

Сжатие

сокращение бит, информация - const

устранение избыточности

виды

- замена часто встречающихся данных короткими кодами (энтропийное сжатие)
- замена повторяющихся фрагментов ссылками на уже имеющиеся данные (указание повторов)
- с потерей информации

Повторяющиеся последовательности символов

1111117777776666644444->[1,6][7,7][6,5][4,5]

Алгоритмы с деревьями

- построение кодового дерева (таблица частот)
- построения отображения символ->код
- архивация

Код Шенона-Фано

- неравномерность символов в документе
- префиксный код

Алгоритм

- таблица частот А-34, Б - 25
- 2 подмножества с одинаковой суммарной частотой

Корень

- m1
 - m11 (00)
 - m12(01)
- m2
 - m21 (10)
 - m22(11)

Алгоритм Хаффмана

- таблица частот P(L) - отсортировали по возрастанию
- 2 узла с минимальной частотой L0 и L1-> новый узел L(1+2)

Адаптивный Хафман

- Инициализация дерева
- увеличение веса считанного символа $P(M_i)++$

- перестановка узлов дерева
 - по упорядоченному (по весу) списку узлов ищем первый узел с большим весом
 - меняете местами найденный узел с M_i
 - повторить для родителя M_i (увеличив вес)

c [ESC,X]

Алгоритм LZW

- заполнить словарь символами ASCII
- $w = M[0]$
- $k = M[i]$
- ? есть $w+k$ есть в словаре, то замена на код (в префикс)
- заносим в словарь

код 0 1 2 3 4 ... 255 | 256 257 258 259 ... 511 (граница 9 бит)
 префикс - - - - -... - - - | 21 70 176 ~~24~~ 256
 суффикс 0 1 2 3 4 ... 255 | 70 176 21 ~~70~~

Разархивация
 21 70 176 256

код 0 1 2 3 4 ... 255 | 256 257 258 259 ... 511 (граница 9 бит)
 префикс - - - - -... - - - | 21 70 176 21
 суффикс 0 1 2 3 4 ... 255 | 70 176 ~~256~~ 21 70

Арифметическое сжатие

- $L=0$ $R=1$
- а с d....
- $L=0$ $R=0,6$
- $L=0,48$ $R=0,54$
- $L=$

a: $P(a)=60\%$ - (0; 0,6]
 b: $P(b)=20\%$ - (0,6; 0,8]
 c: $P(c)=10\%$ - (0,8; 0,9]
 d: $P(d)=10\%$ - (0,9; 1]

c-> $L = R - L$
 $L = L + L * L_i = 0,48 + 0,06 * 0,9 = 0,534$
 $R = L + L * R_i = 0,48 + 0,06 * 1 = 0,54$

код - 0,539 -> a
 $L = R_a - L_a$
 код1 = (код - L_a) / $L = (0,539 - 0) / 0,6 = 0,89$ -> c

$$\text{код2} = (\text{код1} - L_c) / 0,1 = (0,89 - 0,8) / 0,1 = 0,9 \rightarrow d$$