

# **Представление текстовой информации в компьютере**



# Кодирование и декодирование

Язык может быть **естественным** (разговорным) или **формальным** (кодом, мат. Формулы, язык программирования)

**Язык (код)** — определенная система знаков для представления информации

**Кодирование** — процесс представления информации (**сообщения**) в виде кода. Все множество символов, используемых для кодирования, называется *алфавитом кодирования*. Например, в памяти компьютера любая информация кодируется с помощью двоичного алфавита, содержащего всего два символа: 0 и 1.

**Декодирование**- процесс обратного преобразования кода к форме исходной символической системы, т.е. получение исходного сообщения. Например: перевод с азбуки Морзе в письменный текст на русском языке.

В более широком смысле декодирование — это процесс восстановления содержания закодированного сообщения..

При кодировании 1 символа используют количество информации 1 байт,

Т.е.  $I = 1 \text{ байт} = 8 \text{ бит}$

В компьютерном алфавите кол-во символов = 256 . Этого хватает, чтобы выразить весь алфавит.  $N=256$

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

# *США, институт стандартизации (ASCII- Американский стандартный код для обмена информацией)*

*Таблица ASCII разделена на 2 части*

- I. -стандартная(содержит коды от 0 до 127)*
- II. - расширенная (содержит символы с кодами от 128 до 255)*

# Кодовая таблица ASCII

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	

В базовой таблице за кодами с 0 по 32 закреплены специальные функции: нажатие клавиши Enter, Bs, управление курсором, ввод пробела и т.д.

Коды с 33 по 127 соответствуют символам: !, #, \$, цифрам, знакам арифметических операций, знакам препинания и символам латинского алфавита.

Коды с 128 по 255 являются национальными, т.е. в национальных кодировках одному и тому же коду соответствуют различные символы.

В настоящее время существует 5  
различных 8 –ми битовых кодовых  
таблиц для русских букв, поэтому  
тексты созданные в одной  
кодировке, не будут правильно  
отображаться в другой.

## Таблица кодировки символов

Двоичный код	Десятичный код	КОИ8	CP1251	CP866	Mac	ISO
0000 0000	0					
.....						
0000 1000	8	Удаление последнего символа (клавиша Backspace)				
.....						
0000 1101	13	Перевод строки (клавиша Enter)				
.....						
0010 0000	32	Пробел				
0010 0001	33	!				
.....						
0101 1010	90	Z				
.....						
0111 1111	127					
.....	128	-	Ъ	А	А	К
.....						
1100 0010	194	Б	В	-	-	Т
.....						
1100 1100	204	Л	М	:	:	Ь
.....						
1101 1101	221	Щ	Э	-	Ё	Н
.....						
1111 1111	225	Ъ	я	Нераз. пробел	Нераз. пробел	п

## **Стандарты кодировок:**

- 1. КОИ-8 - UNIX**
- 2. CP1251 («CP» означает «Code Page») - Microsoft Windows**
- 3. CP 866 - MS-DOS**
- 4. Mac - Macintosh**
- 5. ISO 8859 – 5**

**1 символ кодируется 1 байтом**

одним из первых стандартов кодирования русских букв на компьютерах был код КОИ – 8 («Код обмена информационный – 8 битный»). Эта кодировка применяется в компьютерах с операционной системой UNIX.

Наиболее распространенная кодировка – это стандартная кириллическая кодировка **Microsoft Windows**, обозначаемая сокращением **CP1251** («CP» означает «Code Page»).

Все Windows – приложения, работающие с русским языком, поддерживают эту кодировку.

Для работы в среде операционной системы **MS-DOS** используется «альтернативная» кодировка, в терминологии фирмы **Microsoft** – кодировка **CP 866**.

**Фирма Apple разработала для компьютеров **Macintosh** свою собственную кодировку русских букв **(Mac)****

**Международная организация по стандартизации (International Standards Organization, ISO) утвердила в качестве стандарта для русского языка еще одну кодировку под названием **ISO 8859 – 5.****

В последнее время широко используют новый международный стандарт **Unicode**, который отводит на каждый символ не один байт, а два, и поэтому с его помощью можно закодировать не 256 символов,  $2^{16}=65\ 536$  различных символов.

**1 символ = 2 байтами**

**Unicode - 16-битном коде**

Эту кодировку поддерживает платформа Microsoft Windows&Office97.

# задачи

- Определите количество информации объема выражения «Жесткий диск»., записанного в кодировке **Unicode**.
- Каждая страница текста состоит из 32 строк, в каждой строке по 64 символа. Определить максимальное количество страниц такого текста, записанного в кодировке КОИ-8 на USB Flash 512 Мб.

# задачи

- Определите количество информации объема выражения «Жесткий диск»., записанного в кодировке Unicode.

Решение:

1 символ кодируется 2 байтами,

$N=12$  символов,  
 $I=12*2=24$  байта

- Каждая страница текста состоит из 32 строк, в каждой строке по 64 символа. Определить максимальное количество страниц такого текста, записанного в кодировке КОИ-8 на USB Flash 512 Мб.

Решение:  $I=KX$

1 символ кодируется 1 байтам,

$K=32*64=2048$

Ёмкость USB

$512\text{Мб}=512*1024*1024=536870912$  байт,

Мах кол-во страниц в кодировке кои-8 на ёмкость 512Мб=  
 $536870912: 2048=262144$

Определите информационный  
объем текста в 8-ми битной  
кодировке.

**Бамбарбия! Керкуду!**

$$I = N * K$$

$$I = 19 * 8 = 152 \text{ бит}$$

# Задание : определите символ по числовому коду.

- Запустите программу БЛОКНОТ
- Нажмите ALT и 0224 (на дополнительной цифровой клавиатуре).
- Появится символ **а**.
- Повторите эту операцию для числовых кодов от 0225 до 0233.
- Появятся символы в кодировке (CP 1251 Windows).
- Запишите их в тетрадь.



## Задание 2: определите символ по числовому коду.

- Запустите программу БЛОКНОТ
- Нажмите ALT и 161 (на дополнительной цифровой клавиатуре).
- Появится символ.
- Повторите эту операцию для числовых кодов 160, 169, 226.
- Появятся символы в кодировке (CP 866 MS-DOS).
- Запишите их в тетрадь.



Оказывается: **1 байт = 8 битов.**

1 Кбайт (один килобайт) = 1024 байт;

1 Мбайт (один мегабайт) = 1024 Кбайт;

1 Гбайт (один гигабайт) = 1024 Мбайт.

# Задание : определите символ по числовому коду.

- Запустите программу БЛОКНОТ
- Нажмите ALT и 0224 (на дополнительной цифровой клавиатуре).
- Появится символ **а**.



## Задание 2: определите символ по числовому коду.

- Запустите программу БЛОКНОТ
- Нажмите ALT и 161 (на дополнительной цифровой клавиатуре).
- Появится символ.

