ИНФОРМАЦИЯ

Тема 2. Измерение количества информации

© К. Поляков, 2006-2011

Как измерить информацию?

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

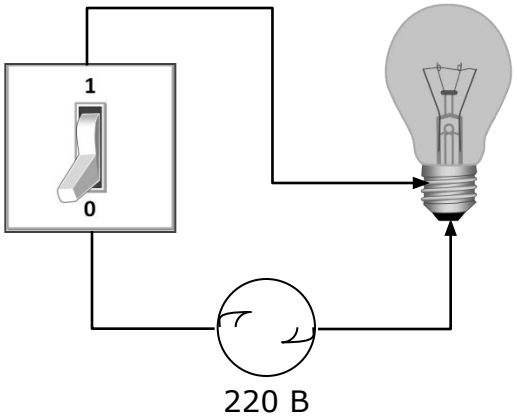
Идея:

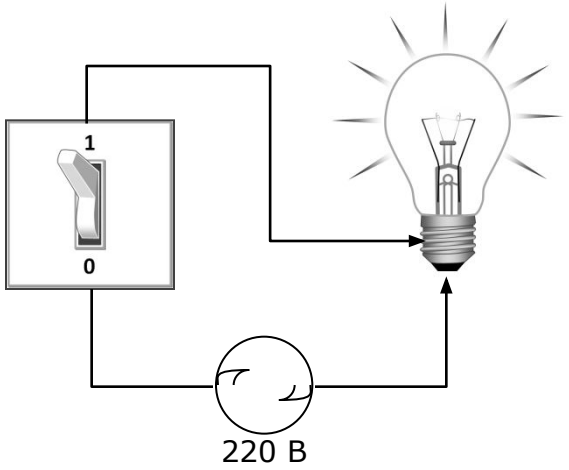
- количество информации определяется временем ее передачи
- количество информации – это длина сообщения, с помощью которого её можно закодировать.

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

Какой код использовать?

Идея: использовать тот код, который применяется в компьютерной технике





«0» «1»

© К. Поляков, 2006-2011

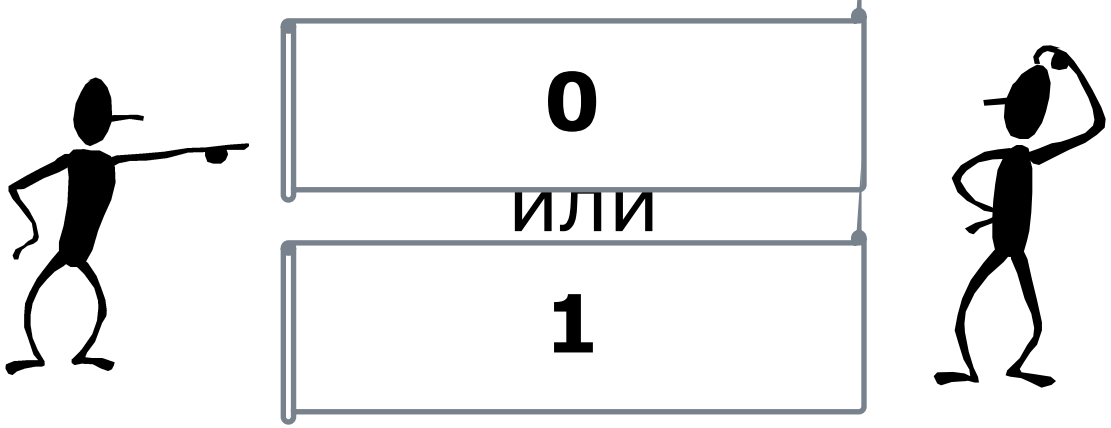
Двоичный код

Код, в котором используются только два знака, называется **двоичным**. В компьютерах применяется двоичный код.

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

*bit = **b**inary dig**it**, 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»

© К. Поляков, 2006-2011

Единицы измерения

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.

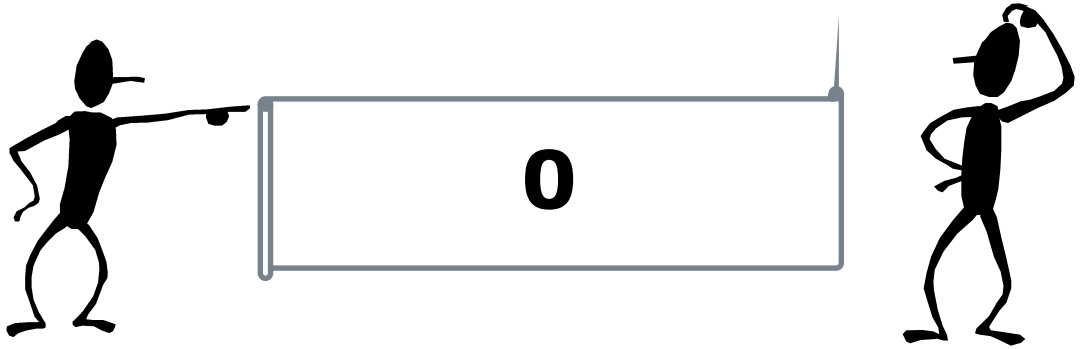
Сегодня выходной? Нет.

Это новый автомобиль? Новый.

Ты будешь чай или кофе? Кофе.

© К. Поляков, 2006-2011

Сколько вариантов?



1 бит: 0 1

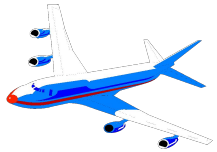
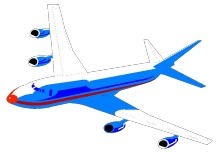
2 бита: 00 01
10 11

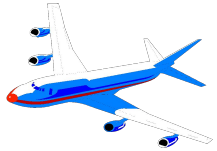
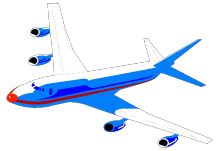
3 бита: 000 001 100 101
010 011 110 111

4 бита: 16 вариантов!

© К. Поляков, 2006-2011

Если вариантов больше...

00   01

10   11

4 варианта – **2** бита

? вариантов – **3** бита

?

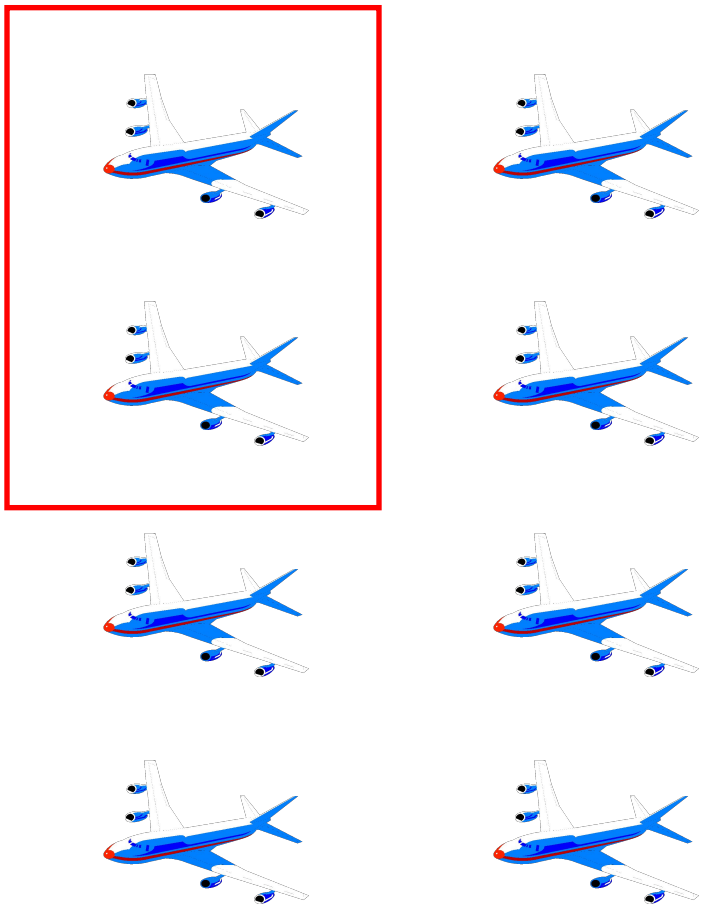
варианта – **5** бит

? варианта – **6** бит

?

© К. Поляков, 2006-2011

Если вариантов больше...



«Да» или «Нет»?

2 варианта – **1** бит

4 варианта – **2** бита

8 вариантов – **3** бита

? Изменится ли количество информации, если сразу указать на нужный самолет?

© К. Поляков

Если вариантов больше...

Количество вариантов	2	4	8	16	32	64	128	256	512	1024
Количество бит информации	1	2	3	4	5	6	7	8	9	10

6 вариантов – между **4** (2 бита) и **8** (3 бита)

Ответ: количество информации между

2 и 3 битами

© К. Поляков, 2006-2011

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}

© К. Поляков, 2006-2011

Единицы измерения (11 класс)

1 байт *(byte)* – это объем компьютерной памяти, который имеет индивидуальный адрес.

Примеры из истории:

1 байт = 4 бита
1 байт = 6 бит
1 байт = 12 бит

Сейчас обычно:

1 байт = 8 бит

© К. Поляков, 2006-2011

Формула Хартли (1928)

$$N = 2^I$$

$$I = \log_2 N$$

I – количество информации в битах

N – количество вариантов

Пример:

В аэропорту стоит 6 самолетов, из них один летит в Москву. Сколько информации в сообщении «В Москву летит второй самолет»?

$$I = \log_2 6 = \frac{\ln 6}{\ln 2} = \frac{\lg 6}{\lg 2} = 2{,}585 \text{ бит}$$

Алфавитный подход

Алфавит – набор знаков, используемых при кодировании информации с помощью некоторого языка.

Примеры:

АБВГДЕЖЗИЙКЛМНОПРСТУФХЦЧШЩЪЫЬЭЮЯ	32
ABCDEFGHIJKLMNOPQRSTUVWXYZ	26
× O	2
0123456789	10

Мощность алфавита – количество символов.

Все символы несут одинаковую информацию:

$$I = \log_2 N$$

информационная емкость символа

мощность алфавита

© К. Поляков, 2006-2011

Алфавитный подход

Задача. Определить объем информации в сообщении

ПРИВЕТВАСЯ

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

Решение:

- считаем все символы (здесь **10** символов)
- мощность алфавита – 32 символа ($32=2^5$)
- 1 символ несет **5 бит** информации

Ответ: 10·5 бит = 50 бит

© К. Поляков, 2006-2011

ИНФОРМАЦИЯ

Тема 3. Задачи

© К. Поляков, 2006-2011

Перевод в другие единицы

5 Кбайт = 5 ·1024 байт = 5120 байт

15 байт = 15 · 8 бит = 120 бит

2048 Кбайт = 2048:1024 Мбайт = 2 Мбайта

1024 Мбайт = 1024:1024 Гбайт = 1 Гбайт

3 Мбайта = 3 · 1024 Кбайт = 3072 Кбайта

© К. Поляков, 2006-2011

Перевод в другие единицы

Сравните (поставьте знак <, > или =):

3 байта = 24 бита

1000 байт < 1 Кбайт

250 байт < 0,25 Кбайт

1 Мбайт > 1000 Кбайт

8192 бита = 1 Кбайт

© К. Поляков, 2006-2011

Перевод в другие единицы

Впишите недостающее число:

8 байт = ? бита

1,5 Кбайт = ? байт

512 бит = ? байта

2 Мбайта = ? Кбайт

4 бита = ? Байта

3072 Кбайта = ? Мбайта

© К. Поляков, 2006-2011

Задачи: текст

Сколько места в памяти надо выделить для хранение предложения

Привет, Вася!

- считаем все символы, включая знаки препинания и пробелы (здесь **13** символов)
- если нет дополнительной информации, то считаем, что 1 символ занимает **1 байт**
- в кодировке UNICODE 1 символ занимает **2 байта**

Ответ: 13 байт или 104 бита

(в UNICODE: 26 байт или 208 бит)

© К. Поляков, 2006-2011

Задачи: текст

Решение:

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

Ответ:

- 20480 байт или …
- 20480·8 бит или …
- 20480:1024 Кбайт = 20 Кбайт

© К. Поляков, 2006-2011

Задачи: кодирование

Сколько бит нужно выделить для хранения текста

МУНСА УРЕ КАМУКА

при использовании алфавита племени МУМУКА: буквы МУКАЕНРС и пробел?

Решение:

- в алфавите 9 символов (8 букв и пробел)
- $2^3 < 9 < 2^4$, поэтому на 1 символ нужно выделить 4 бита
- в тексте 16 символов (считая пробелы)

Ответ: 4·16 бит = 64 бита = 8 байт

Если в алфавите 25 символов?

© К. Поляков, 2006-2011

Задачи: кодирование

Объем сообщения, содержащего 1024 символов, составил 1/512 часть Мбайта. Какова мощность алфавита, с помощью которого записано сообщение?

Решение:

- объем сообщения в байтах:
  1024 Кбайт / 512 = 2 Кбайта = 2048 байт
- на 1 символ приходится
  2048 / 1024 = 2 байта = 16 бит
- мощность алфавита 2^{16} = 65536 символов

Ответ: 65536 символов (кодировка UNICODE)

© К. Поляков, 2006-2011