**Двоичные числа; двоичное кодирование текста в памяти компьютера. Информационный объем текста.**

**Кодирование информации**

Представление информации происходит в различных формах в процессе восприятия окружающей среды живыми организмами и человеком, в процессах обмена информацией между человеком и человеком, человеком и компьютером, компьютером и компьютером и так далее. Преобразование информации из одной формы представления (знаковой системы) в другую называется **кодированием.**

Средством кодирования служит таблица соответствия знаковых систем, которая устанавливает взаимно однозначное соответствие между знаками или группами знаков двух различных знаковых систем.

В процессе обмена информацией часто приходится производить операции кодирования и декодирования информации. При вводе знака алфавита в компьютер путем нажатия соответствующей клавиши на клавиатуре происходит кодирование знака, то есть преобразование его в компьютерный код. При выводе знака на экран монитора или принтер происходит обратный процесс - декодирование, когда из компьютерного кода знак преобразуется в его графическое изображение.

**Двоичное кодирование информации**

В компьютере для представления информации используется двоичное кодирование, так как удалось создать надежно работающие технические устройства, которые могут со стопроцентной надежностью сохранять и распознавать не более двух различных состояний (цифр).

Все виды информации в компьютере кодируются на машинном языке, в виде логических последовательностей нулей и единиц.

Цифры двоичного кода можно рассматривать как два равновероятных состояния (события). При записи двоичной цифры реализуется выбор одного из двух возможных состояний (одной из двух цифр) и, следовательно, она несет количество информации, равное 1 биту.

Даже сама единица измерения количества информации бит (bit) получила свое название от английского словосочетания BInary digiT (двоичная цифра).

Важно, что каждая цифра машинного двоичного кода несет информацию в 1 бит. Таким образом, две цифры несут информацию в 2 бита, три цифры — в 3 бита и так далее. Количество информации в битах равно количеству цифр двоичного машинного кода.

**Представление числовой информации с помощью систем счисления**

Для записи информации о количестве объектов используются числа. Числа записываются с использованием особых знаковых систем, которые называются системами счисления. Алфавит систем счисления состоит из символов, которые называются цифрами. Например, в десятичной системе счисления числа записываются с помощью десяти всем хорошо известных цифр: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

Все системы счисления делятся на две большие группы: позиционные и непозиционные системы счисления. В позиционных системах счисления значение цифры зависит от ее положения в числе, а в непозиционных — не зависит.

*Римская непозиционная система счисления.* Самой распространенной из непозиционных систем счисления является римская. В качестве цифр в ней используются: I (1), V (5), X (10), L (50), С (100), D (500), М (1000).
Значение цифры не зависит от ее положения в числе. Например, в числе XXX (30) цифра X встречается трижды и в каждом случае обозначает одну и ту же величину - число 10, три числа по 10 в сумме дают 30.

*Позиционные системы счисления.*

В позиционных системах счисления количественное значение цифры зависит от ее позиции в числе: в числе 15 цифра «5» означает 5 единиц, а в числе 51 означает 50, т.е. 5 десятков.

**Двоичное кодирование текстовой информации**

Начиная с конца 60-х годов, компьютеры все больше стали использоваться для обработки текстовой информации и в настоящее время большая часть персональных компьютеров в мире (и наибольшее время) занято обработкой именно текстовой информации.

Традиционно для кодирования одного символа используется количество информации, равное 1 байту, то есть 8 битов.

*Для кодирования одного символа требуется 1 байт информации.*

Если рассматривать символы как возможные события, то можно вычислить, какое количество различных символов можно закодировать:
N = 28 = 256.

Такое количество символов вполне достаточно для представления текстовой информации, включая прописные и строчные буквы русского и латинского алфавита, цифры, знаки, графические символы и пр.

Кодирование заключается в том, что каждому символу ставится в соответствие уникальный десятичный код от 0 до 255 или соответствующий ему двоичный код от 00000000 до 11111111. Таким образом, человек различает символы по их начертаниям, а компьютер - по их кодам.

При вводе в компьютер текстовой информации происходит ее двоичное кодирование, изображение символа преобразуется в его двоичный код. Пользователь нажимает на клавиатуре клавишу с символом, и в компьютер поступает определенная последовательность из восьми электрических импульсов (двоичный код символа). Код символа хранится в оперативной памяти компьютера, где занимает один байт.

В процессе вывода символа на экран компьютера производится обратный процесс — декодирование, то есть преобразование кода символа в его изображение.

Важно, что присвоение символу конкретного кода — это вопрос соглашения, которое фиксируется в кодовой таблице. Первые 33 кода (с 0 по 32) соответствуют не символам, а операциям (перевод строки, ввод пробела и так далее).
Коды с 33 по 127 являются интернациональными и соответствуют символам латинского алфавита, цифрам, знакам арифметических операций и знакам препинания.
Коды с 128 по 255 являются национальными, то есть в национальных кодировках одному и тому же коду соответствуют различные символы.

К сожалению, в настоящее время существуют пять различных кодовых таблиц для русских букв (КОИ8, СР1251, СР866, Mac, ISO), поэтому тексты, созданные в одной кодировке, не будут правильно отображаться в другой.

В настоящее время широкое распространение получил новый международный стандарт Unicode, который отводит на каждый символ не один байт, а два, поэтому с его помощью можно закодировать не 256 символов, а N = 216 = 65536 различных символов.

*С развитием IBM PC международным стандартом стала таблица кодов ASCII:*



Сегодня очень многие люди для подготовки писем, документов, статей, книг и пр. используют компьютерные текстовые редакторы. Компьютерные редакторы, в основном, работают с алфавитом размером 256 символов.

В этом случае легко подсчитать объем информации в тексте. Если 1 символ алфавита несет 1 байт информации, то надо просто сосчитать количество символов, полученное число даст информационный объем текста в байтах.

Пусть небольшая книжка, сделанная с помощью компьютера, содержит 150 страниц; на каждой странице - 40 строк, в каждой строке - 60 символов. Значит, страница содержит 40 х 60 = 2400 байт информации. Объем всей информации в книге: 2400 х 150 = 360 000 байт.

*Обратите внимание!* Цифры кодируются по стандарту ASCII в двух случаях - при вводе-выводе и когда они встречаются в тексте. Если цифры участвуют в вычислениях, то осуществляется их преобразование в другой двоичный код.

Возьмем число 57.

При использовании в тексте каждая буква будет представлена своим кодом в соответсвии с таблицей ASCII. В двоичной системе - это 00110101 00110111.

При использовании в вычислениях, код этого числа будет получен по правилам перевода в двоичную систему и получим - 00111001.