

Кодирование текстовой информации

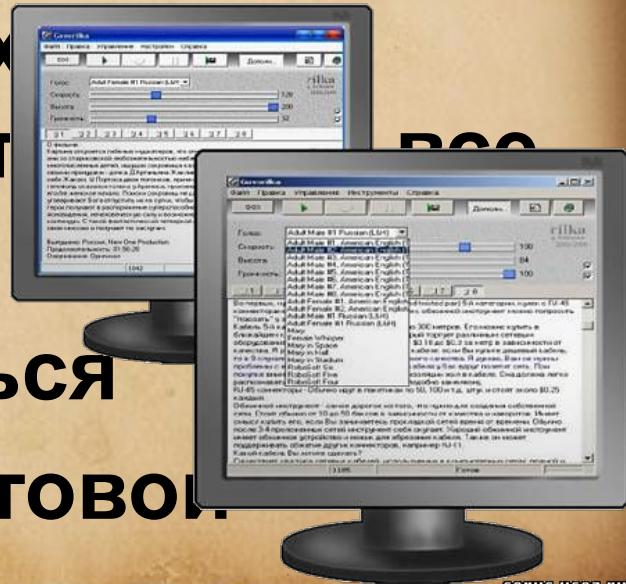


АЗБУКА МОРЗЕ

А	.-	И	..	Р	..	Ш	---
Б	-...	Й	---	С	...	Щ	---.
В	.-.	К	-.-	Т	-	Ъ	-.-..
Г	--.	Л	..	У	.-	Ы	.-.
Д	-..	М	--	Ф	-..	Ь	-..-
Е	.	Н	-..	Х	...	Э	..-.
Ж	...-	О	---	Ц	-..	Ю	..--
З	--.	П	-..	Ч	-..	Я	-.-..

Информация, выраженная с помощью естественных и формальных языков в письменной форме, называется текстовой информацией.

Начиная с конца 60-х прошлого века, компьютеры стали использоваться для обработки текстовой информации.



Историческая справка

Пайцике тсут т "камащамлтой чмароке" — кайтонили, нмирепяшеейля т Молли ися цинсоракигелтой неменилти

Криптография – это тайнопись, система изменения письма с целью сделать текст непонятным для непосвященных лиц

Знаки кода Морзе	Буквы		Знаки кода Морзе	Цифры	Знаки кода Морзе	Знаки препинания и служебные сигналы
	Рус.	Анг.				
·-·-·	А	Aa	·-·-·	1	·-·-·	(.) запятая
·-·-·	Б	Bb	·-·-·	2	·-·-·	(.) Точка
·-·-·	В	Vv	·-·-·	3	·-·-·	(;) точка с запятой
·-·-·	Г	Gg	·-·-·	4	·-·-·	(:) двоеточие
·-·-·	Д	Dd	·-·-·	5	·-·-·	(?) вопросит. знак
·-·-·	Е	Ee	·-·-·	6	·-·-·	(N#) номер
·-·-·	Ж	Vv	·-·-·	7	·-·-·	(') ковычки
·-·-·	З	Zz	·-·-·	8	·-·-·	(') апостроф
·-·-·	И	Ii	·-·-·	9	·-·-·	() скобки
·-·-·	К	Kk	·-·-·	0	·-·-·	(!) восклицательный знак
·-·-·	Л	Ll	·-·-·		·-·-·	(-) тире
·-·-·	М	Mm	·-·-·		·-·-·	Ждать
·-·-·	Н	Nn	·-·-·		·-·-·	Понял
·-·-·	О	Oo	·-·-·		·-·-·	(/) дробная черта
·-·-·	П	Pp	·-·-·		·-·-·	Знак раздела
·-·-·	Р	Rr	·-·-·		·-·-·	Перебой (исправление ошибки)
·-·-·	С	Ss	·-·-·		·-·-·	Сигнал о начале передачи (НП)
·-·-·	Т	Tt	·-·-·		·-·-·	Сигнал о готовности к приёму (ПО)
·-·-·	У	Uu	·-·-·		·-·-·	Начало действия
·-·-·	Ф	Ff	·-·-·		·-·-·	Знак окончания передачи
·-·-·	Х	Hh	·-·-·		·-·-·	
·-·-·	Ц	Cc	·-·-·		·-·-·	
·-·-·	Ч		·-·-·		·-·-·	
·-·-·	Ш		·-·-·		·-·-·	
·-·-·	Щ	Qq	·-·-·		·-·-·	
·-·-·	Ы	Yy	·-·-·		·-·-·	
·-·-·	Ю		·-·-·		·-·-·	
·-·-·	Я	Jj	·-·-·		·-·-·	
·-·-·	Ь,ъ	Xx	·-·-·		·-·-·	
·-·-·	Э	Ee	·-·-·		·-·-·	

Азбука Морзе или неравномерный телеграфный код, в котором каждая буква или знак представлены своей комбинацией коротких элементарных посылок электрического тока (точек) и элементарных посылок утроенной продолжительности (тире)

Сурдожесты – язык жестов, используемый людьми с нарушениями слуха



Кодирование текстовой информации -

преобразование входной информации в форму, воспринимаемую компьютером, т.е. данные (двоичный код).

Декодирование текстовой информации -

преобразование данных из двоичного кода в форму, понятную человеку.

**Для кодирования одного
символа требуется один
байт информации**

1 символ – 1 байт (8 бит)

Учитывая, что каждый бит принимает значение 1 или 0, получаем, что с помощью 1 байта можно закодировать 256 различных СИМВОЛОВ.

$$N=2^I$$

N – мощность алфавита

$$2^8=256$$

I – информационный вес

$$N=2^I \Rightarrow 256=2^I \Rightarrow 2^8=2^I \Rightarrow I=8 \text{ битов} = 1 \text{ байт}$$

Кодирование и декодирование текстовой информации

Кодирование заключается в том, что каждому символу ставится в соответствие уникальный десятичный код от 0 до 255 или соответствующий ему двоичный код от 00000000 до 11111111. Таким образом, человек различает символы по их начертанию, а компьютер - по их коду.

Процессор

010000

01

Оперативная память

0100000

1



Таблица кодов

**таблица, в которой всем символам
компьютерного алфавита поставлены в
соответствие порядковые номера (коды)**

Таблица кодировки ASCII

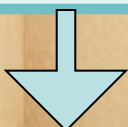
**является стандартной, и ее понимают
абсолютно все программы, работающие с
текстами.**

Кодовая таблица *ASCII*

American Standard Code for Information Interchange

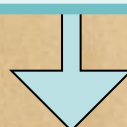
sp 32	! 33	« 34	# 35	\$ 36	% 37	& 38	' 39	(40) 41	* 42	+ 43	, 44	- 45	. 46	/ 47
0 48	1 49	2 50	3 51	4 52	5 53	6 54	7 55	8 56	9 57	: 58	; 59	< 60	= 61	> 62	? 63
@ 64	A 65	B 66	C 67	D 68	E 69	F 70	G 71	H 72	I 73	J 74	K 75	L 76	M 77	N 78	O 79
P 80	Q 81	R 82	S 83	T 84	U 85	V 86	W 87	X 88	Y 89	Z 90	[91	\ 92] 93	^ 94	_ 95
` 96	a 97	b 98	c 99	d 100	e 101	f 102	g 103	h 104	i 105	j 106	k 107	l 108	m 109	n 110	o 111
p 112	q 113	r 114	s 115	t 116	u 117	v 118	w 119	x 120	y 121	z 122	{ 123	 124	} 125	~ 126	ف 127

коды от 0 до 31



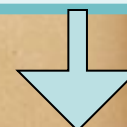
функциональные
клавиши

коды от 32 до 127



буквы английского алфавита,
знаки математических операций и т.д.

коды от 128 до 255



национальный
алфавит

Кодовая таблица в системе Windows

Символ Десятичный код Двоичный код Символ Десятичный код Двоичный код

Пробел	32	0010 0000	0	48	0011 0000
!	33	0010 0001	1	49	0011 0001
*	42	0010 1010	2	50	0011 0010
+	43	0010 1011	3	51	0011 0011
,	44	0010 1100	4	52	0011 0100
-	45	0010 1101	5	53	0011 0101
.	46	0010 1110	6	54	0011 0110
/	47	0010 11110	7	55	0011 0111
=	61	0011 1101	8	56	0011 1000
?	63	0011 1111	9	57	0011 1001

А	192	1100 0000	Р	208	1101 0000
Б	193	1100 0001	С	209	1101 0001
В	194	1100 0010	Т	210	1101 0010
Г	195	1100 0011	У	211	1101 0011
Д	196	1100 0100	Ф	212	1101 0100
Е	197	1100 0101	Х	213	1101 0101
Ж	198	1100 0110	Ц	214	1101 0110
З	199	1100 0111	Ч	215	1101 0111
И	200	1100 1000	Ш	216	1101 1000
Й	201	1100 1001	Щ	217	1101 1001
К	202	1100 1010	Ъ	218	1101 1010
Л	203	1100 1011	Ы	219	1101 1011
М	204	1100 1100	Ь	220	1101 1100
Н	205	1100 1101	Э	221	1101 1101
О	206	1100 1110	Ю	222	1101 1110
П	207	1100 1111	Я	223	1101 1111

Кодировки русского

алфавита

В настоящее время широкое распространение получил Новый международный стандарт Unicode, который отводит на каждый символ не один байт, а два, и потому с его помощью можно закодировать не 256 символов, а $N = 2^{16} = 65\,536$ различных символов. Такого количества символов достаточно, чтобы закодировать не только русский и латинский алфавиты, цифры, знаки и математические символы, но и греческий, арабский, иврит и другие алфавиты.



Таблица кодировки

Unicode

Стандарт кодирования Unicode

**отводит на каждый символ 2 байта, что
позволяет закодировать многие
алфавиты в одной таблице.**

$$N=2^I=2^{16}=65\ 536$$

**В настоящее время существует
5 кодовых таблиц для русских букв
(Windows, MS-DOS, КОИ-8, Mac,
ISO), поэтому тексты, созданные в
одной кодировке, не будут правильно
отображаться в другой.**

Таблицы кодировки русскоязычных символов KOI8-R

128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

CP1251

128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

CP866

128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

Mac

128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

ISO

128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

СИМВОЛ	Windows	MS-DOS	КОИ-8	Mac	ISO	Unicode
А	192	128	225	128	176	1040
В	194	130	247	130	178	1042
М	204	140	237	140	188	1052
Э	221	157	252	157	205	1069
Я	255	239	241	223	239	1103

Декодировать текст с помощью кодовой
таблицы **ASCII**:

99 111 109 112 117 116 101 114

computer

Работа в текстовом редакторе

Запустите текстовый редактор

MS Word

MS Word. Удерживая клавишу «ALT»,
наберите коды на дополнительной
цифровой клавиатуре:

152 170 174 171 160

Школа

Домашнее задание

п. 3.1.

Произвести кодирование

стихотворения из 4-х строк

(до 100 символов)