
Кодирование текстовой информации



Двоичное кодирование текстовой информации в компьютере

Для представления текстовой информации достаточно 256 различных символов.

$N = 2^l$, $256 = 2^8$, $2^8 = 2^l$, $l = 8$ битов

Для кодирования каждого знака требуется количество информации, равное 8 битам.

Для представления текста в памяти компьютера необходимо представить его в двоичной знаковой системе.

Каждому знаку необходимо поставить в соответствие уникальный двоичный код в интервале от 00000000 до 11111111 (в десятичном коде от 0 до 255)

Кодовая таблица

Для представления символов и соответствующих им кодов используется *кодовая таблица*.

В качестве стандарта во всем мире принята таблица ASCII (American Standard Code for Information Interchange – Американский стандартный код для обмена информацией). Условно таблица разделена на части:

- от 0 до 32 коды соответствуют операциям;
 - с 33 по 127 соответствуют символам латинского алфавита, цифрам, знакам арифметических операций и знакам препинания;
 - со 128 по 255 являются национальными.
-

32	00100000	56	8	00111000	80	P	01010000	04	h	01101000
33	! 00100001	57	9	00111001	81	Q	01010001	05	i	01101001
34	" 00100010	58	:	00111010	82	R	01010010	06	j	01101010
35	# 00100011	59	;	00111011	83	S	01010011	07	k	01101011
36	\$ 00100100	60	<	00111100	84	T	01010100	08	l	01101100
37	% 00100101	61	=	00111101	85	U	01010101	09	m	01101101
38	& 00100110	62	>	00111110	86	V	01010110	10	n	01101110
39	' 00100111	63	?	00111111	87	W	01010111	11	o	01101111
40	(00101000	64	@	01000000	88	X	01011000	12	p	01110000
41) 00101001	65	A	01000001	89	Y	01011001	13	q	01110001
42	* 00101010	66	B	01000010	90	Z	01011010	14	r	01110010
43	+ 00101011	67	C	01000011	91	[01011011	15	s	01110011
44	, 00101100	68	D	01000100	92	\	01011100	16	t	01110100
45	- 00101101	69	E	01000101	93]	01011101	17	u	01110101
46	. 00101110	70	F	01000110	94	^	01011110	18	v	01110110
47	/ 00101111	71	G	01000111	95	_	01011111	19	w	01110111
48	0 00110000	72	H	01001000	96	'	01100000	20	x	01111000
49	1 00110001	73	I	01001001	97	a	01100001	21	y	01111001
50	2 00110010	74	J	01001010	98	b	01100010	22	z	01111010
51	3 00110011	75	K	01001011	99	c	01100011	23	{	01111011
52	4 00110100	76	L	01001100	00	d	01100100	24		01111100
53	5 00110101	77	M	01001101	01	e	01100101	25	}	01111101
54	6 00110110	78	N	01001110	02	f	01100110	26	~	01111110
55	7 00110111	79	O	01001111	03	g	01100111	27	□	11111111

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
Б	128	10000000		160	10100000	А	192	11000000	а	224	11100000
Г	129	10000001	Ÿ	161	10100001	Б	193	11000001	б	225	11100001
,	130	10000010	ÿ	162	10100010	В	194	11000010	в	226	11100010
г	131	10000011	Ј	163	10100011	Г	195	11000011	г	227	11100011
„	132	10000100	о	164	10100100	Д	196	11000100	д	228	11100100
...	133	10000101	Ѓ	165	10100101	Е	197	11000101	е	229	11100101
†	134	10000110	ј	166	10100110	Ж	198	11000110	ж	230	11100110
‡	135	10000111	§	167	10100111	З	199	11000111	з	231	11100111
€	136	10001000	Е	168	10101000	И	200	11001000	и	232	11101000
‰	137	10001001	©	169	10101001	Й	201	11001001	й	233	11101001
Љ	138	10001010	€	170	10101010	К	202	11001010	к	234	11101010
<	139	10001011	«	171	10101011	Л	203	11001011	л	235	11101011
Њ	140	10001100	¬	172	10101100	М	204	11001100	м	236	11101100
К	141	10001101	-	173	10101101	Н	205	11001101	н	237	11101101
Ћ	142	10001110	®	174	10101110	О	206	11001110	о	238	11101110
Ц	143	10001111	Ї	175	10101111	П	207	11001111	п	239	11101111
Ђ	144	10010000	°	176	10110000	Р	208	11010000	р	240	11110000
‘	145	10010001	±	177	10110001	С	209	11010001	с	241	11110001
’	146	10010010	І	178	10110010	Т	210	11010010	т	242	11110010
“	147	10010011	і	179	10110011	У	211	11010011	у	243	11110011
”	148	10010100	г	180	10110100	Ф	212	11010100	ф	244	11110100
•	149	10010101	μ	181	10110101	Х	213	11010101	х	245	11110101
–	150	10010110	¶	182	10110110	Ц	214	11010110	ц	246	11110110
—	151	10010111	·	183	10110111	Ч	215	11010111	ч	247	11110111
□	152	10011000	ë	184	10111000	Ш	216	11011000	ш	248	11111000
™	153	10011001	№	185	10111001	Щ	217	11011001	щ	249	11111001
љ	154	10011010	€	186	10111010	Ъ	218	11011010	ъ	250	11111010
>	155	10011011	»	187	10111011	Ы	219	11011011	ы	251	11111011
њ	156	10011100	ј	188	10111100	Ь	220	11011100	ь	252	11111100
ќ	157	10011101	ѕ	189	10111101	Э	221	11011101	э	253	11111101
ћ	158	10011110	ѕ	190	10111110	Ю	222	11011110	ю	254	11111110
џ	159	10011111	ї	191	10111111	Я	223	11011111	я	255	11111111

Принцип последовательного кодирования алфавита:

В кодовой таблице ASCII латинские буквы (прописные и строчные) располагаются в алфавитном порядке. Расположение цифр также упорядочено по возрастанию значений.

Данное правило соблюдается и в других таблицах кодировки. Благодаря этому и в машинном представлении для символьной информации сохраняется понятие «алфавитный порядок».

Различные кодировки знаков

В настоящее время существуют пять различных кодировок для русских букв

(Windows, MS-DOS, КОИ-8, Mac, ISO), поэтому тексты созданные в одной кодировке не будут правильно отображаться в другой.

В последние годы широкое распространение получил новый международный стандарт кодирования текстовых символов Unicode, который отводит на каждый символ 2 байта (16 битов), что позволило закодировать многие алфавиты в одной таблице.

$N = 2^1$, $N = 2^{16}$, $N = 65536$

Пример 1

Буква «i» в таблице кодировки символов имеет десятичный код 105.

Что зашифровано последовательностью десятичных кодов:
108 105 110 107?

При расшифровке данной последовательности кодов не нужно обращаться к таблице кодировки символов.

Необходимо учесть принцип последовательного кодирования алфавитов и вспомнить порядок букв в латинском алфавите (...i, j, k, l, m, n, o, ...)

Закодировано: «link»

Пример 2

С помощью последовательности десятичных кодов:

99 111 109 112 117 116 101 114 закодировано слово «computer». Какая последовательность десятичных кодов будет соответствовать этому же слову, записанному заглавными буквами?

При шифровке слова не пользоваться таблицей кодировки символов. Необходимо учесть, что разница между десятичным кодом строчной буквы латинского алфавита и десятичным кодом соответствующей заглавной буквы равна 32.

Если «с» имеет код 99, то «С» имеет код $99 - 32 = 67$.

COMPUTER – 67 79 77 80 85 84 69 82

Пример 3

- Слова «компьютер»?
 - Слова «алфавит»?
-