
Введение в информатику

§2. Данные в компьютере

Данные в компьютере

Как хранятся данные?



Как можно знания, находящиеся у вас «в голове», передать другим людям или сохранить для потомков?

Кодирование – это представление информации в форме, удобной для её хранения, передачи и обработки.

Двоичный код – это код, в котором используются только два знака(обычно – 0 и1).

Как хранятся данные?

Для примера посмотрим, как можно представить символы текста в двоичном коде. Каждой букве ставится в соответствие цепочка нулей и единиц (она называется кодовое слово), например, так:

А	Б	К	О
00	01	10	11

Как хранятся данные?

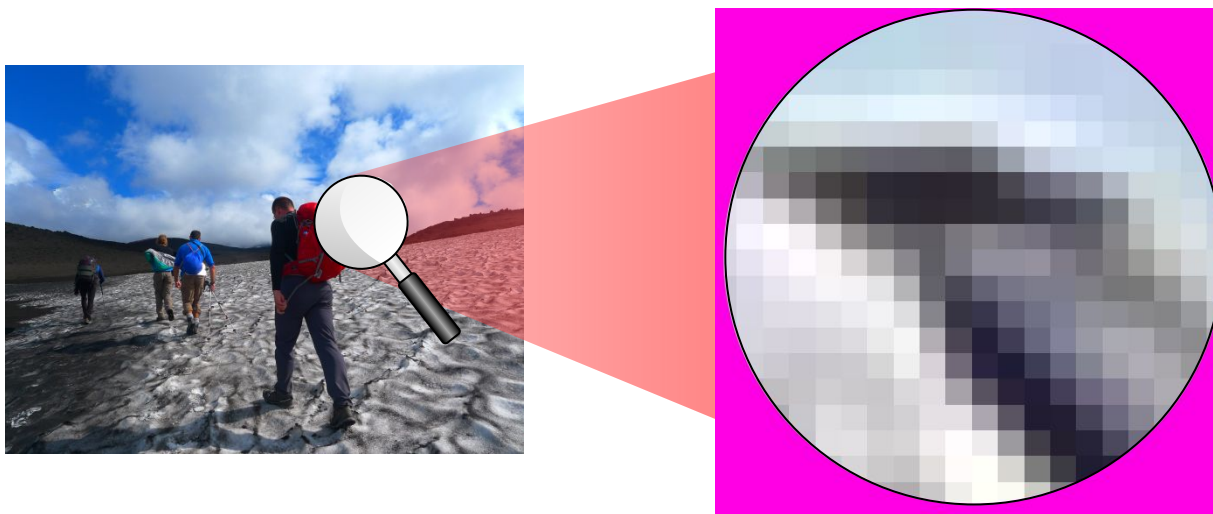
Эту таблицу, которая задает правила преобразования текста в двоичный код, называют **кодовой таблицей**.

Код, в котором все кодовые слова имеют одинаковую длину, называется **равномерным**.

Цепочку символов, полученную в результате кодирования, называют **сообщением**.

Кодирование рисунков

При работе с графическими редакторами вы, возможно, видели, что получается, если очень сильно увеличить рисунок:



Изображение состоит из отдельных квадратиков, и перекрасить часть такого элемента невозможно, можно только закрасить одним цветом весь элемент.

Кодирование рисунков

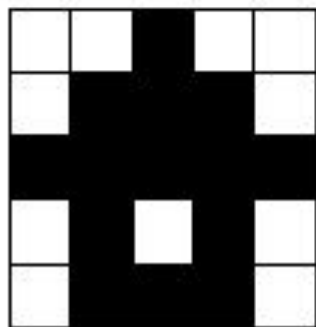
Пиксель – это наименьший элемент цифрового рисунка, для которого можно задать свой цвет независимо от других.

Растровый рисунок – это рисунок, который представлен в памяти как множество точек разного цвета (пикселей).

*Английское слово **pixel** – это сокращение от слов *pictures element* (элемент рисунков) по правилам, принятым в бытовом английском языке.*

Кодирование рисунков

Для того чтобы закодировать рисунок, нужно для каждого цвета определить кодовое слово (цепочку нулей и единиц). Для простоты рисунок будет чёрно-белый. Тогда пиксель чёрного цвета можно обозначить единицей, а пиксель белого цвета – нулём (или наоборот!).



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

Тогда двоичным кодом рисунка можно считать данные из таблицы выписанные по строкам:
0010001110111110101001110.

Измерение данных

Для человека информация – это смысл сообщения, который зависит от его предыдущих знаний.

Для компьютера информация – это просто набор нулей и единиц, и только человек может придать им какой-то смысл.

Количество информации с точки зрения компьютера определяет только «затраты» на её хранение и передачу, независимо от значения этих данных.

Измерение данных

Бит – это одна двоичная цифра.

Таким образом, для того чтобы измерить количество информации в битах, нужно просто измерить длину цепочки нулей и единиц, в виде которой записаны данные.

Измерение данных

1 байт = 8 битов = 2^3 битов

1 Кбайт (килобайт) = 1024 байта = 2^{10} байтов = 2^{13} битов

1 Мбайт (мегабайт) = 1024 Кбайт = 2^{10} Кбайт = 2^{20} байтов
= 2^{23} битов

1 Гбайт (гигабайт) = 1024 Мбайт

1 Тбайт (терабайт) = 1024 Гбайт

1 Пбайт (петабайт) = 1024 Тбайт

В этих единицах измеряется память компьютера, как оперативная, так и долговременная.

Измерение данных

Строго говоря, нужно называть такие кило-(мега-, гига-,...) байты двоичными, поскольку множитель 1024 – это 2^{10} . Стандарт Международной электротехнической комиссии (МЭК) предлагает называть их «кибибайт», «мебибайт», «гибибайт», «тебибайт» и «пебибайт», **но эти названия на практике не прижились.**

Данные в долговременной памяти

Системные программы (точнее – операционная система) позволяют нам воспринимать наборы данных «по-человечески», как нам удобно: как картинки, тексты, звук, видео. Для этого в долговременной памяти данные хранятся в виде файлов.

Файл – это набор данных, расположенный в долговременной памяти компьютера и имеющий имя.

Перевод количества информации в другие единицы

Сколько битов

*в 1 байте?
в 2 байтах?
в 3 байтах?
в 4 байтах?
в 10 байтах?
в 1/2 байта?
в 1/4 байта?
в 1/8 байта?*

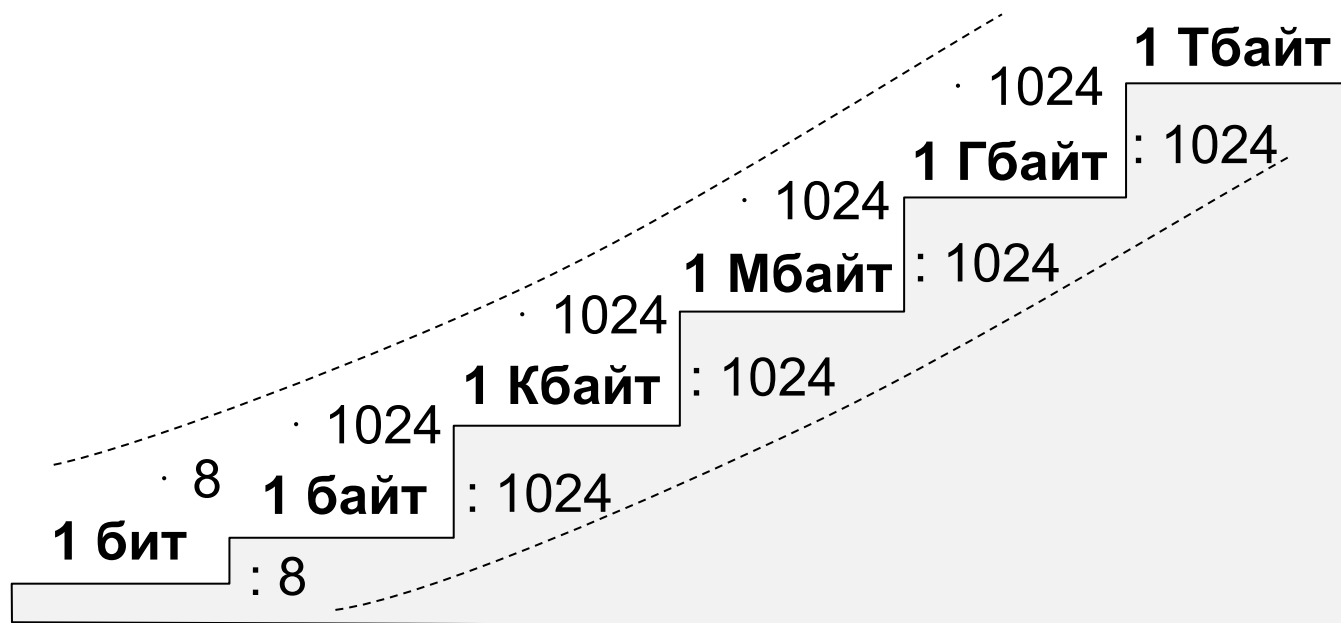
Сколько байтов

*в 16 битах?
в 32 битах?
в 64 битах?
в 4 битах?*

Сколько байтов

*в 1 Кбайте?
в 2 Кбайтах?
в 3 Кбайтах?
в 4 Кбайтах?
в 5 Кбайтах?
в 1/2 Кбайта?
в 1/4 Кбайта?
в 1/8 Кбайта?*

Перевод количества информации в другие единицы



«Лесенка» показывает действия, которые нужно выполнить для перевода количества информации из одних единиц в другие.

Выводы:

- Кодирование – это представление информации в форме, удобной для её хранения, передачи и обработки.
- Двоичный код – это код, в котором используются только два знака
- Код, в котором все кодовые слова имеют одинаковую длину, называется равномерным.
- Пиксель – это наименьший элемент цифрового рисунка, для которого можно задать свой цвет независимо от других.
- Растровый рисунок – это рисунок, который представлен в памяти как множество точек разного цвета (пикселей).
- Бит – это одна двоичная цифра.
- Файл – это набор данных, расположенный в долговременной памяти компьютера и имеющий имя.
- Размеры файлов измеряются в тех же единицах, что и объём памяти – в байтах, Кбайтах, Мбайтах и т.д.

Интеллект-карта
