Кодирование информации

Кодирование числовой информации

Кодирование числовой информации осуществляется с помощью систем счисления.

Существуют позиционные и непозиционные системы счисления.

В позиционных системах счисления вес каждой цифры изменяется в зависимости от её положения (позиции) в последовательности цифр, изображающих число. Например, в числе 757,7 первая семёрка означает 7 сотен, вторая – 7 единиц, а третья – 7 десятых долей единицы.

Любая позиционная система счисления характеризуется своим основанием. За основание системы можно принять любое натуральное число – два, три, четыре и т.д. Следовательно, возможно бесчисленное множество позиционных систем: двоичная, троичная, четверичная и т.д.

В непозиционных системах счисления вес цифры (т.е. тот вклад, который она вносит в значение числа) не зависит от её позиции в записи числа. Так, в римской системе счисления в числе XXXII (тридцать два) вес цифры X в любой позиции равен десяти.

Существует несколько форм записи римских чисел: классическая, 3 варианта более наглядной записи и упрощённый вариант.

Так, например, число 999 в классической записи имеет вид CMXCIX, более наглядно его можно записать, как LMVLIV или XMIX или VMIV, а в упрощённом виде – IM.

Кодирование текстовой информации

В 60-х годах XX века в *американском национальном институте стандартизации (ANSI)* была разработана таблица кодирования символов, которая впоследствии была использована во всех операционных системах. Эта таблица называется **ASCII (American Standard Code for Information Interchange – американский стандартный код для обмена информацией)**. Чуть позже появилась **расширенная версия ASCII**.

В соответствие с таблицей кодирования ASCII для представления одного символа выделяется 1 байт (8 бит). Набор из 8 ячеек может принять 2^8 = 256 различных значений. Первые 128 значений (от 0 до 127) постоянны и формируют так называемую основную часть таблицы, куда входят десятичные цифры, буквы латинского алфавита (заглавные и строчные), знаки препинания (точка, запятая, скобки и др.), а также пробел и различные служебные символы (табуляция, перевод строки и др.). Значения от 128 до 255 формируют дополнительную часть таблицы, где принято кодировать символы национальных алфавитов.

Поскольку национальных алфавитов огромное множество, то расширенные ASCII-таблицы существуют во множестве вариантов. Даже для русского языка существует несколько таблиц кодирования (распространены Windows-1251 и Koi8-r). Всё это создает дополнительные трудности.

Есть и другая проблема. В алфавитах некоторых языков слишком много символов и они не помещаются в отведённые им позиции со 128 до 255 однобайтовой кодировки.

Третья проблема – что делать, если в тексте используется несколько языков (например, русский, английский и французский)? Нельзя же использовать две таблицы сразу…

Чтобы решить эти проблемы одним разом для кодирования на компьютере всех возможных алфавитов в начале 90-х был разработан стандарт кодирования символов, получивший название **Unicode**. Данный стандарт позволяет использовать в тексте почти любые языки и символы.

В Unicode для кодирования символов предоставляется 31 бит (4 байта за вычетом одного бита). Количество возможных комбинаций даёт запредельное число: 2^{31} = 2 147 483 684 (т.е. более двух миллиардов). Поэтому Unicode описывает алфавиты всех известных языков, даже «мёртвых» и выдуманных, включает многие математические и иные специальные символы. Однако информационная ёмкость 31-битового Unicode всё равно остаётся слишком большой. Поэтому чаще используется сокращённая 16-битовая версия (2^{16} = 65 536 значений), где кодируются все современные алфавиты.

В Unicode первые 128 кодов совпадают с таблицей ASCII.

Кодирование звуковой информации

При преобразовании звуковой информации в цифровую форму её подвергают **дискретизации** и **квантованию**. Дискретизация заключается в замерах величины аналогового сигнала огромное множество раз в секунду. Полученной величине аналогового сигнала сопоставляется определённое значение из заранее выделенного диапазона: 256 (8 бит) или 65536 (16 бит). Установление соответствия уровня сигнала с определённой величиной диапазона и есть квантование.

*Например: Для записи качественной музыки аналоговый звуковой сигнал измеряют более 44 000 раз в секунду и квантуют 2 байтами (16 бит даёт диапазон из 65536 значений). Т.е. за одну секунду записывается 88 000 байт информации. Это равно (88 000 / 1024) примерно 86 Кбайт. Минута обойдётся уже в 5168 Кбайт (86*60), что немного больше 5 Мб.*

Итак, чтобы найти объём звукового файла (моно аудиофайла), используют формулу:

$$\boldsymbol{I = L \cdot N \cdot H}, \text{ где}$$

L – это длительность звучания (в секундах);

N – битовая глубина (т.е. сколько бит используют для квантования);

H – частота дискретизации (т.е. сколько раз в секунду измеряют звуковой сигнал) – выражается в герцах.

Если требуется найти объём стерео аудиофайла, то значение предыдущей формулы надо умножить на 2, т.к. стереозапись идёт по двум дорожкам (каналам).

Кодирование графической информации

При кодировании графической информации изображение так же подвергается дискретизации и квантованию.

Дискретизация – это разделение пространства на фиксированные по размеру области (точки, которые точками, по сути, не являются) или отрезки. Так описываемое двумерное изображение разбивается на маленькие плоскости. **В пределах каждой такой плоскости характеристики изображения считаются одинаковыми.** Основными характеристиками являются цвет и яркость.

Разделение непрерывного ряда значений какой-либо характеристики на ограниченное количество диапазонов называют **квантованием**. В компьютере сохраняется лишь номер диапазона, в который попало конкретное значение свойства.

Если при дискретизации разделяется пространство, то при квантовании этому подвергаются возможные значения свойств.

Изображение на экране компьютера представляет картину, состоящую из массива точек, имеющих такие атрибуты как координаты и цвет.

Пиксель – наименьший элемент изображения на экране компьютера. Размер экранного пикселя приблизительно равен 0,0018 дюйма.

Качество изображения определяется разрешающей способностью экрана и глубиной цвета.

Число цветов (***K***), воспроизводимых на экране дисплея, зависит от числа бит (***N***), отводимых в видеопамяти под каждый пиксель: $K = 2^N$. Величину ***N*** называют *битовой глубиной*.

Для получения богатой палитры цветов базовым цветам могут быть заданы различные интенсивности. Например, при глубине цвета в 24 бита на каждый из цветов выделяется по 8 бит (RGB), т.е. для каждого из цветов возможны $K = 2^8 = 256$ уровней интенсивности. Один бит видеопамяти занимает информация об одном пикселе на чёрно-белом экране (без полутонов).

Страница – раздел видеопамяти, вмещающий информацию об одном образе экрана (одной «картинке» на экране). В видеопамяти одновременно могут размещаться несколько страниц.

Пример: *Если на экране с разрешающей способностью 800 × 600 высвечиваются только двухцветные изображения, то битовая глубина двухцветного изображения равна 1, а объём видеопамяти на одну страницу изображения равен 800 × 600 × 1 = 480000 бит = 60000 байт.*

Для хранения двух страниц такого изображения полученное значение надо умножить на 2.

Итак, объём видеопамяти находится по формуле:

$$I = x \cdot y \cdot N, \text{ где}$$

x – это ширина экрана;

y – высота экрана;

N – битовая глубина.

Если изображение состоит из нескольких страниц, то полученное значение надо умножить на количество страниц.