

Представление информации

Петрова Ирина Александровна, учитель
информатики МОУ Лицей №6.



Цель работы

Изучить процесс и методы представления текстовой, числовой, графической и музыкальной информации.



Содержание

- Введение в тему кодирования
- Понятия
- Кодирование:
 - Текстов
 - Чисел
 - Изображений
 - Музыки



Кодирование - форма представления информации.

Компьютер может обрабатывать:

- числовую
- текстовую
- графическую
- звуковую информацию.



Информация кодируется последовательностью электрических импульсов:

есть импульс (1), нет импульса (0),
т.е. *последовательностью нулей и единиц.*

Такое кодирование называется *двоичным*,
а последовательность нулей и единиц — *машинным языком.*



Алфавит - набор знаков, в котором определен порядок.

Примеры некоторых алфавитов:

- алфавит русских букв (А, Б, В, Г, Д...);
- алфавит латинских букв (А, В, С, D, Е, F...);
- алфавит десятичных цифр (0, 1, 2, 3, 4, 5, 6, 7, 8, 9);

Двоичный алфавит - алфавит, состоящий только из двух знаков.

Примеры двоичных алфавитов:

- пара знаков «+» и «-»;
- пара знаков «точка» и «тире» (•, —);
- пара цифр (0, 1);
- пара ответов «да» и «нет».



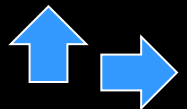
Представление и измерение текстовой информации

Для представления текстовой информации обычно используют алфавит из 256 различных символов:

- прописных и заглавных букв русского и латинского алфавита
 - цифр
 - знаков препинания
- математических и графических символов.

Следовательно, для двоичного кодирования 1 символа необходимо 8 бит информации или 8 двоичных разрядов, т.е. 1 байт.

1 символ = 8 бит = 1 байт

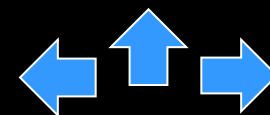


Каждому символу соответствует свой уникальный двоичный код
(последовательность из 8 нулей и единиц), например:

А	11100001	А	01000001
Б	11100010	В	01000010
В	11100011	С	01000011

Присвоение символу конкретного двоичного кода — это вопрос соглашения, которое фиксируется в *кодовой таблице*.

Во всем мире в качестве стандарта принята таблица *ASCII*
(American Standard Code for Information Interchange -
американский стандартный код для информационного обмена).



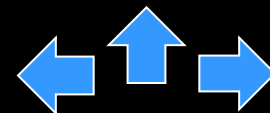
Все символы компьютерного алфавита пронумерованы от 0 до 255. *Стандартными в этой таблице являются только первые 128 символов:*

- буквы латинского алфавита
- цифры
- знаки препинания
- скобки и некоторые другие символы.

Латинские буквы располагаются в алфавитном порядке.
Расположение цифр также упорядочено по возрастанию.

Вторая половина кодовой страницы (с кодами от 128 до 255) не определена американскими стандартами и используется для кодировки букв национальных алфавитов, символов псевдографики и научных символов.

Пример



ЗАДАЧА:

Сколько дискет объемом 1,44 Мб надо для хранения энциклопедии из 60 томов по 500 страниц, на каждой из которых по 2 полосы из 80 строк по 45 символов?

РЕШЕНИЕ:

1 символ несет 1 байт информации, следовательно для хранения 1 страницы текста нужно $80 \text{ строк} * 45 \text{ символов} * 2 \text{ полосы} * 1 \text{ байт} = 7200 \text{ байт}$.

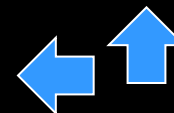
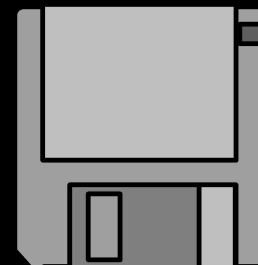
Следовательно, 1 том энциклопедии будет занимать $7200 \text{ байт} * 500 \text{ страниц} = 3600000 \text{ байт}$.

Вся же энциклопедия будет занимать

$3600000 \text{ байт} * 60 \text{ томов} = 216000000 \text{ байт}$.

Поскольку 1 дискета имеет объем 1,44 Мб=1509949,44 байт, то окончательно получаем, электронная энциклопедия будет занимать

$216000000 \text{ байт} : 1509949,44 \text{ байт} = 144 \text{ дискеты}$

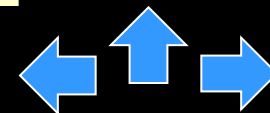


На разных моделях компьютеров, в разных операционных системах
могут использоваться

и разные варианты второй половины кодовой таблицы.

К сожалению, существует 5 различных кодировок русских букв,
поэтому тексты, созданные в одной кодировке, не будут
правильно отображаться в другой:

двоичный код	КОИ-8 UNIX	CP1251 Microsoft Windows	CP866 MS DOS	MAC Apple	ISO8859-5 ISO
11100100	Д	Д	Ф	Д	Ф
11100101	Е	е	Х	е	Х

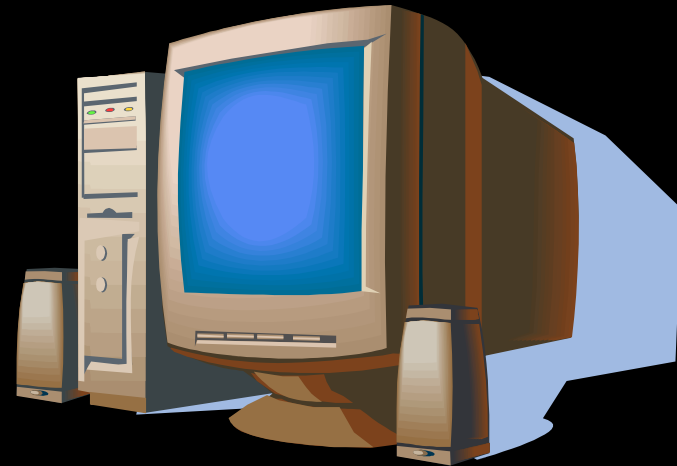


Появился новый международный стандарт Unicode,
который отводит на каждый символ 2 байта.

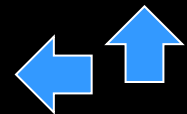
Это позволяет включить в код символа информацию о том,
какому языку принадлежит символ и как его надо воспроизводить.

Эту кодировку поддерживают

- *Windows 98*
- пакеты *Microsoft Office 97*
- *Microsoft Office 2000*



В них используются все известные миру символы.



ПРЕДСТАВЛЕНИЕ ЧИСЛОВОЙ ИНФОРМАЦИИ

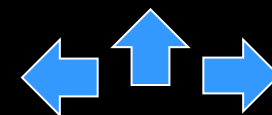
Целые числа могут занимать в памяти компьютера один, два или четыре байта.

Целые числа без знака в двухбайтовом формате (16 двоичных знаков) могут принимать значения: от 0 до 65 535 ($2^{16} - 1$), а со знаком: от -32768 (-2^{15}) до +32767 ($+2^{15}$).



Знак «+» кодируется как «0»,
знак «-» кодируется как «1».

Пример



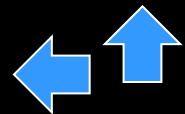
ЗАДАЧА:

представить десятичное число (-18) в двухбайтовом формате.

РЕШЕНИЕ:

1. Переведем модуль числа - (18_{10}) в двоичную систему счисления:
 $18_{10} = 10010_2$.
2. Запишем в старший (15) разряд знак числа — 1.
3. Полученные цифры запишем в разрядную сетку, начиная с младшего разряда.
4. Недостающие цифры заполним нулями.

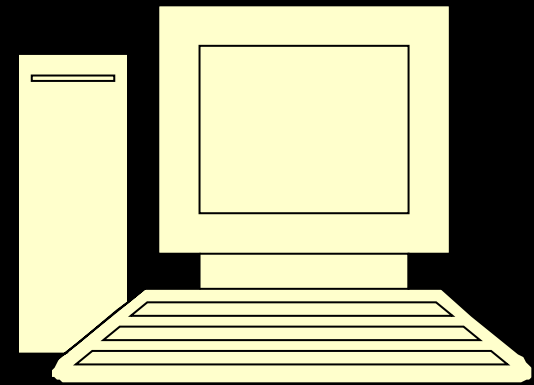
15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0



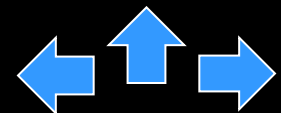
Вещественные числа в компьютере могут занимать 4, 6, 8 или 10 байт.

Вещественные числа представляются в компьютере в форме:

- с фиксированной точкой, например, 24.43;
- с плавающей точкой, например, $0.2543 \cdot 10^2$; $2.543 \cdot 10^1$; $254.3 \cdot 10^{-1}$



С плавающей точкой очень удобно представлять числа, очень близкие к нулю или очень большие по абсолютной величине.



Любое число с плавающей точкой можно представить в виде:

$$N = m * q^p,$$

где m - мантисса, представляющая собой правильную дробь ($0.1 < m < 1$);

q - основание системы счисления;

p - порядок числа, указывающий положение запятой в числе.

Например, $25.324 = 0.25324 * 10^2$.

Пример:

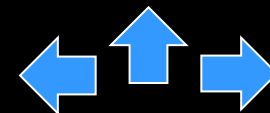
$$n = -100,1_2 = -0,1001 * 2^{11}$$

$$m = -0.1001$$

$$q = 2$$

$$p = 11 (3_{10} = 11^2).$$

Данное число также можно представить в четырёхбайтовом формате (32 бита).

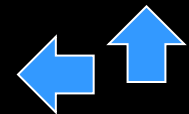


На мантиссу отводится 23 бита, поэтому максимальная величина мантиссы равна $2^{23} - 1 = 8\,388\,607$, то есть 7 десятичных цифр.

Компьютер при вычислениях отбрасывает лишние цифры в мантиссе, поэтому все вычисления с вещественными числами всегда выполняются приближенно (с ошибкой).



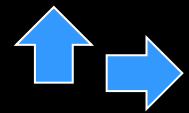
Вещественные числа обрабатываются в компьютере медленнее, чем целые.



ПРЕДСТАВЛЕНИЕ И ИЗМЕРЕНИЕ ГРАФИЧЕСКОЙ ИНФОРМАЦИИ

Растровое представление

- Графическая информация на экране монитора представляется в виде изображения, которое формируется из точек (*пикселей*).
- В видеопамяти компьютера хранится *битовая карта*, являющаяся двоичным кодом изображения, которая считывается процессором (не реже 50 раз в секунду) и отображается на экран.



Растр - прямоугольная сетка пикселей на экране монитора.

Компьютерные программы

Paintbrush; Paint;
PhotoFinish;
Adobe Photoshop

Пример:

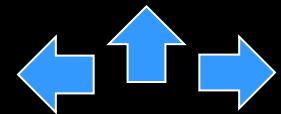


Форматы файлов

bmp; psx; gif; tiff;
tga; img; jpeg

Применение

в художественной графике (рисунки, фотографии...),
в реставрационных работах...

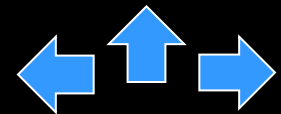


Качество изменяется при увеличении или уменьшении
размеров изображения.

Цветные изображения имеют
большой объем (до нескольких Мбайт).

Для хранения растровых изображений используют *сжатие*, т.е.
уменьшение размера файла.

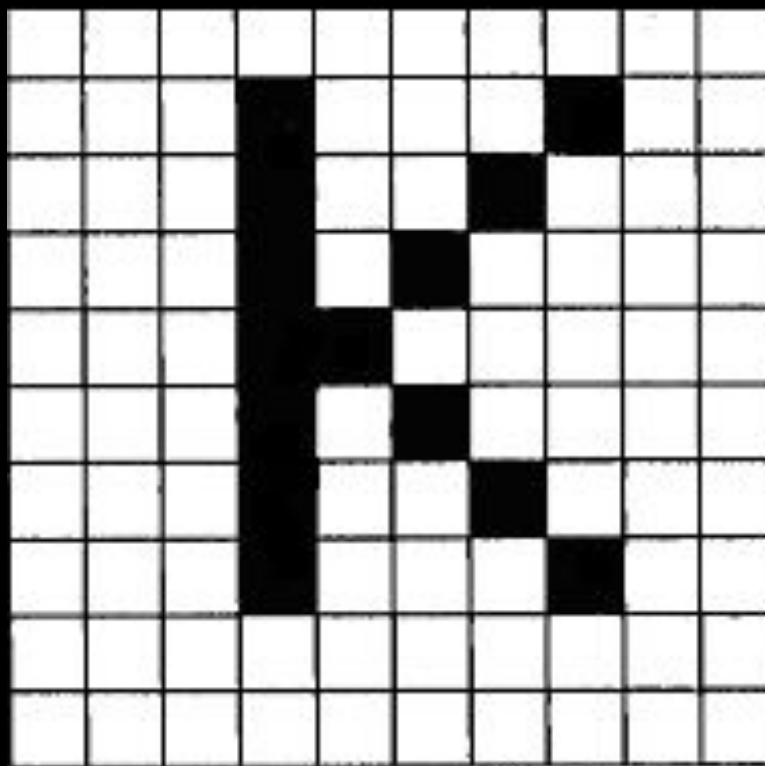
Наиболее распространенный *метод*
сжатия JPEG позволяет уменьшать размеры
файла *в несколько десятков раз*.



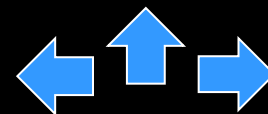
Пример: Рассмотрим растровое черно-белое изображение размером 10x10 пикселей.

Для кодирования изображения потребуется 100 бит видеопамяти. Так будет выглядеть битовая карта:

```
0 0 0 0 0 0 0 0 0 0
0 0 0 1 0 0 0 1 0 0
0 0 0 1 0 0 1 0 0 0
0 0 0 1 0 1 0 0 0 0
0 0 0 1 1 0 0 0 0 0
0 0 0 1 0 1 0 0 0 0
0 0 0 1 0 0 1 0 0 0
0 0 0 1 0 0 0 1 0 0
0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0
```



"1" - точка светится; "0" - точка не светится

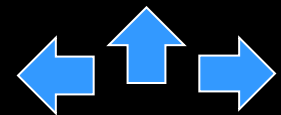


Красный (R), зеленый (G) и синий (B) - основные цвета; при их смешивании можно получить любой цвет.

Каждый цвет имеет свой уникальный двоичный код!

Код цвета пикселя содержит информацию о доле каждого базового цвета. Если все три составляющие имеют одинаковую интенсивность (яркость), то из их сочетания можно получить 8 различных цветов (2^3).

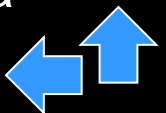
Двоичный код восьмицветной палитры



Двоичный код восьмицветной палитры

R	G	B	Цвет
0	0	0	черный
0	0	1	синий
0	1	0	зеленый
0	1	1	голубой
1	0	0	красный
1	0	1	розовый
1	1	0	коричневый
1	1	1	белый

"1" - наличие базового цвета; "0" - отсутствие базового цвета

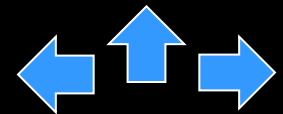


Палитра из 16 цветов получается при использовании 4-х разрядной кодировки пикселя: к трем битам базовых цветов добавляется 1 бит интенсивности (он управляет яркостью трех цветов одновременно).

Например:

- 0100 — красный;
- 1100 — ярко-красный;
- 0110 — коричневый;
- 1110 — ярко-коричневый (желтый).

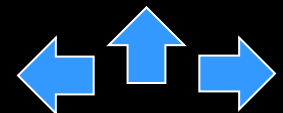
Таким образом, количество цветов **K** и количество битов для их кодировки **b** связаны формулой: **$K=2^b$** .



Количество цветов **K** и количество битов для их кодировки **b**
связаны формулой: **$K=2^b$**

Число цветов, воспроизводимых на экране монитора (**K**),
и число бит, отводимых в видеопамяти под каждый пиксель:

Кодирование цветов	1 пиксель
2 (черно-белое)	1 бит
16	4 бит
256	8 бит
65 536 (high color)	16 бит
16777216 (true color)	24 бита



ЗАДАЧА:

Пусть видеопамять компьютера имеет объем 512 Кбайт. Разрешение монитора 800*600 пикселей. Сколько страниц экрана разместится в видеопамяти при палитре из 256 цветов?

РЕШЕНИЕ:

Из формулы $K=2^b$, где K - количество цветов,

b - количество битов для их кодировки, получаем, что $256=2^b$.

Следовательно, $b=8$ бит – требуется на кодирование одного пикселя экрана.

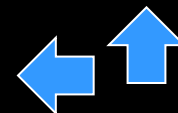
Тогда одна страница экрана будет занимать

$$800*600*8=3840000 \text{ бит.}$$

$$3840000 \text{ бит} : 8 : 1024 = 468,75 \text{ Кбайт}$$

Значит в видеопамяти компьютера разместится

$$512 : 468,75 \approx 1 \text{ страница экрана.}$$



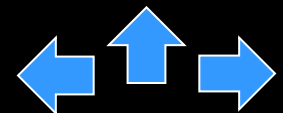
Векторное представление

Графическое изображение на экране монитора состоит из объектов - линий, прямоугольников, окружностей, дуг, закрасок и пр., которые называются графическими примитивами.

Качество изображения не меняется при увеличении или уменьшении размеров изображения.

Положение и форма графических примитивов задаются в системе координат, связанной с экраном (начало координат расположено в верхнем левом углу, ось X направлена слева направо, ось Y - сверху вниз).

Растровая сетка (пикселей) совпадает с координатной сеткой.



Представление и измерение звуковой информации



Сигнал - способ передачи информации.

Аналоговый (непрерывный)

(могут изменяться
в любой момент
времени)

заранее определенные

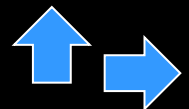
- речь человека
- влажность воздуха
- пение птиц
- кардиограмма

Дискретный (цифровой)

(могут изменяться только
в определенный момент
времени и принимать

значения)

- сигналы светофора
- телеграфная азбука Морзе
- сигналы, несущие текстовую и символную информацию

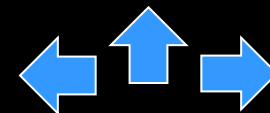


При двоичном кодировании аналогового звукового сигнала непрерывный сигнал *дискретизируется (оцифровывается)*, т.е. заменяется последовательностью отдельных во времени отсчетов этого сигнала (*выборок*).

После такого преобразования непрерывный сигнал представляется последовательностью чисел: 4-5-3-5-2-3-1-2-3-6-5, которые затем с помощью *звуковой платы* преобразуются в *числовой двоичный код*: 100-101-011--101-010-011-001-010-011-110-101.

Качество преобразования непрерывного сигнала в цифровой зависит от двух параметров:

- разрешения звуковой карты;
- частоты дискретизации.



Разрешение (R) - это количество распознаваемых дискретных уровней сигналов.

Чем больше уровней сигналов, тем больше разрядов у двоичных чисел.

Различные звуковые карты могут обеспечить 8- или 16-разрядные выборки.

Частота дискретизации (v) - это количество выборок в секунду. Одно измерение в секунду соответствует частоте 1 Гц (Герц). 1000 измерений в секунду - 1 кГц.

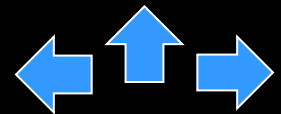
Частота, с которой происходит выборка сигналов, может принимать значения: 5,5 кГц, 11 кГц, 22,05 кГц, 44,1 кГц.

Качество звука в дискретной форме может быть *очень плохим* (при 8 битах и 5,5 кГц) и *очень высоким* (при 16 битах и 44,1 кГц), так же как радиотрансляция и аудиоCD.

Формула для расчета размера цифрового моноаудиофайла:

$$N = V * t * R.$$

Пример



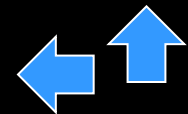
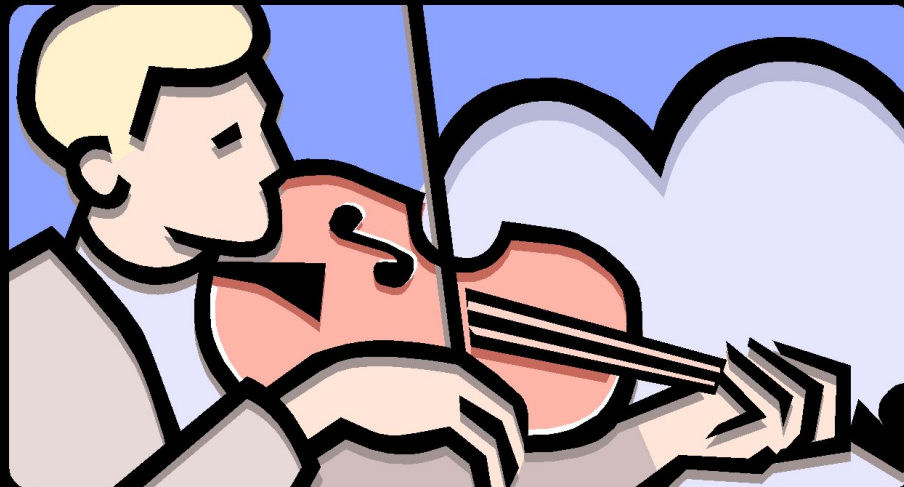
ЗАДАЧА:

Определите объем памяти для хранения цифрового моноаудио - файла, время звучания которого составляет 2 минуты при частоте дискретизации 44,1 КГц и разрешении звуковой карты 16 бит.

РЕШЕНИЕ:

$$N = V * t * R = 44100 \text{ Гц} * 120 \text{ с} * 16 \text{ бит} = 84672000 \text{ бит}$$

$\approx 10 \text{ Мбайт}$ требуется для хранения данного моноаудиофайла.



Литература

- Петцольд Ч. - Код. - М.: Издательско-торговый дом «Русская Редакция», 2001. - 512с.:ил.
- Информатика. Базовый курс. 7-9 классы-М.: Лаборатория Базовых Знаний, 2001 - 384с.:ил.
- Шафрин Ю.А. - Информационные технологии в 2ч., Ч.1 - Основы информатики и информационных технологий - М.:Бином, Лаборатория знаний, 2003 - 320с.
- БЭКМ - 2004.

