# FAQ. Компрессия данных

**Q.** Задачка на алгоритм Хаффмана

На вход алгоритма Хаффмана подается n частот кодируемых символов. Какова наибольшая длина кодов символов в худшем случае? В лучшем?

**A.** Две экстремальные ситуации - это полностью сбалансированное дерево (минимальной высоты) и дерево максимальной высоты. Эти деревья двойственны (дуальны) и, в вырожденном случае, совпадают. По этой причине, попытка сжатия "алфавитной" последовательности (0,1,2..n) или любой ее пермутации бессмысленна. А максимальное "сжатие" получается при инверсии - замене дерева максимальной высоты на сбалансированное.

Высота сбалансированного дерева: h = Log2(n) // округляем вверх до целого

Высота дерева максимальной высоты: h = n - 1

Это очевидно, если просто нарисовать их рядом.

В любом случае, необходимый и достаточный критерий "сжимаемости" - незаполненность пространства состояний (наличие "запрещенных" кодовых комбинаций в исходном сообщении).

Исходный и "сжатый" текст можно рассматривать как инвариантный объект в исходных и штрихованных координатах (системах кодовых последовательностей). "Сжатие" - на самом деле, (пере)кодирование - при этом эквивалентно повороту к собственной (симметрированной) системе координат - к системе наименьшего объема (максимальной информационной энтропии). Уже симметричный объект ("шар"), естественно, несжимаем. Способ поворота (метод "сжатия") - LZ, Huffman или что угодно другое, неважен, это детали реализации.