколько дуг, то это дерево называется двоичным. Как раз таким и является дерево Хаффмана. Особенностью узлов данного построения является то, что вес каждого родителя равен сумме веса всех его узловых детей.

Алгоритм построения дерева по Хаффману

Построение кода Хаффмана делается из букв входного алфавита. Образуется список тех узлов, которые свободны в будущем кодовом дереве. Вес каждого узла в этом списке должен быть таким же, как и вероятность возникновения буквы сообщения, соответствующей этому узлу. При этом среди нескольких свободных узлов будущего дерева выбирается тот, который весит меньше всего. При этом если минимальные показатели наблюдаются в нескольких узлах, то можно свободно выбирать любую из пар.

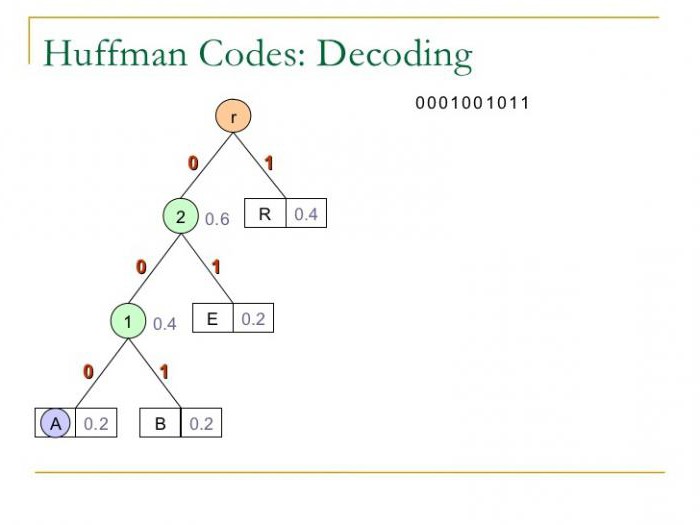


построение кода хаффмана

После чего происходит создание родительского узла, который должен весить столько же, сколько весит сумма этой пары узлов. После этого родителя отправляют в список со свободными узлами, а дети удаляются. При этом дуги получают соответствующие показатели, единицы и нули. Этот процесс повторяется ровно столько, сколько нужно, чтобы оставить только один узел. После чего выписываются двоичные цифры по направлению сверху вниз.

Заключение

Коды Хаффмана - простой и давно созданный алгоритм, который до сих пор используется многими известными программами и компаниями. Его простота и понятность позволяют добиться эффективных результатов сжатия файлов любых объемов и значительно уменьшить занимаемое ими место на диске хранения. Иными словами, алгоритм Хаффмана – давно изученная и проработанная схема, актуальность которой не уменьшается по сей день.



кодирование по коду хаффмана

А благодаря возможности уменьшить размер файлов, их передача через сеть или другими способами становится более простой, быстрой и удобной. Работая с алгоритмом, можно сжать совершенно любую информацию без вреда для ее структуры и качества, но с максимальным эффектом уменьшения веса файла. Иными словами, кодирование по коду Хаффмана было и остается самым популярным и актуальным методом сжатия размера файла.