

Урок 2. Представление информации в ПК

Кодирование информации

На прошлом уроке мы с вами изучили понятие «информация» и ее свойства. Информация во всех сферах предполагает создание, передачу, обработку и хранение. Все эти процессы называются информационными. Когда человек или какой-либо другой живой организм или какое-то устройство участвуют в информационном процессе, то все они представляют информацию в той или иной форме. Мы с вами на прошлом уроке представляли информацию в различных формах.

Когда мы представляем информацию в разных формах или преобразуем ее из одной формы в другую, мы информацию кодируем.

Код – это система условных знаков для представления информации.

Кодирование – операция преобразования символов или группы символов одного кода в символы или группы символов другого кода.

С помощью двух цифр 0 и 1 можно закодировать любое сообщение. Это явилось причиной того, что в компьютере обязательно должно быть организовано два важных процесса:

Кодирование – преобразование входной информации в форму, воспринимаемую компьютером, т.е. двоичный код.

Декодирование – преобразование данных из двоичного кода в форму, понятную человеку.

С точки зрения технической реализации использование двоичной системы счисления для кодирования информации оказалось намного более простым, чем применение других способов. Действительно, удобно кодировать информацию в виде последовательности нулей и единиц, если представить эти значения как два возможных устойчивых состояния электронного элемента: 0 – отсутствие электрического сигнала; 1 – наличие электрического сигнала. Эти состояния легко различать. Недостаток двоичного кодирования – длинные коды. Но в технике легче иметь дело с большим количеством простых элементов, чем с небольшим числом сложных. Вам приходится постоянно сталкиваться с устройством, которое может находиться только в двух устойчивых состояниях: включено/выключено. Конечно же, это хорошо знакомый всем выключатель. А вот придумать выключатель, который мог бы устойчиво и быстро переключаться в любое из 10 состояний, оказалось невозможным. В результате после ряда неудачных попыток разработчики пришли к выводу о невозможности построения компьютера на основе десятичной системы счисления. И в основу представления чисел в компьютере была положена именно двоичная система счисления. Способы кодирования и декодирования информации в компьютере, в первую очередь, зависят от вида информации, а именно, что должно кодироваться: числа, текст, графические изображения или звук. Рассмотрим основные способы двоичного кодирования информации в компьютере.

Представление чисел

Для записи информации о количестве объектов используются числа. Числа записываются с использованием особых знаковых систем, которые называют системами счисления. Система счисления – совокупность приемов и правил записи чисел с помощью определенного набора символов. Все системы счисления делятся на две большие группы: **ПОЗИЦИОННЫЕ** и **НЕПОЗИЦИОННЫЕ**. **Позиционные** – количественное значение каждой цифры числа зависит от того, в каком месте (позиции или разряде) записана та или иная цифра. **Непозиционные** – количественное значение цифры числа не зависит от того, в каком месте (позиции или разряде) записана та или иная цифра. Самой распространенной из непозиционных систем счисления является римская. В качестве цифр используются: I(1), V(5), X(10), L(50), C(100), D(500), M(1000). Величина числа

определяется как сумма или разность цифр в числе. Первая позиционная система счисления была придумана еще в Древнем Вавилоне, причем вавилонская нумерация была шестидесятеричная, т.е. в ней использовалось шестьдесят цифр! В XIX веке довольно широкое распространение получила двенадцатеричная система счисления. В настоящее время наиболее распространены десятичная, двоичная, восьмеричная и шестнадцатеричная системы счисления. Количество различных символов, используемых для изображения числа в позиционных системах счисления, называется **основанием системы счисления**.

Двоичное кодирование текстовой информации

Начиная с 60-х годов, компьютеры все больше стали использовать для обработки текстовой информации и в настоящее время большая часть ПК в мире занята обработкой именно текстовой информации.

Традиционно для кодирования одного символа используется количество информации = 1 байту (1 байт = 8 битов).

Для кодирования одного символа требуется один байт информации.

Учитывая, что каждый бит принимает значение 1 или 0, получаем, что с помощью 1 байта можно закодировать 256 различных символов. ($2^8=256$)

Кодирование заключается в том, что каждому символу ставится в соответствие уникальный двоичный код от 00000000 до 11111111 (или десятичный код от 0 до 255).

Важно, что присвоение символу конкретного кода – это вопрос соглашения, которое фиксируется кодовой таблицей.

Таблица, в которой всем символам компьютерного алфавита поставлены в соответствие порядковые номера (коды), называется таблицей кодировки.

Для разных типов ЭВМ используются различные кодировки. С распространением IBM PC международным стандартом стала таблица кодировки ASCII (American Standard Code for Information Interchange) – Американский стандартный код для информационного обмена.

Стандартной в этой таблице является только первая половина, т.е. символы с номерами от 0 (00000000) до 127 (01111111). Сюда входят буква латинского алфавита, цифры, знаки препинания, скобки и некоторые другие символы.

Остальные 128 кодов используются в разных вариантах. В русских кодировках размещаются символы русского алфавита.

В настоящее время существует 5 разных кодовых таблиц для русских букв (КОИ8, CP1251, CP866, Mac, ISO).

В настоящее время получил широкое распространение новый международный стандарт Unicode, который отводит на каждый символ два байта. С его помощью можно закодировать 65536 ($2^{16}=65536$) различных символов.

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 цифра будет представлена своим кодом в соответствии 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.

Кодирование графической информации.

Создавать и хранить графические объекты в компьютере можно двумя способами – как растровое или как векторное изображение. Для каждого типа изображений используется свой способ кодирования. Растровое изображение представляет собой совокупность точек (пикселей) разных цветов. Для черно-белого изображения

информационный объем одной точки равен одному биту (либо черная, либо белая – либо 1, либо 0). Для четырех цветного – 2 бита. Для 8 цветов необходимо – 3 бита. Для 16 цветов – 4 бита. Для 256 цветов – 8 бит (1 байт). Цветное изображение на экране монитора формируется за счет смешивания трех базовых цветов: красного, зеленого, синего. Т.н. модель RGB. Для получения богатой палитры базовым цветам могут быть заданы различные интенсивности. 4 294 967 296 цветов (True Color) – 32 бита (4 байта).

Кодирование векторных изображений. Векторное изображение представляет собой совокупность графических примитивов (точка, отрезок, эллипс...). Каждый примитив описывается математическими формулами. Кодирование зависти от прикладной среды.

Кодирование звуковой информации.

Звук – волна с непрерывно изменяющейся амплитудой и частотой. Чем больше амплитуда, тем он громче для человека, чем больше частота, тем выше тон. В процессе кодирования звукового сигнала производится его временная дискретизация – непрерывная волна разбивается на отдельные маленькие временные участки. Качество двоичного кодирования звука определяется глубиной кодирования и частотой дискретизации.