

Измерение информации

Лекция 3



Как измерить информацию?

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

объёмный
подход

Идея:

- количество информации определяется временем ее передачи
- количество информации определяется длиной сообщения.

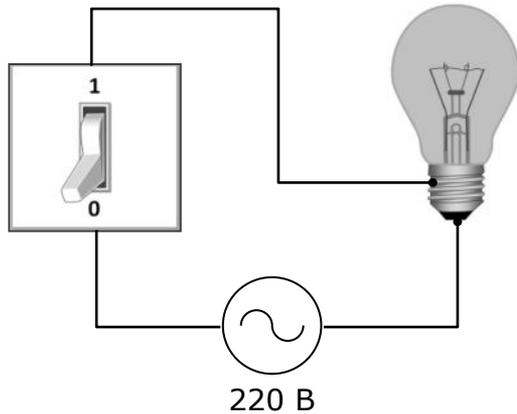


Как именно закодировать?

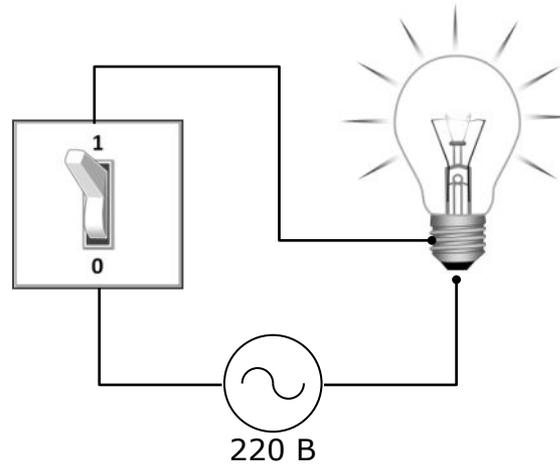


Какой код использовать?

Идея: использовать тот код, который применяется в компьютерной технике



«0»



«1»



Двоичный код

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

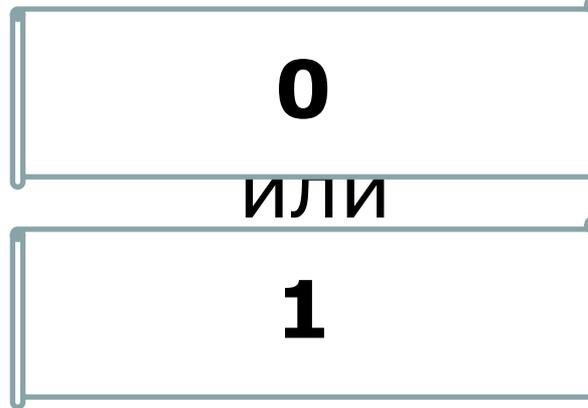
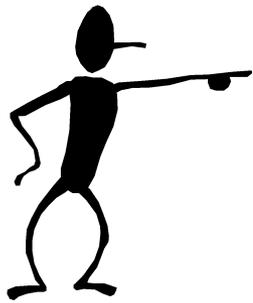
1 бит – это количество информации, которое можно передать с помощью одного знака в двоичном коде («0» или «1»).

К. Шеннон, 1948:

bit = binary digit, двоичная цифра



1 бит



Что можно сообщить с помощью 1 знака (1 бита)?

выбрать один из двух вариантов, если заранее договориться, что означают «0» и «1»



1 бит

1 бит – это количество информации, которое мы получаем при выборе одного из двух возможных вариантов (вопрос: «Да» или «Нет»?)

Примеры:

Эта стена – зеленая? Да.

Дверь открыта? Нет.

Сегодня выходной? Нет.

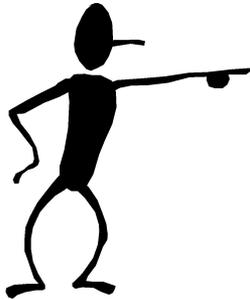
Это новый автомобиль? Новый.

Ты будешь чай или кофе? Кофе.



Сколько информации?

Определите количество информации:



01

2 бита



10101

5 битов

1010111

7 битов

1010101001

10 битов



Единицы измерения

1 байт (*byte*) = 8 бит

КиБ (кибибайт)

2^{10}

1 Кбайт (килобайт) = 1024 байта

МиБ (мебибайт)

1 Мбайт (мегабайт) = 1024 Кбайт

ГиБ (гибибайт)

1 Гбайт (гигабайт) = 1024 Мбайт

ТиБ (тебибайт)

1 Тбайт (терабайт) = 1024 Гбайт

ПиБ (пебибайт)

1 Пбайт (петабайт) = 1024 Тбайт



Перевод в другие единицы

5 Кбайт = $5 \cdot 1024$ байтов = 5120 байтов

15 байтов = $15 \cdot 8$ битов = 120 битов

2048 Кбайт = $2048 : 1024$ Мбайт = 2 Мбайта

1024 Мбайт = $1024 : 1024$ Гбайт = 1 Гбайт

3 Мбайта = $3 \cdot 1024$ Кбайт = 3072 Кбайта



Формула Хартли (1928)

$$N = 2^I$$

$$I = \log_2 N$$

I – количество информации в битах
 N – количество вариантов



Ральф
Хартли

Пример:

В аэропорту стоит 10 самолетов, из них один летит в Санкт-Петербург. Оценить количество информации в сообщении «В Санкт-Петербург летит второй самолет»?

$$I = \log_2 10 = \frac{\ln 10}{\ln 2} = \frac{\lg 10}{\lg 2} = 3,322 \text{ бита}$$



Алфавитный подход

N – мощность алфавита

Информационный объём

символа:

$$i = \log_2 N$$

вверх до
целого числа

сообщения длиной L :

$$I = L \cdot \log_2 N$$



Пример: сообщение длиной 100 символов закодировано с помощью алфавита из 50 знаков.

$$i = \log_2 50 \approx 5,644 \text{ бита}$$

6 битов

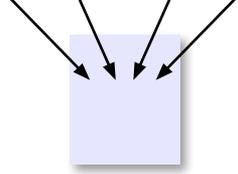
$$I = 100 \cdot \log_2 50 \approx 564,4 \text{ бита}$$

600 битов

Количество различных сообщений

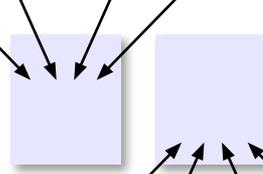
алфавит: А, Б, В, Г

А, Б, В, Г



всего: 4

А, Б, В, Г



всего: $4 \cdot 4 = 4^2 = 16$

А, Б, В, Г для **каждого** варианта

N – мощность алфавита

L – длина сообщения

Q – количество различных сообщений

$$Q = N^L$$

Вероятность и информация

$f(p) = -K \cdot \log_2 p$ при $K = 1 \Rightarrow$ информация в битах

Если событие имеет вероятность p , то количество информации в битах, полученное в сообщении об этом событии, равно

$$I = -\log_2 p = \log_2 \frac{1}{p}$$

$$p = 1 \Rightarrow I = \log_2 1 = 0$$

$$p \rightarrow 0 \Rightarrow I \rightarrow \log_2 \infty = \infty$$



Связь с формулой Хартли

N равновероятных событий $\Rightarrow p = \frac{1}{N}$

$$I = \log_2 \frac{1}{p} = \log_2 N$$

совпадает с
формулой Хартли

Если вероятности разные:



«**Васе достался зелёный шарик**».

$$p = \frac{6}{8} = \frac{3}{4}$$

$$I = \log_2 \frac{4}{3} \approx 0,415 \neq 0,5$$



Формула Шеннона

Количество полученной информации равно уменьшению неопределенности.

$$I = \Delta H = H_{\text{нач}} - H_{\text{кон}}$$



Как вычислить H ?

Неопределённость знаний об источнике данных (N событий, вероятности p_i):

$$H = \sum_{i=1}^N p_i \cdot \log_2 \frac{1}{p_i} = p_1 \cdot \log_2 \frac{1}{p_1} + \dots + p_N \cdot \log_2 \frac{1}{p_N}$$



Клод Шеннон

информационная **энтропия**

