

Кодирование информации



Кодирование информации в компьютере

- ▶ Современный компьютер может обрабатывать числовую, текстовую, графическую, звуковую и видео информацию. Все эти виды информации в компьютере представлены в двоичном коде, т. е. используется алфавит мощностью два (всего два символа 0 и 1). Связано это с тем, что удобно представлять информацию в виде последовательности электрических импульсов: импульс отсутствует (0), импульс есть (1). Такое кодирование принято называть двоичным, а сами логические последовательности нулей и единиц - машинным языком.

Кодирование текстовой информации

- ▶ В текстах мы используем: прописные и строчные русские буквы Аа Бб Вв ... прописные и строчные латинские буквы Аa Bb Cc ... знаки препинания !, ?. ... цифры 1 2 3 ... знаки арифметических операций + - ... другие символы ([\ ... Достаточно 256 различных символов, следовательно нужно 256 различных двоичных кодов для кодирования текстовой информации Именно эта идея положена в основу двоичного кодирования текстовой информации Компьютеры могут обрабатывать информацию, представленную только в числовой форме. При вводе информации в память компьютера каждый символ русского и латинского алфавита, цифра, знак пунктуации или знак действия – кодируется определенной последовательностью двоичных цифр.

Кодовые таблицы.

- ▶ Соответствие символов и кодов задается с помощью специальных кодовых таблиц. В кодовых таблицах каждому символу ставится в соответствие уникальная цепочка из восьми нулей и единиц.

Международный стандарт.

- ▶ В качестве международного стандарта принята кодовая таблица ASCII (American Standard Code for Information Interchange). В этой таблице представлены коды от 33 до 127 (буквы английского алфавита, знаки математических операций, служебные символы и т.д.), коды от 0 до 32 отведены функциональным клавишам. Коды с 128 по 255 выделены для национальных стандартов каждой страны. В настоящее время для кодировки русских букв используют пять различных кодовых таблиц альтернативная кодовая таблица CP-866 (OS MS DOS 6.22 для IBM PC) международный стандарт ISO 8859 (OS UNIX на компьютерах DEC и SUN) кодовая таблица фирмы Microsoft CP-1251 (кодировка Windows) кодовая таблица, применяемая в KOI 8-(в большинстве OS UNIX) Mac (Macintosh, Apple) причем тексты, закодированные при помощи одной таблицы не будут правильно отображаться в другой кодировке. Наглядно это можно представить в виде фрагмента объединенной таблицы кодировки символов.

Unicode

- ▶ В мире существует примерно 6800 различных языков. Чтобы буквы любой страны можно было читать на любом компьютере, для их кодировки стали использовать два байта (16 бит), что позволяет кодировать 65536 символов. Такая кодировка называется Unicode и обозначается как UCS-2. Этот код включает в себя все существующие алфавиты мира, а также множество математических, музыкальных, химических символов и многое другое. Существует кодировка и UCS-4, где для кодирования используют 4 байта, то есть можно кодировать более 4 млрд. символов.

Графическая информация.

- ▶ Графическую информацию можно представлять в двух формах: аналоговой или дискретной. Живописное полотно, цвет которого изменяется непрерывно - это пример аналогового представления, а изображение, напечатанное при помощи струйного принтера и состоящее из отдельных точек разного цвета - это дискретное представление. Графическая информация из аналоговой формы в дискретную преобразуется путем дискретизации т. е. разбиения непрерывного графического изображения на отдельные элементы, каждому из которых присваивается определенный код. Качество кодирования зависит от размера точки, чем меньше размер точки, тем большее количество точек составляет изображение. От количества цветов, чем больше цветов может принимать точка, тем качественнее изображение. Существуют два основных способа кодирования графической информации: векторный и растровый.

РАСТОВЫЙ

- ▶ Расчет объема растрового изображения Растровый формат характеризуется тем, что все изображение по вертикали и горизонтали разбивается на достаточно мелкие прямоугольники - так называемые элементы изображения, или пиксели (минимальная единица изображения). Разрешающая способность – количество точек по горизонтали и вертикали на единицу длины изображения. Чем меньше размер точки, тем больше разрешающая способность и тем выше качество. Разрешающая способность выражается в dpi (кол. Точек на дюйм 1дюйм=2,54 см) Глубина цвета определяет количество битов, или разрядов, с помощью которых составляются коды потенциальных значений тона или цвета. $N=2^I$ где N – количество цветов палитры I – глубина цвета (количество битов используемых для кодирования информации)

Системы кодирования граф. информации.

- ▶ Применяют несколько систем кодирования: HSB, RGB и CMYK. Первая цветовая модель проста и интуитивно понятна, т. е. удобна для человека, вторая наиболее удобна для компьютера, а последняя модель CMYK-для типографий. Модель HSB характеризуется тремя компонентами: оттенок цвета (Hue), насыщенность цвета (Saturation) и яркость цвета (Brightness). Можно получить большое количество произвольных цветов, регулируя эти компоненты. Эту цветовую модель лучше применять в тех графических редакторах, в которых изображения создают сами, а не обрабатывают уже готовые.

Векторный.

- ▶ При векторном, рисунок представляется в виде комбинации простых геометрических фигур - точек, отрезков прямых и кривых, окружностей, прямоугольников и т. п. При этом для полного описания рисунка необходимо знать вид и базовые координаты каждой фигуры, например, координаты двух концов отрезка, координаты центра и диаметр окружности и т. д. Этот способ кодирования идеально подходит для рисунков, которые легко представить в виде комбинации простейших фигур, например, для технических чертежей

Кодирование звука.

- ▶ Кодирование звука Звук представляет собой звуковую волну с непрерывно меняющейся амплитудой и частотой. Чем больше амплитуда сигнала, тем он громче для человека, чем больше частота сигнала, тем выше тон. В процессе кодирования непрерывного звукового сигнала, производится его дискретизация по времени, непрерывная звуковая волна разбивается на отдельные маленькие временные участки, причем для каждого такого участка устанавливается определенная величина амплитуды. Качество двоичного кодирования звука определяется глубиной кодирования звука и частотой дискретизации (Частота дискретизации - это количество измерений громкости звука в единицу времени). Глубина звука – это количество информации, которое необходимо для кодирования дискретных уровней громкости цифрового звука $N=2^I$ где N – количество уровней звука I – глубина кодирования звука