

ГБОУ Школа № 1394

Кодирование и обработка текстовой информации

Автор: учитель информатики Гилева Елена Евгеньевна

Начиная с 60-х годов, компьютеры все больше стали использовать для обработки текстовой информации. В настоящее время большая часть ПК в мире занято обработкой именно текстовой информации.



«ПАЗЛЫ» УРОКА

ВСПОМНИМ

ПОВТОРИМ

ПОДУМАЕМ

ЗАПОМНИМ

УЗНАЕМ

ВЫПОЛНИМ

ПРОВЕРИМ

«ПАЗЛЫ» УРОКА

У

ПОВТОРИМ

ПОДУМАЕМ

ЗАПОМНИМ

УЗНАЕМ

ВЫПОЛНИМ

ПРОВЕРИМ

«ПАЗЛЫ» УРОКА

У

ЗАПОМНИМ

И

УЗНАЕМ

ПОДУМАЕМ

ВЫПОЛНИМ

ПРОВЕРИМ

«ПАЗЛЫ» УРОКА

У

Н

І

ЗАПОМНИМ

УЗНАЕМ

ВЫПОЛНИМ

ПРОВЕРИМ

«ПАЗЛЫ» УРОКА

У

Н

І

С

УЗНАЕМ

ВЫПОЛНИМ

ПРОВЕРИМ



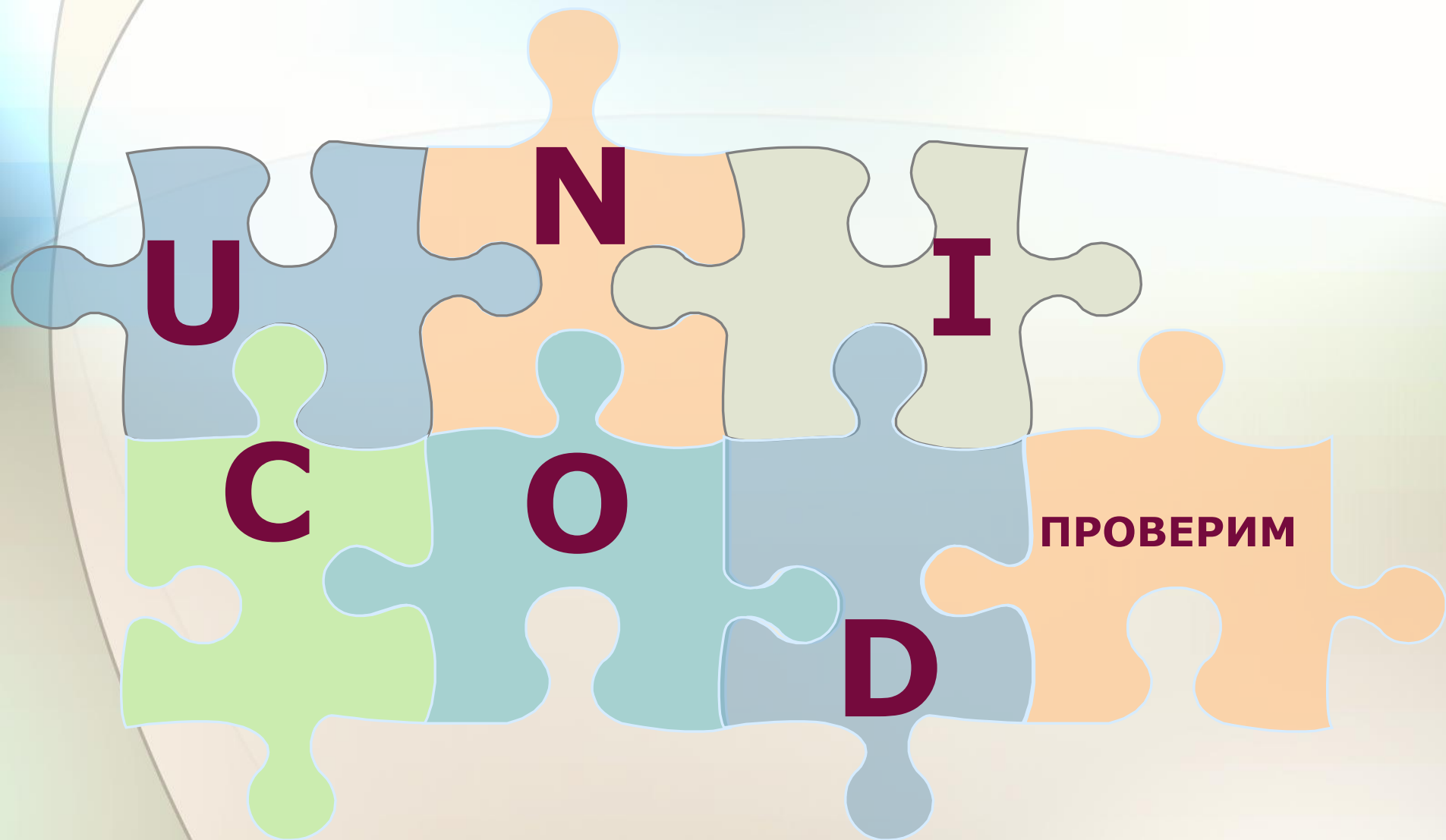
«ПАЗЛЫ» УРОКА



«ПАЗЛЫ» УРОКА



«ПАЗЛЫ» УРОКА



U

N

I

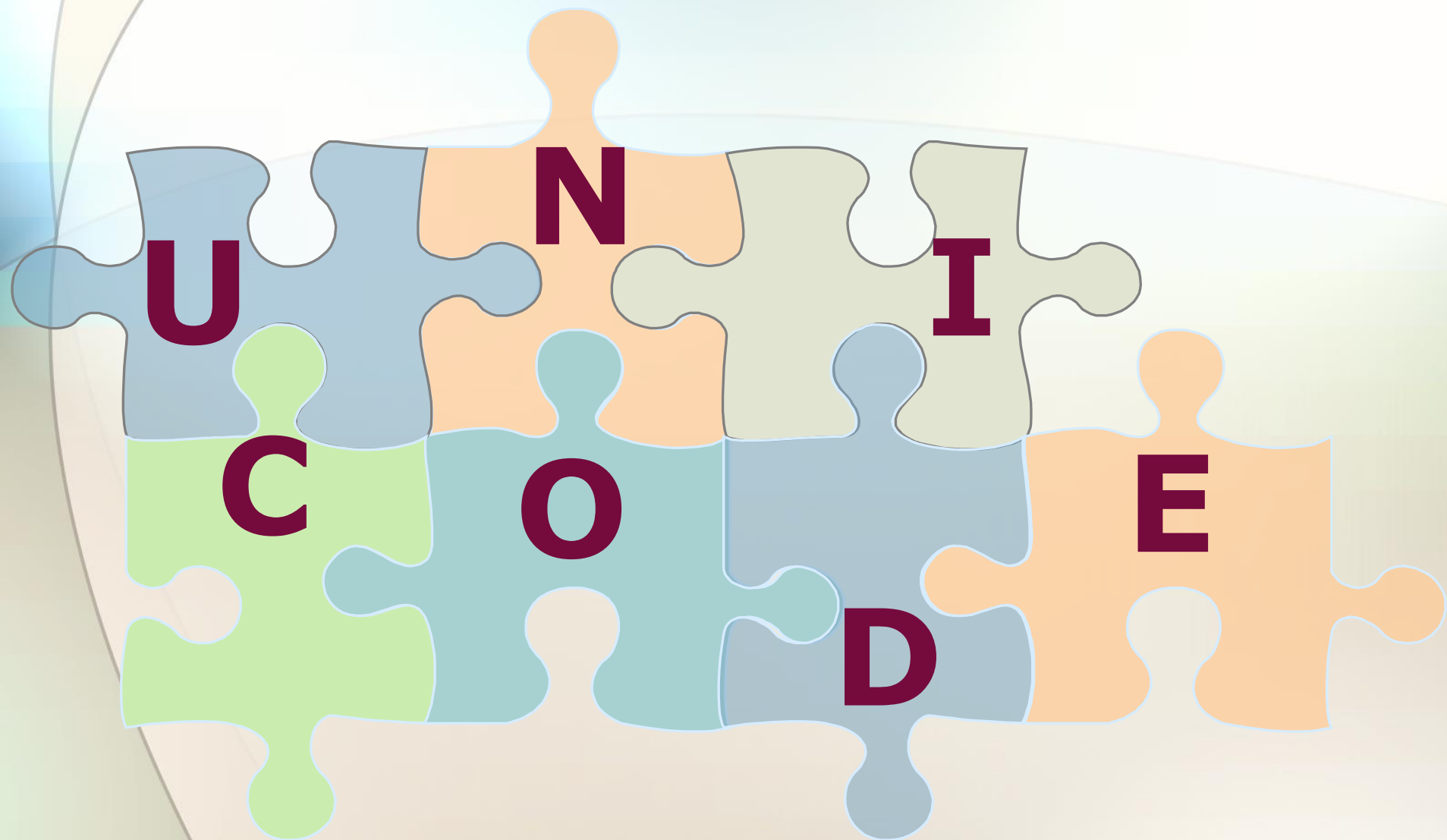
C

O

D

ПРОВЕРИМ

«ПАЗЛЫ» УРОКА





UNICODE

В конце 90-ых годов появился новый международный стандарт **Unicode**, который отводит под один символ не один байт, а два (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

<http://fcior.edu.ru/card/28666/predstavlenie-teksta-v-razlichnyh-kodirovках.html>

Представление текста в различных кодировках, ЭОР

САМОЕ ГЛАВНОЕ

- Для кодирования одного символа используется количество информации, **равное 1 байту.**
- Таблица, в которой всем символам компьютерного алфавита поставлены в соответствие порядковые номера (коды), называется **таблицей кодировки.**
- Существуют различные кодовые таблицы: **ASC II, КОИ8, CP1251** и др.
- **Unicode** международный стандарт, который **отводит под один символ два байта.**

ДОМАШНЕЕ ЗАДАНИЕ

1. п.1.1.1, учебник Угринович Н.Д. Информатика и ИКТ, для 10 класса,- М., 2010

2. ЭОР, Представление текста в различных кодировках.

<http://fcior.edu.ru/card/28666/predstavlenie-teksta-v-razlichnyh-kodirovkah.html>

3. ЭОР, Представление текста в различных кодировках, проверь себя.

<http://fcior.edu.ru/card/28605/predstavlenie-teksta-v-razlichnyh-kodirovkah.html>

Вспомним известные факты

Процесс преобразования информации в форму, воспринимаемую компьютером **называется...**

Процесс обратный кодированию **называется ...**

Каким шифром закодировано словосочетание?

з	е	л	ё	н	а	я		ё	л	к	а
---	---	---	---	---	---	---	--	---	---	---	---

к	з	о	и	р	г	в		и	о	н	г
---	---	---	---	---	---	---	--	---	---	---	---

**Шифр
Цезаря**

Каждая буква исходного текста заменяется третьей после нее буквой в алфавите.

Любая информация кодируется в
компьютере с помощью
последовательностей двух цифр - 0 и 1.

Последовательности из 0 и 1

называются ...,

а цифры 0 и 1 - ... или

Такое кодирование информации на
компьютере называется

Повторим основные понятия

- Множество символов, с помощью которых записывается текст, **называется...**
- Число символов в алфавите - **это его ...**
- Формула определения количества информации: $N = 2^b$,
где **N - это ...** **b - это ...**
- Единице измерения 8 бит присвоили **название ...**
- С помощью 8 бит можно **закодировать ...**
СИМВОЛОВ.



**Какие символы
можно закодировать с
помощью 8 бит ?**

Подсчитаем количество символов

- для русского алфавита 33 строчные буквы + 33 прописные буквы = 66;
- для английского алфавита 26 + 26 = 52;
- цифры от 0 до 9.

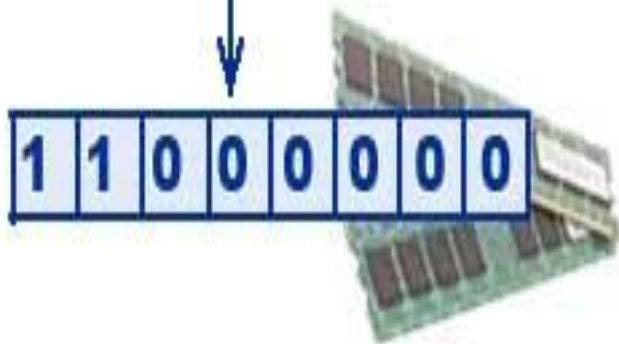
Итого - 128 символов.

- Остается 128 значений, которые можно использовать для обозначения ...

Человек различает символы по их начертанию, а компьютер - по их коду.



уникальный десятичный код от 0 до 255



СИМВОЛ



**соответствующий двоичный
КОД**

от 00000000 до 11111111

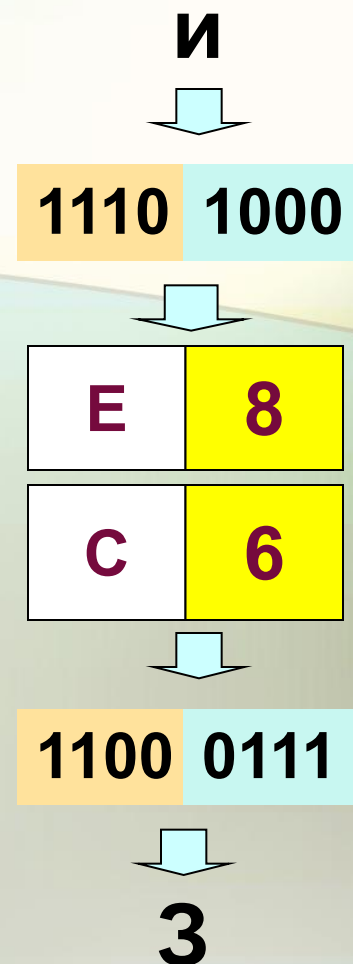
Таблица кодировки -

таблица, в которой устанавливается соответствие между символами и их порядковыми номерами в компьютерном алфавите.

	00	01	02	03
00		☐	⊖	♥
10	▶	◀	‡	!!
20		‡	"	#
30	0	1	2	3

СИМВОЛ - КОД

	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
1000	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143
	Ъ	Ґ	,	ѓ	„	…	†	‡	€	%	Љ	<	Њ	Ќ	Ѓ	Ѕ
1001	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
	ђ	‘	’	“	”	•	–	—	□	™	љ	>	њ	ќ	ћ	џ
1010	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
		Ў	ў	Ј	Ѡ	Ґ	І	§	Ё	©	Є	«	¬		®	ї
1011	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191
	°	±	І	і	ѓ	µ	¶	·	ё	№	є	»	ј	Ѕ	ѕ	ї
1100	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
1101	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
1110	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239
	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
1111	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255
	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я



ASC II

(American Standard Code for Information Interchange) – стандартный код информационного обмена США.

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	

Основы построения таблицы

1. Символы с номерами от нуля (двоичный код 00000000) до 127 (01111111), буквы латинского алфавита, цифры, знаки препинания, скобки и некоторые другие символы.
2. От 128 (двоичный код 10000000) до 255 (11111111), используются для кодировки букв национальных алфавитов, символов псевдографики и научных символов (например \leq , \geq , \approx).

Принцип последовательного кодирования алфавита

В кодовой таблице ASCII располагаются в алфавитном порядке - прописные, строчные с повтором через 32 знака.

Расположение цифр также упорядочено по возрастанию значений.

Благодаря этому и в машинном представлении для символьной информации сохраняется понятие «алфавитный порядок».

КОИ8

Первый стандарт кодирования русских букв на компьютерах -
«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	┌	○	2	•	÷
144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159
=		F	ё	П	Г	Г	П	П	Е	Ц	┌	┌	┌	┌	┌
160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
┌	┌	┌	Ё	┌	┌	┌	┌	┌	┌	┌	┌	┌	┌	┌	©
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

CP1251

Наиболее распространенная в настоящее время кодировка Microsoft Windows ("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	‘	’	“	”	•	–	—	è	™	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	!	§	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



Физкультминутка



Упражнение первое:

резко зажмурить глаза на 2-3 секунды: и широко открыть на 2-3 секунды, повторить упражнение 10 раз.

Упражнение второе:

часто-часто моргать глазами, повторить 10 раз.

Упражнение третье:

поднять глаза вверх, при этом голова остается в одном положении, задержать взгляд на 2-3 секунды, затем опустить глаза вниз и задержать взгляд на 2-3 секунды повторить упражнение 10 раз .

ЗАДАНИЯ



1. В таблице ниже представлена часть кодовой таблицы ASCII.

Каков шестнадцатеричный код символа “q”?

Символ	1	5	A	B	Q	a
Десятичный код	49	53	65	66	81	97
Шестнадцатеричный код	31	35	41	42	51	61

1) 71

2) 83

3) A1

4) B3

2. Буква «**i**» в таблице кодировки символов имеет десятичный код **105**. Что зашифровано последовательностью десятичных кодов: **108 105 110 107**?

3. С помощью последовательности десятичных кодов:

99 111 109 112 117 116 101 114

зашифровано слово «**computer**». Какая последовательность десятичных кодов будет соответствовать этому же слову, записанному заглавными буквами?

4. Представьте в форме десятичного кода слово «ЭВМ» в КОИ8 и CP1251 кодировках.



Ответ

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

КОИ8-Р: 252 247 237

CP1251: 221 194 204

Если перевести последовательности кодов из десятичной системы в шестнадцатеричную:

КОИ8-Р: FC F7 ED

CP1251: DD C2 CC

ВЫПОЛНИ САМОСТОЯТЕЛЬНО **КОНТРОЛЬНЫЕ ЗАДАНИЯ**

<http://fcior.edu.ru/card/10902/predstavlenie-teksta-v-razlichnyh-kodirovках.html>
Представление текста в различных кодировках, ЭОР