МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ  
ФЕДЕРАЦИИ ФЕДЕРАЛЬНОЕ ГОСУДРАСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**«Санкт-Петербургский национальный исследовательский университет  
информационных технологий, механики и оптики»**

**Факультет информационных технологий и программирования**

**Кафедра информационных систем**

**Лабораторная работа №9**

**Эффективное кодирование**

Выполнил студент группы №M3109  
Певнев Георгий Владиславович

Проверил:  
Иванов Роман Владимирович

Исходное изображение



Черно-белая версия изображения (Grayscale)



Исходная последовательность символов: 185,185,181,184,181,179,246,247,245,188,190,184,187,189,205,205,206,207,205,208,204,206,204,207,208,206,207,206,180,180,181,180,181,180,180,180,180,181,180,180,181,181,181,181,180,180,181,180,180,180,180,180,180,180,181,181,180,180,180,180,180,180,180,180,180,180,211,177,179,184,178,181,180,181,181,180,181,181,181,181,180,180,180,181,181,181,180,181,179,183,181,180,207,206,204,208,203,206,207,206,207,207,207,208,204,205,189,165,159,160,164,185,245,250,251,255,253,253,253,255,254,255,254,254,254,254,254,255.

Последовательность после операции квантования: 180, 180, 180, 180, 180, 160, 240, 240, 240, 180, 180, 180, 180, 180, 200, 200, 200, 200, 200, 200, 200, 200, 200, 200, 200, 200, 200, 200, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 200, 160, 160, 180, 160, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 180, 160, 180, 180, 180, 200, 200, 200, 200, 200, 200, 200, 200, 200, 200, 200, 200, 200, 200, 180, 160, 140, 160, 160, 180, 240, 240, 240, 240, 240, 240, 240, 240, 240, 240, 240, 240, 240, 240, 240, 240.

**Ошибка! Переменная документа не указана.**

**Равномерное кодирование (перевод символов в двоичную систему) и первичный алфавит**

Закодированное сообщение: 0000000100000001000000010000000100000001000000100000001100000011000000110000000100000001000000010000000100000001000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000010000000010000000100000000100000010000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000010000000010000000100000001000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000100000001000000010000000001000000100000010100000010000000100000000100000011000000110000001100000011000000110000001100000011000000110000001100000011000000110000001100000011000000110000001100000011

Длина кода: 1024

Значение энтропии ПА: 3,84574363220496

Информационный объем сообщения ПА: 492,255184922234

Количество символов ПА: 31

Средняя длина кода: 50464

Минимальная длина кода: 8

Частота появления символов и их код: 140(1); 160(8); 180(71); 200(29); 240(19);

**Кодирование по методу Хаффмана**

Закодированное сообщение: 00100100100100100011110010010010010010101010101010101010101010101001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001010001000100100010010010010010010010010010010010010010010010010010010001001001001010101010101010101010101010100100010000000100010011111111111111111

Длина сообщения:326

Информационный объем сообщения: 50136,1531242006

Степень сжатия: 227,315315315315

Количество символов алфавита: 2

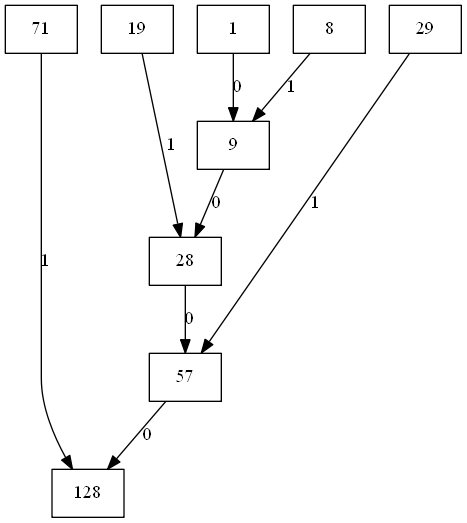
Средняя длина кода: 222

Минимальная длина кода: 1

Избыточность: 0,989627615296182

Символы и их код: Коды Хаффмана: '140(1)' - 0000; '160(8)' - 0001; '180(19)' - 001; '200(29)' - 01; '240(71)' - 1;Средняя длина кодовой комбинации = 222.

**Кодовое дерево Хаффмана:**

**Кодирование по методу Шеннона-Фано**

Закодированное сообщение: 00100100100100100011110010010010010010101010101010101010101010101001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001001010001000100100010010010010010010010010010010010010010010010010010010001001001001010101010101010101010101010100100010000000100010011111111111111111

Длина сообщения: 326

Информационный объем сообщения: 50136,1531242006

Степень сжатия: 137,879781420765

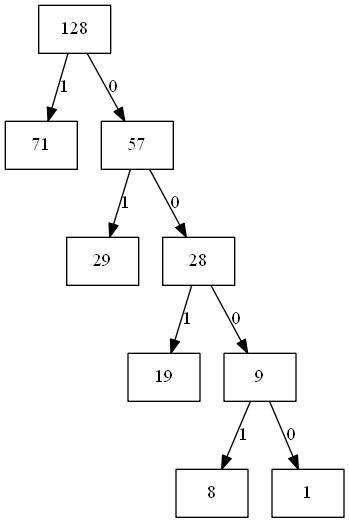
Количество символов алфавита: 2

Средняя длина кода: 366

Минимальная длина кода: 1

Избыточность: 0,993708553540307

Символы и их код: Коды Шеннона-Фано: '240(71)' - 1; '200(29)' - 01; '180(19)' - 001; '160(8)' - 0001; '140(1)' - 0000;Средняя длина кодовой комбинации: 366.

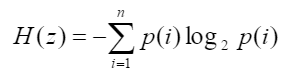
**Кодовое дерево Фано:** 

**Вывод**

Кодирование по методу Шеннона-Фано является менее эффективным в сравнении с методом Хаффмана, так как допускает погрешность при делении символов на группы. Таким образом, может быть сгенерирован не самый удачный код, однако в отдельных случаях коды Шеннона-Фано и Хаффмана совпадают, и фактически коды Хаффмана являются подмножеством всех возможных кодов Шеннона-Фано. Равномерное кодирование же является наименее эффективным из всех перечисленных так как не учитывает частоту появления символа в последовательности. Также при равномерном кодировании длина кодовой комбинации часто заметно превышает длины кодовых комбинаций, полученных методами эффективного кодирования.

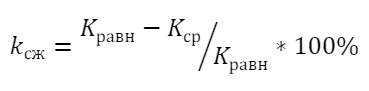
**Справочный материал**

**Энтропия —** это количество информации, приходящейся на одно элементарное сообщение источника, вырабатывающего статистически независимые сообщения.

****

**Информационный объем сообщения –** равно произведению энтропии(кол-во информации на один символ) на длину сообщения. I = H \* count

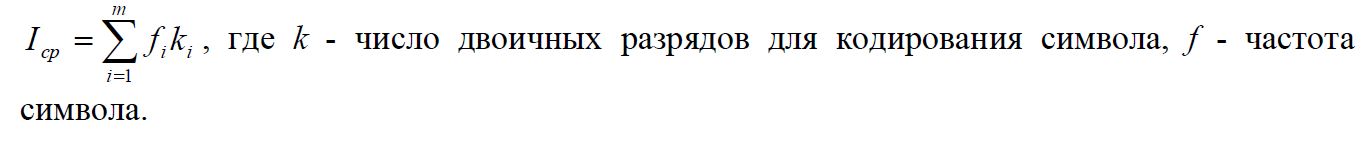
**Степень сжатия** – алгоритмическое преобразование данных, производимое с целью уменьшения занимаемого ими объёма.



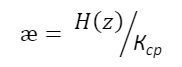
Если k < 1, то алгоритм порождает при сжатии сообщение большего размера, нежели несжатое, то есть, совершает „вредную“ работу

K = Ia / Ib, где Ia – информационный объем сообщения в алфавите А, Ib – инофрмационный объем сообщения в алфавите B

**Средняя длина кода -** эффективность кода определяется средним числом двоичных разрядов для кодирования одного символа



**Эффективность кода** определяется по следующему принципу: чем ближе значение Iср к энтропии Н(z), тем более эффективно кодирование.



Эффективность является безразмерной величиной и всегда меньше либо равна 1, т.е. æ ≤ 1. Чем ближе этот показатель к единице, тем эффективнее код.

**Избыточность**

Первичный алфавит A содержит N знаков со средней информацией на знак, определенной с учетом вероятностей их появления, I1(A). Вторичный алфавит B содержит M знаков со средней информационной емкостью I1(B). Исходное сообщение, представленное в первичном алфавите, содержит n знаков, а закодированное сообщение – m знаков. Отношение m/n характеризует среднее число знаков вторичного алфавита, которое приходится использовать для кодирования одного знака первичного алфавита – длина кода K(B).

