Повторение

1. Код однознач. декод если f – инъекция
2. разделимые коды $\Leftrightarrow$ $f(c_1 \dots c_{|s|}) = f(c_1) \dots f(c_{|s|})$
3. префиксные коды $\Leftrightarrow$ $f(c_i)$ не префикс $f(c_j)$ ← код Хаффмана
4. оптимальный код $\Leftrightarrow$ $\sum p_i L_i$ – мин

p_i – кол-во c_i в S

L_i – длина кода $f(c_i)$

$\exists$ однознач декод код с длинами $L_1 \dots L_k$

$\Updownarrow$

$\sum 2^{-l_i} \le 1$ $\quad$ * $\Uparrow$ $\exists$ префиксный код

Код f $\sum p_i L_i$ – min $\overset{(\Downarrow)}{\Rightarrow}$ $\sum 2^{-l_i} \le 1$ $\overset{(\Uparrow) + *}{\Rightarrow}$ $\exists$ преф код с длинами $L_1 \dots L_k$

$\Rightarrow$ Хаффман – опт. префикс

$\triangleright$ $\boxed{\Leftarrow}$ $\sum 2^{-l_i} \le 1 \Rightarrow \exists$ преф код с длинами $L_1 \dots L_k$

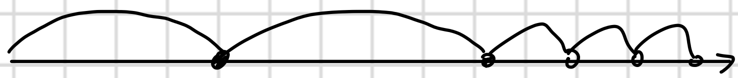
$K=1$: $L_1 \ge 1$ $\quad c_1 = \underbrace{0 \quad 0}_{L_1}$

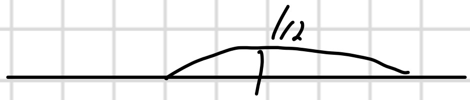
$K=2$: $L_1 > L_2$

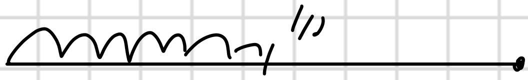
$f(c_1)\ \underbrace{00000}_{L_2-1}\ 0 \qquad f(c_2) = \underbrace{00000}_{L_2-1}\ \underbrace{1111}_{L_1-L_2+1}$

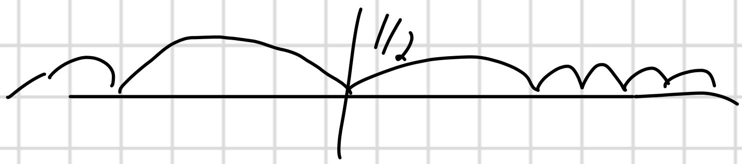
] $l_1 \le l_2 \le \ldots \le l_k$

$2^{-l_1} \ge 2^{-l_2} \ge \ldots \ge 2^{-l_k}$



нет: 

есть: 



∢ 

$$\sum_{j=1}^{i} 2^{-l_j} < \frac{1}{2}$$

$$\sum_{j=1}^{i+1} 2^{-l_j} > \frac{1}{2} \qquad \times 2^{l_{i+1}}$$

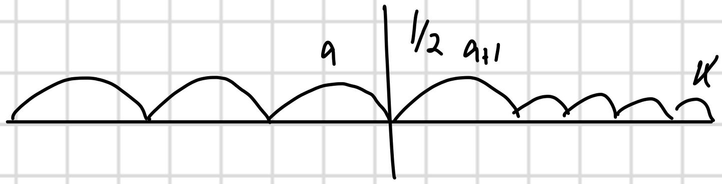
$$c = \sum_{j=1}^{i} 2^{l_{i+1}-l_j} < 2^{l_{i+1}-1} = d$$

$$\sum_{j=1}^{i+1} 2^{l_{i+1}-l_j} + 1 > 2^{l_{i+1}-1}$$

$c < d$

$c+1 > d$ т.к. c - целое т.к. это сумма степеней двоек, то такого не бывает

$L_1 \dots L_k$



$$\sum_{i=1}^{a} 2^{-l_i} \le \frac{1}{2} \iff \sum_{i=1}^{a} 2^{-(l_i-1)} \le 1$$

$L_1-1, L_2-1, \dots, L_a-1$

$\Downarrow$

$\exists$ преф. код

$L_{a+1}-1, L_{a+2}-1, \dots L_k-1$

тоже удвл нер-ву Мак Мил.

0	0101
0	001
$\vdots$	$\vdots$
0	1111

$\longleftrightarrow$ не префиксы

1	$\dots$
1	$\dots$
1	$\dots$
1	$\dots$

$\boxed{\Rightarrow}$ однозн. дек код (с длин $L_1 \dots L_k$) $\Rightarrow \sum 2^{-l_i} \le 1$

$f(c_1) = 000$ $\quad 0 = a$ $\quad aaa$

$f(c_2) = 001$ $\quad 1 = b$ $\quad aab$

$f(c_3) = 11$ $\quad bb$

$$(aaa + aab + bb)^n = \underbrace{aaa\,aaa \dots aaa}_{n \text{ троек}} + \underbrace{aaa \dots aaa}_{n-1 \text{ тройке}}aab + \dots +$$

$$+ \underbrace{bb\,bb\,bb \dots bb}_{n}$$

$$f(\underbrace{c_1 c_1 \dots c_1}_{n}) = \underbrace{aaa\,aac \dots aaa}_{n \text{ троек}}$$

$$f(\underbrace{c_1 \dots c_1}_{n-1} c_2) = \underbrace{aaa \dots aaa}_{n-1 \text{ тройка}} aab$$

всего слагаемых 3^n

$$\left(f(c_1) + \dots f(c_k)\right)^n \qquad k^n \text{ слаг}$$

т.к. f - инъекция все слагаемые различны

$$(aba + ab + aa + bbb)^k = \underbrace{aba\;aba \dots aba}_{k} + \underbrace{abc\;abc \dots abc}_{k-1}\,ab + \dots$$

$$+\underbrace{bbbb \dots bbb}_{3k \text{ длин } b} = \Big(\text{слова длины } 1\Big)_{\leq 1} + \Big(\text{слова дл. } 2\Big)_{\leq 1} + \Big(\,\cdot\,\Big)_{\leq 1} + \Big(\text{слова } KL\Big)_{\leq 1}$$

длина i

Код однознач

n^k слагаемых

$L = \max l_i$

KL — max длина длины: от 1 до KL

$\exists\; a = 1/2 \quad b = 1/2$

$$\boxed{\left(\sum 2^{-l_i}\right)^k \leq KL \quad \forall k}$$

т.к. если $2^{-l_i} > 1$ то возраст экспонента от тот линию

Имеем: $\exists$ однозн. декод с $l_1 \dots l_n$

$$\Downarrow$$

$$\sum 2^{-l_i} \leq 1$$

$$\Updownarrow$$

$\exists$ префиксный с длинами l_i

$\Rightarrow$ Код Хаффмана – оптимальный префиксный $\Rightarrow$

Оптимальный однозначно декодируемый

$\exists$ a b a – очень много b – редко

0 — $\underbrace{a\,a\,a \dots a}_{10}$

10 — a

11 — b

Арифметическое кодирование

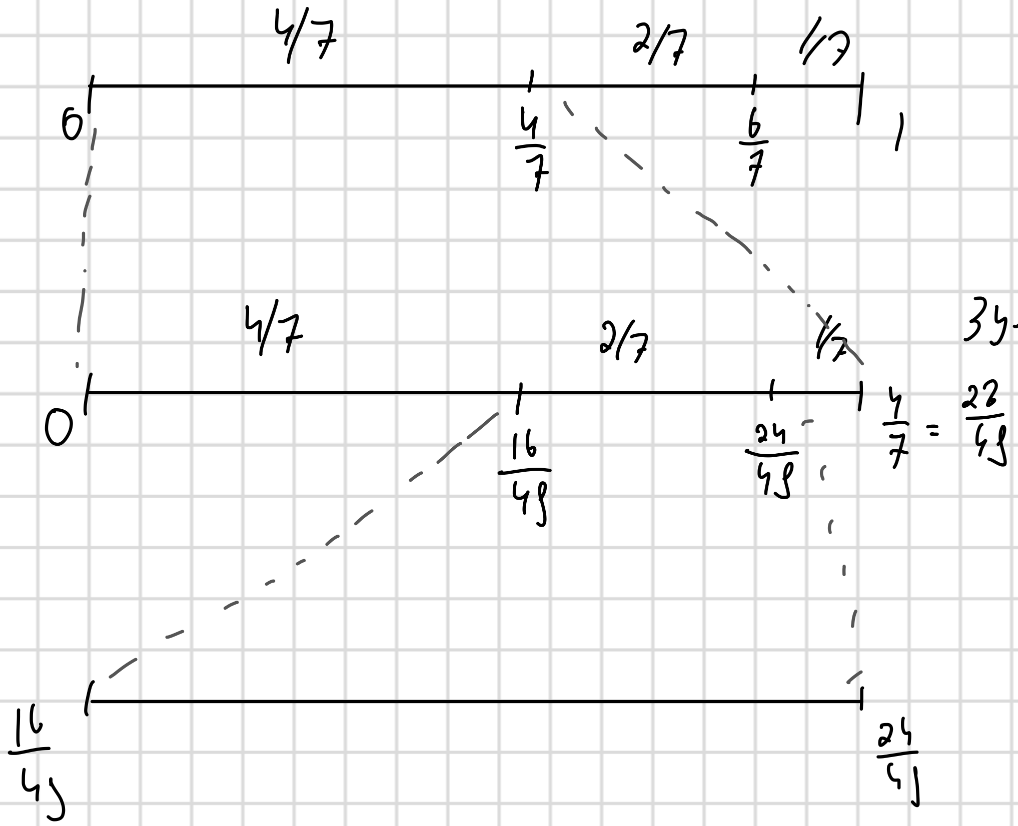
$\exists$ $c_1, c_2 \dots c_n$ символы
$f_1\ f_2 \dots f_n$ кол-во вхождений

$\sum f_i = L$ – длина текста

$$p_i = \frac{f_i}{L}$$

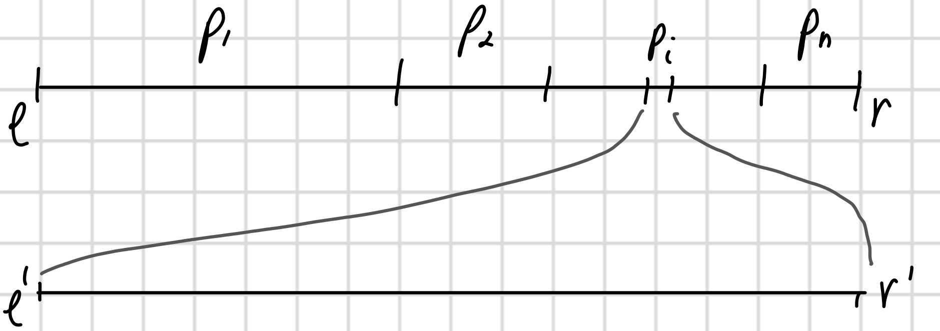
аbа саbа

$p_a = \frac{4}{7} \quad p_b = \frac{2}{7} \quad p_c = \frac{1}{7}$



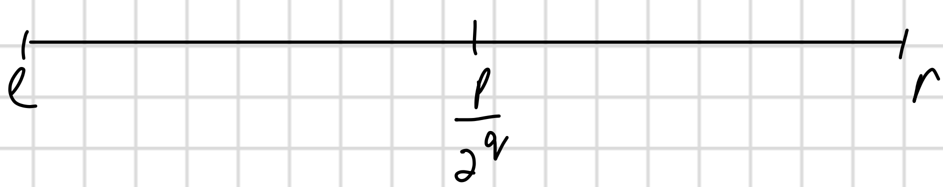
Зум в „а"

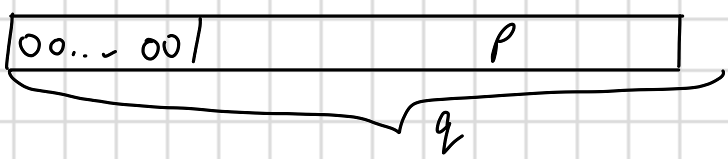
Зум в „b"



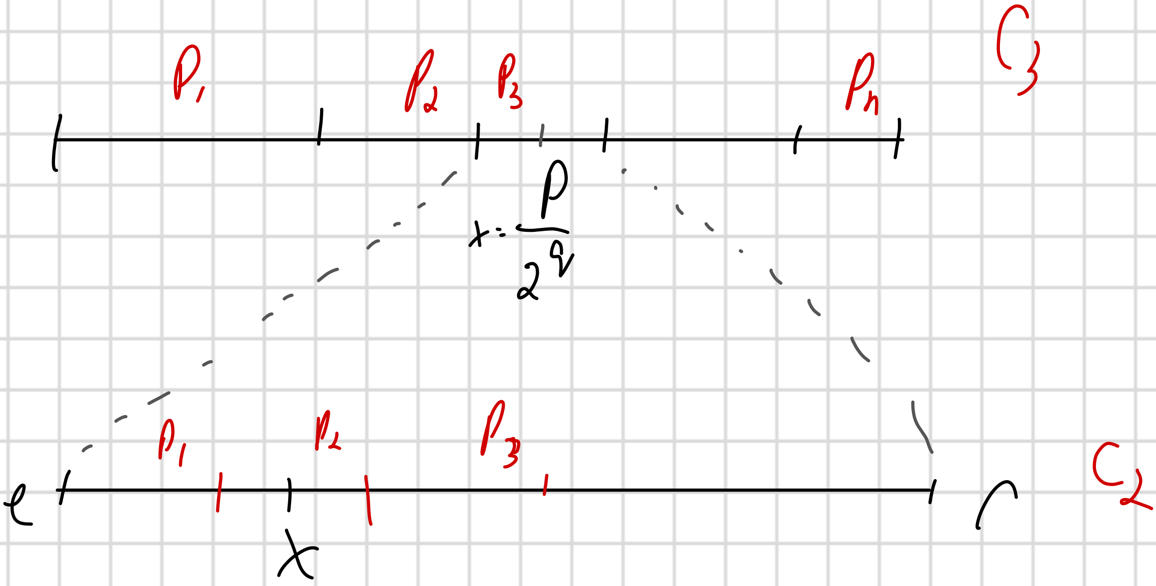
$$l' = l + \sum_{j=1}^{i-1} p_j (r - l)$$

$$r' = l' + p_i (r - l)$$

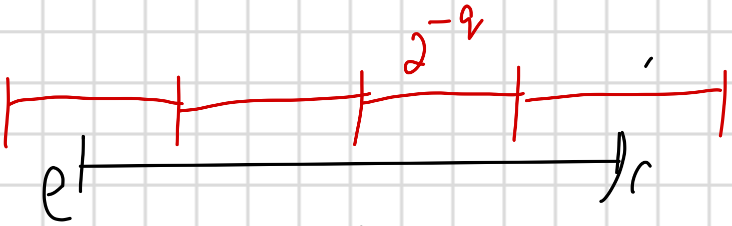




← код текста



L раз

1) 

$$2^{-q} \le r-l \Rightarrow -q \le \log_2(r-l)$$

$$\Uparrow$$

$$q \ge -\log_2(r-l)$$

$$q \le \lceil -\log_2(r-l) \rceil$$

$$r' - l' = (r-l)\, p_i$$

$$r-l = \prod_{i=1}^{L} p_{c_i} = \left(\prod_{i=1}^{n} p_i^{p_i}\right)^{L}$$

произведение

$$-\log_2(r-l) = -\log_2\left(\prod p_i^{p_i}\right)^{L} = -L\log_2\left(\prod p_i^{p_i}\right) =$$

$$= -L\sum_{i=1}^{n} p_i \log_2 p_i$$

$$q \le -L\sum_{i=1}^{n} p_i \log_2 p_i \leftarrow H(p_1, p_2 \dots p_n)$$

RLE Кодирование

a b bbb aaaa bbbbb

1a 4b 5a 5b

MTF Кодирование

a	b	c
0	1	2

a bbbb aaaa bbbbb ccccccccc aa

0 1000 1000 10000 200000000 20

0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
b	a	c	a	b	c	b	a	c	c	b	a	a	c	b

BWT Кодирование

abacaba\$ \$ – min лек

\$abacaba

a\$abacab

aba\$abac

abacaba\$

acaba\$ab

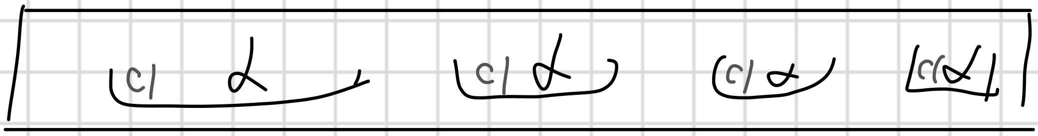
ba\$abaca

bacaba\$a

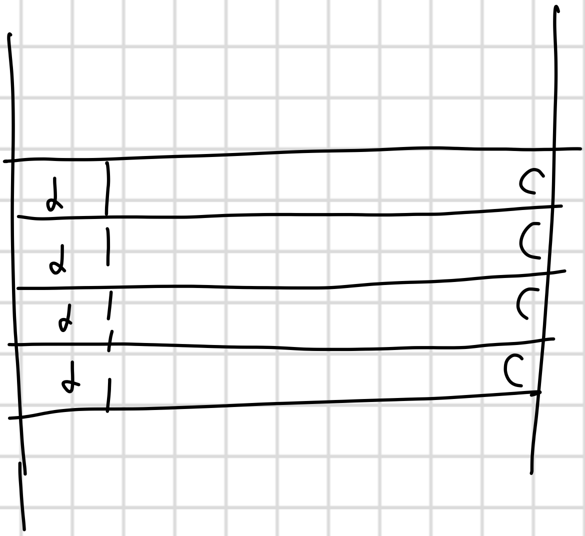
caba\$aba

abc\$baaa

FS



$S \xrightarrow{BWT} \circ \xrightarrow{MTF} \circ \xrightarrow{AE/Huff} t$



LZ алгоритмы

LZ 77-78

abacaba

abac(4,3)

ababababc

ab(2,2)(4,4)c

ab(2,6)c

LZW

aba caba

0 1 0 2 3 0

a 0

b 1

c 2

ab 3

ba 4

ac 5

ca 6

aba 7

a$ 8