

# Двоичное кодирование текстовой информации



Информация, выраженная с помощью  
естественных и формальных языков в письменной  
форме, называется **текстовой информацией**

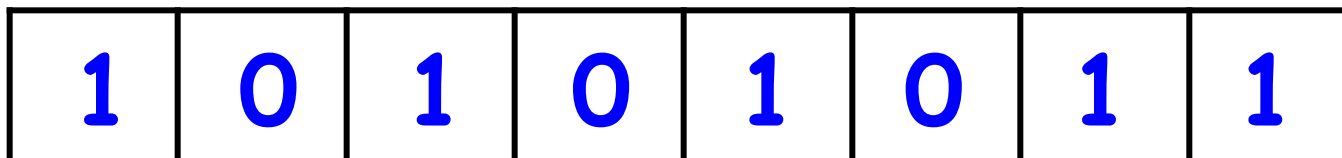
Для представления текстовой информации (прописные и строчные буквы русского и латинского алфавитов, цифры, знаки и математические символы) достаточно 256 различных знаков

а-я	А-Я	а-z	A-Z	0-9	!@#\$%^&·()_ -+»№; V:?=	всего
33	33	26	26	10	25	153

$$128 (2^7) < \mathbf{153} < 256 (2^8)$$

Для обработки текстовой информации на компьютере необходимо представить ее в двоичной знаковой системе

Длина двоичного кода знака составляет восемь двоичных знаков



**I = 8 битов**

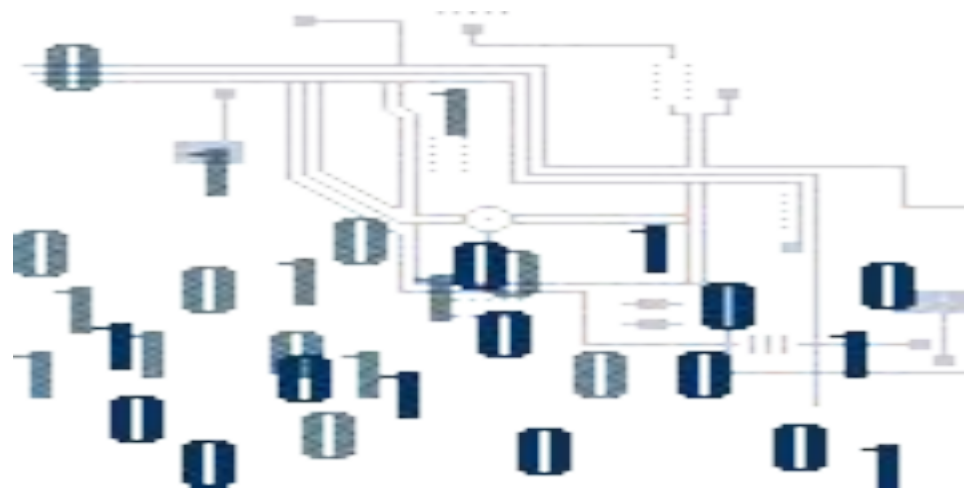
*Каждому знаку необходимо поставить в соответствие уникальный двоичный код в интервале от 00000000 до 11111111*

192	А	10000000	208	Р	10010000	224	а	11100000	240	р	11110000
193	Б	10000001	209	С	10010001	225	б	11100001	241	с	11110001
194	В	10000010	210	Т	10010010	226	в	11100010	242	т	11110010
195	Г	10000011	211	У	10010011	227	г	11100011	243	у	11110011
196	Д	10000100	212	Ф	10010100	228	д	11100100	244	ф	11110100
197	Е	10000101	213	Х	10010101	229	е	11100101	245	х	11110101
198	Ж	10000110	214	Ц	10010110	230	ж	11100110	246	ц	11110110
199	З	10000111	215	Ч	10010111	231	з	11100111	247	ч	11110111
200	И	10001000	216	Ш	10011000	232	и	11101000	248	ш	11111000
201	Й	10001001	217	Щ	10011001	233	й	11101001	249	щ	11111001
202	К	10001010	218	Ъ	10011010	234	к	11101010	250	ъ	11111010
203	Л	10001011	219	Ы	10011011	235	л	11101011	251	ы	11111011
204	М	10001100	220	Ь	10011100	236	м	11101100	252	ь	11111100
205	Н	10001101	221	Э	10011101	237	н	11101101	253	э	11111101
206	О	10001110	222	Ю	10011110	238	о	11101110	254	ю	11111110
207	П	10001111	223	Я	10011111	239	п	11101111	255	я	11111111

Для каждой цифры, буквы, символа, а так же пробела существует своя комбинация из 8 бит

«М» русская большая - 11101101

«M» латинская большая - 01001101





## Задание:

Посчитайте количество бит и байт в следующих выражениях:

Мир

3 байта = 24 бит

Миру мир!

9 байт = 72 бит

Vile, vide, vice

16 байт = 128 бит

# Система кодирования

**КОИ-7** (код обмена информацией, семизначный),  
действовавшая в СССР

# Система кодирования

Кодировка **Windows-1251** получила широкое применение на компьютерах компании **Microsoft**

## Система кодирования

Кодировка **ГОСТ** — альтернативная. Действует на компьютерах в операционных системах **MS-DOS**

# Система кодирования

Кодировка **ISO** (International Standard Organization — Международный институт стандартизации) — содержит символы русского алфавита, но на практике используется редко

## Система кодирования

Институт стандартизации в США ввел в действие таблицу кодов **ASCII** (American Standard Code for Information Interchange — стандартный код информационного обмена США).

# Система кодирования

## UNICODE

Диапазон значений кодов увеличится в два раза  
с 256 до 65536 символов

# Система кодирования

## UNICODE

Двухбайтовое описание кодов символов позволяет закодировать очень большое число символов из различных письменностей



# Система кодирования

## UNICODE

В документах Unicode могут соседствовать русские, латинские, греческие буквы, китайские иероглифы и математические символы

## Решение задач

Стр. 39, п.2.1, задача 2.1

В текстовом режиме экран монитора компьютера обычно разбивается на 25 строк по 80 символов в строке.

Определите объём текстовой информации, занимающий весь экран монитора, в кодировке Unicode.

$$25 \text{ строк} * 80 \text{ символов} = 2\,000 \text{ символов}$$

$$2\,000 \text{ символов} * 16 \text{ бит} = 32\,000 \text{ бит}$$

$$32\,000 \text{ бит} / 8 \text{ бит} / 1024 \text{ байт} = \mathbf{3,9 \text{ К байт}}$$

## Решение задач

Стр. 39, п.2.1, задача 2.2

Пользователь компьютера, хорошо владеющий навыками ввода информации с клавиатуры, может вводить в минуту 100 знаков.

Какое количество информации может ввести пользователь в компьютер за одну минуту в кодировке Windows? В кодировке Unicode?

$8 \text{ бит} * 100 \text{ знаков (символов)} = 800 \text{ бит} \Rightarrow \mathbf{100 \text{ байтов}}$

$16 \text{ бит} * 100 \text{ знаков (символов)} = 1\,600 \text{ бит} \Rightarrow \mathbf{200 \text{ байтов}}$