



# ОЦЕНКА КОЛИЧЕСТВЕННЫХ ПАРАМЕТРОВ ТЕКСТОВЫХ ДОКУМЕНТОВ

## ОБРАБОТКА ТЕКСТОВОЙ ИНФОРМАЦИИ

**7 класс**



ИЗДАТЕЛЬСТВО

**БИНОМ**

# Ключевые слова

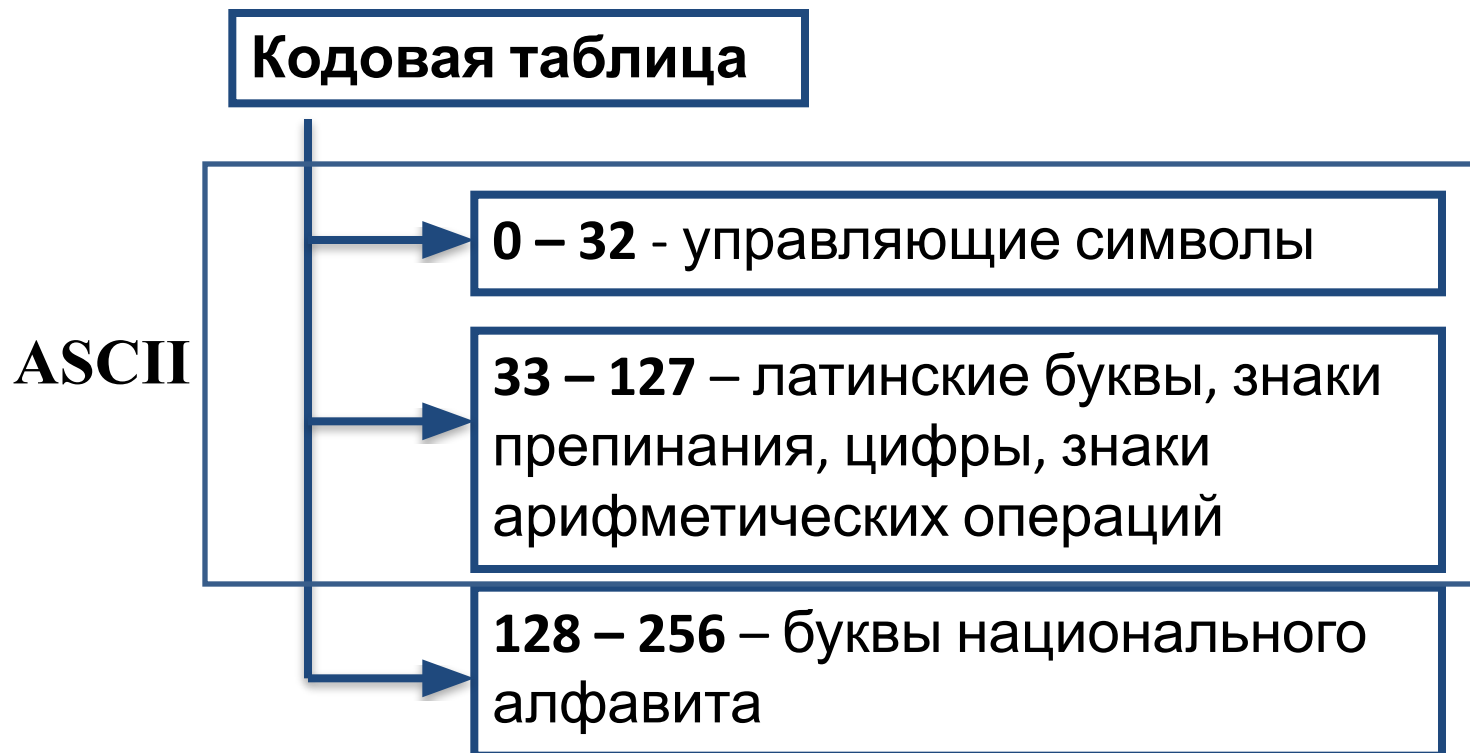
- кодовая таблица
- восьмиразрядный двоичный код
- информационный объём текста



# 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	
,	44	00101100	6	54	00110110
-	45	00101101	7	55	00110111
/	46	00101110	8	56	00111000
	47	00101111	9	57	00111001
<b>А</b>	65	01000001	<b>Н</b>	78	01001110
<b>В</b>	66	01000010	<b>О</b>	79	01001111
<b>С</b>	67	01000011	<b>Р</b>	80	01010000

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) - **восьмиразрядная 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) 16 битов 2) 16 байт 3) 16 символов 4) 16 бит

Сообщение, информационный объем которого равен

1000 бит, записано в кодировке Юникод. Сколько

5 Кбайт, занимает 4 страницы по 32 строки, в каждой

2) строке 2 символа. Сколько байт занимает этот текст?

3) 1000 байт 4) 1000 бит

На экране монитора, в кодировке Unicode

3) 1000 байт 4) 1000 бит

1) 512 битов 2) 304 байта

3) 608 битов 4) 456 байт

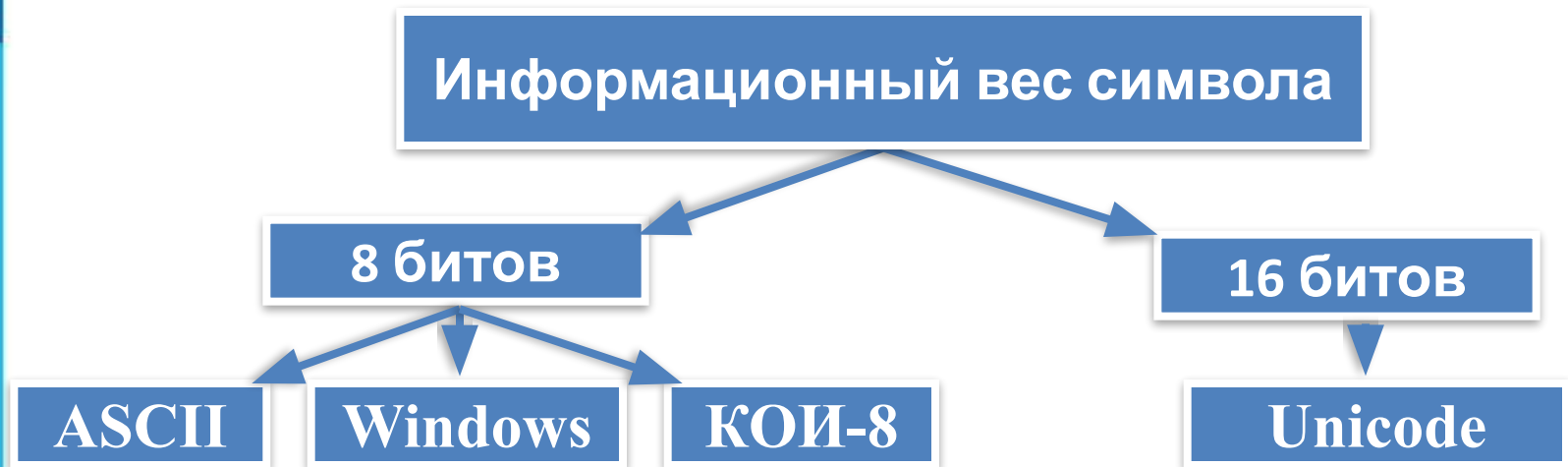
3) 8 Кбайт 4) 704 байта

4) 123 байта

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

# Домашнее задание

- Просмотреть презентацию, опорный конспект записать в тетрадь
- Решить в тетради задачу

Средняя скорость передачи данных по некоторому каналу связи равна 29 Кбит/с. Сколько секунд потребуется для передачи по этому каналу 50 страниц текста, если считать, что один символ кодируется одним байтом и на каждой странице в среднем 96 символов? (1Кбит= 1024 Бит)

Скан или фото конспекта и решенной задачи прислать на почту:

Группа Солопова С.Г. – [distant-ikt@mail.ru](mailto:distant-ikt@mail.ru)

Группа Кислица П.А. - [polinakislitsa@yandex.ru](mailto:polinakislitsa@yandex.ru)