**Тема 1.3. Представление, передача и хранение информации**

**Информационные процессы**

Человек постоянно участвует в разнообразных процессах дома, на работе, на улице, в общественных местах.

**Под процессом понимают ход, развитие какого-нибудь явления, последовательную смену состояния объекта.**

Одни процессы характерны для общества, другие — для живой природы.

В одних ситуациях человек активно участвует в процессе: например, школьник в процессе обучения, водитель — при управлении автомобилем, рабочий — при строительстве дома и т. д.

В других случаях он пассивен и занимает позицию наблюдателя: например, во время грозы, во время экскурсии, на просмотре спектакля, у экрана телевизора.

Особую роль среди всего существующего разнообразия процессов занимает процесс, называемый информационным.

**Информационный процесс — это процесс сбора (приёма), передачи (обмена), хранения, обработки (преобразования) информации.**

Информационные процессы протекают в человеческом обществе, в растительном и животном мире. С помощью органов чувств люди воспринимают информацию, осмысливают её и на основании своего опыта, имеющихся знаний, интуиции принимают определённые решения. Эти решения воплощаются в реальные действия, которые воздействуют на окружающий мир.

В середине XX века многократно увеличилась интенсивность информационных процессов. Лавинообразный поток информации, хлынувший на человека, уже не воспринимается в полном объёме; ориентироваться в нем становится всё труднее и труднее.

Подчас оказывается проще создать новый материальный или интеллектуальный продукт, нежели разыскать сделанный ранее аналог.

Для более эффективного участия в информационных процессах человек создавал и создаёт различные устройства, которые помогают ему воспринимать, преобразовывать, хранить и использовать информацию.

Основные типы информационных процессов

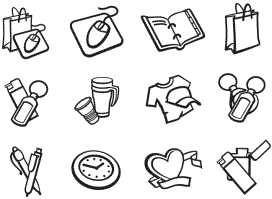
|  |  |
| --- | --- |
| **Получение** информации | Это сбор сведений из каких либо источников (извлечение данных из хранилища/источника данных, наблюдение за событиями и явлениями, общение, СМИ и масс-медиа). Получение информации основано на отражении различных свойств процессов, объектов и явлений окружающей среды. Этот процесс выражается в восприятии с помощью органов чувств. Для улучшения восприятия информации человек придумал различные индивидуальные приспособления и приборы — очки, бинокль, микроскоп, стетоскоп, различные датчики и т. д. |
| **Хранение**информации | Хранение информации имеет большое значение для многократного использования информации и передачи информации во времени. Для долговременного хранения используются книги, в настоящее время — компьютерные носители, устройства внешней памяти и др. Информация чаще всего хранится для неоднократной дальнейшей работы с ней. В этом случае для ускорения поиска информация должна быть как-то упорядочена. В библиотеках — это картотеки, при хранении с использованием компьютера — размещение информации в определенных папках, в более сложных случаях — это базы данных, информационно-поисковые системы и т. д. |
| **Обработка** информации | Обработка информации подразумевает преобразование ее к виду, отличному от исходной формы или содержания информации. Процесс изменения информации может включать в себя, например, такие действия как численные расчёты, редактирование, упорядочивание, обобщение, систематизация и т. д. Результаты обработки информации в дальнейшем используются в тех или иных целях, например: получение новой информации из уже известной путем логических рассуждений или математических вычислений (например, решение геометрической задачи); изменение формы представления информации без изменения ее содержания (например, перевод текста с одного языка на другой); упорядочение (сортировка) информации (например, упорядочение расписания движения поездов по времени их отправления). |
| **Передача**информации | Передача информации необходима для её распространения. Основными устройствами для быстрой передачи информации на большие расстояния в настоящее время являются телеграф, радио, телефон, телевизионный передатчик, телекоммуникационные сети на базе вычислительных систем. Такие средства связи принято называть каналами передачи информации. Следует отметить, что в процессе передачи информации, она может искажаться или теряться. Это происходит в тех случаях, когда информационные каналы плохого качества или на линии связи присутствуют помехи. Передача информации — это всегда двусторонний процесс, в котором есть источник и есть приемник информации. Источник передает информацию, а приемник её получает. |

**Представление информации.**

Полученную информацию человек должен зафиксировать каким-нибудь образом в материальной форме для сохранения и передачи другому человеку. Для этого использует разнообразные знаки.

**Знак представляет собой заменитель предмета, явления, действия, свойства или отношения.**

Если форма знака обозначает понятие или даже целое сообщение, то знаки называют **пиктограммами** (рис. 1). Пиктограммы пришли к нам от первобытных людей. Пиктографическая символика часто используется и в наше время. Например, дорожные знаки.

  
Рис. 1

Если связь между формой и значением знака устанавливается по доверенности, то такие знаки называют **символами** (рис. 2). Например, при письменной форме человек использует буквы, цифры, точки, запятые или другие знаки.

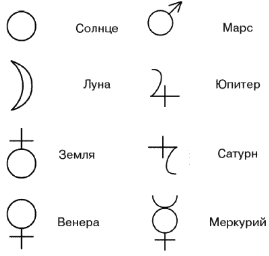


Рис. 2

Люди используют не только знаки, но и **знаковые системы**. Например, такой знаковой системой является **язык**, который человек использует для выражения своих мыслей, при общении с другими людьми.

Общение между людьми проходит в **устной** или **письменной форме**при использовании различных знаков.

**Устная речь** складывается у нас из звуковых знаков, которые называются **фонемами**. Из них складываются слоги, затем слова, а в конечном результате — фразы.

**Письменность** складывается с помощью отдельной буквы или сочетанием букв. Такая письменность называется **буквенно-звуковой**.

Также ещё встречаются **слоговая** и **идеографическая** письменность. **Слоговая письменность** используется в Японии (один значок обозначает целый слог). А **идеографическая письменность** — в Китае (отдельные значки используются для обозначения слов).

Человеческая речь и письменность очень связаны со словом «язык». Существуют два типа языков:

* естественные языки;
* формальные языки.

К **естественным языкам**относят китайский, английский, русский, французский и другие языки.  То есть языки, которые используются для общения с людьми.

**Формальные языки**применяются специалистами в профессиональной деятельности. К ним относятся математические системы, химические символы, нотная грамота, языки программирования и многое другое.

Каждый человек представляет информацию разными способами: в **знаковой**или **образной** форме.

В **знаковой форме** человек воспринимает информацию в виде текста на естественном языке в устной или письменной форме, а также символов формального языка (ноты, числа, дорожные знаки и другие).

В образной **форме** человек воспринимает информацию в виде графической формы (чертежи, рисунки, диаграммы и т. д.).

**1. Представление текстовой информации в компьютере**

Нажатие любой алфавитно-цифровой клавиши на клавиатуре приводит к тому, что в компьютер посылается сигнал в виде двоичного числа, представляющего собой одно из значений кодовой таблицы.

**Кодовая таблица — это внутреннее представление символов в компьютере.**

В качестве стандарта долгое время использовалась таблица ASCII(American Standard Code for Informational Interchange — Американский стандартный код информационного обмена).

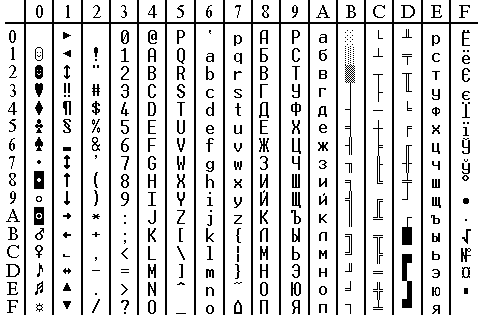
Для хранения двоичного кода одного символа выделен 1 байт = 8бит. Учитывая, что каждый бит принимает значение 1 или 0, количество возможных сочетаний единиц и нулей равно 28 = 256.

Значит, с помощью 1 байта можно получить 256 разных двоичных кодовых комбинаций и отобразить с их помощью 256 различных символов.

Эти коды и составляют таблицу ASCII.

Для сокращения записи и удобства пользования этими кодами символов в таблице используют шестнадцатеричную систему счисления, состоящую из 16 символов — 10 цифр и 6 латинских букв: A, B, C, D, E, F. При кодировании символов сначала записывается цифра столбца, а затем строки, на пересечении которых находится данный символ.

ASCII-коды



Например, латинская буква S в таблице ASCII представлена шестнадцатеричным кодом — 53. При нажатии клавиши с буквой S в память компьютера записывается код 01010011, представляющий собой двоичный эквивалент шестнадцатеричного числа 53. Этот код может быть получен путем замены каждой шестнадцатеричной цифры её двоичным представлением. В данном случае цифра 5 заменена кодом 0101, а цифра 3 — кодом 0011. При выводе буквы S на экран, компьютер выполняет декодирование: на основании этого двоичного кода строится изображение символа.

*Обрати внимание!*

Любой символ в таблице ASCII кодируется с помощью 8 двоичных разрядов или 2 шестнадцатеричных разрядов.

Стандарт ASCII кодирует первые 128символов от 0 до 127: цифры, буквы латинского алфавита, управляющие символы. Таблица выше отображает кодировку символов в шестнадцатеричной системе счисления.

Первые 32 символа являются управляющими и предназначены в основном для передачи команд управления. Их назначение может варьироваться в зависимости от программных и аппаратных средств. Вторая половина кодовой таблицы (от 128 до 255) американским стандартом не определена и предназначена для символов национальных алфавитов, псевдографических и некоторых математических символов. В разных странах могут использоваться различные варианты второй половины кодовой таблицы.

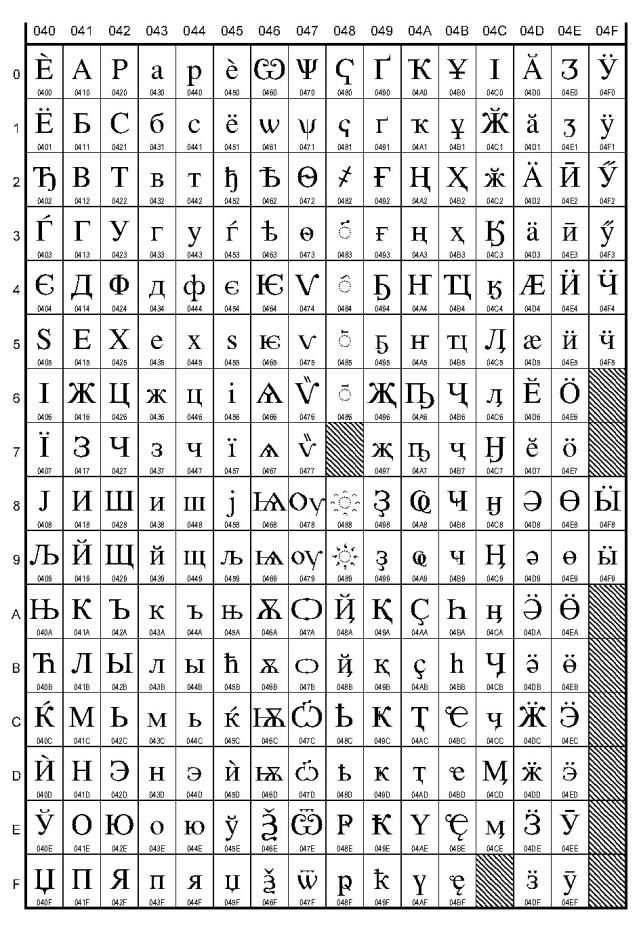
В настоящее время широко распространен код Unicode. Эта кодировка поддерживается в большинстве операционных систем, во всех современных браузерах и многих программах.

Стандарт Unicode явился результатом сотрудничества Международной организации по стандартизации (ISO) с ведущими производителями компьютеров и программного обеспечения. В мире существует 6700 живых языков, но только 50 из них являются официальными языками государств. Письменностей используется около 25, что делает возможным создание универсального стандарта.

Для кодирования этих письменностей достаточно 16-битового диапазона (2 байта на символ), то есть диапазона от 0000 до FFFF. Стандарт ASCII занимает в кодовом пространстве свое почетное место в диапазоне от 0000 до 00FF.

Каждой письменности выделен свой блок кодов. На сегодняшний день кодирование всех живых официальных письменностей считается завершенным: распределено около 29000 позиций из 65535 возможных.

Кодовая таблица Unicode



В последнее время консорциум Unicode приступил к кодированию остальных письменностей нашей планеты, которые представляют какой-либо интерес: письменности мёртвых языков, выпавших из современного обихода, китайские иероглифы, искусственно созданные алфавиты и т. п.

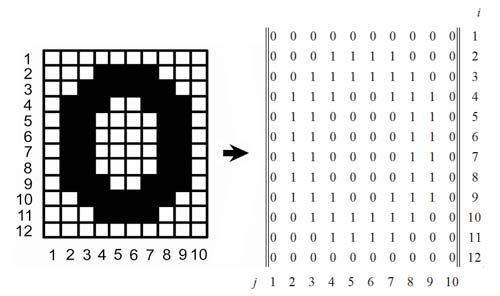
Для представления такого разнообразия языков 16-битового кодирования уже недостаточно, и сегодня Unicode уже приступил к освоению 21-битового пространства кодов (000000-10FFFF), которое разбито на 16 зон, названных плоскостями.

**2. Представление графической информации в компьютере**

Создавать и хранить графические объекты в компьютере можно двумя способами: как **растровое** или как **векторное** изображение. Для каждого типа изображения используется свой способ кодирования.

**Растровое изображение представляет собой совокупность точек, используемых для его отображения на экране монитора.**

Объём растрового изображения определяется как произведение количества точек и информационного объёма одной точки, который зависит от количества возможных цветов. Для черно-белого изображения информационный объём одной точки равен 1 биту, так как точка может быть либо чёрной, либо белой, что можно закодировать одной из двух цифр — 0 или 1.

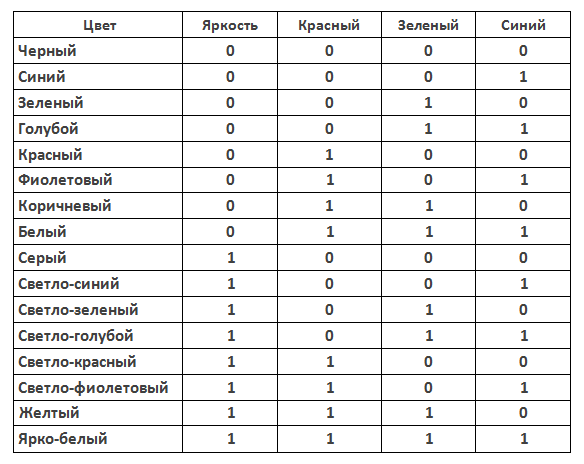


Информационный объём растрового изображения (V) определяется как произведение числа входящих в изображение точек (N) на информационный объём одной точки (q), который зависит от количества возможных цветов, т. е. V=N⋅q.

При чёрно-белом изображении q = 1 бит (например, 1 — точка подсвечивается и 0 — точка не подсвечивается). Поэтому для хранения чёрно-белого (без оттенков) изображения размером 100x100 точек требуется 10000бит.

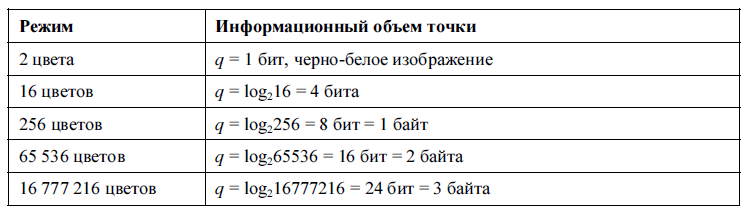
Если между чёрным и белым цветами имеется ещё шесть оттенков серого (всего 8), то информационный объём точки равен 3 бита (log28 = 3). Информационный объём такого изображения увеличивается в три раза: V = 30000бит.

Рассмотрим, сколько потребуется бит для отображения цветной точки: для 8 цветов необходимо 3 бита; для 16 цветов — 4 бита; для 256 цветов — 8 битов (1 байт). В таблице ниже представлено кодирование цветовой палитры из 16 цветов.



Разные цвета и их оттенки получаются за счёт наличия или отсутствия трёх основных цветов (красного, синего, зеленого) и степени их яркости. Каждая точка на экране кодируется с помощью 4 битов.

Цветные изображения могут отображаться в различных режимах, соответственно изменяется и информационный объём точки (см. таблицу ниже).



 Описание цвета пикселя является кодом цвета.

**Количество бит, отводимое на каждый пиксель для представления цвета, называют глубиной цвета (англ. color depth). От количества выделяемых бит зависит разнообразие палитры.**

Наиболее распространенными значениями глубины цвета являются 8, 16, 24 или 32 бита.

Чем больше глубина цвета, тем больше объем графического файла.

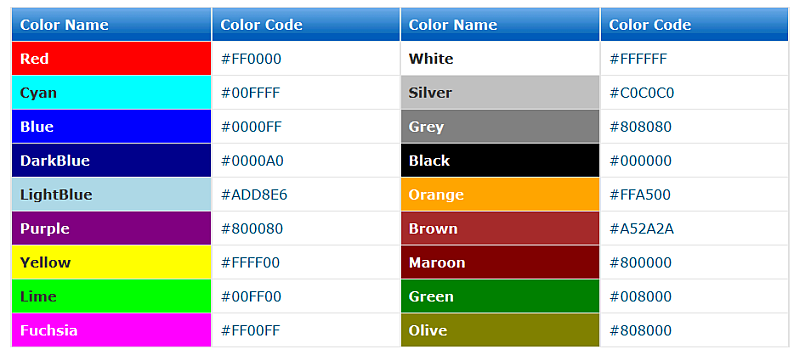
Для хранения растрового изображения размером 32x32пикселя отвели 512 байтов памяти.

Каково максимально возможное число цветов в палитре изображения?

**Решение**. Число точек изображения равно 32⋅32 = 1024. Мы знаем, что 512 байтов = 512⋅8=4096 бит. Найдём глубину цвета 4096÷1024=4. Число цветов равно 24 = 16.

Цвет на Web-страницах кодируется в виде **RGB**-кода в шестнадцатеричной системе: **#RRGGBB**, где **RR**, **GG** и **BB** — яркости красного, зеленого и синего, записанные в виде двух шестнадцатеричных цифр; это позволяет закодировать 256 значений от 0 (0016) до 255 (FF16) для каждой составляющей.

Коды некоторых цветов:



При обозначении цветов в HTML-документах вначале ставят знак номера **#**. В HTML: #FF0000 — интенсивно красный цвет, **#00FF00**— зелёный цвет, **#0000FF**— синий цвет. Отсутствие цветов (**#000000**) даёт чёрный цвет, а самое интенсивное сочетание всех трёх каналов (**#FFFFFF**) даёт белый цвет.

**FF** — наибольшая яркость цветовой компоненты, для получения различных оттенков одного и того же цвета изменяют яркость.

Чтобы получить светлый оттенок какого-то «чистого» цвета, нужно одинаково увеличить нулевые составляющие; например, чтобы получить светло-красный цвет, нужно сделать максимальной красную составляющую и, кроме этого, одинаково увеличить остальные — синюю и зелёную: **#FF9999**(сравните с красным: **#FF0000**).

Чтобы получить тёмный оттенок чистого цвета, нужно одинаково уменьшить все составляющие, например, **#660066**— это тёмно-фиолетовый цвет (сравните с фиолетовым **#FF00FF**).

Заметим, что если старший бит в коде (первая, третья или пятая цифра) находится в диапазоне от 0 до 3, то можно считать, что эта цветовая компонента отсутствует в цвете, то есть **#0F0F0F** — это чёрный цвет.

Также следует отметить, что равное или почти равное сочетание цветовых компонент обозначает серый цвет разной интенсивности.

**Векторное изображение представляет собой совокупность графических примитивов. Каждый примитив состоит из элементарных отрезков кривых, параметры которых (координаты узловых точек, радиус кривизны и пр.) описываются математическими формулами.**

Для каждой линии указываются её тип (сплошная, пунктирная, штрих-пунктирная), толщина и цвет, а замкнутые фигуры дополнительно характеризуются типом заливки.

Рассмотрим, например, такой графический примитив, как окружность радиуса r. Для её построения необходимо и достаточно следующих исходных данных:

- координаты центра окружности;  
- значение радиуса r;  
- цвет заполнения (если окружность не прозрачная);  
- цвет и толщина контура (в случае наличия контура).

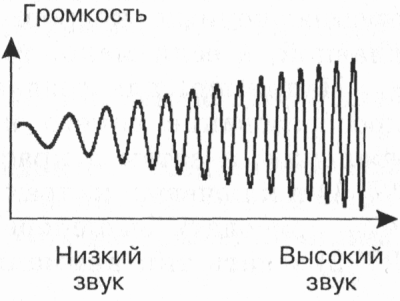
Информация о векторном рисунке кодируется обычным способом, как хранятся тексты, формулы, числа, т. е. хранится не графическое изображение, а только координаты и характеристики изображения его деталей. Поэтому для хранения векторных изображений требуется существенно меньше памяти, чем растровых изображений.

**3. Представление звуковой информации в компьютере**

Звук представляет собой непрерывный сигнал — звуковую волну с меняющейся амплитудой и частотой.

Чем больше амплитуда сигнала, тем он громче для человека.

Чем больше частота сигнала, тем выше тон.



Частота звуковой волны выражается числом колебаний в секунду и измеряется **в герцах** (Гц, Hz).

Человеческое ухо способно воспринимать звуки в диапазоне от 20 Гц до 20 кГц, который называют **звуковым**.

**Количество бит, отводимое на один звуковой сигнал, называют глубиной кодирования звука.**

Современные звуковые карты обеспечивают 16-, 32- или 64-битную глубину кодирования звука.

При кодировании звуковой информации непрерывный сигнал заменяется **дискретным**, то есть превращается в последовательность электрических импульсов (двоичных нулей и единиц).

**Процесс перевода звуковых сигналов от непрерывной формы представления к дискретной, цифровой форме называют оцифровкой.**

Важной характеристикой при кодировании звука является **частота дискретизации** — количество измерений уровней сигнала за 1секунду:

- 1 (одно) измерение в секунду соответствует частоте 1 Гц;

- 1000 измерений в секунду соответствует частоте 1 кГц.

**Частота дискретизации звука — это количество измерений громкости звука за одну секунду.**

Количество измерений может лежать в диапазоне от 8 кГц до 48 кГц (от частоты радиотрансляции до частоты, соответствующей качеству звучания музыкальных носителей).

Чем больше частота и глубина дискретизации звука, тем более качественным будет звучание оцифрованного звука. Самое низкое качество оцифрованного звука, соответствующее качеству телефонной связи, получается при частоте дискретизации 8000 раз в секунду, глубине дискретизации 8 битов и записи одной звуковой дорожки (режим «моно»). Самое высокое качество оцифрованного звука, соответствующее качеству аудио-CD, достигается при частоте дискретизации 48000раз в секунду, глубине дискретизации 16 битов и записи двух звуковых дорожек (режим «стерео»).

Необходимо помнить, что чем выше качество цифрового звука, тем больше информационный объем звукового файла.

Оценить информационный объём моноаудиофайла (V) можно следующим образом: V = N⋅f⋅k, где N — общая длительность звучания (секунд), f — частота дискретизации (Гц), k — глубина кодирования (бит).

Например, при длительности звучания 1 минуту и среднем качестве звука (16 бит, 24 кГц):

V = 60⋅24000⋅16 бит = 23040000 бит = 2880000 байт = 2812,5 Кбайт = 2,75 Мбайт.

При кодировании стереозвука процесс дискретизации производится отдельно и независимо для левого и правого каналов, что, соответственно, увеличивает объём звукового файла в два раза по сравнению с монозвуком.

Например, оценим информационный объём цифрового стереозвукового файла длительностью звучания 1секунда при среднем качестве звука (16 битов, 24000 измерений в секунду). Для этого глубину кодирования необходимо умножить на количество измерений в 1секунду и умножить на 2 (стереозвук):

V=16 бит ⋅24000⋅2 = 768000 бит = 96000 байт = 93,75 Кбайт.

Звуковые файлы имеют несколько форматов. Наиболее популярные из них **MIDI**, **WAV**, **МРЗ**.

Формат **MIDI** (Musical Instrument Digital Interface) изначально был предназначен для управления музыкальными инструментами. В настоящее время используется в области электронных музыкальных инструментов и компьютерных модулей синтеза.

Формат аудиофайла **WAV** (waveform) представляет произвольный звук в виде цифрового представления исходного звукового колебания или звуковой волны. Все стандартные звуки Windows имеют расширение **WAV**.

Формат **МРЗ** (MPEG-1 Audio Layer 3) — один из цифровых форматов хранения звуковой информации. Он обеспечивает более высокое качество кодирования.

**4. Представление видеоинформации в компьютере**

Когда говорят о видеозаписи, прежде всего имеют в виду движущееся изображение на экране телевизора или монитора.

Преобразование оптического изображения в последовательность электрических сигналов осуществляется видеокамерой. Эти сигналы несут информацию о яркости и цвете отдельных участков изображения. Они сохраняются на носителе в виде изменения намагниченности видеоленты (аналоговая форма) или в виде последовательности кодовых комбинаций электрических импульсов (цифровая форма).

**Процесс превращения непрерывного сигнала в набор кодовых слов называется аналого-цифровым преобразованием.**

Это сложный процесс, состоящий из:

- **дискретизации**, когда непрерывный сигнал заменяется последовательностью мгновенных значений через равные промежутки времени;

- **квантования**, когда величина каждого отсчёта заменяется округлённым значением ближайшего уровня;

- **кодирования**, когда каждому значению уровней квантования, полученных на предыдущем этапе, сопоставляются их порядковые номера в двоичном виде.

По своей сути видеофайл — это набор статичных изображений, меняющих друг друга с определенной частотой. Каждое статичное изображение является отдельным кадром видео. Это действительно так, если мы говорим о несжатом видео. Однако в таком формате никто не хранит фильмы.

Дело в том, что несжатое видео занимает на диске очень много места. Кадр видео формата **PAL** состоит из 720 точек по горизонтали и 576 по вертикали. То есть один кадр состоит из 414720 точек.

Для хранения цвета каждой точки в памяти отводится 24 бита (по 8 бит для каждой из составляющих RGB).

Следовательно, для хранения одного кадра понадобится 9953280 бит (или примерно 1,2 Мбайт).

То есть секунда несжатого видео в формате **PAL** будет занимать почти 30 Мбайт. А один час такого видео — более 100Гбайт.

Каким же образом полнометражный фильм (а то и несколько) умещается на одном компакт-диске или флеш-накопителе?

Дело в том, что, в основном, видео хранят в видеофайлах, в которых применены различные алгоритмы сжатия информации. Благодаря этим технологиям видеофайл можно сжимать в десятки и сотни раз практически без потери качества картинки и звука.

**AVI (Audio Video Interleave) — это контейнерный формат, что означает, что в нём могут содержаться аудио/видео, сжатые различными комбинациями кодирования.**

**AVI** файл может содержать различные виды компрессированных данных (например, **DivX** для видеоинформации и **MP3** для аудио), в зависимости от того, какой кодек используется для кодирования/декодирования. В файле с расширением **AVI** может храниться несжатое видео, видео в форматах **DV**, **MPEG-4**, **DivX**, **Xvid** и даже**MPEG-1** и **MPEG-2**. Кроме того, файл формата **AVI** может, например, содержать в себе только звук. То есть файлы формата **AVI** являются контейнером для хранения данных различного типа.

**Контейнер** — это файл с каким либо расширением, служащий для хранения в цифровом виде преобразованной аналоговой информации. Контейнер является файлом некоего стандарта, в котором одновременно может содержаться несколько различных типов информации.

**DivX** — технология видеозаписи, позволяющая создавать и просматривать медиа файлы с высокой степенью сжатия. Это технология сжатия, которая делает фильм в 8-12 раз меньшим с небольшой потерей качества.

**DivX** широко используется для сжатия компьютерных видеофайлов и файлов **DVD**, чтобы они помещались на стандартный CD.

**MPEG (Moving Picture Expert Group) — формат, предназначенный для сжатия звуковых и видеофайлов для загрузки или пересылки, например, через Интернет.**

Разработан Экспертной группой кинематографии, которая занимается разработкой стандартов кодирования и сжатия видео- и аудиоданных. Существуют разные стандарты **MPEG**: **MPEG-1**, **MPEG-2**, **MPEG-3**, **MPEG-4**, **MPEG-7**. Например, **MPEG-4** — стандарт, благодаря которому фильмы можно кодировать в хорошем качестве при низкой скорости передачи данных. Стандарт **MPEG-4** в основном используется для вещания (потоковое видео), записи фильмов на компакт-диски, видеотелефонии (видеотелефон) и широковещания, в которых активно используется сжатие цифровых видео и звука.