



5: кодирование и декодирование данных

9-1: кодирование графической информации

9-2: кодирование звуковой информации

9-1, 9-2: скорость передачи информации

10: кодирование, комбинаторика

13: вычисление количества информации

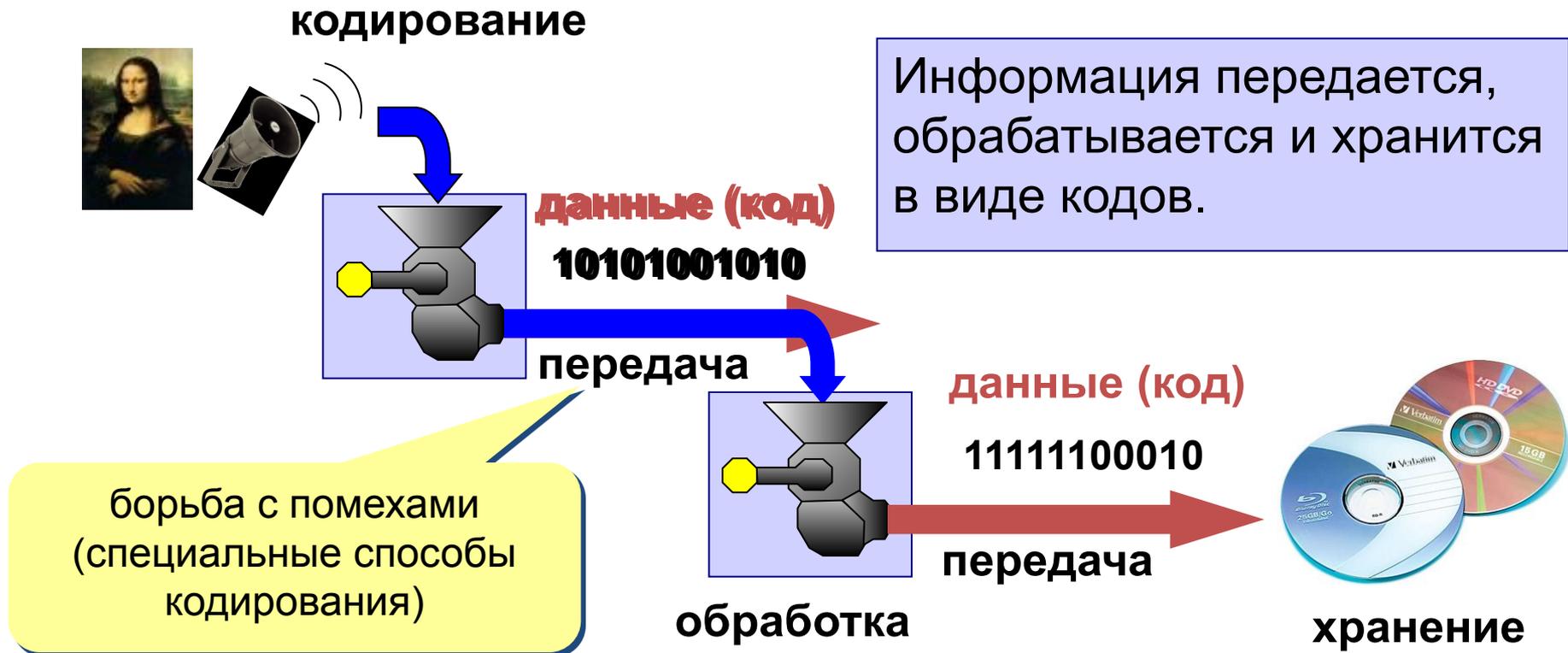
Информационные процессы

- Получение (через органы чувств)
- Хранение
- Обработка
- Кодирование – изменение формы, запись в некоторой знаковой системе (в виде кода)
- Поиск
- Сортировка
- Передача



Кодирование информации

Кодирование – это запись информации с помощью некоторой знаковой системы (языка).



ЯЗЫКИ

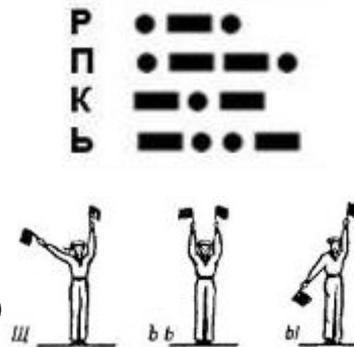
Язык – знаковая система, используемая для хранения и передачи информации.

естественные (русский, английский, ...)

есть правила и исключения

формальные (строгие правила)

$$E = mc^2$$



```
program qq;
begin
writeln("Привет!");
end.
```

Грамматика – правила, по которым строятся слова.

Синтаксис – правила, по которым из слов строятся предложения.

з алфавита строятся

Кодирование

Задача 1. Закодируйте свое имя с помощью азбуки Морзе.

А	● —	П	● — — — ●	Ь	— ● ● —
Б	— ● ● ●	Р	● — ●	Ы	— ● — — —
В	● — — —	С	● ● ●	Й	● — — — —
Г	— — — ●	Т	—		
Д	— ● ●	У	● ● —	1	● — — — — —
Е	●	Ф	● ● — ●	2	● ● — — — —
Ж	● ● ● —	Х	● ● ● ●	3	● ● ● — — —
З	— — — ● ●	Ц	— ● — — ●	4	● ● ● ● —
И	● ●	Ч	— — — — ●	5	● ● ● ● ●
К	— ● — —	Ш	— — — — —	6	— ● ● ● ●
Л	● — — ● ●	Щ	— — — ● —	7	— — — ● ● ●
М	— — —	Э	● ● — — ● ●	8	— — — — ● ●
Н	— ●	Ю	● ● — — —	9	— — — — — ●
О	— — — —	Я	● — — ● —	0	— — — — — —

ВАСЯ



Код неравномерный, нужен разделитель!

Кодирование

Задача 2. Закодируйте свое имя с помощью кодовой таблицы (*Windows-1251*):

	0	1	2	3	4	5	6	7	8	9	А	В	С	Д	Е	Ф
С	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
Д	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я

В А С Я

ВАСЯ

С2 С0 D1 DF



Код равномерный, разделитель **НЕ** нужен!



-
- Что такое «много информации» и «мало информации»?
 - Как определить, в каком сообщении больше информации?

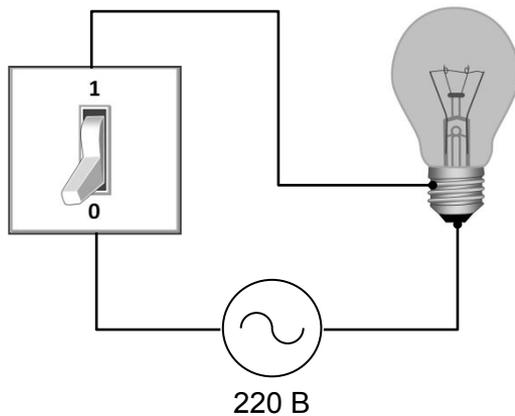
Идея:

количество информации – это длина сообщения, с помощью которого её можно закодировать.

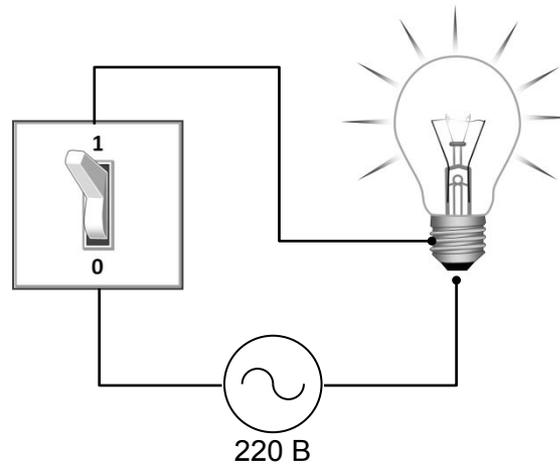
- От чего зависит длина сообщения?
от алфавита!
- Какой алфавит выбрать?
абвг...эюя?
abcd...хуz?

Какой код использовать?

Идея: использовать тот код, который применяется в компьютерной технике



«0»



«1»

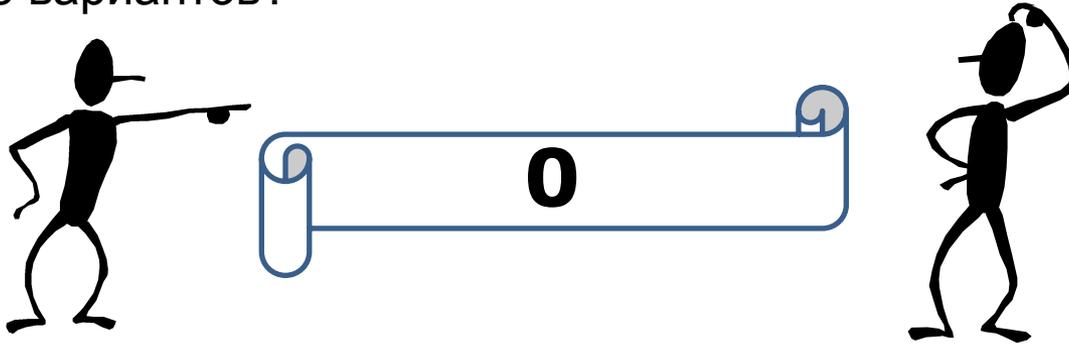
Код, в котором используются только два знака, называется **двоичным**. Все виды информации в компьютерах кодируются в двоичном коде.

1 бит – это количество информации, которое можно передать с помощью одного знака в двоичном коде («0» или «1»).

*bit = **binary digit**, двоичная цифра*



Сколько вариантов?



1 бит:

0 1

2 бита:

00 01
10 11

3 бита:

000 001 100 101
010 011 110 111

4 бита: 16 вариантов!

Если вариантов больше...

Количество вариантов	2	4	8	16	32	64	128	256	...	2^k
Количество бит информации	1	2	3	4	5	6	7	8	...	k

6 вариантов – ? $2^2 < 6 < 2^3$

Ответ: 3 бита



Единицы измерения

1 байт (*byte*) = 8 бит

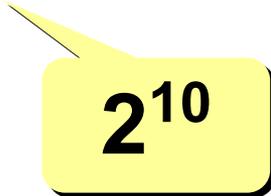
1 Кбайт (килобайт) = 1024 байта

1 Мбайт (мегабайт) = 1024 Кбайт

1 Гбайт (гигабайт) = 1024 Мбайт

1 Тбайт (терабайт) = 1024 Гбайт

1 Пбайт (петабайт) = 1024 Тбайт



2^{10}

Задача. Определить объем информации в сообщении

ПРИВЕТВАСЯ

для кодирования которого используется русский алфавит (только заглавные буквы).

$$I = N \cdot k, \quad k = \log_2 M$$

Решение:

- считаем все символы $N = \mathbf{10}$ (символов)
- мощность алфавита $M = 32$ символа ($32 = 2^5$)
- 1 символ несет **5 бит** информации
 $I = N \cdot k = 10 \cdot 5$ бит

Ответ: 50 бит

Вероятностный подход

Как посчитать информацию, если варианты не равновероятны?

Клод Шеннон (1916 —2001)

американский математик и электротехник, один из создателей математической теории информации и криптографии.



Идея: если случается менее вероятное событие, мы получаем больше информации.

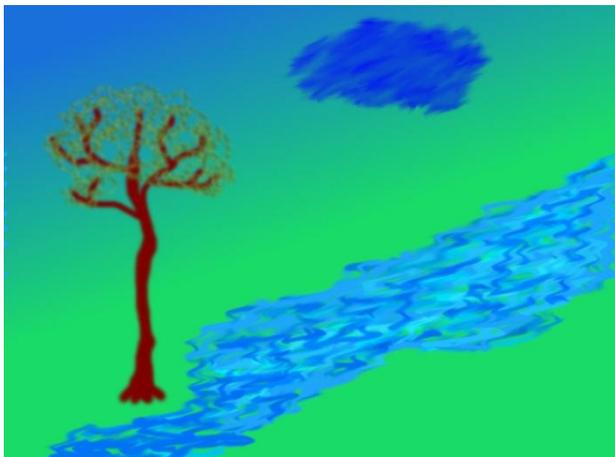
$0 \leq p_i \leq 1$ – вероятность выбора i -ого варианта ($i=1, \dots, N$)

Если произошло событие i , мы получаем информацию

$$I_i = \log_2 \frac{1}{p_i}$$



m пикселей



n пикселей

Цветовое разрешение (бит на пиксель) k

Количество цветов в палитре $M = 2^k$

Количество информации $I = m n k$



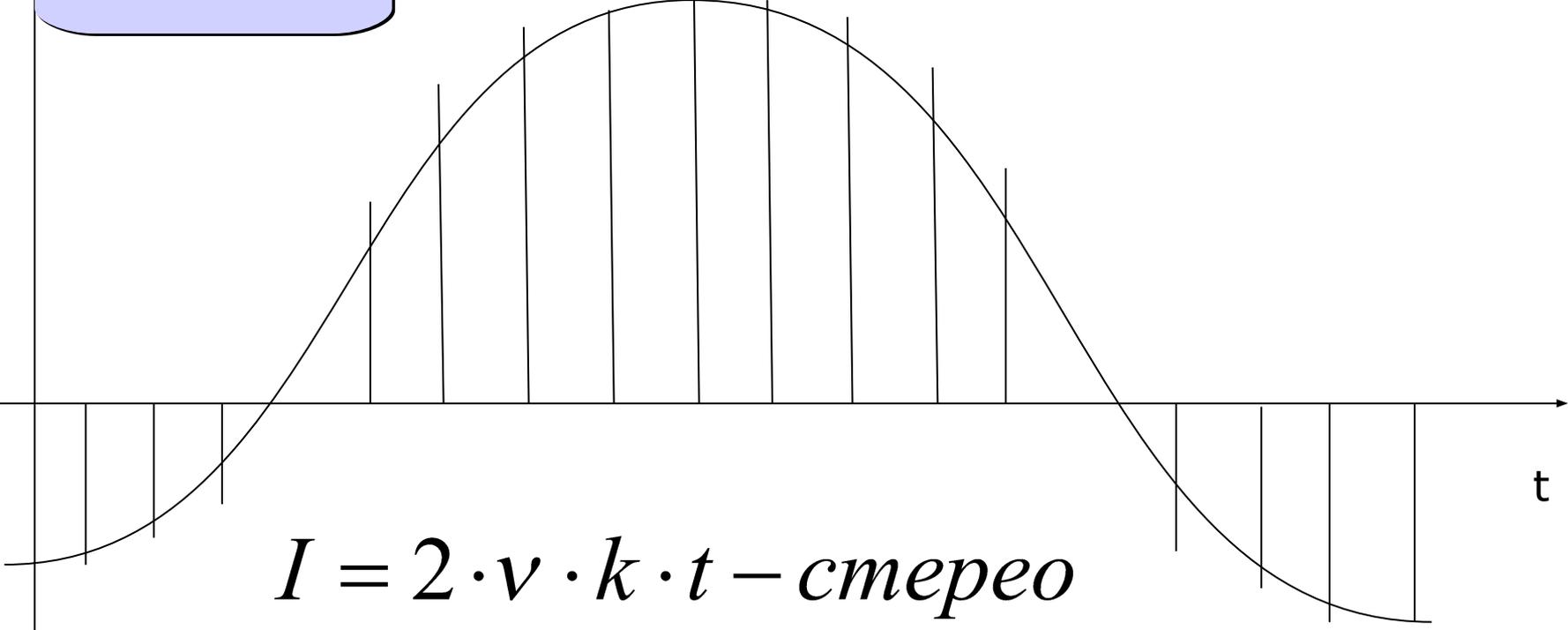
$$I = v \cdot k \cdot t$$

$$M = 2^k$$

Частота
дискретизации
(Гц, 1/с)

Звуковое
разрешение
(бит)

Количество уровней
дискретизации


$$I = 2 \cdot v \cdot k \cdot t - \text{стерео}$$



$$I = V t$$

Скорость

Время передачи!

$$I = v k t$$

Время звучания!



-
- на 1 странице $32 \cdot 64 = 2048$ символов
 - на 10 страницах $10 \cdot 2048 = 20480$ символов
 - каждый символ занимает 1 байт

 - 20480 байт или ...
 - $20480 \cdot 8$ бит или ...
 - $20480 : 1024$ Кбайт = 20 Кбайт



-
- мощность алфавита $25 + 10 = 35$ символов
 - на кодирование 1 символа нужно 6 бит:
 - $2^5 = 32 < 35 \leq 2^6 = 64$
 - для хранения номера нужно $7 \cdot 6 = 42$ бита
 - ... или 6 байт (целое число байт!)
 $5 \cdot 8 = 40 < 42 \leq 6 \cdot 8 = 48$
 - для 50 номеров нужно $50 \cdot 6 = 300$ байт

300 байт

Какой минимальный объём памяти (в Кбайт) нужно зарезервировать, чтобы можно было сохранить любое растровое изображение размером 64 на 64 пикселов при условии, что в изображении могут использоваться 256 различных цветов? В ответе запишите только целое число, единицу измерения писать не нужно.

Решение:

1. находим количество пикселей, используя для вычисления степени числа 2:
2. $N = 64 \cdot 64 = 2^6 \cdot 2^6 = 2^{12}$
3. $256 = 2^8$, поэтому для кодирования одного из 256 вариантов цвета нужно выделить в памяти $8 = 2^3$ бит на пиксель
4. объём файла в битах $2^{12} \cdot 2^3 = 2^{15}$
5. объём файла в Кбайтах $2^{15} : 2^{13} = 2^2 = 4$

Ответ: 4.

Музыкальный фрагмент был оцифрован и записан в виде файла без использования сжатия данных. Получившийся файл был передан в город А по каналу связи за 30 секунд. Затем тот же музыкальный фрагмент был оцифрован повторно с разрешением в 2 раза выше и частотой дискретизации в 1,5 раза меньше, чем в первый раз. Сжатие данных не производилось. Полученный файл был передан в город Б; пропускная способность канала связи с городом Б в 4 раза выше, чем канала связи с городом А. Сколько секунд длилась передача файла в город Б? В ответе запишите только целое число, единицу измерения писать не нужно.

$$I = v \cdot k \cdot t$$

Решение :

- объём музыкального файла вычисляется по формуле
- Где – V частота дискретизации, k – разрешение (глубина кодирования), t – время звучания
- при повышении разрешения в 2 раза объём файла увеличивается в 2 раза, поэтому время тоже увеличится в 2 раза
- при снижении частоты дискретизации в 1,5 раза объём файла уменьшается в 1,5 раза, поэтому время тоже уменьшится в 1,5 раза
- при увеличении пропускной способности канала связи в 4 раза время передачи уменьшится в 4 раза

Поэтому исходное время передачи файла нужно

- а) умножить на 2
- б) разделить на 1,5
- в) разделить на 4

получается $30 \cdot 2 / 1,5 / 4 = 10$ секунд

Ответ: 10.

- время передачи: $2 \cdot 60 \text{ сек} = 120 \text{ сек}$
 - передано информации
 - $256 \cdot 1000 \cdot 120 \text{ бит}$
 - $= 2^8 \cdot 2^3 \cdot 125 \cdot 2^2 \cdot 30 \text{ бит} = \frac{2^{13} \cdot 125 \cdot 3}{2^3} \text{ Кбайт}$
- объем файла 3750 Кбайт