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

«ПАЗЛЫ» УРОКА

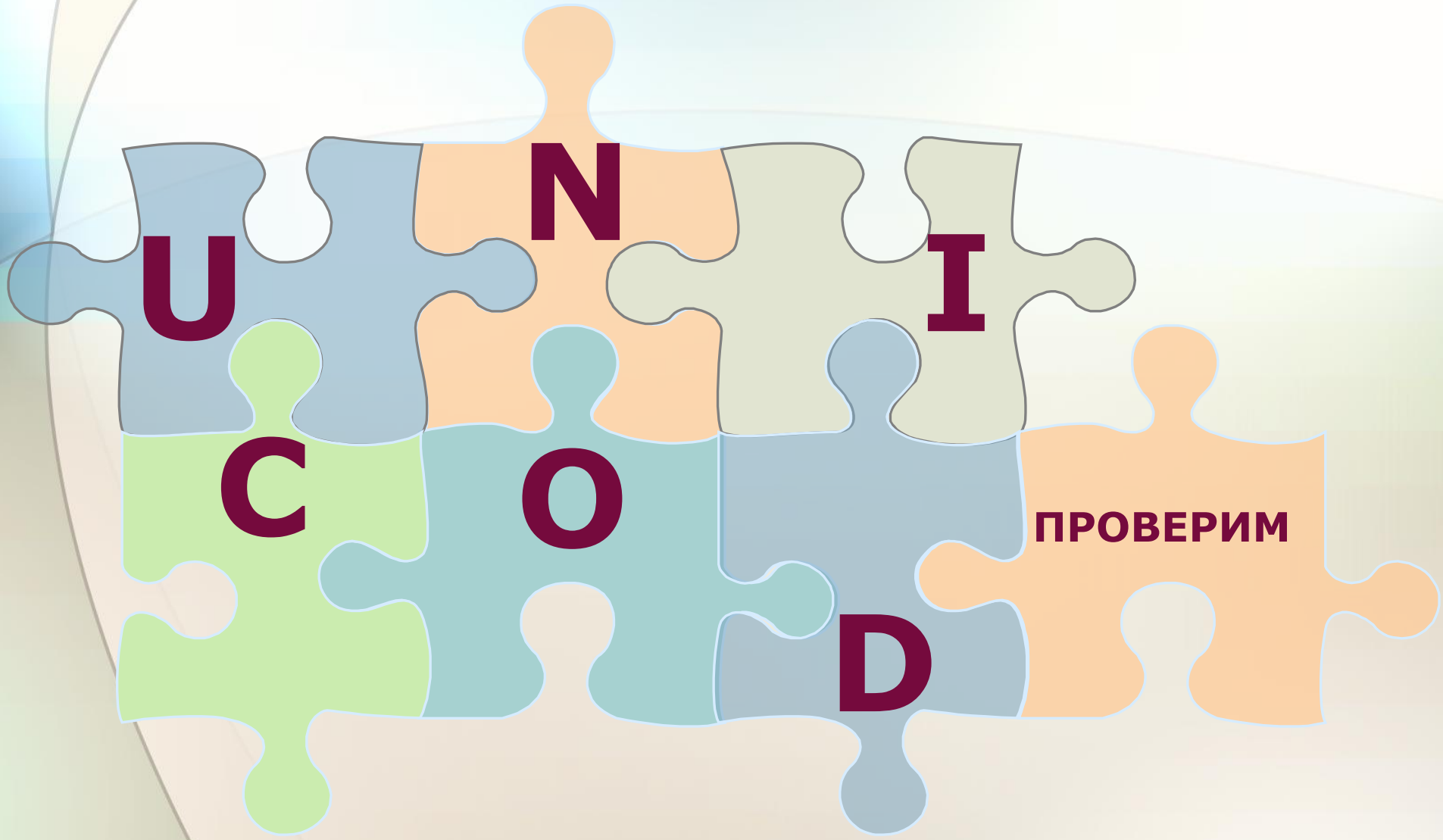
«ПАЗЛЫ» УРОКА

«ПАЗЛЫ» УРОКА

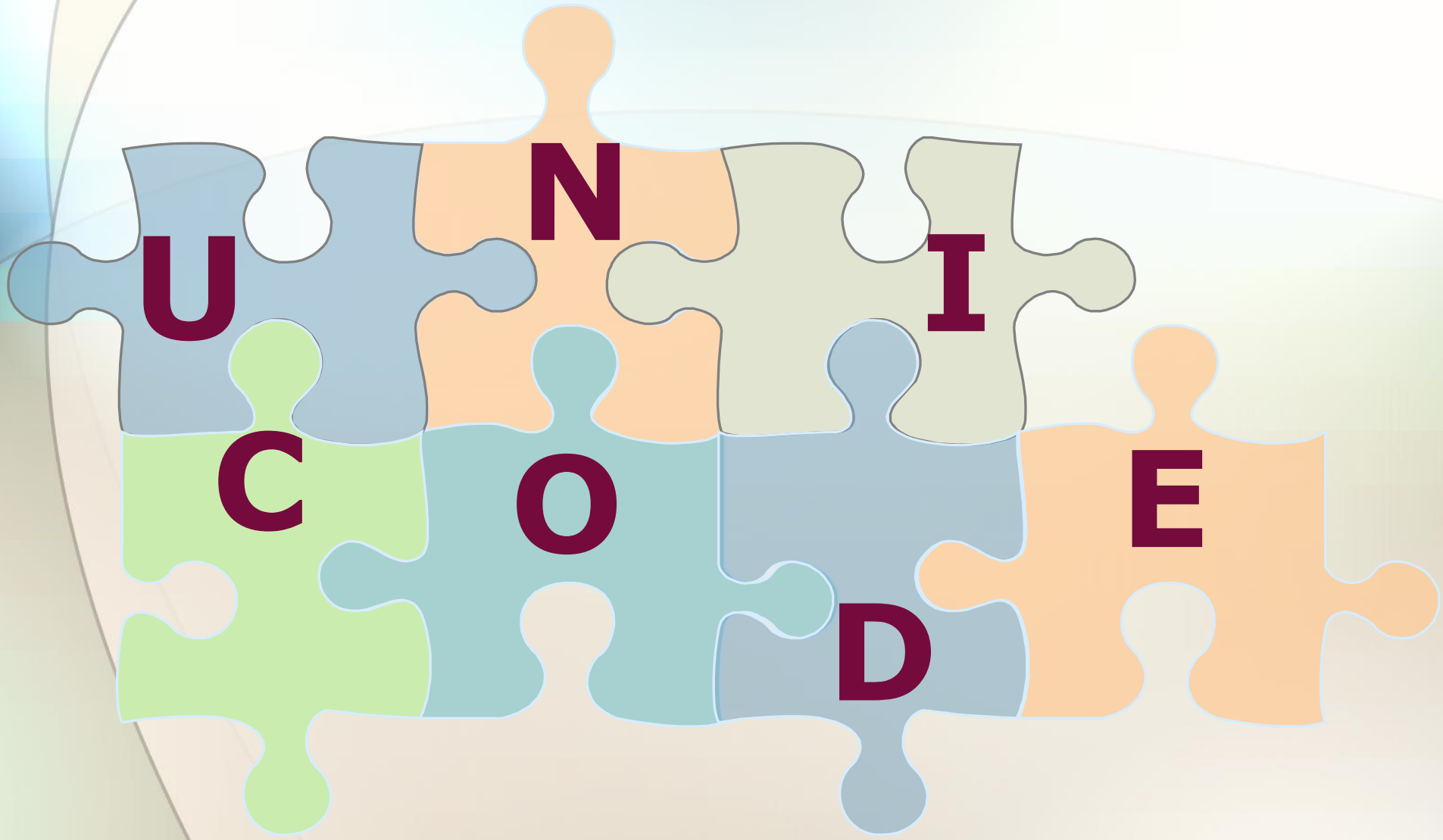
«ПАЗЛЫ» УРОКА

«ПАЗЛЫ» УРОКА

«ПАЗЛЫ» УРОКА



«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-kodirovkah.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, СР1251 и др.
- Unicodeмеждународный 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.

символ

↓

уникальный десятичный код от 0 до 255

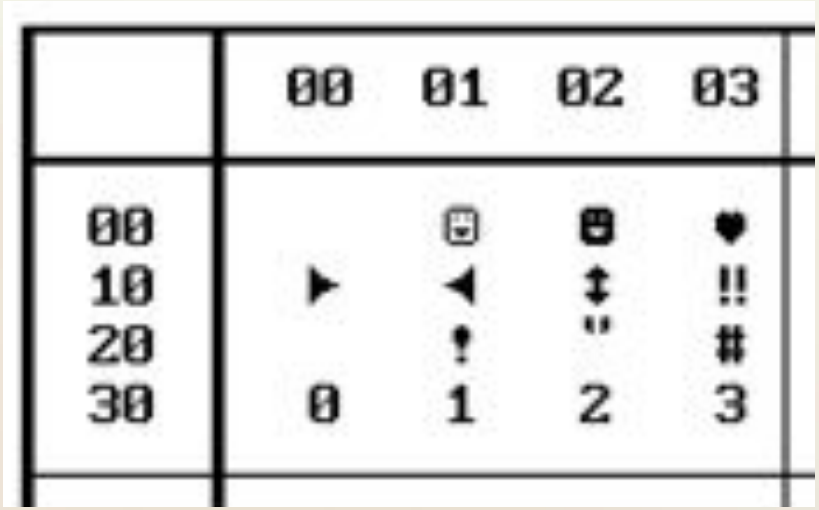
↓

соответствующий двоичный код

от 00000000 до 11111111

Таблица кодировки -

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



СИМВОЛ - КОД

	0000 0000	0000 0001	0000 0010	0000 0011	0001 0000	0001 0001	0001 0010	0001 0011	0001 0100	0001 0101	0001 0110	0001 0111	0001 1000	0001 1001	0001 1010	0001 1011
1000	128 Ђ	129 Ѓ	130 ‚	131 ѓ	132 „	133 …	134 †	135 ‡	136 €	137 ‰	138 Љ	139 ‹	140 Њ	141 Ќ	142 Ћ	143 Џ
1001	144 ђ	145 ‘	146 ’	147 “	148 ”	149 •	150 –	151 —	152	153 ™	154 љ	155 ›	156 њ	157 ќ	158 ћ	159 џ
1010	160	161 Ў	162 ў	163 Ј	164 ¤	165 Ґ	166 ¦	167 §	168 Ё	169 ©	170 Є	171 «	172 ¬	173	174 ®	175 Ї
1011	176 °	177 ±	178 І	179 і	180 ґ	181 µ	182 ¶	183 ·	184 ё	185 №	186 є	187 »	188 ј	189 Ѕ	190 ѕ	191 ї
1100	192 А	193 Б	194 В	195 Г	196 Д	197 Е	198 Ж	199 З	200 И	201 Й	202 К	203 Л	204 М	205 Н	206 О	207 П
1101	208 Р	209 С	210 Т	211 У	212 Ф	213 Х	214 Ц	215 Ч	216 Ш	217 Щ	218 Ъ	219 Ы	220 Ь	221 Э	222 Ю	223 Я
1110	224 а	225 б	226 в	227 г	228 д	229 е	230 ж	231 з	232 и	233 й	234 к	235 л	236 м	237 н	238 о	239 п
1111	240 р	241 с	242 т	243 у	244 ф	245 х	246 ц	247 ч	248 ш	249 щ	250 ъ	251 ы	252 ь	253 э	254 ю	255 я

и

1110 1000

Е 8

С 6

1100 0111

З

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 национальнх алфавитов, символов псевдографики и научных символов (например ≤, ≥, ≈).

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 стандартов кодирования русских букв на компьютерах -
«Код обмена информацией, 8-битный».

─ 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	nbsp 154	⌡ 155	° 156	² 157	· 158	÷ 159
═ 160	║ 161	╒ 162	ё 163	╓ 164	╔ 165	╕ 166	╖ 167	╗ 168	╘ 169	╙ 170	╚ 171	╛ 172	╜ 173	╝ 174	╞ 175
╟ 176	╠ 177	╡ 178	Ё 179	╢ 180	╣ 181	╤ 182	╥ 183	╦ 184	╧ 185	╨ 186	╩ 187	╪ 188	╫ 189	╬ 190	© 191
ю 192	а 193	б 194	ц 195	д 196	е 197	ф 198	г 199	х 200	и 201	й 202	к 203	л 204	м 205	н 206	о 207
п 208	я 209	р 210	с 211	т 212	у 213	ж 214	в 215	ь 216	ы 217	з 218	ш 219	э 220	щ 221	ч 222	ъ 223
Ю 224	А 225	Б 226	Ц 227	Д 228	Е 229	Ф 230	Г 231	Х 232	И 233	Й 234	К 235	Л 236	М 237	Н 238	О 239
П 240	Я 241	Р 242	С 243	Т 244	У 245	Ж 246	В 247	Ь 248	Ы 249	З 250	Ш 251	Э 252	Щ 253	Ч 254	Ъ 255

CP1251

Наиболее распространенная в настоящее время кодировка Microsoft Windows ("CP" означает "Code Page", "кодовая страница").

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

Физкульминутка

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
Десятич-ный код	49	53	65	66	81	97
Шест-надцате-ричный 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 СР1251 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-kodirovkah.html- Представление текста в различных кодировках, ЭОР