



Кодирование информации. Особенности кодирования различных видов информации: текстовой, графической,

**ЛЕКТОР: доцент МАЙОРОВ
ЕВГЕНИЙ ЕВГЕНЬЕВИЧ**

- 
- **Кодирование информации** - это представление информации в той или иной стандартной форме.
 - Информация всегда хранится и передается в закодированном виде. При разговоре информация кодируется с помощью звуков, комбинации из которых образуют слова, а также с помощью жестов, мимики. При записи слова могут быть закодированы с помощью букв, числа с помощью цифр.

- 
- Одна и та же информация может быть закодирована в различных видах. Количество учеников в классе может быть закодировано в виде рисунка, диаграммы, буквенной или числовой записи. При этом сама информация остается неизменной, меняются лишь способы кодирования. Выбор способа кодирования информации зависит от целей кодирования.

- 
- Различные языки служат средством для кодирования информации. Человек в своей практике общения использует много различных языков. Прежде всего это языки устной и письменной речи. Это языки жестов и мимики.
 - Кроме того, человек использует ряд языков профессионального назначения. Сюда относятся языки математических формул, обозначений электроники

- 
- Возникновение целого ряда языков было продиктовано необходимостью привлечения технических средств для передачи информации. Примером такого языка является азбука Морзе, изобретенная для передачи телеграфных сообщений. В нем каждый символ обычного алфавита кодируется набором точек и тире (что соответствует передаче коротких и длинных электромагнитных импульсов).

- 
- Вообще использование двухсимвольного алфавита оказалось столь же естественным в различных технических средствах связи, как десятисимвольного для записи чисел. Дело в том, что технически двухсимвольный алфавит легко реализуется: есть электрический импульс или нет его, есть намагниченность или она отсутствует, проходит свет или не проходит

- 
- Так же, как 10 пальцев руки послужили основой для возникновения десятичной нумерации, различимость двух состояний той или иной технической системы легла в основу всех современных средств автоматической передачи информации.
 - Рассмотрим более подробно, как кодируются числа, тексты, музыка и графика на компьютерах

Кодирование числовой информации

- **Системой счисления** называется способ записи (кодирования) чисел. Все системы счисления можно разделить на две группы: позиционные и непозиционные.
- **Позиционной** называется такая система счисления, в которой величина цифры зависит от позиции (места), занимаемой этой цифрой в записи числа.

- 
- Если же величина цифры не зависит от места, занимаемого этой цифрой, то такая система счисления называется **непозиционной**. Непозиционные системы счисления первичны по своему происхождению; но поскольку они имеют ряд недостатков по сравнению с позиционными системами счисления, то постепенно они потеряли свое значение.

- Римская система счисления, где для обозначения цифр используются латинские буквы:

I	V	X	L	C	D	M
1	5	10	50	100	500	1000

- Числа в римской системе счисления записываются по определенным правилам. Вот они:
 - 1) если большая цифра стоит перед меньшей, они складываются, например:
 $VI=6$;

- 
- 2) если меньшая цифра стоит перед большей, то из большей вычитается меньшая, причем в этом случае меньшая цифра уже повторяться не может, например: XL=40, XXL-нельзя;
- 3) цифры M, C, X, I могут повторяться в записи числа не более трех раз подряд;
- 4) цифры L, D, V могут использоваться в записи числа только по одному разу.

- 
- Например число 1996 будет записано в римской система счисления как MCMXCVI.
 - Самое большое число, которое можно записать в этой системе счисления, это число 3999 (MMMCMXCIX). Для записи еще больших чисел пришлось бы вводить еще новые обозначения.

- 
- Недостатки непозиционных система счисления:
 - а) в них нельзя записывать сколь угодно большие числа;
 - б) запись чисел обычно громоздка и неудобна;
 - в) математические операции над ними крайне затруднены.
 - Именно поэтому, хотя римская и очень "красивая" система счисления, она не получила широкого распространения.

- Можно рассматривать системы счисления и с другими основаниями. Если p - основание системы счисления, то любое число N в этой системе счисления может быть представлено в виде

$$N = a(n) \cdot p^n + a(n-1) \cdot p^{(n-1)} + \dots + a(1) \cdot p + a(0),$$

где коэффициенты - цифры p -ичной системы счисления

- 
- Использование двух символьного алфавита оказалось очень удобным в различных технических средствах связи, т.к. технически двух символьный алфавит легко реализуется. Поэтому самое широкое распространение получила двоичная система счисления.
 - Вопреки распространенному заблуждению, двоичная система счисления была придумана не инженерами - конструкторами ЭВМ, а математиками и философами задолго до появления компьютеров, еще в XVII - XIX веках.

- 
- Великий немецкий ученый Лейбниц (1646 - 1716) считал:

“Вычисление с помощью двоек ... является для науки основным и порождает новые открытия При сведении чисел к простым началам, каковы 0 и 1, везде появляется чудесный порядок”.

- Позже двоичная система была забыта, и только в 1936-1938 гг. американский инженер и математик Клод Шеннон нашел замечательные применения двоичной системы при конструировании электронных схем.

Кодирование текстовой информации

- В большинстве современных компьютеров для хранения символа отводится 8-разрядная ячейка (1байт). В байт можно записать 256 различных чисел, что позволяет закодировать 256 разных символов. Соответствие символов и их кодов задается специальными таблицами. Существует несколько систем кодировки, т.е. несколько различных таблиц соответствий.

- Самая распространенная система кодировки – ASCII, что означает *американский стандартный код для обмена информацией*. Этот стандарт закрепляет за первыми 128 кодами латинские буквы, цифры и различные специальные знаки (запятая, точка, скобки, знаки арифметических действий и др.). Значение же остальных 128 кодов зависит от того, какой язык используется при работе с компьютером.

Кодирование графической информации

- Если очень внимательно рассмотреть рисунки или фотографии, напечатанные в газете или книге, то можно увидеть, что они состоят из отдельных точек. Дело в том, что человеческий глаз изображение, составленное из большого числа мелких точек, воспринимает как непрерывное.

- Разобьем картинку вертикальными и горизонтальными линиями на маленькие прямоугольники. Полученный двумерный массив прямоугольников называется растром, а сами прямоугольники - пикселями (это слово произошло от английского picture's element - элемент картинки). Теперь, чтобы закодировать изображение, надо закодировать числами цвета каждого пиксела.

- Чем меньше прямоугольники, тем точнее будет закодировано наше изображение. Информация о графическом изображении хранится в специальном разделе оперативной памяти компьютера, который называется видеопамятью. В видеопамати содержится информация о состоянии каждого пиксела экрана

- 
- Если каждый пиксел может принимать только два состояния: светится - не светится (белый - черный), то для кодирования одного пиксела достаточно одного бита памяти (1 - белый, 0 - черный). Если надо закодировать большее количество состояний пиксела (различную яркость свечения или различные цвета), то одного бита на пиксел будет недостаточно.

- 
- Цвет точки на экране формируется из трех основных цветов: красного, синего, зеленого. Различные цвета получаются в результате наложения цветовых пятен, возникающих под действием лучей трех электронных пушек. Например, сиреневый цвет получается путем наложения красного и синего пятен, желтый цвет - красного и зеленого пятен. Количество цветов увеличивается, если имеется возможность управлять яркостью основных цветов.

- Например, оранжевый цвет тоже получается из красного и зеленого, но другой яркости. Поэтому цвет пиксела можно закодировать тремя числами - яркостью его красной, зеленой и синей составляющей. Этот способ кодирования называется RGB - по первым буквам английских слов Red, Green, Blue - красный, зеленый, синий.

- Обычно для построения качественного цветного изображения достаточно яркость каждого из цветов разбить на 16 градаций (уровней). Тогда для кодирования яркости одного цвета потребуется 4 бита информации, а для кодирования цвета точки $4 \times 3 = 12$ битов.

Кодирование звуков

- Из курса физики вы знаете, что звук представляет собой колебания воздуха. Амплитуда этого колебания непрерывно меняется во времени. По своей природе звук является непрерывным (аналоговым) сигналом. Для кодирования звука этот аналоговый сигнал превращают в дискретный, а затем кодируют с помощью нулей и единиц.

- 
- Этот процесс называется дискредитацией. С помощью микрофона звук превращают в колебания электрического тока. Аналогово-цифровой преобразователь (АЦП) измеряет электрическое напряжение в каком-то диапазоне и выдает ответ в виде многоразрядного двоичного числа.

- 
- Воспроизведение закодированного таким образом звука производится при помощи цифро-аналогового преобразователя (ЦАП). Полученный на выходе ЦАП ступенчатый сигнал сначала сглаживается при помощи аналогового фильтра, а затем преобразуется в звук при помощи усилителя и динамика.

Виды информации и способы представления ее в компьютере.

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

Вид информации	Двоичный код
Числовая	
Текстовая	
Графическая	
Звуковая	
Видео	

- 
- Информация в компьютере представлена в двоичном коде, алфавит которого состоит из двух цифр (0 и 1)
 - Каждая цифра машинного двоичного кода несет количество информации, равное 1 бит.
 - Например. Латинская буква A представлена в двоичном коде – 01000001.
 - Русская буква А представлена в двоичном коде - 10000000.
 - 0 - 00110000
 - 1 – 00110001