



# ОЦЕНКА КОЛИЧЕСТВЕННЫХ ПАРАМЕТРОВ ТЕКСТОВЫХ ДОКУМЕНТОВ

## ОБРАБОТКА ТЕКСТОВОЙ ИНФОРМАЦИИ

**7 класс**



ИЗДАТЕЛЬСТВО

**БИНОМ**

# Ключевые слова

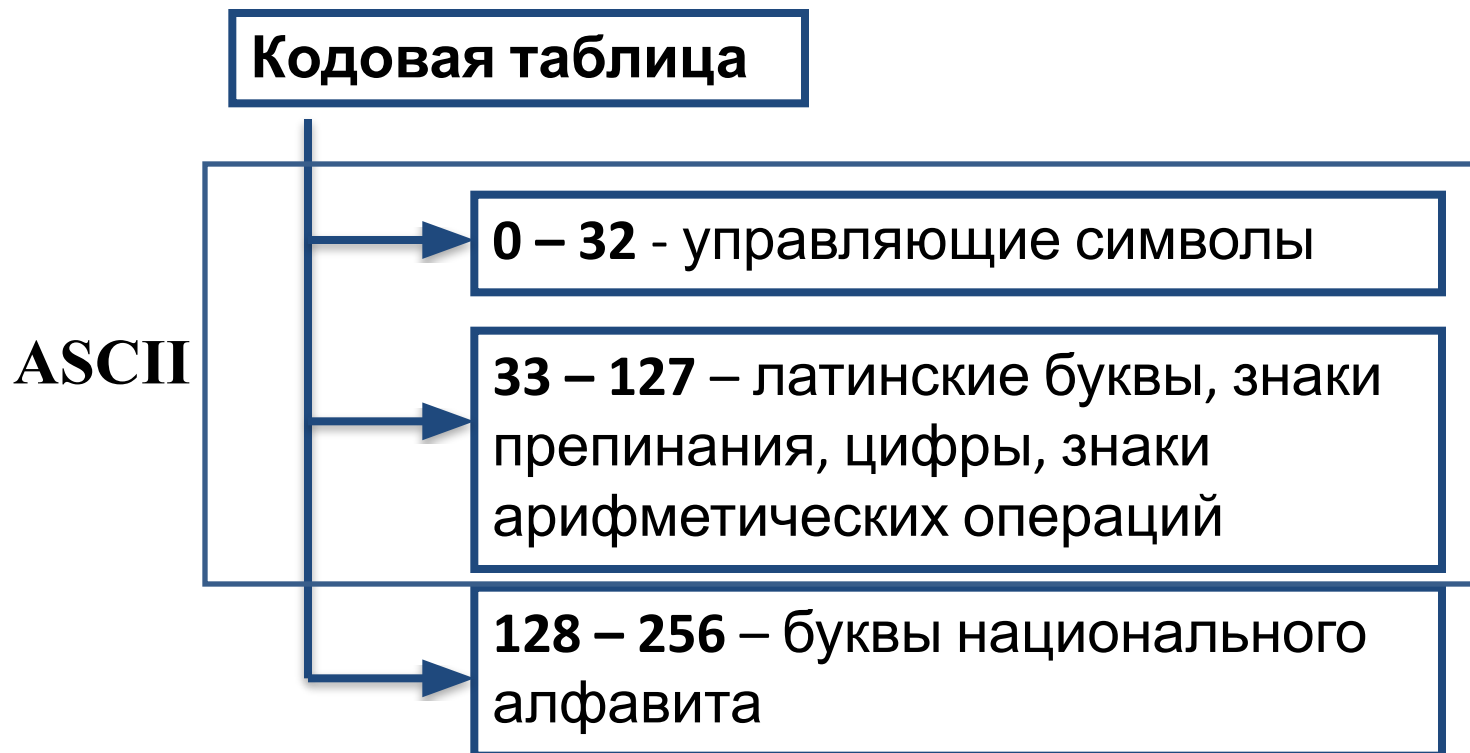
- кодовая таблица
- восьмиразрядный двоичный код
- информационный объём текста



# Представление текстовой информации в памяти компьютера

Текст состоит из символов - букв, цифр, знаков препинания и т. д., которые компьютер различает по их **двоичному коду**.

Соответствие между изображениями символов и кодами символов устанавливается с помощью **кодовых таблиц**.



# Представление текстовой информации в памяти компьютера

## Коды русских букв в таблицах КОИ-8

Символ	Кодировка			
	Windows		КОИ-8	
	десятичный код	двоичный код	десятичный код	двоичный код
<b>А</b>	192	11000000	225	11100001
<b>Б</b>	193	11000001	226	11100010
<b>В</b>	194	11000010	247	11110111
<b>Г</b>	195	11000011	248	11110110
<b>Д</b>	196	11000100	249	11110101
<b>Е</b>	197	11000101	250	11110100
<b>Ё</b>	198	11000110	251	11110011
<b>Ж</b>	199	11000111	252	11110010
<b>З</b>	200	11001000	253	11110001
<b>И</b>	201	11001001	254	11110000
<b>Й</b>	202	11001010	255	11110000
<b>К</b>	203	11001011	256	11110000
<b>Л</b>	204	11001100	257	11110000
<b>М</b>	205	11001101	258	11110000
<b>Н</b>	206	11001110	259	11110000
<b>О</b>	207	11001111	260	11110000
<b>П</b>	208	11010000	261	11110000
<b>Р</b>	209	11010001	262	11110000
<b>С</b>	210	11010010	263	11110000
<b>Т</b>	211	11010011	264	11110000
<b>У</b>	212	11010100	265	11110000
<b>Ф</b>	213	11010101	266	11110000
<b>Х</b>	214	11010110	267	11110000
<b>Ц</b>	215	11010111	268	11110000
<b>Ч</b>	216	11011000	269	11110000
<b>Ш</b>	217	11011001	270	11110000
<b>Щ</b>	218	11011010	271	11110000
<b>Ъ</b>	219	11011011	272	11110000
<b>Ы</b>	220	11011100	273	11110000
<b>Э</b>	221	11011101	274	11110000
<b>Ю</b>	222	11011110	275	11110000
<b>Я</b>	223	11011111	276	11110000

Стандарт кодирования символов Unicode позволяет пользоваться более чем двумя языками. В Unicode каждый символ кодируется шестнадцатиразрядным двоичным кодом. Такое количество разрядов позволяет закодировать 65 536 различных символов:  $2^{16} = 65\ 536$ .

# Информационный объём фрагмента текста

$I$  - информационный объём сообщения

$K$  – количество символов

$i$  – информационный вес символа

$$I = K \times i$$

В зависимости от разрядности используемой кодировки информационный вес символа текста, создаваемого на компьютере, может быть равен:

- 8 битов (1 байт) - **восемьразрядная кодировка;**
- 16 битов (2 байта) - **шестнадцатиразрядная кодировка.**

**Информационный объём** фрагмента текста - это количество битов, байтов (килобайтов, мегабайтов), необходимых для записи фрагмента оговорённым способом кодирования.

# Информационный объём фрагмента текста

**Задача 1.** Считая, что каждый символ кодируется одним байтом, определите, чему равен информационный объём следующего высказывания Жан-Жака Руссо:

**Тысячи путей ведут к заблуждению, к истине - только один.**

**Решение**

В данном тексте 57 символов (с учётом знаков препинания и пробелов). Каждый символ кодируется одним байтом. Следовательно, информационный объём всего текста - 57 байтов.

**Ответ:** 57 байтов.

# Информационный объём фрагмента текста

**Задача 2.** В кодировке Unicode на каждый символ отводится два байта. Определите информационный объём слова из 24 символов в этой кодировке.

**Решение.**

$$I = 24 \times 2 = 48 \text{ (байтов).}$$

**Ответ:** 48 байтов.

# Информационный объём фрагмента текста

**Задача 3.** Автоматическое устройство осуществило перекодировку информационного сообщения на русском языке, первоначально записанного в 8-битовом коде, в 16-битовую кодировку **Unicode**. При этом информационное сообщение увеличилось на 2048 байтов. Каков был информационный объём сообщения до перекодировки?

## **Решение**

Информационный вес каждого символа в 16-битовой кодировке в два раза больше информационного веса символа в 8-битовой кодировке. Поэтому при перекодировании исходного блока информации из 8-битовой кодировки в 16-битовую его информационный объём должен был увеличиться вдвое, другими словами, на величину, равную исходному информационному объёму. Следовательно, информационный объём сообщения до перекодировки составлял 2048 байтов = 2 Кб.



# Информационный объём фрагмента текста

**Задача 4.** Выразите в мегабайтах объём текстовой информации в «Современном словаре иностранных слов» из 740 страниц, если на одной странице размещается в среднем 60 строк по 80 символов (включая пробелы). Считайте, что при записи использовался алфавит мощностью 256 символов.

## Решение

$$K = 740 \times 80 \times 60$$

$$N = 256$$

$$I - ?$$

$$I = K \times i$$

$$N = 2^i$$

$$256 = 2^i = 2^8, i = 8$$

$$K = 740 \times 80 \times 60 \times 8 = 28\,416\,000 \text{ бит} = 3\,552\,000 \text{ байтов} = \\ = 3\,468,75 \text{ Кбайт} \approx 3,39 \text{ Мбайт.}$$

**Ответ:** 3,39 Мбайт.

# Самое главное

Текст состоит из символов - букв, цифр, знаков препинания и т. д., которые человек различает по начертанию. Компьютер различает вводимые символы по их двоичному коду. Соответствие между изображениями и кодами символов устанавливается с помощью **кодовых таблиц**.

В зависимости от разрядности используемой кодировки информационный вес символа текста, создаваемого на компьютере, может быть равен:

- 8 битов (1 байт) - **восемьразрядная кодировка**;
- 16 битов (2 байта) - **шестнадцатиразрядная кодировка**.

Информационный объём фрагмента текста - это количество битов, байтов (килобайтов, мегабайтов), необходимых для записи фрагмента оговорённым способом кодирования.



# Вопросы и задания

Сообщение занимает 6 страниц по 40 строк, в каждой

в какой кодировочной таблице можно закодировать

в строке записано по какому количеству символов

65 536 различных символов?

1) 1 байт 2) 2 байта 3) 4 байта 4) 8 байт

Сообщение, информационный объем которого равен

1000 бит, занимает 4 страницы по 32 строки, в каждой

2) строке записано по какому количеству символов?

1) 10 2) 15 3) 20 4) 25

3) Символы в алфавите языка кодировке Unicode

1) 1 байт 2) 2 байта 3) 4 байта 4) 8 байт

4) Привычка свинье нам дана! Замена е частью оа.

1) 8 бит 2) 16 бит 3) 32 бит 4) 64 бит

1) 512 битов 2) 608 битов 3) 704 байта 4) 8 Кбайт

1) 512 битов 2) 608 битов 3) 704 байта 4) 8 Кбайт

1) 512 битов 2) 608 битов 3) 704 байта 4) 8 Кбайт

1) 512 битов 2) 608 битов 3) 704 байта 4) 8 Кбайт

1) 512 битов 2) 608 битов 3) 704 байта 4) 8 Кбайт

1) 512 битов 2) 608 битов 3) 704 байта 4) 8 Кбайт

1) 512 битов 2) 608 битов 3) 704 байта 4) 8 Кбайт

1) 512 битов 2) 608 битов 3) 704 байта 4) 8 Кбайт

1) 512 битов 2) 608 битов 3) 704 байта 4) 8 Кбайт

1) 512 битов 2) 608 битов 3) 704 байта 4) 8 Кбайт

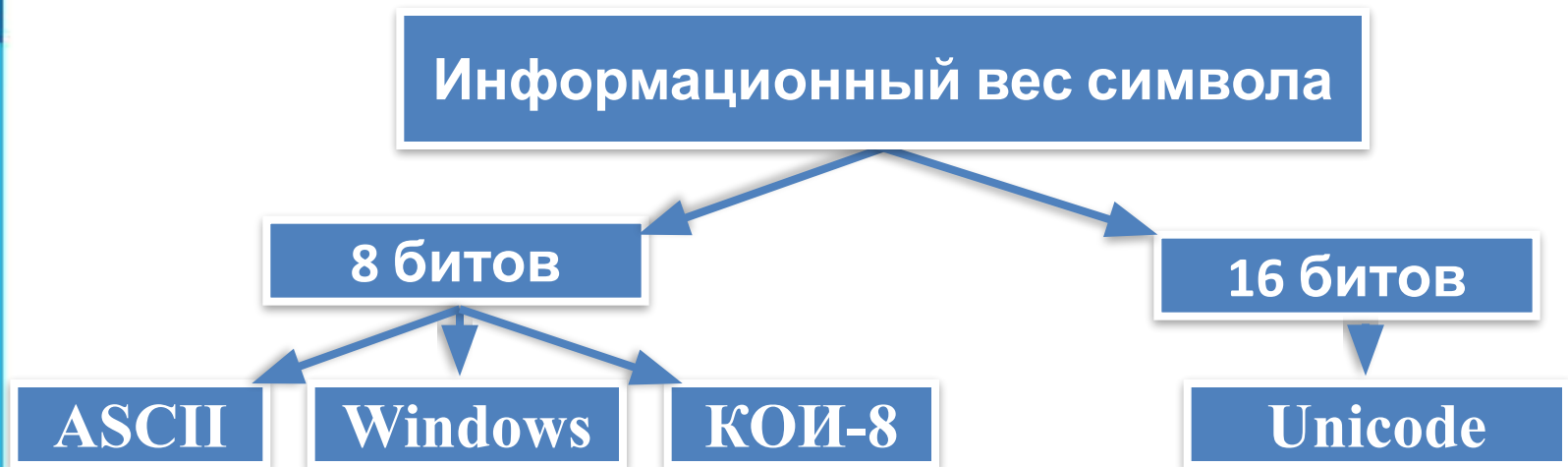
1) 512 битов 2) 608 битов 3) 704 байта 4) 8 Кбайт

1) 512 битов 2) 608 битов 3) 704 байта 4) 8 Кбайт

есть его основная ошибка.

# Опорный конспект

Компьютер различает вводимые символы по их двоичному коду. Соответствие между изображениями и кодами символов устанавливается с помощью **кодовых таблиц**.



$$I = K \times i$$

*I* - информационный объём сообщения

*K* - количество символов

*i* - информационный вес символа