

Введение в информатику

§ 2. Данные в компьютере

Данные

Данные — это записанная (зафиксированная) информация. Компьютеры работают только с данными.



текст



картинка



звук



видео



Как записать на одном языке?

Определения

Кодирование — это представление информации в форме, пригодной для её хранения, передачи и автоматической обработки.

Код — это правило, по которому сообщение преобразуется в цепочку знаков.

Двоичный код — это код, в котором используются только два знака (обычно 0 и 1).



Двоичное кодирование

Кодовая таблица

А	Г	Р
000	010	100

КОДОВОЕ СЛОВО

ГАГАРА: 010 000 010 000 100 000

Равномерный код — это код, в котором все кодовые слова имеют одинаковую длину.



Сколько существует кодовых слов длиной N в двоичном коде?

2^N

Декодирование

Кодовая таблица

А	Г	Р
000	010	100

?: 100000010100000

Декодирование — это восстановление исходного сообщения из кода.



Сколько символов было в сообщении?

5



Как разбить на кодовые слова?

по 3

100 000 010 100 000

Р А Г Р А

Декодирование. Задание 1

Слово АРКА закодировано числовой последовательностью **0100100010**, причем коды согласных и гласных букв имеют различную длину.

Какое слово по этому коду соответствует последовательности: **0100001 ?**

А	Р	К

Декодирование. Задание 2

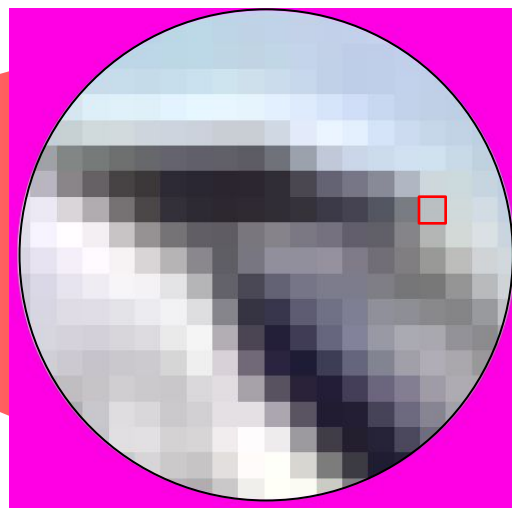
Одно из слов, приведенных ниже, закодировано следующим образом: **2 + X = 2X**

Найдите это слово.

2	+	X	=

- 1) сервер
- 2) курсор
- 3) модем
- 4) ресурс

Кодирование рисунков

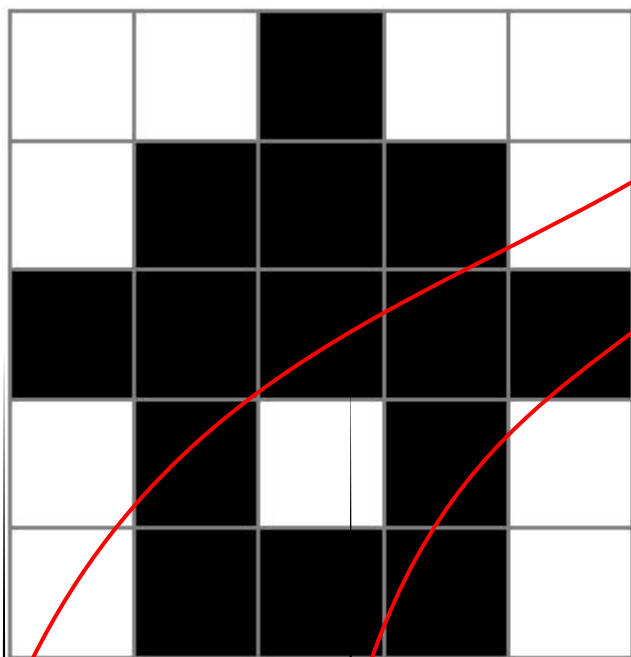


пиксель

Пиксель – это наименьший элемент рисунка, для которого можно задать свой цвет.

Растровый рисунок – это рисунок, которое кодируется как множество пикселей.

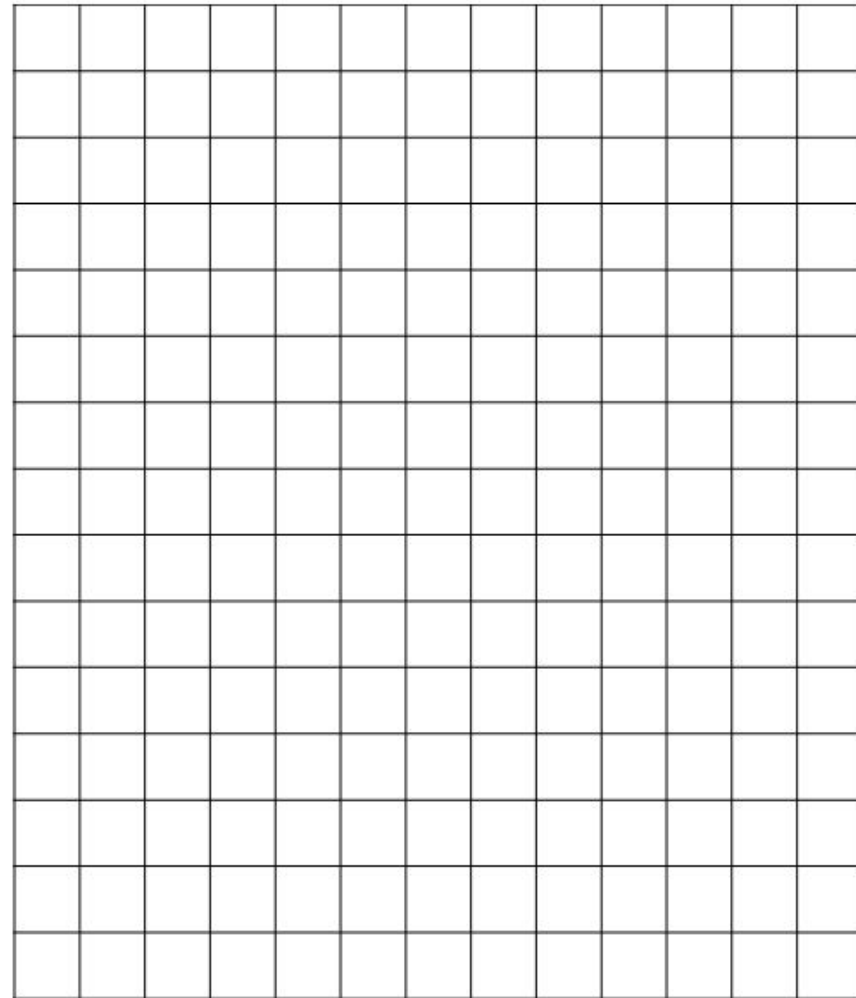
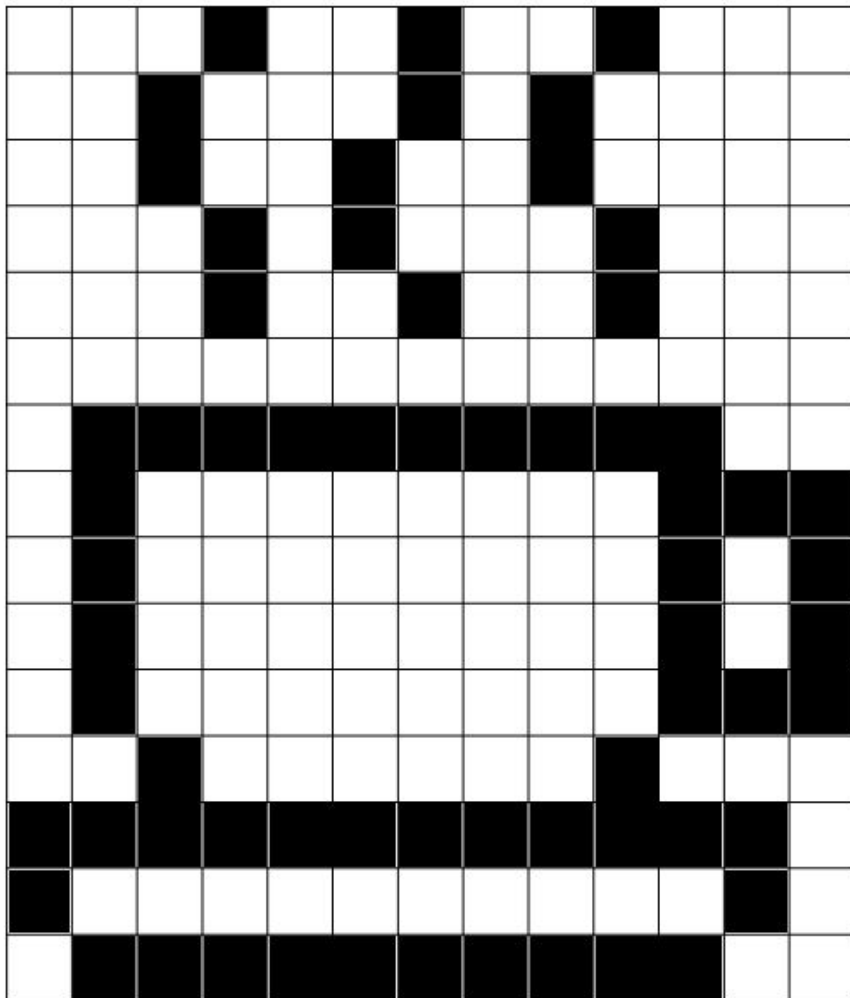
Кодирование рисунков



0	0	1	0	0
0	1	1	1	0
1	1	1	1	1
0	1	0	1	0
0	1	1	1	0

0010001110111110101001110

Кодирование рисунков. Задание 3

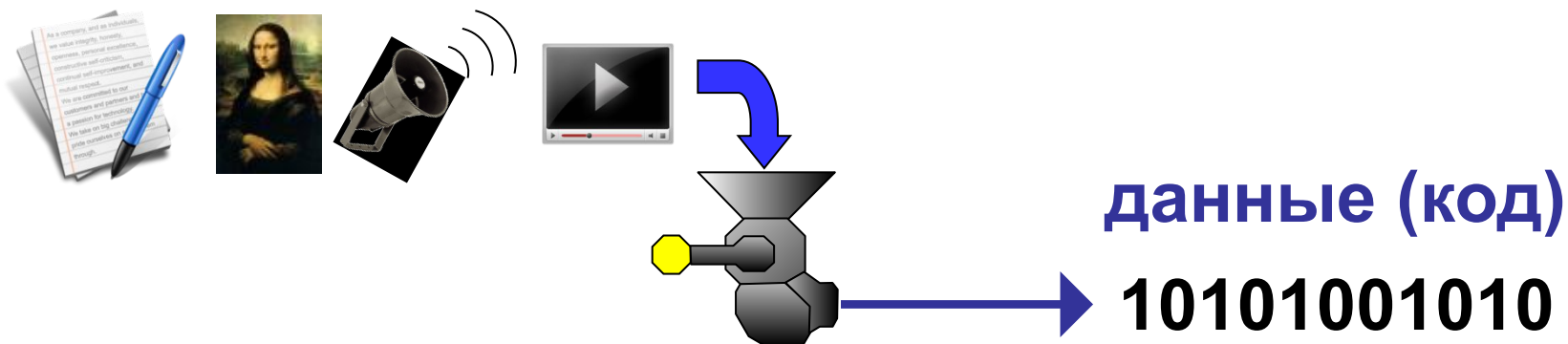


Какое слово закодировано на слайде?

	1	2	3
A	И	Е	З
B	Р	М	Н
C	Е	И	Е

A1 A3 B2 C1 B1 C3 B3 C2 A2 - ?

Как измерить информацию?



Алфавитный подход к измерению информации – это способ измерения информационного объема текста, не связанного с его содержанием.

Алфавитный подход

Каждый символ некоторого сообщения имеет определённый ***информационный вес*** – несёт ***фиксированное количество информации***.

Все символы одного алфавита имеют один и тот же вес, зависящий от мощности алфавита.

Алфавитный подход

Мощность алфавита:

- набор букв;
- знаков препинания;
- цифр;
- скобок и других символов, используемых в тексте;
- пробел между словами.

А Б В Г Д Е Ё
Ж З И К Л М
Н О П Р С Т У
Ф Х Ц Ч Ш Щ
Ъ Ы Ь Э Ю Я

0 1 2 3 4
5 6 7 8 9

A B C D E F
G H I J K L M
N O P Q R S T
U V W X Y Z

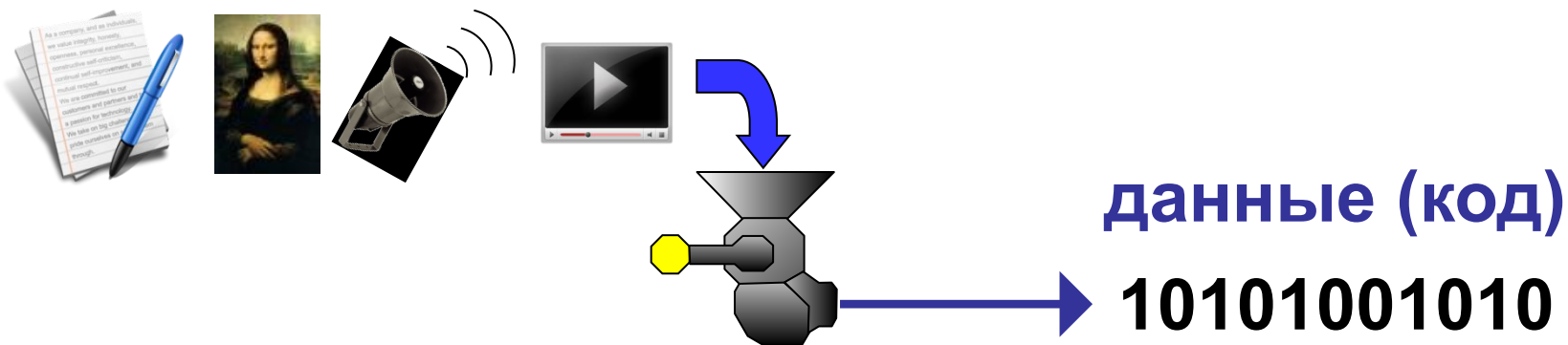
Формула Хартли

$$2^i = N$$

где **N** – количество символов знаковой системы (**мощность алфавита**);
 i – информационный вес символа.

Информационный вес символа двоичного алфавита принят за единицу информации и называется **1 БИТ**.

Как измерить информацию?



Количество информации в битах определяется длиной сообщения в двоичном коде.

10101100

8 битов

Бит – это одна двоичная цифра.
Наименьшая единица измерения информации.

Формула Хартли

Алфавит, из которого составляется «компьютерный текст», содержит 256 символов.

$$N = 2^i$$
$$256 = 2^8$$

1 символ компьютерного алфавита «весит» 8 битов.

Данные и носители информации

Роман «Евгений Онегин» (187 стр): **350 Кбайт**

фото высокого качества (сжатие): **~ 5 Мбайт**

1 минута звука (сжатие): **~ 1 Мбайт**

1 минута HD-video (сжатие): **~ 200 Мбайт**



1-16 Гбайт



до 17 Гбайт



до 512 Гбайт



до 10 Тбайт

Данные Google: **~ 15 000 000 Тбайт**

Единицы измерения

1 **байт** = 8 бит

1 **Кбайт** (килобайт) = 1024 байта

1 **Мбайт** (мегабайт) = 1024 Кбайт

1 **Гбайт** (гигабайт) = 1024 Мбайт

1 **Тбайт** (терабайт) = 1024 Гбайт



Можно записать через степени двойки?

Через степени числа 2:

1 **байт** = 2^3 бит

1 **Кбайт** = 2^{10} байта = $2^{10} \cdot 2^3$ бит = 2^{13} бит

1 **Мбайт** = 2^{10} Кбайт = $2^{10} \cdot 2^{13}$ бит = 2^{23} бит