



Кодирование текстовой информации



保

*Защита
(входная дверь,
детская комната)*

春

*Весна
(предприятия)*

美

*Красота
(ванная комната)*

道

*Путь
(решения)*

光

*Свет
(депрессии)*

財

*Благосостояние
(кошелек, банков-
ские документы)*

Кол

(по

не

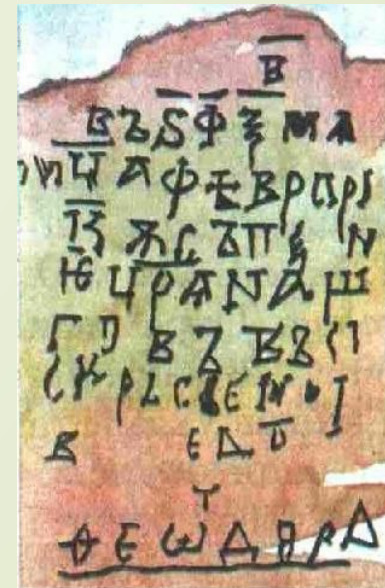
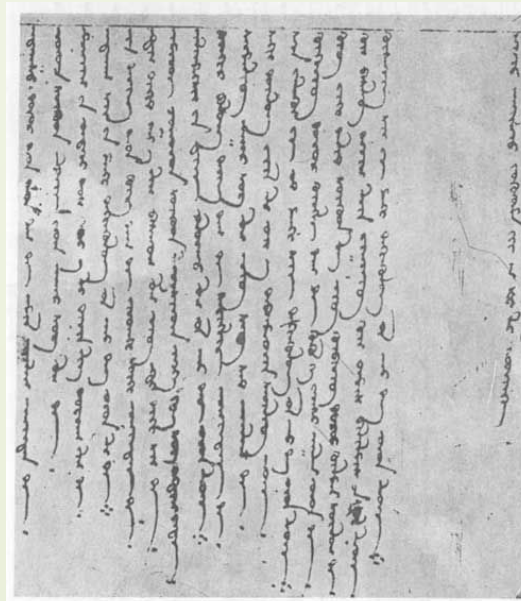


ов

ей

а,

Письменность – способ кодирования устной речи на естественном языке





**СПОСОБ КОДИРОВАНИЯ ЗАВИСИТ ОТ
НАЗНАЧЕНИЯ КОДА**

**ПРАВИЛО – КАЖДЫЙ СИМВОЛ АЛФАВИТА
ИСХОДНОГО ТЕКСТА ЗАМЕНЯЕТСЯ НА
КОМБИНАЦИЮ СИМВОЛОВ АЛФАВИТА
КОДИРОВАНИЯ**

00	01	02	03	04	05	06	07
NUL	E 3	LF	A -	SP	S ' I 8	U 7	
08	09	0A	0B	0C	0D	0E	0F
CR	D [NO]	R 4	J BEL	N ,	F !	C :	K <
10	11	12	13	14	15	16	17
T 5	Z +	L >	W 2	H £	Y 6	P 0	Q 1
18	19	1A	1B	1C	1D	1E	1F
0 9	B ?	G &	FIGS	M .	X /	V ;	LTRS
Letters			Figures		Control Chars.		

Буквы

знаки

СИМВОЛЫ

**Режим
ввода
букв**

Телеграфный код ИТА2

Кодировка ASCII

(American Standard Code for Information Interchang) –

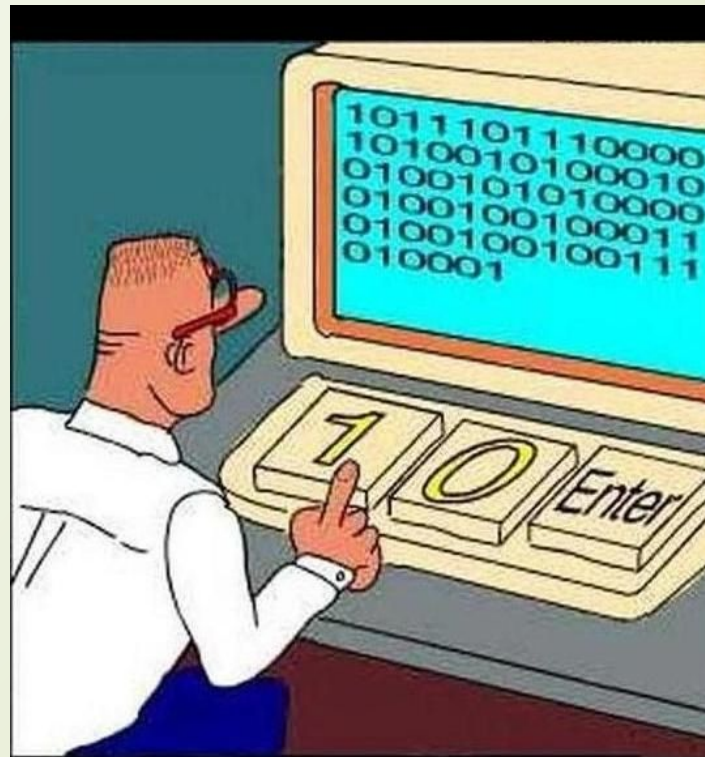
1963 год – для компьютерной обработки текстовой информации

кодирующая первую половину символов с числовыми кодами от 0 до 127

	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9	.A	.B	.C	.D	.E	.
0.	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	TAB	LF	VT	FF	CR	SO	SI
1.	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2.		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3.	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	:
4.	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5.	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6.	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7.	p	q	r	s	t	u	v	w	x	y	z	{		}	DE	

(коды от 0 до 32 отведены не символам, а функциональным клавишам).

- Код символа – порядковый номер
- Первые 32 символа – управляющие. На экране не отражаются, определяют некоторое действие.



Расширение кода

ASCII

	80	90	A0	B0	C0	D0	E0	F0
0	А	Р	а	⋮	Л	Ц	Р	⋮
1	Б	С	б	▀	⊥	⊟	с	±
2	В	Т	в	▁	⊓	⊠	т	>
3	Г	У	г		⊔	⊡	у	<
4	Д	Ф	д	⊕	—	⊢	ф	⊗
5	Е	Х	е	⊖	+	⊣	х	⊘
6	Ж	Ц	ж	⊗	⊤	⊤	ц	÷
7	З	Ч	з	⊘	⊥	⊥	ч	≈
8	И	Ш	и	⊙	⊦	⊦	ш	⊚
9	Й	Щ	й	⊚	⊧	⊧	щ	⋅
А	К	Ь	к	⊛	⊨	⊨	ь	⋆
В	Л	Ы	л	⊜	⊩	■	ы	⊞
С	М	Ъ	м	⊝	⊪	■	ъ	⊟
Д	Н	Э	н	⊞	=	◐	э	⊠
Е	О	Ю	о	⊠	⊫	◑	ю	⊡
Ф	П	Я	п	⊡	⊬	■	я	

1 – 127 совпадают с ASCII
128 – 225 – кодовая страница.
Размещаются нелатинские
алфавиты, символы
псевдографики...

Наиболее распространенной в настоящее время является кодировка Microsoft Windows, обозначаемая сокращением

CP1251

("CP" означает "Code Page" "кодовая


Á	à	,	è	„	…	†	‡	€	%	É	<	й	Й	Ó	ú
128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
á	‘	’	“	”	•	–	—	è	™	é	>	ò	й	ó	ú
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
nbsp	ỳ	Ы	Э	н	ы	!	§	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

Хронологически одним из первых стандартов кодирования русских букв на компьютерах был **КОИ8** ("Код обмена информацией, 8-битный"). Unix

—		Г	Г	Л	Л	Т	Т	Т	Т	Т	■	■	■	■	■
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
=		Р	ё	П	Г	Э	П	П	Е	Ц	Ц	Ј	Ш	Ш	Т
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
		Э	Ё			Т	П	П	±	Ц	Ц	Т			©
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

#154 МЕПЮГПШБМШИ ОПНАЕК.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	—		Г	Г	└	└	┌	┌	┐	┐	■	■	■	■	■	■
9	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
A	▒	▒	▒	“	■	●	”	—	№	™	nbsp	»	®	«	•	¤
B	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
C	=		F	ё	€	Г	і	ї	Г	Е	Ц	Ц	Г	Ў	Г	
D	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
E	Г	Г	Г	Ё	Є	Г	І	Ї	Г	Е	Ц	Ц	Г	Ў	©	
F	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
Т	Ю	а	б	ц	д	е	ф	г	х	и	й	к	л	м	н	о
Д	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207
Е	п	я	р	с	т	у	ж	в	ь	ы	з	ш	э	щ	ч	ъ
С	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223
О	Ю	А	Б	Ц	Д	Е	Ф	Г	Х	И	Й	К	Л	М	Н	О
Р	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
Т	П	Я	Р	С	Т	У	Ж	В	Ь	Ы	З	Ш	Э	Щ	Ч	Ъ
Е	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255



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

	040	041	042	043	044	045	046	047	048	049	04A	04B	04C	04D	04E	04F
0	È	А	Р	а	р	è	Ѧ	Ѣ	Ѥ	Г	К	Ѹ	І	Ǻ	З	Ў
1	Ë	Б	С	б	с	ë	Ѧ	Ѣ	Ѥ	Г	к	Ѹ	Ж	ǻ	з	ў
2	Ђ	В	Т	в	т	ђ	Ђ	Ө	Ѧ	Н	Х	Ж	Ǻ	Й	Ў	
3	Ѓ	Г	У	г	у	ѓ	Ѧ	ө	Ѧ	н	х	Ђ	ǻ	й	ў	
4	Є	Д	Ф	д	ф	є	Ю	Ѹ	Ѧ	Н	Ц	Ђ	Æ	Й	Ї	
5	Ɔ	Е	Х	е	х	Ɔ	Ю	Ѹ	Ђ	н	ц	Д	æ	й	ї	
6	І	Ж	Ц	ж	ц	і	Ѓ	Ѹ	Ѧ	Ц	Ч	Д	Ё	Ö		
7	Ї	З	Ч	з	ч	ї	Ѓ	Ѹ		ж	ц	ч	Н	ё	ö	
8	Ј	И	Ш	и	ш	ј	Ѓ	Оу	☀	З	Ѧ	Ч	Н	Э	Ө	Ї
9	Љ	Й	Щ	й	щ	љ	Ѓ	оу	☀	з	Ѧ	ч	Н	э	ө	ї
A	Њ	К	Ъ	к	ъ	њ	Ж	О	Й	К	Ѥ	Һ	ң	Ӗ	Ӗ	
B	Ћ	Л	Ы	л	ы	ќ	ж	О	й	к	Ѥ	Һ	Ч	ӗ	ӗ	
C	Ќ	М	Ь	м	ь	ќ	Ж	О	Ь	К	Т	Ѧ	ч	Ж	Ӗ	
D	Ў	Н	Э	н	э	ў	Ж	О	Ь	к	т	Ѧ	М	ж	Ӗ	
E	Ў	О	Ю	о	ю	ў	Ж	Ѧ	Р	К	У	Ѧ	М	Ӗ	Ў	
F	Ѹ	П	Я	п	я	Ѹ	Ѧ	р	к	у	Ѧ			Ӗ	ў	

Фрагмент спецификации и UNICODE 4.0 для кириллицы



Пример 1.

Представьте в форме шестнадцатеричного кода слово «ЭВМ» во всех пяти кодировках.

Воспользуйтесь компьютерным калькулятором для перевода чисел из десятичной в шестнадцатеричную систему счисления

Ответ

Последовательности десятичных кодов слова «ЭВМ» в различных кодировках составляем на основе кодировочных таблиц:

КОИ8-Р: 252 247 237

CP1251: 221 194 204

CP866: 157 130 140

Mac: 157 130 140

ISO: 205 178 188

Переводим с помощью калькулятора последовательности кодов из десятичной системы в шестнадцатеричную:



КОИ8-Р: FC F7 ED

CP1251: DD C2 CC


CP866: 9D 82 8C

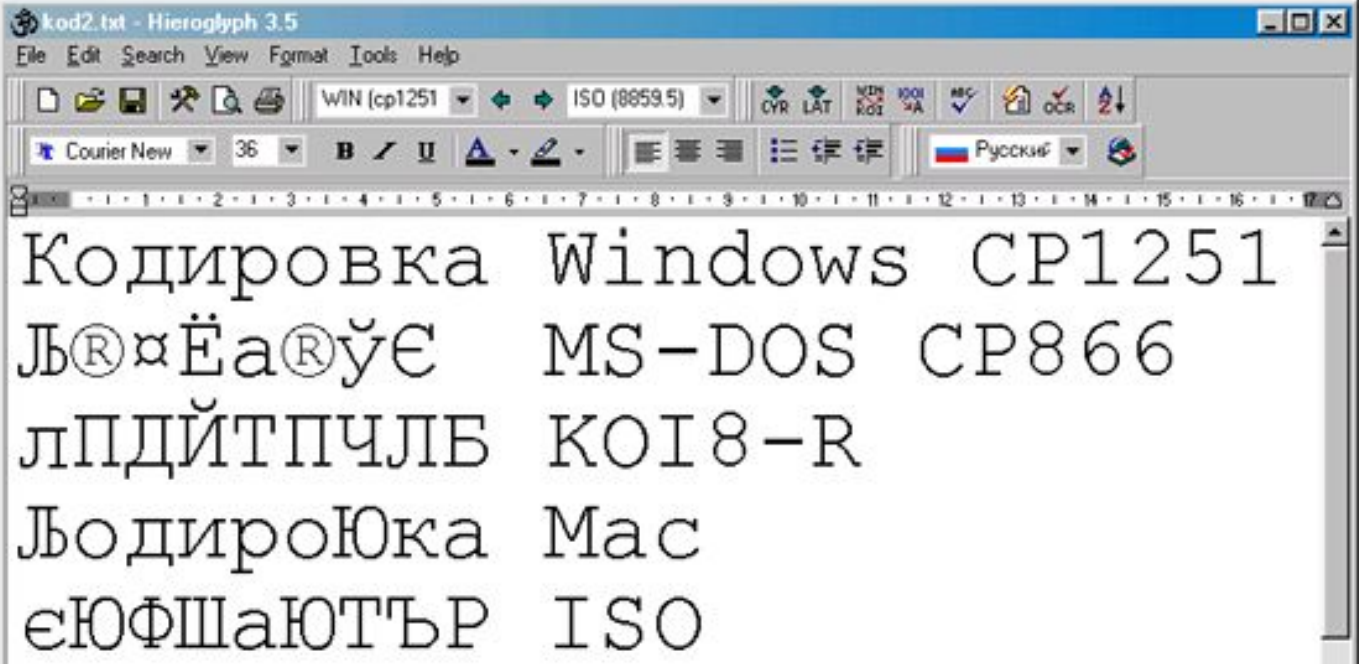
Mac: 9D 82 8C

ISO: CD B2 BC

- 
- 
- Для преобразования русскоязычных текстовых документов из одной кодировки в другую используются специальные программы-конверторы.
 - Одной из таких программ является текстовый редактор **Hieroglyph**, который позволяет осуществлять перевод набранного текста из одной кодировки в другую и даже использовать различные кодировки в одном тексте

Пример 2.47. Представить в пяти различных кодировках слово «Кодировка». Выполним это задание с использованием текстового редактора Hieroglyph.

	Перекодирование текста.
1	Запустить текстовый редактор Hieroglyph.
2	В раскрывающемся списке исходных кодировок выбрать кодировку WIN(cp1251) и ввести текст: «Кодировка Windows CP1251».
3	Скопировать текст четыре раза и, выделяя строки, последовательно выбрать в раскрывающемся списке конечные кодировки (DOS, KOI8-R, Mac и ISO), каждый раз нажимая кнопку перекодирования. Для каждой кодировки отредактировать ее название.
4	В результате текст будет состоять из пяти строк, записанных в различных кодировках.



The screenshot shows the Hieroglyph 3.5 editor window titled 'kod2.txt - Hieroglyph 3.5'. The menu bar includes File, Edit, Search, View, Format, Tools, and Help. The toolbar shows the current encoding as WIN (cp1251) and the target encoding as ISO (8859 5). The text area contains five lines of text, each representing the word 'Кодировка' encoded in a different character set:

- Кодировка Windows CP1251
- Љ®αЁа®ўЄ MS-DOS CP866
- лпдйтпчљб KOI8-R
- ЉодироЮка Mac
- єЮФШаЮТЬР ISO

Учимся программировать

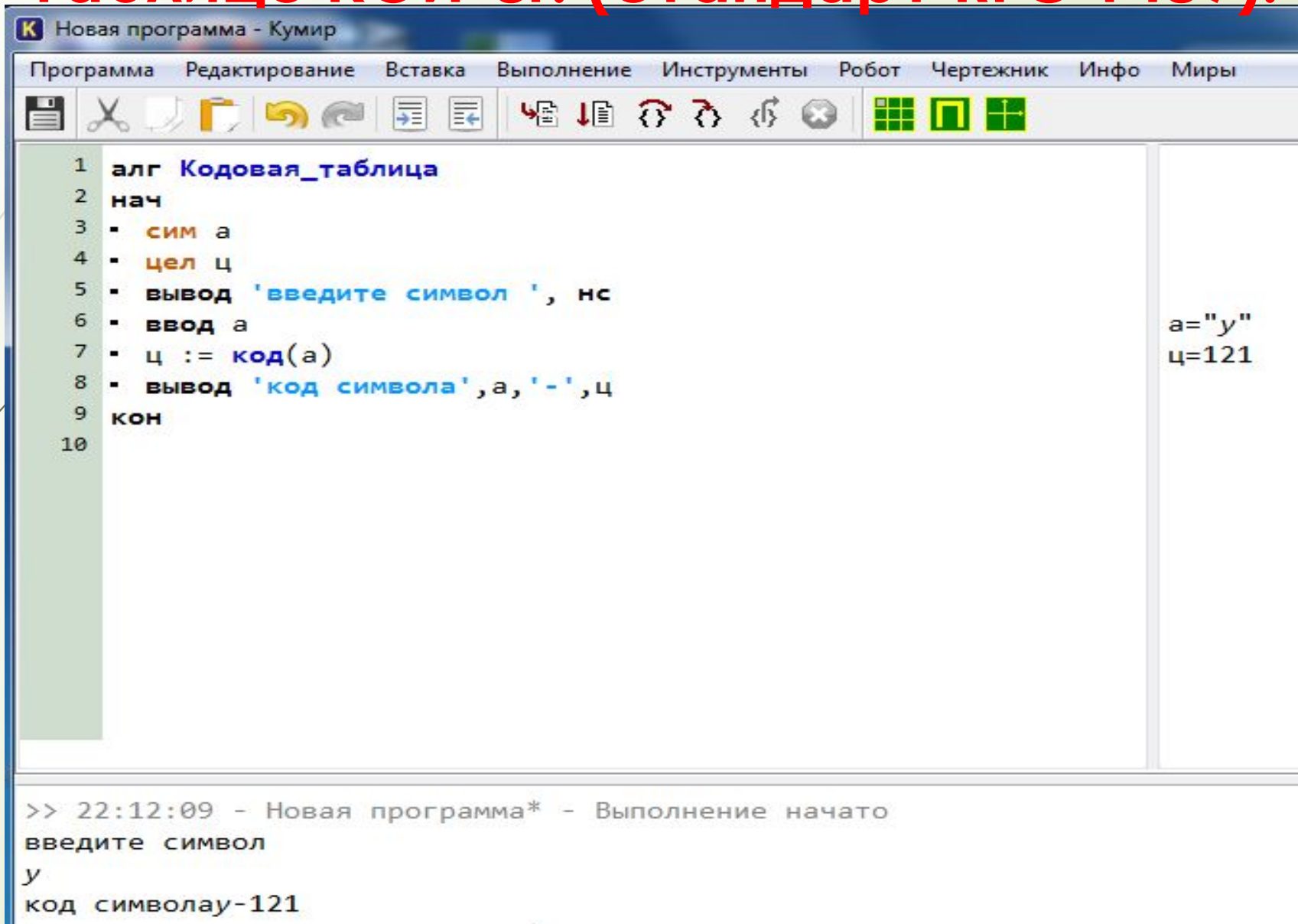
```
Программа  Редактирование  Вставка  Выполнение  Инструменты  Робот  Чертежник  Инфо  Миры
[Icons]
1  алг Кодовая_таблица
2  нач
3  ▪ цел а
4  ▪ сим б
5  ▪ нц для а от 32 до 127
6  ▪ ▪ б := символ2(а)
7  ▪ ▪ вывод б, ' '
8  ▪ кц
9  кон
10 |
```

a=127
б=""

**Возвращает символ,
соответствующий номеру N в
таблице Юникода.**

```
>> 22:06:03 - Новая программа* - Выполнение начато
! " # $ % & ' ( ) * + , - . / 0 1 2 3 4 5 6 7 8 9 :
; < = > ? @ A B C D E F G H I J K L M N O P Q R S T U
V W X Y Z [ \ ] ^ _ ` a b c d e f g h i j k l m n o
p q r s t u v w x y z { | } ~ ☐
>> 22:06:17 - Новая программа* - Выполнение завершено
```


Возвращает номер символа в таблице КОИ-8r. (стандарт RFC 1489).



```
1  алг Кодовая_таблица
2  нач
3  - сим а
4  - цел ц
5  - вывод 'введите символ ', нс
6  - ввод а
7  - ц := код(а)
8  - вывод 'код символа', а, '-', ц
9  кон
10
```

a="y"
ц=121

>> 22:12:09 - Новая программа* - Выполнение начато
введите символ
у
код символау-121



```
К Новая программа - Кумир
Программа Редактирование Вставка Выполнение Инструменты Робот Чертежник Инфо Миры
[Icons]
1 алг Кодовая_таблица
2 нач
3   ▪ сим а
4   ▪ цел с
5   ▪ вывод 'введите любой символ с клавиатуры ', нс
6   ▪ ввод а
7   ▪ с:=юникод(а)
8   ▪ вывод 'код символа ',а,' в таблице Unicod -',с
9 кон
10

a="&"
с=38

>> 22:24:11 - Новая программа* - Выполнение начато
введите любой символ с клавиатуры
&
код символа & в таблице Unicod -38
>> 22:24:29 - Новая программа* - Выполнение завершено
```