



Кодирование текстовой информации

В традиционных кодировках для кодирования одного символа используется 8 бит. Легко подсчитать, что такой 8-разрядный код позволяет закодировать 256 различных СИМВОЛОВ.

Присвоение символу определенного числового кода – это вопрос соглашения.

В качестве международного стандарта принята кодовая таблица ASCII (American Standard Code for Information Interchange), кодирующая первую половину символов с числовыми кодами от 0 до 127 (коды от 0 до 32 отведены не символам, а функциональным клавишам).

Кодовая таблица ASCII

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

Во второй половине содержатся коды национальных алфавитов, символы псевдографики и некоторые математические знаки. К сожалению, в настоящее время существуют пять различных кодировок кириллицы (КОИ8-Р, Windows, MS-DOS, Macintosh и ISO), что вызывает дополнительные трудности при работе с русскоязычными документами.

Хронологически одним из первых стандартов кодирования русских букв на компьютерах был КОИ8 .

Эта кодировка применялась еще в 70-ые годы на компьютерах серии ЕС ЭВМ, а с середины 80-х стала использоваться в первых русифицированных версиях операционной системы UNIX.

КОИ8 («Код обмена информацией, 8-битный»)

—		Г	Г	└	└	┌	┌	┐	┐	┘	■	■	■	■	■
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
▒	▒	▒	Г	■	●	√	≈	≤	≥	nbsp	Ј	◦	²	•	÷
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
=		F	ё	П	П	Э	П	П	Е	Ц	Ц	Ц	Ц	Ц	┌
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
┌	┌	┌	Ё			Т	П	П	±	Ц	Ц	┌	┌	┌	©
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
Ю	а	б	ц	д	е	ф	г	х	и	й	к	л	м	н	о
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
п	я	р	с	т	у	ж	в	ь	ы	з	ш	э	щ	ч	ъ
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
Ю	А	Б	Ц	Д	Е	Ф	Г	Х	И	Й	К	Л	М	Н	О
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
П	Я	Р	С	Т	У	Ж	В	Ь	Ы	З	Ш	Э	Щ	Ч	Ъ
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

Наиболее распространенной в настоящее время является кодировка Microsoft Windows, обозначаемая сокращением CP1251 ("CP" означает "Code Page", "кодовая страница").

Á	à	,	è	„	...	†	‡	€	‰	É	<	Й	Ї	Ó	Ú
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
á	‘	’	“	”	•	–	—	ë	™	é	>	ò	í	ó	ú
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
nbsp	ÿ	Ы	Э	И	Ы	І	§	Ё	©	Ю	«	¬	shy	®	Я
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
°	±	Ы	Э	’	µ	¶	•	ё	№	ю	»	Э	ю	я	я
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

От начала 90-ых годов, времени господства операционной системы MS DOS, остается кодировка CP866.

А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
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
р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
Ё	ё	Є	є	Ї	ї	Ў	ў	°	•	•	√	N°	¤	■	nbsp
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

Компьютеры фирмы Apple, работающие под управлением операционной системы Mac OS, используют свою собственную кодировку Mac.

А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
†	°	Ы	£	§	•	¶	Ы	®	©	™	Á	á	ě	à	è
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
Ë	±	≤	≥	э	μ	г	Э	Ю	ю	Я	я	É	é	Й	ò
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
Э	Ю	¬	√	f	≈	Δ	«	»	...	nbsp	Ó	ó	Й	й	я
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
–	—	“	”	‘	’	÷	„	ù	Ы	ú	ý	Nº	Ë	ë	я
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

Международная организация по стандартизации (International Standards Organization, ISO) утвердила в качестве стандарта для русского языка еще одну кодировку под названием ISO 8859-5.

І	І	І	І	І	І	І	І	І	І	І	І	І	І	І	І
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
І	І	І	І	І	І	І	І	І	І	І	І	І	І	І	І
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
nbsp	Ё	Ъ	Ѓ	Є	Ѕ	І	Ї	Ј	Љ	Њ	Ћ	Ќ	shy	Ў	Џ
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
№	ё	ђ	ѓ	є	ѕ	і	ї	ј	љ	њ	ћ	ќ	ѕ	ў	џ
240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

Unicode

В конце 90-ых годов появился новый международный стандарт Unicode, который отводит под один символ **не один байт, а два**, и поэтому с его помощью можно закодировать не 256, а 65536 различных символов. Полная спецификация стандарта Unicode включает в себя все существующие, вымершие и искусственно созданные алфавиты мира, а также множество математических, музыкальных, химических и прочих символов.

Пример

Представьте в форме шестнадцатеричного кода слово «ЭВМ».

- CP1251: 221 194 204
- CP866: 157 130 140 .

В 16 с/с

- CP1251: DD C2 CC
- CP866: 9D 82 8C

Задания

1. Закодируйте с помощью кодировочной таблицы ASCII следующие тексты:

а) Password; б) Windows.

2. Декодируйте с помощью кодировочной таблицы ASCII следующие тексты:

а) 54 6F 72 6E 61 64 6F;

б) 49 20 6C 6F 76 65 20 79 6F 75.