



Двоичное

кодирование

текстовой

информации

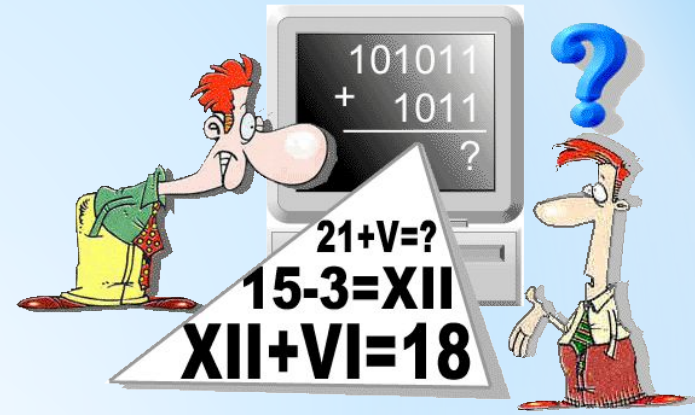


В 40-е годы прошлого столетия было положено начало созданию вычислительной машины.

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



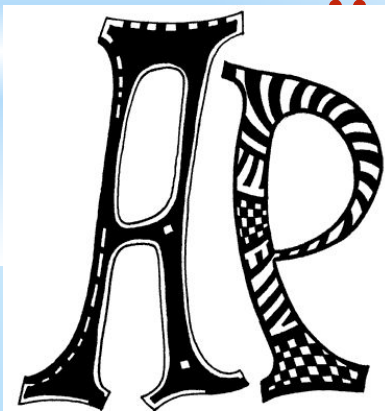
*** Двоичное кодирование текстовой**



+ ||
12

++ ||||
24

+++ |||
33

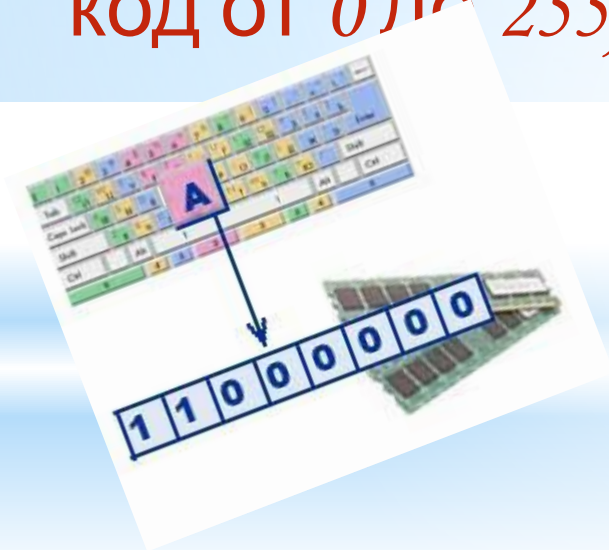


Таинственные знаки

* В памяти компьютера любой текст представляется последовательностью кодов символов, т. е. вместо самой буквы хранится ее номер в кодовой таблице. Изображение же букв и символов сформируется только в момент их вывода на экран или бумагу.

* **Двоичное
кодирование
текстовой**

Кодирование заключается в том, что каждому символу ставится в соответствие уникальный двоичный код от 00000000 до 11111111 (или десятичный код от 0 до 255).



*** Двоичное
кодирование
текстовой**

Важно, что присвоение символу конкретного кода – это вопрос соглашения, которое фиксируется кодовой таблицей.

Для разных типов ЭВМ
используются различные
кодировки

* Кодовые таблицы

* По началу применялось 7-битная кодировка, которая могла представить 128 символов. С распространением *IBM PC* международным стандартом стала таблица кодировки *ASCII (American Standart Code for Information Interchange)* – Американский стандартный код для информационного обмена.

* Кодовые таблицы

Позже она была расширена до 8 бит (256 символов) и дошла в таком виде практически до сегодняшнего дня. При этом первая половина (символы 0-127) были всегда одни и те же, соответствующие стандарту *ASCII*, а вторая половина таблицы (символы 128-255) менялась в зависимости от страны, где она использовалась.

* Таблица кодировки

ASCII

символ	10- й код	2-й код	символ	10- й код	2-й код	символ	10-й код	2-й код	символ	10-й код	2-й код
	32	00100000	8	56	00111000	P	80	01010000	h	104	01101000
!	33	00100001	9	57	00111001	Q	81	01010001	i	105	01101001
"	34	00100010	:	58	00111010	R	82	01010010	j	106	01101010
#	35	00100011	;	59	00111011	S	83	01010011	k	107	01101011
\$	36	00100100	<	60	00111100	T	84	01010100	l	108	01101100
%	37	00100101	=	61	00111101	U	85	01010101	m	109	01101101
&	38	00100110	>	62	00111110	V	86	01010110	n	110	01101110
'	39	00100111	?	63	00111111	W	87	01010111	o	111	01101111
(40	00101000	@	64	01000000	X	88	01011000	p	112	01110000
)	41	00101001	A	65	01000001	Y	89	01011001	q	113	01110001
*	42	00101010	B	66	01000010	Z	90	01011010	r	114	01110010
+	43	00101011	C	67	01000011	[91	01011011	s	115	01110011
,	44	00101100	D	68	01000100	\	92	01011100	t	116	01110100
-	45	00101101	E	69	01000101]	93	01011101	u	117	01110101
.	46	00101110	F	70	01000110	^	94	01011110	v	118	01110110
/	47	00101111	G	71	01000111	_	95	01011111	w	119	01110111
0	48	00110000	H	72	01001000	`	96	01100000	x	120	01111000
1	49	00110001	I	73	01001001	a	97	01100001	y	121	01111001
2	50	00110010	J	74	01001010	b	98	01100010	z	122	01111010
3	51	00110011	K	75	01001011	c	99	01100011	{	123	01111011
4	52	00110100	L	76	01001100	d	100	01100100		124	01111100
5	53	00110101	M	77	01001101	e	101	01100101	}	125	01111101
6	54	00110110	N	78	01001110	f	102	01100110	~	126	01111110
7	55	00110111	O	79	01001111	g	103	01100111	□	127	01111111

символ	10-Б код	2-Б код	символ	10-Б код	2-Б код	символ	10-Б код	2-Б код	символ	10-Б код	2-Б код
Ъ	128	10000000		160	10100000	А	192	11000000	а	224	11100000
Г	129	10000001	Ў	161	10100001	Б	193	11000001	б	225	11100001
,	130	10000010	ў	162	10100010	В	194	11000010	в	226	11100010
г	131	10000011	Ј	163	10100011	Г	195	11000011	г	227	11100011
„	132	10000100	о	164	10100100	Д	196	11000100	д	228	11100100
...	133	10000101	Г	165	10100101	Е	197	11000101	е	229	11100101
†	134	10000110	‡	166	10100110	Ж	198	11000110	ж	230	11100110
‡	135	10000111	§	167	10100111	З	199	11000111	з	231	11100111
€	136	10001000	Е	168	10101000	И	200	11001000	и	232	11101000
‰	137	10001001	©	169	10101001	Й	201	11001001	й	233	11101001
Љ	138	10001010	€	170	10101010	К	202	11001010	к	234	11101010
<	139	10001011	«	171	10101011	Л	203	11001011	л	235	11101011
Њ	140	10001100	¬	172	10101100	М	204	11001100	м	236	11101100
Ќ	141	10001101	-	173	10101101	Н	205	11001101	н	237	11101101
Ћ	142	10001110	®	174	10101110	О	206	11001110	о	238	11101110
џ	143	10001111	Ї	175	10101111	П	207	11001111	п	239	11101111
ђ	144	10010000	°	176	10110000	Р	208	11010000	р	240	11110000
‘	145	10010001	±	177	10110001	С	209	11010001	с	241	11110001
’	146	10010010	І	178	10110010	Т	210	11010010	т	242	11110010
“	147	10010011	і	179	10110011	У	211	11010011	у	243	11110011
”	148	10010100	г	180	10110100	Ф	212	11010100	ф	244	11110100
•	149	10010101	μ	181	10110101	Х	213	11010101	х	245	11110101
—	150	10010110	¶	182	10110110	Ц	214	11010110	ц	246	11110110
—	151	10010111	·	183	10110111	Ч	215	11010111	ч	247	11110111
□	152	10011000	ë	184	10111000	Ш	216	11011000	ш	248	11111000
™	153	10011001	№	185	10111001	Щ	217	11011001	щ	249	11111001
љ	154	10011010	€	186	10111010	Ъ	218	11011010	ъ	250	11111010
›	155	10011011	»	187	10111011	Ы	219	11011011	ы	251	11111011
њ	156	10011100	ј	188	10111100	Ь	220	11011100	ь	252	11111100
ќ	157	10011101	š	189	10111101	Э	221	11011101	э	253	11111101
ћ	158	10011110	s	190	10111110	Ю	222	11011110	ю	254	11111110
џ	159	10011111	ï	191	10111111	Я	223	11011111	я	255	11111111

* Таблица
расширенного
кода
ASCII
Кодировка
Windows-125
1 (CP1251)

* В Советском Союзе различные организации и сети, имевшие большое влияние на компьютерный и программный рынок тех времен, создавали свои кодировки (т.е. вторые половины таблицы), содержащие русские символы.

* Кодовые таблицы

В настоящее время существует 5 разных кодовых таблиц для русских букв (KOI8, CP 1251, CP 866, Mac, ISO).

Широкое распространение получил новый международный стандарт *Unicode*, который отводит на каждый символ два байта. С его помощью можно закодировать 65536 ($2^{16} = 65536$) различных СИМВОЛОВ.



Кодовые таблицы для русских букв

* Проблемы с кодировками делятся на несколько типов. Первый тип - это отсутствие информации о кодировке.

*** Проблемы с кодировками**

* Проблемы второго типа - это когда кодировка в файле указана, но конечная программа такой кодировки не знает.

*** Проблемы с кодировками**

* Третий тип проблем, наоборот, связан с избытком информации о кодировках. Это актуальная в настоящее время проблема (например, для веб-страниц).

*** Проблемы с кодировками**

Цифры кодируются по стандарту *ASCII* в двух случаях – при вводе-выводе и когда они встречаются в тексте. Если цифры участвуют в вычислениях, то осуществляется их преобразование в другой двоичных код.



***Обратите внимание!**

Возьмем число *57*.

При использовании в тексте каждая цифра будет представлена своим кодом в соответствии с таблицей *ASCII*. В двоичной системе это – *0011010100110111*.

При использовании в вычислениях, код этого числа будет получен по правилам перевода в двоичную систему и получим – *00111001*.

*** Обратите внимание!**

Сегодня очень многие люди для подготовки писем, документов, статей, книг и пр. используют **компьютерные текстовые редакторы**. Компьютерные редакторы, в основном, работают с **алфавитом размером 256 символов**.

В этом случае легко подсчитать объем информации в тексте. Если **1 символ алфавита несет 1 байт информации**, то надо просто сосчитать количество символов; полученное число даст **информационный объем текста в байтах**.

**Информационный
объем текста**

$$I = K \times i, \text{ где}$$

I -информационный объем
сообщения

K - количество символов в тексте

i - информационный вес одного
символа

$$2^i = N$$

* N мощность алфавита

**Формулы для расчета
информационного
объема текста**

* Мощность алфавита равна 256. Сколько Кбайт памяти потребуется для сохранения 160 страниц текста, содержащего в среднем 192 символа на каждой странице?

* **Задание 1**

* Мощность алфавита равна 64.
Сколько Кбайт памяти
потребуется, чтобы сохранить
128 страниц текста,
содержащего в среднем 256
символов на каждой
странице?

* **Задание 1'**

* Объем сообщения – 7,5 Кбайт.
Известно, что данное
сообщение содержит 7680
символов. Какова мощность
алфавита?

* **Задание 2**

* Объем сообщения равен 11 Кбайт. Сообщение содержит 11264 символа. Какова мощность алфавита?

* **Задание 2'**

Племя Мумбу-Юмбу использует алфавит из букв: α β γ δ ϵ ζ η θ λ μ ξ σ φ ψ , точки и для разделения слов используется пробел.

Сколько информации несет свод законов племени, если в нем *12* строк и в каждой строке по *20* символов?

***Задание 3**

* Для кодирования секретного сообщения используются 12 специальных значков-символов. При этом символы кодируются одним и тем же минимально возможным количеством бит. Чему равен информационный объем сообщения длиной в 256 символов?

Задание 3'

- * В чем заключается кодирование текстовой информации в компьютере?
- * Закодируйте с помощью *ASCII*-кода свою фамилию, имя, номер класса.
- * Считая, что каждый символ кодируется одним байтом, оцените информационный объем следующего предложения из пушкинского четверостишия:

Певец-Давид был ростом мал, Но повалил ~~же~~ Голиафа!

Вопросы и задания:

- Учебник Угринович Н.Д. Информатика и ИКТ базовый курс 9 класс;
- [Яндекс-картинка](#)
- [Изображение 2](#)
- [Изображение 3](#)
- [Изображение 4](#)
- [Изображение 5](#)
- [Изображение 6](#)
- <http://inn.h1.ru/topic.shtml?h1=16&h2=7>
- <http://www.galaktionoff.ru/unpub/TTF.htm>
- http://www.infospir.ru/articles/chto_takoe_kodirovka_2.php
- http://gym1.pupils.ru/img_school/gym1/Ekzamen10/variant1.pdf



Используемая

литература