

# ЛЕКЦИЯ 2

ТЕХНОЛОГИИ ОБРАБОТКИ ИНФОРМАЦИИ

# ЭНТРОПИЯ ШЕННОНА

Энтропия, или количество информации вычисляется по формуле

$$I(X, X) = \sum_i p_i \log_2 \frac{p_i}{p_i p_i} = - \sum_i p_i \log_2 p_i$$

# СЖАТИЕ ИНФОРМАЦИИ

- Сжатие без потерь

  - Арифметическое кодирование

  - Кодирование Хаффмана

- Сжатие с потерями

  - Преобразование Фурье

  - Дискретное косинусное преобразование

  - Вейвлет-преобразование

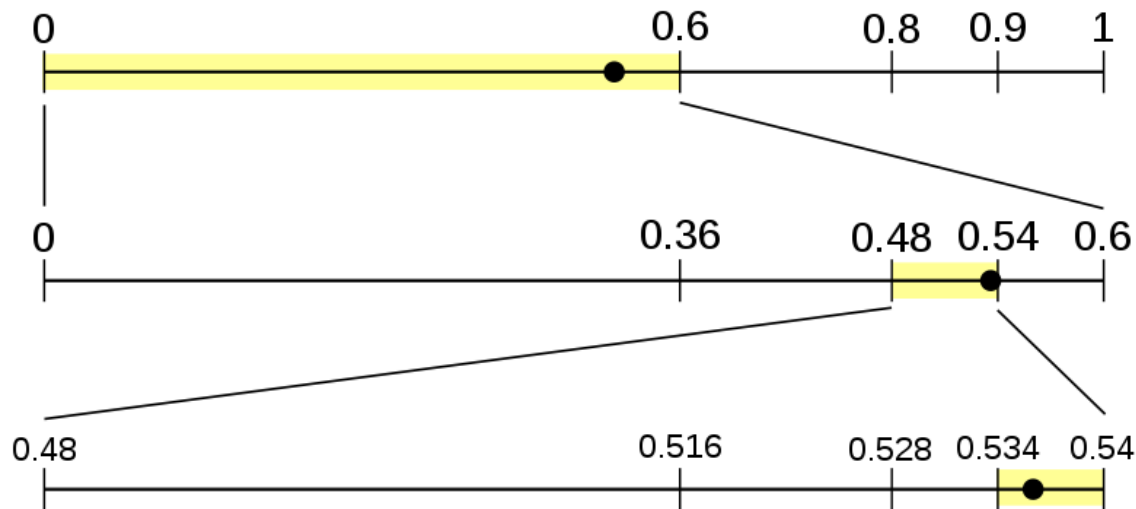
# КОДИРОВАНИЕ ХАФФМАНА

$\vec{X}$	$p$
00	9/16
01	3/16
10	3/16
11	1/16

$\vec{X}$	00	01	10	11
$p$	$\frac{9}{16}$	$\frac{3}{16}$	$\frac{3}{16}$	$\frac{1}{16}$
<div> <div>0 ↘</div> <div>↙ 1</div> <div><math>\frac{4}{16}</math></div> </div>				
<div> <div>0 ↘</div> <div>↙ 1</div> <div><math>\frac{7}{16}</math></div> </div>				
<div> <div>0 ↘</div> <div>↙ 1</div> <div>1</div> </div>				
$code(\vec{X})$	0	10	110	111

# АРИФМЕТИЧЕСКОЕ КОДИРОВАНИЕ

х	р
00	0.6
01	0.2
10	0.1
11	0.1



# АРИФМЕТИЧЕСКОЕ КОДИРОВАНИЕ

Пусть случайная величина  $X$  может принимать два значения – 0 и 1 с вероятностями  $2/3$  и  $1/3$

Интервалы и коды			Вероятность	Код Хаффмена
		$111[\frac{26}{27}, 1] \ni \frac{31}{32} = 0.11111$	$\frac{1}{27}$	0000
	$11[\frac{8}{9}, 1]$	$110[\frac{8}{9}, \frac{26}{27}] \ni \frac{15}{16} = 0.1111$	$\frac{2}{27}$	0001
		$101[\frac{22}{27}, \frac{8}{9}] \ni \frac{7}{8} = 0.111$	$\frac{2}{27}$	010
$1[\frac{2}{3}, 1]$	$10[\frac{2}{3}, \frac{8}{9}]$	$100[\frac{2}{3}, \frac{22}{27}] \ni \frac{3}{4} = 0.11$	$\frac{4}{27}$	001
		$011[\frac{16}{27}, \frac{2}{3}] \ni \frac{5}{8} = 0.101$	$\frac{2}{27}$	011
	$01[\frac{4}{9}, \frac{2}{3}]$	$010[\frac{4}{9}, \frac{16}{27}] \ni \frac{1}{2} = 0.1$	$\frac{4}{27}$	100
		$001[\frac{8}{27}, \frac{4}{9}] \ni \frac{3}{8} = 0.011$	$\frac{4}{27}$	101
$0[0, \frac{2}{3}]$	$00[0, \frac{4}{9}]$	$000[0, \frac{8}{27}] \ni \frac{1}{4} = 0.01$	$\frac{8}{27}$	11.

# МЕТОДЫ ЛЕМПЕЛА-ЗИВА (LZ77)

## КРАСНАЯ КРАСКА

СЛОВАРЬ (8)	БУФЕР (5)	КОД
" . . . . . "	"КРАСН"	<0,0,'К'>
" . . . . . К"	"РАСНА"	<0,0,'Р'>
" . . . . . КР"	"АСНАЯ"	<0,0,'А'>
" . . . . . КРА"	"СНАЯ "	<0,0,'С'>
" . . . . . КРАС"	"НАЯ К"	<0,0,'Н'>
" . . . . . КРАСН"	"АЯ КР"	<5,1,'Я'>
" . . . . . КРАСНАЯ"	" КРАС"	<0,0,' ' >
"КРАСНАЯ "	"КРАСК"	<0,4,'К'>
"АЯ КРАСК"	"А . . . . "	<0,0,'А'>

# ПРЕОБРАЗОВАНИЕ БАРРОУЗА-УИЛЛЕРА

SIX.MIXED.PIXIES.SIFT.SIXTY.PIXIE.DUST.BOXES

TEXYDST.E.XIIXIXXSMPPSS.B...S.EEUSFXDIOIIIIIT

Трансформация			
Вход	Все Перестановки	Сортировка Строк	Выход
.BANANA.	.BANANA. ..BANANA A..BANAN NA..BANA ANA..BAN NANA..BA ANANA..B BANANA..	ANANA..B ANA..BAN A..BANAN BANANA.. NANA..BA NA..BANA .BANANA. ..BANANA	BNN.AA.A



# ПРЕОБРАЗОВАНИЕ MOVE-TO-FRONT

bananaaaa

словарь – {a:0, b:1, c:2 ...}

b - 1

словарь – {b:0, a:1, c:2 ...}

a - 1

словарь – {a:0, b:1, c:2 ...}

n - 13

словарь – {n:0, a:1, b:2 ...}

a - 1

словарь – {a:0, n:1, b:2 ...}

n - 1

словарь – {n:0, a:1, b:2 ...}

a - 1

словарь – {a:0, n:1, b:2 ...}

a - 0

словарь – {a:0, b:1, c:2 ...}

a - 0

словарь – {a:0, b:1, c:2 ...}

# АЛГОРИТМ BZIP

- несжатые данные делятся на блоки фиксированного размера;
- выполняется преобразование Барроуза — Уилера для превращения последовательностей многократно чередующихся символов в строки одинаковых символов;
- применяется преобразование MTF;
- используется кодирование Хаффмана.