

Повторение

- 1) Код однознач. декод. если есть f -инъекция
- 2) Различные 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 $\Leftrightarrow \sum p_i L_i - \min$
 p_i - кол-во c_i в S
 L_i - длина кода $f(c_i)$

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

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

* $\Uparrow \exists$ префиксный код

$$\text{Код } f \sum p_i L_i - \min \xRightarrow{(\Downarrow)} \sum 2^{-l_i} \leq 1 \xRightarrow{(\Uparrow)^*} \exists \text{ преф. код с длинами } L_1 \dots L_k$$

\Rightarrow Хаффман - опт. префикс

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

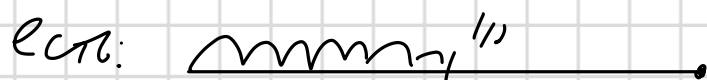
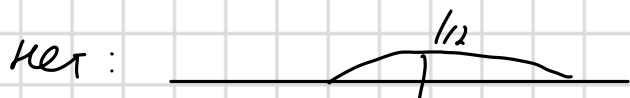
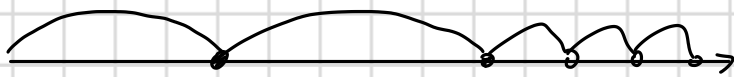
$$k=1: L_1 \geq 1 \quad c_1 = \underbrace{0 \quad 0}_{L_1}$$

$$k=2: L_1 > L_2$$

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

$$\exists L_1 \leq L_2 \leq \dots \leq L_k$$

$$2^{-L_1} \geq 2^{-L_2} \geq \dots \geq 2^{-L_k}$$



$$\sum_{j=1}^d 2^{-L_j} < \frac{1}{2}$$

$$\sum_{j=1}^{i+1} 2^{-L_j} > \frac{1}{2} \quad \times 2^{L_{i+1}}$$

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

$$\sum_{j=1}^{i+1} 2^{L_{i+1} - L_j} + 1 > 2^{L_{i+1} - 1}$$

$$c < d$$

$$c+1 > d$$

т.к. c - целое т.к. это сумма целых чисел, то
такого не должно

$L_1 \dots L_k$



$$\sum_{i=1}^a 2^{-L_i} \leq \frac{1}{2} \Leftrightarrow \sum_{i=1}^a 2^{-(L_i-1)} \leq 1$$

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

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

\Downarrow

тоже уже не-бз макс мин.

\exists пер. ког

0 0101
0 001
:
:
0 1111

не перемещать

1 |
1 |
1 |
1 |
1 |
1 |

\Rightarrow от нз. фк ког (с гнн $L_1 \dots L_k$) $\Rightarrow \sum 2^{-L_i} \leq 1$

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

$f(c_2) = 001 \quad \exists 1 = b \quad a a b$

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

$(a a a + a a b + b b)^n = \underbrace{a a a a a c \dots a a c}_n + \underbrace{a a a \dots a a a a a b + \dots}_n$

$+ \underbrace{b b b b \dots b b}_n$

$$f(\underbrace{c_1 c_1 \dots c_1}_n) = \underbrace{aaa aac \dots aas}_n \text{ Троек}$$

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

всех символов 3^n

$$(f(c_1) + \dots + f(c_k))^n \quad k^n \text{ слов}$$

т.к. f - линейная без символов разрыва

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

$$\underbrace{bbbbb \dots bbb}_{3k \text{ букв}} = \binom{\text{слова } j \text{ букв } 1}{\leq 1} + \binom{\text{слова } j \text{ букв } 2}{\leq 1} + \dots + \binom{\text{слова } KL}{\leq 1}$$

коэф. отбрасываем

n^k символов

$$L = \max p_i$$

KL - max букв букв: от 1 до KL

$$\square a = 1/2 \quad b = 1/2$$

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

т.к. если $2^{-p_i} > 1$ то возростает экспонента от суммы букв

Имеем: \exists гужина букв с $L_1 \dots L_n$

$$\begin{array}{c} \Downarrow \\ \sum 2^{-l_i} \leq 1 \\ \Uparrow \\ \exists \text{ префиксы с гужинами } l_i \end{array}$$

\Rightarrow Код Хаффмана - оптимальный префиксный \Rightarrow

Оптимальный гужинистый префиксный

\exists a b a - очень много b - редко

0 — $\underbrace{000 \dots 0}_b$

10 — a

11 — b

Арифметическое кодирование

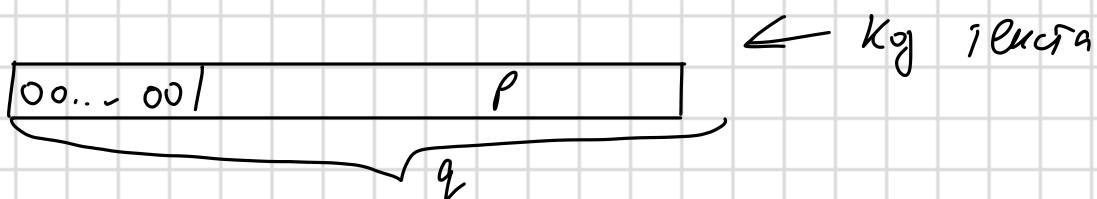
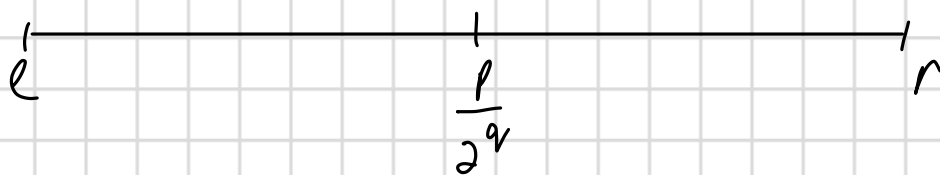
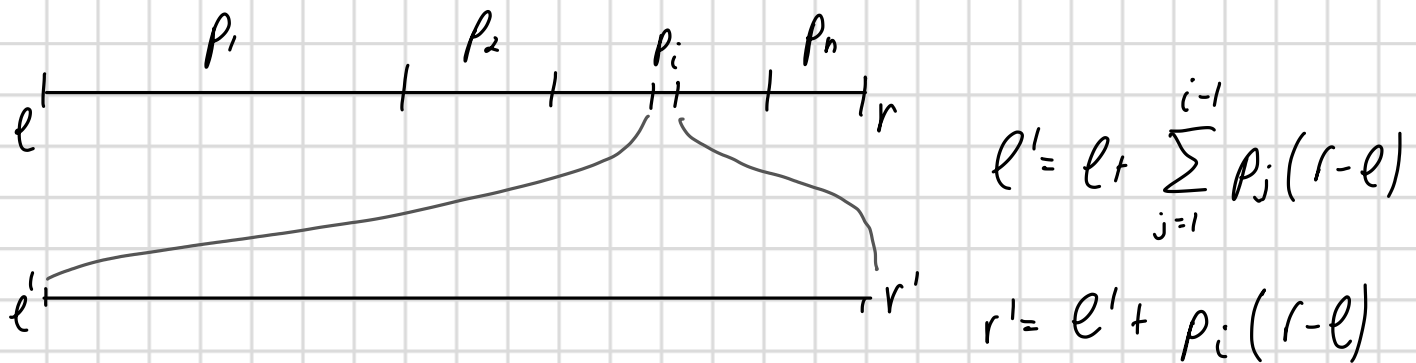
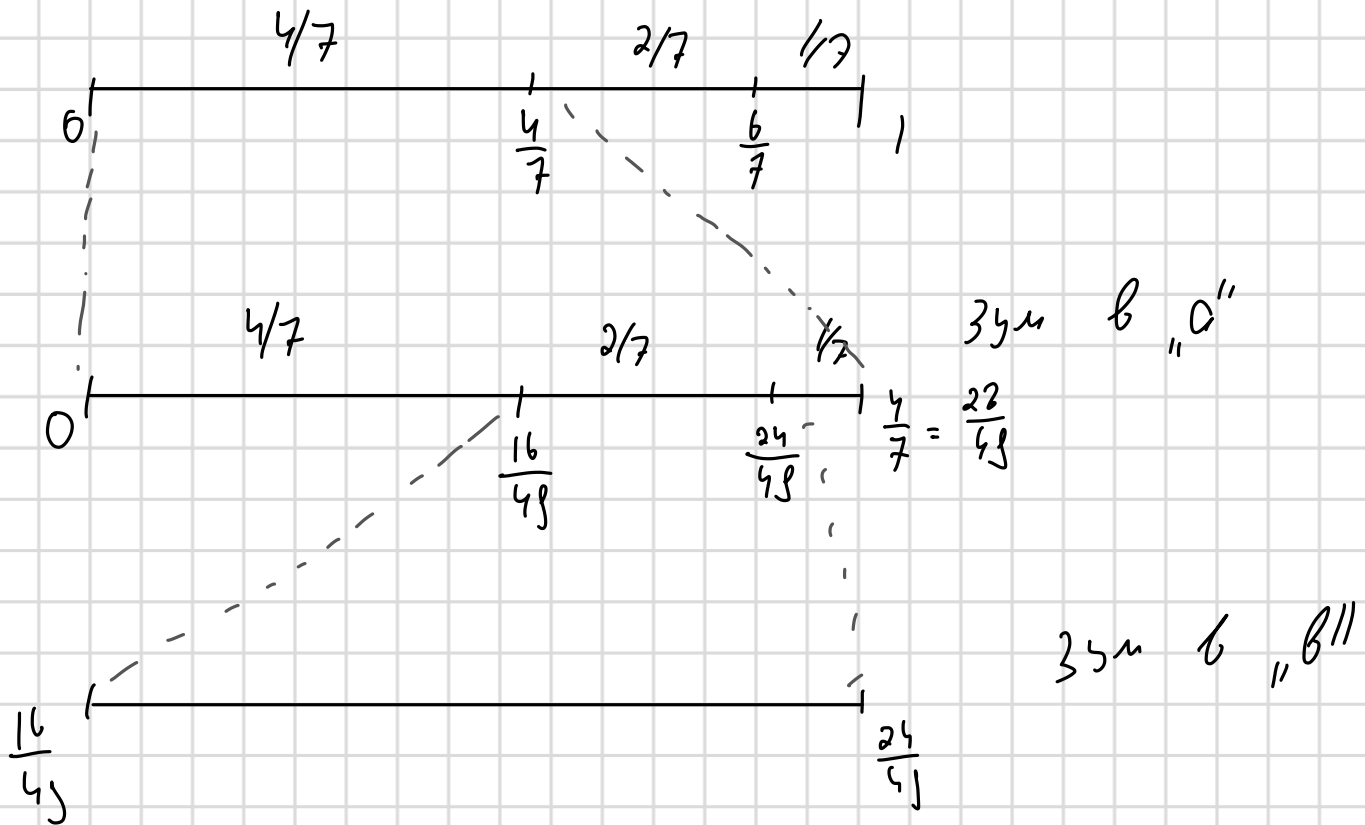
\exists $c_1, c_2 \dots c_n$ символы
 $f_1, f_2 \dots f_n$ кол-во вхождений

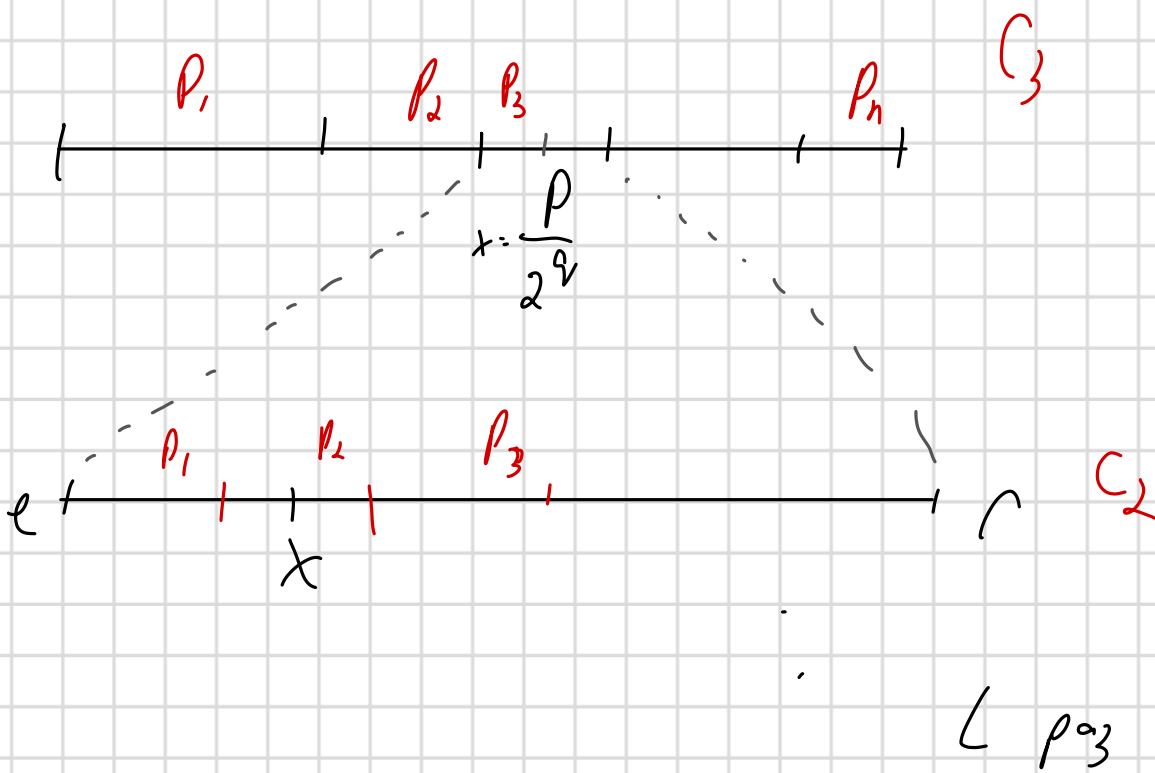
$$\sum f_i = L \quad - \text{длина текста}$$

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

аба саба

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





1) $2^{-q} \leq r-l \Rightarrow -q \leq \log_2(r-l)$
 $q \geq -\log_2(r-l)$
 $q \leq \lceil -\log_2(r-l) \rceil$

$r' - l' = (r-l) p_i$
 $r-l = \prod_{i=1}^L p_{c_i} = \left(\prod_{i=1}^n p_i^{p_i} \right)^L$
 (Note: $\prod_{i=1}^L p_{c_i}$ is labeled as "produkt")

$-\log_2(r-l) = -\log_2 \left(\prod_{i=1}^n p_i^{p_i} \right)^L = -L \log_2 \left(\prod_{i=1}^n p_i^{p_i} \right) =$
 $= -L \sum_{i=1}^n p_i \log_2 p_i$

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

RLE

Кодирование

a b bbb aaaa bbbbbb

1a 4b 5a 5b

MTF

Кодирование

a	b	c
0	1	2

a	bbb	aaaa	bbbbb	ccccccc	aa
0	1000	1000	10000	20000000	20
↓	↓	↓	↓	↓	↓
0 1 2	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 lex

\$abacab
a\$abacab
ab\$abac
abacaba\$
acaba\$ab
ba\$abac
bacaba\$a
caba\$ab

abc\$baaa

∇S

$c d$	d	$c d$	d	$c d$	d	$c d$	d
-------	-----	-------	-----	-------	-----	-------	-----

S
 \downarrow BWT
 \circ
 \downarrow MTF
 \circ
 \downarrow AE/Huff
 t

d	c
d	c
d	c
d	c

LZ $anzapn\ n61$

$LZ\ 77-78$

\swarrow $abacaba$
 \swarrow $abac(4,3)$
 \swarrow $ababababc$
 \swarrow $ab(2,2)(4,4)c$
 \swarrow $ab(2,6)c$

LZW

aba caba
ا ب ا ا ب ا

a 0

010230

b 1

c 2

ab 3

ba 4

ac 5

ca 6

aba 7

ab 8

