



# КОДИРОВАНИЕ ТЕКСТОВОЙ ИНФОРМАЦИИ

ПРЕДСТАВЛЕНИЕ ИНФОРМАЦИИ В КОМПЬЮТЕРЕ

**10 класс**

# Компьютерное представление текстовой информации

Для компьютерного представления текстовой информации достаточно:



...	...
64	01000000
65	01000001
66	01000010
67	01000011
68	01000100

Определить алфавит  
(множество всех  
символов)

Присвоить каждому  
символу алфавита  
порядковый номер

Перевести номер  
символа в двоичную  
систему счисления

# Кодировка ASCII

**American Standard Code for Information Interchange** – американский стандартный код для обмена информацией, разработанный в 1960-х годах в США.

		0	0	0	0	0	0	0	0	5						
0		NUL	SOH	STX	ETX	EOT	ENC									
1		0	0	1	0	0	0	0	0		NAI					
2																
3	0															
4	@															
5	P															
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	{		}	~	DEL

**Изображаемые символы**  
(буквы латинского алфавита, цифры, знаки препинания и арифметических операций, скобки и некоторые специальные символы)

Первые 32 символа и 128-й – управляющие  
(при выводе текста они не отображаются графически)

**A**

0 1 0 0 0 0 0 1

0 0 0 1 1 1 1 1

0 1 1 1 1 1 1 0

# Расширение кодировки ASCII

	0	0	0	0	0	0	0	0	5										
0	NUL	SOH	STX	ETX	EOT	ENC													
1	DLE	DC1	DC2	DC3	DC4	NAK													
2		!	"	#	\$	%													
3	0	1	2	3	4	5													
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							u	v	w	x	y	z							
		1	0	0	0	0	0	0	0										
8	Ђ	Ѓ	Ѕ	Ї	Љ	Њ	Ћ	Ќ	Ў	Ў	Ў	Ў	Ў	Ў	Ў	Ў	Ў	Ў	Ў
9	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ
A	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ
B	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ
C	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ
D	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ
E	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ
F	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ	Ђ

Стандартная часть кода (0 ... 127)

**КОИ-8**

**Расширение ASCII (128 ... 255)**  
 (буквы национального алфавита,  
 символы национальной валюты и т.п.)

1 1 1 1 1 1 1 1

# Расширение кодировки ASCII

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	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	Windows-1251					KOI-8				
8	Ъ -	ѓ	, ґ	ѓ ґ	„ ґ	... ґ	† ґ	‡ ґ	€ ґ	% ґ	Љ †	‹ ґ	Њ ґ	Ќ ґ	Ќ ґ	Љ ґ
9	ђ ґ	' ґ	' ґ	" ґ	" ґ	• ґ	- ґ	- ґ	≤ ґ	™ ґ	љ ґ	› ґ	њ ґ	ќ ґ	ћ ґ	џ ґ
A	= ґ	ђ ґ	ђ ґ	ђ ґ	ђ ґ	ђ ґ	ђ ґ	ђ ґ	ђ ґ	ђ ґ	ђ ґ	ђ ґ	ђ ґ	ђ ґ	ђ ґ	ђ ґ
B	° ґ	± ґ	І ґ	і ґ	ґ ґ	μ ґ	¶ ґ	· ґ	ё ґ	№ ґ	є ґ	» ґ	ј ґ	ѕ ґ	ѕ ґ	ї ґ
C	А ю	Б а	В б	Г ц	Д д	Е е	Ж ф	З г	И х	Й и	К й	Л к	М л	Н м	О н	П о
D	Р п	С я	Т р	У с	Ф т	Х у	Ц ж	Ч в	Ш ь	Щ ы	Ъ з	Ы ш	Ь э	Э щ	Ю ч	Я ь
E	а ю	б а	в б	г ц	д д	е е	ж ф	з г	и х	й и	к й	л к	м л	н м	о н	п о
F	р п	с я	т р	у с	ф т	х у	ц ж	ч в	ш ь	щ ы	ъ з	ы ш	ь э	э щ	ю ч	я ь

# Стандарт Unicode

---



**Unicode** — это «уникальный код для любого символа, независимо от платформы, независимо от программы, независимо от языка» ([www.unicode.org](http://www.unicode.org)).

---

Стандарт Unicode был разработан в 1991 году и описывает алфавиты всех известных, в том числе и «мертвых», языков. Для языков, имеющих несколько алфавитов или вариантов написания (японского и индийского), закодированы все варианты. В кодировку Unicode внесены все математические и иные научные символные обозначения и даже некоторые придуманные языки (язык эльфов из трилогии Дж. Р. Р. Толкина «Властелин колец»).



# Клавиатуры некоторых стран мира



РУССКАЯ



АМЕРИКАНСКАЯ



АРАБСКАЯ



АРМЯНСКАЯ



ЯПОНСКАЯ

# Кодировки стандарта Unicode

Для представления символов в памяти компьютера в стандарте Unicode имеется несколько кодировок.



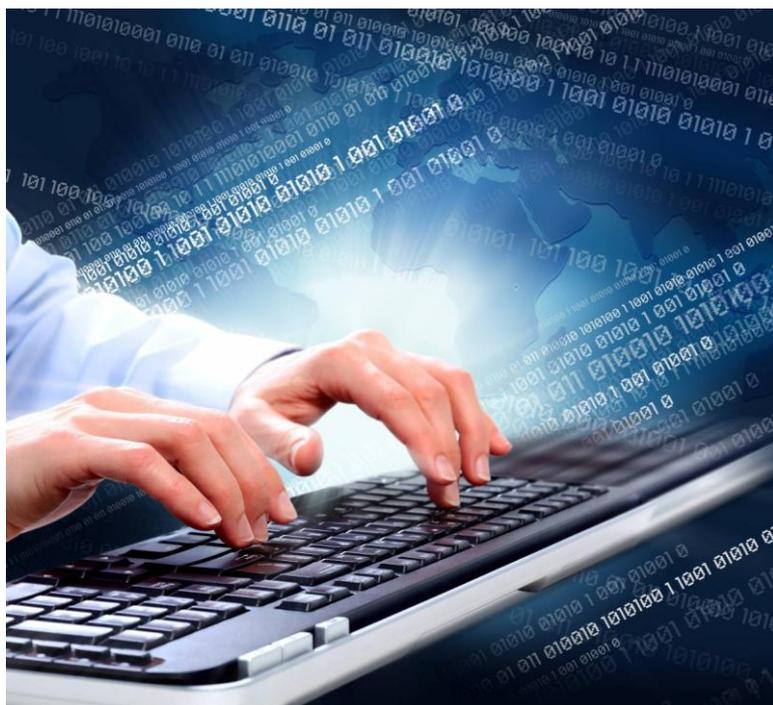
Кодировки Unicode позволяют включать в один документ символы самых разных языков, но их использование ведёт к увеличению размеров текстовых файлов.



# Информационный объем сообщения



Информационным объёмом текстового сообщения называется количество бит (байт, килобайт, мегабайт и т. д.), необходимых для записи этого сообщения путём заранее оговоренного способа двоичного кодирования.



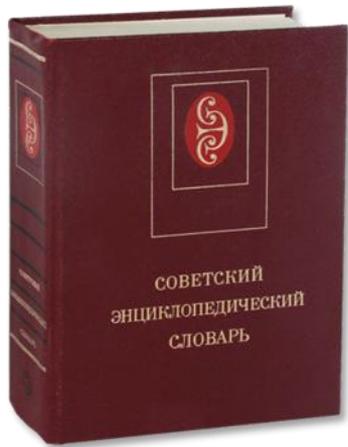
Количество символов  
в сообщении

$$I = K \cdot i$$

ASCII, KOI-8,  
Windows-1251, ...  
**1 символ = 1 байт**

Unicode  
**1 символ = 2 байта**

# Вопросы и задания



В Советском энциклопедическом словаре (1983 года издания) 1600 страниц. На одной странице размещается в среднем 100 строк по 140 символов (включая пробелы) в каждой. Найдите объем (в Мбайтах) текстовой информации в словаре, если при записи используется кодировка «*один символ — один байт*».

**Дано:**

$$i = 1 \text{ байт}$$

$$K = 1600 \cdot 100 \cdot 140$$

$I - ?$

$$I = K \cdot i$$

$$I = \frac{1600 \cdot 100 \cdot 140}{1024 \cdot 1024} \text{ Мб} \approx 21,36 \text{ Мб}$$

**Ответ:** 21,36 Мбайта

# Вопросы и задания



Задание 1. Представьте в кодировке ASCII текст  
*Happy New Year!*

а) шестнадцатеричным кодом

48 61 70 70 79 20 4E 65 77 20 59 65 61 72 21

б) десятичным кодом

72 97 112 112 121 32 78 101 119 32 89 101 97 114 33

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	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	{		}	~	DEL

ОТВЕТ

# Вопросы и задания



**Задание 3.** В 15-м издании энциклопедии Britannica 32 тома, в каждом из которых порядка 1000 страниц. На одной странице размещается в среднем 70 строк по 120 символов (включая пробелы) в каждой. Найдите объем текстовой информации в энциклопедии, если при записи используется кодировка Unicode («один символ — два байта»).

**Дано:**

$i = 2$  байта

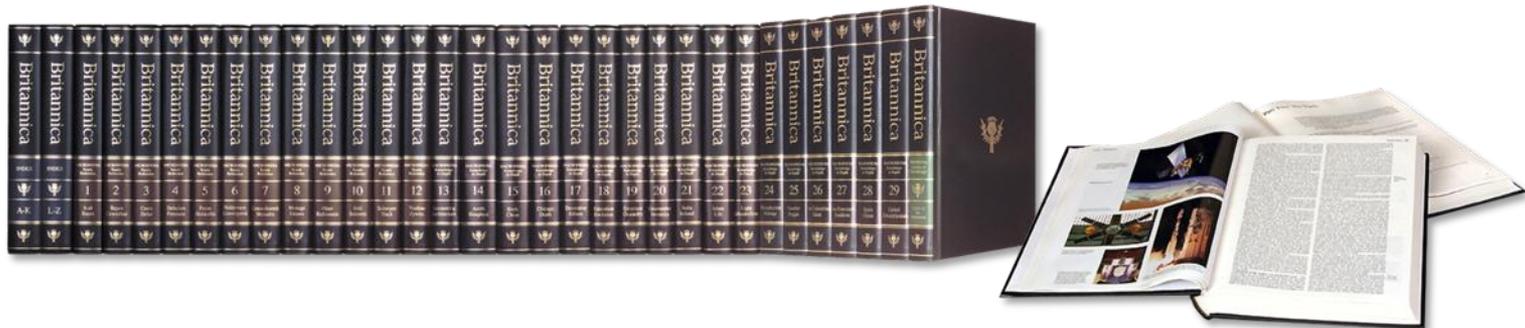
$K = 32 \cdot 1000 \cdot 70 \cdot 120$

$I - ?$

$I = K \cdot i$

$$I = \frac{32 \cdot 1000 \cdot 70 \cdot 120 \cdot 2}{1024 \cdot 1024} \text{ Мб} \approx 513 \text{ Мб}$$

**Ответ:** 513 Мбайт





В кодировке КОИ-8 каждый символ кодируется 8 битами.

Игорь скачал текст (в нём нет лишних пробелов):

«Названия оттенков розового: Коралловый, Сакура, Фламинго, Лососёвый, Мексика, Танго, Амарантовый, Бело-лиловый, Ультра-розовый»

Ученик вычеркнул из списка название одного цвета. Заодно он вычеркнул ставшие лишними запятые и пробелы – два пробела не должны идти подряд.

При этом размер нового предложения в данной кодировке оказался на 10 байт меньше, чем размер исходного предложения. Напишите в ответе вычеркнутое название цвета.

В кодировке КОИ-8 каждый символ кодируется 8 битами.

Петя написал текст (в нём нет лишних пробелов):

«Уж, эму, лиса, выдра, барсук, опоссум, дикобраз, орангутанг — дикие животные».

Ученик решил добавить в список название ещё одного животного — волк. При этом он добавил в текст необходимую запятую и пробел.

На сколько байт при этом увеличился размер нового предложения в данной кодировке? В ответе укажите только одно число — количество байт.

В одной из кодировок Unicode каждый символ кодируется 16 битами. Кирилл написал текст (в нем нет лишних пробелов):

**«Близнецы, дева, рак, телец, стрелец - знаки Зодиака».**

Ученик вычеркнул из списка название одного из знаков Зодиака. Заодно он вычеркнул ставшие лишними запятые и пробелы - два пробела не должны идти подряд. При этом размер нового предложения в данной кодировке оказался на 14 байт меньше, чем размер исходного предложения. Напишите в ответе количество букв вычеркнутого названия знака Зодиака.

1. В одной из кодировок Unicode каждый символ кодируется 16 битами. Петя написал текст (в нем нет лишних пробелов, символы переноса слов не учитываются):

**«Минск, Москва, Киев, Вильнюс, Рига, София, Будапешт, Бухарест, Варшава, Осло, Стокгольм, Париж, Берлин, Амстердам, Рим, Мадрид, Афины, Прага, Лондон - столицы стран Европы».**

Ученик вычеркнул из списка название одной столицы. Заодно он вычеркнул ставшие лишними запятые и пробелы - два пробела не должны идти подряд. При этом размер нового предложения в данной кодировке оказался на 18 байт меньше, чем размер исходного предложения. Среди столиц, имеющих одинаковое количество букв, Петя вычеркивает первую по порядку. Напишите в ответе вычеркнутое название столицы.



В одной из кодировок Unicode каждый символ кодируется 16 битами. Петя написал текст (в нём нет лишних пробелов):

**«Ель, кедр, сосна, кипарис, лиственница, можжевельник — хвойные растения».**

Ученик вычеркнул из списка название одного из растений. Заодно он вычеркнул ставшие лишними запятые и пробелы — два пробела не должны идти подряд.

При этом размер нового предложения в данной кодировке оказался на 26 байт меньше, чем размер исходного предложения. Напишите в ответе вычеркнутое название хвойного растения.



В одной из кодировок Unicode каждый символ кодируется 16 битами. Ваня написал текст (в нём нет лишних пробелов):

**«D, Io, Ada, Java, Swift, Python, ColdFusion — языки программирования».**

Ученик вычеркнул из списка название одного из языков программирования. Заодно он вычеркнул ставшие лишними запятые и пробелы — два пробела не должны идти подряд.

При этом размер нового предложения в данной кодировке оказался на 10 байт меньше, чем размер исходного предложения. Напишите в ответе вычеркнутое название языка программирования.



В одной из кодировок Unicode каждый символ кодируется 16 битами. Ваня написал текст (в нём нет лишних пробелов):

**«Уфа, Азов, Пермь, Белово, Вологда, Камбарка, Соликамск — города России».**

Ученик вычеркнул из списка название одного из городов. Заодно он вычеркнул ставшие лишними запятые и пробелы — два пробела не должны идти подряд.

При этом размер нового предложения в данной кодировке оказался на 22 байта меньше, чем размер исходного предложения. Напишите в ответе вычеркнутое название города России.



В кодировке КОИ-8 каждый символ кодируется 8 битами. Лена написала текст (в нём нет лишних пробелов):

**«Ява, Куба, Лусон, Маражо, Суматра, Сулавеси, Эспаньола — острова».**

Ученица вычеркнула из списка название одного из островов. Заодно она вычеркнула ставшие лишними запятые и пробелы — два пробела не должны идти подряд.

При этом размер нового предложения в данной кодировке оказался на 9 байтов меньше, чем размер исходного предложения. Напишите в ответе вычеркнутое название острова.