



**Презентация**  
по учебной дисциплине  
*«Теория Информации»*  
**на тему:**  
*«Статистическое кодирование  
сообщений»*

Подготовили курсанты  
224 уч.гр.  
Синикин Л.  
Власов А.

**Кодирование информации** – это процесс формирования определенного представления информации.

В общем смысле кодирование информации можно определить как перевод информации, представленной сообщением в первичном алфавите, в последовательность кодов.

В более узком смысле под термином «кодирование» часто понимают переход от одной формы представления информации к другой, более удобной для хранения, передачи или обработки.

символ	10- в код	2-в код	символ	10- в код	2-в код	символ	10-в код	2-в код	символ	10-в код	2-в код
	32	00100000	8	56	00111000	P	80	01010000	h	104	01101000
!	33	00100001	9	57	00111001	Q	81	01010001	i	105	01101001
"	34	00100010	:	58	00111010	R	82	01010010	j	106	01101010
#	35	00100011	;	59	00111011	S	83	01010011	k	107	01101011
\$	36	00100100	<	60	00111100	T	84	01010100	l	108	01101100
%	37	00100101	=	61	00111101	U	85	01010101	m	109	01101101
&	38	00100110	>	62	00111110	V	86	01010110	n	110	01101110
'	39	00100111	?	63	00111111	W	87	01010111	o	111	01101111
(	40	00101000	@	64	01000000	X	88	01011000	p	112	01110000
)	41	00101001	A	65	01000001	Y	89	01011001	q	113	01110001
*	42	00101010	B	66	01000010	Z	90	01011010	r	114	01110010
+	43	00101011	C	67	01000011	[	91	01011011	s	115	01110011
,	44	00101100	D	68	01000100	\	92	01011100	t	116	01110100
-	45	00101101	E	69	01000101	]	93	01011101	u	117	01110101
.	46	00101110	F	70	01000110	^	94	01011110	v	118	01110110
/	47	00101111	G	71	01000111	_	95	01011111	w	119	01110111
0	48	00110000	H	72	01001000	`	96	01100000	x	120	01111000
1	49	00110001	I	73	01001001	a	97	01100001	y	121	01111001
2	50	00110010	J	74	01001010	b	98	01100010	z	122	01111010
3	51	00110011	K	75	01001011	c	99	01100011	{	123	01111011
4	52	00110100	L	76	01001100	d	100	01100100		124	01111100
5	53	00110101	M	77	01001101	e	101	01100101	}	125	01111101
6	54	00110110	N	78	01001110	f	102	01100110	~	126	01111110
7	55	00110111	O	79	01001111	g	103	01100111	□	127	01111111

Пример кодирования буквенной информации

**Таблица 1.1. Базовая таблица кодировки ASCII**

32 пробел	48 0	64 @	80 P	96 `	112 p
33 !	49 1	65 A	81 Q	97 a	113 q
34 "	50 2	66 B	82 R	98 b	114 r
35 #	51 3	67 C	83 S	99 c	115 s
36 \$	52 4	68 D	84 T	100 d	116 t
37 %	53 5	69 E	85 U	101 e	117 u
38 &	54 6	70 F	86 V	102 f	118 v
39 `	55 7	71 G	87 W	103 g	119 w
40 (	56 8	72 H	88 X	104 h	120 x
41 )	57 9	73 I	89 Y	105 i	121 y
42 *	58 :	74 J	90 Z	106 j	122 z
43 +	59 ;	75 K	91 [	107 k	123 {
44 ,	60 <	76 L	92 \	108 l	124
45 -	61 =	77 M	93 ]	109 m	125 }
46 .	62 >	78 N	94 ^	110 n	126 ~
47 /	63 ?	79 O	95 _	111 o	127

Пример кодирования информации в ASCII

**Статистическое кодирование** - Метод кодирования, базирующийся на использовании кодов переменной величины. Для передачи наиболее часто встречающихся символов (или их комбинаций) применяются короткие коды. Редко встречающиеся символы передаются с помощью длинных кодов.

Статистическое кодирование используется для исключения, точнее существенного уменьшения избыточности сообщений, обусловленной неравновероятностью и зависимостью символов, вырабатываемых источником. Суть статистического кодирования сводится к кодированию символов источника неравномерным двоичным кодом по следующему правилу: для часто встречающихся символов присваиваются короткие двоичные кодовые комбинации, а для редко встречающихся – длинные кодовые комбинации.

## кодирования методом Шеннона-Фано

1. Символы первичного алфавита  $m_1$  выписывают по убыванию вероятностей.
2. Символы полученного алфавита делят на две части, суммарные вероятности символов которых максимально близки друг другу.
3. В префиксном коде для первой части алфавита присваивается двоичная цифра «0», второй части — «1».
4. Полученные части рекурсивно делятся и их частям назначаются соответствующие двоичные цифры в префиксном коде.

Когда размер подалфавита становится равен нулю или единице, то дальнейшего удлинения префиксного кода для соответствующих ему символов первичного алфавита не происходит, таким образом, алгоритм присваивает различным символам префиксные коды разной длины. На шаге деления алфавита существует неоднозначность, так как разность суммарных вероятностей может быть одинакова для двух вариантов деления (учитывая, что все символы первичного алфавита имеют вероятность больше нуля)

Символы	Вероятность	Кодовое слово
$a_1$	0,30	00
$a_2$	0,25	01
$a_3$	0,25	10
$a_4$	0,10	110
$a_5$	0,10	111

$$L_c = (2) [0,3 + 0,25 + 0,25] + (3) [0,1 + 0,1] = 2,2 \text{ дв. ед.},$$

$$H = - (0,3) \log_2 (0,3) - (2) \log_2 (0,25) - (2) (0,1) \log_2 (0,1),$$

$$H = 2,185 \text{ дв. ед.},$$

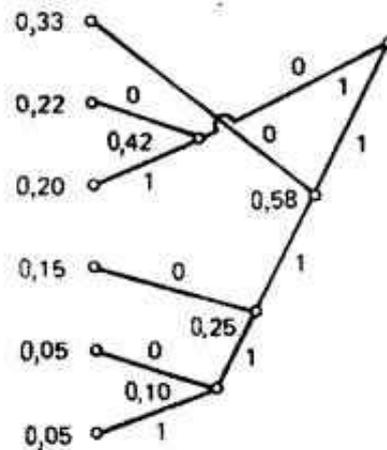
$$E_F = L_c / H = 0,993.$$

Рис. 1. Пример кодирования методом Шеннона-Фано.

Алгоритм построения сжимающего кода Хаффмена включает в себя следующие действия.

1. Все символы дискретного источника располагаются в таблице в порядке убывания вероятностей.
2. Два символа, имеющих наименьшие вероятности, объединяются в один блок, а их вероятности суммируются.
3. Ветви скобки, идущей к большей вероятности, присваивается символ «1», а идущей к меньшей – символ «0».
4. Операции 2 и 3 повторяются до тех пор, пока не сформируется один блок с вероятностью единица.
5. Записывается код для каждого символа источника; при этом считывание кода осуществляется справа налево.

<u>Кодовое слово</u>	<u>Символы</u>	<u>Вероятность</u>
10	$a_1$	0,33
00	$a_2$	0,22
01	$a_3$	0,20
110	$a_4$	0,15
1110	$a_5$	0,05
1111	$a_6$	0,05



$$L_c = (2) (0,33 + 0,22 + 0,20) + (3) (0,15) + (4) (0,05 + 0,05),$$

$$L_c = 2,35 \text{ дв. ед.},$$

$$H = -(0,33) \log_2 (0,33) - (0,22) \log_2 (0,22) - (0,20) \log_2 (0,20) -$$

$$- (0,15) \log_2 (0,15) - (2)(0,05) \log_2 (0,05).$$

$$H = 2,32 \text{ дв. ед.},$$

$$E_F = L_c / H = 0,897.$$

Рис. 2. Пример кодирования методом Хаффмена.