Дискретная форма представления информации. Единицы измерения информации.

Вся информация, которую обрабатывает компьютер должна быть представлена двоичным кодом с помощью двух цифр 0 и 1. Эти два символа принято называть двоичными цифрами или битами. С помощью двух цифр 0 и 1 можно закодировать любое сообщение. Это явилось причиной того, что в компьютере обязательно должно быть организованно два важных процесса: кодирование и декодирование.

Таблица 1.1. Базовая таблица кодировки ASCII

32	пробел	48	0	64	@	80	P	96	`	112	p
33	!	49	1	65	A	81	Q	97	a	113	q
34	"	50	2	66	B	82	R	98	b	114	r
35	#	51	3	67	C	83	S	99	c	115	s
36	$	52	4	68	D	84	T	100	d	116	t
37	%	53	5	69	E	85	U	101	e	117	u
38	&	54	6	70	F	86	V	102	f	118	v
39	'	55	7	71	G	87	W	103	g	119	w
40	(	56	8	72	H	88	X	104	h	120	x
41	)	57	9	73	I	89	Y	105	i	121	y
42	*	58	:	74	J	90	Z	106	j	122	z
43	+	59	;	75	K	91	[	107	k	123	{
44	,	60	<	76	L	92	\	108	l	124	|
45	-	61	=	77	M	93	]	109	m	125	}
46	.	62	>	78	N	94	^	110	n	126	~
47	/	63	?	79	O	95	_	111	o	127	

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

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

Азбука Морзе

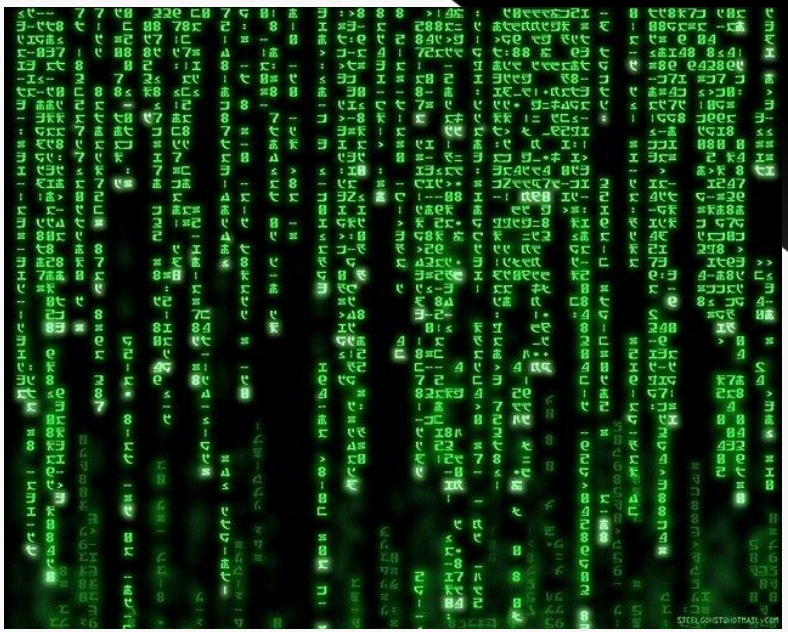
A •−	J •−−−	S •••	2 ••−−−
B −•••	K −•−	T −	3 •••−−
C −•−•	L •−••	U ••−	4 ••••−
D −••	M −−	V •••−	5 •••••
E •	N −•	W •−−	6 −••••
F ••−•	O −−−	X −••−	7 −−•••
G −−•	P •−−•	Y −•−−	8 −−−••
H ••••	Q −−•−	Z −−••	9 −−−−•
I ••	R •−•	1 •−−−−	0 −−−−−

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;

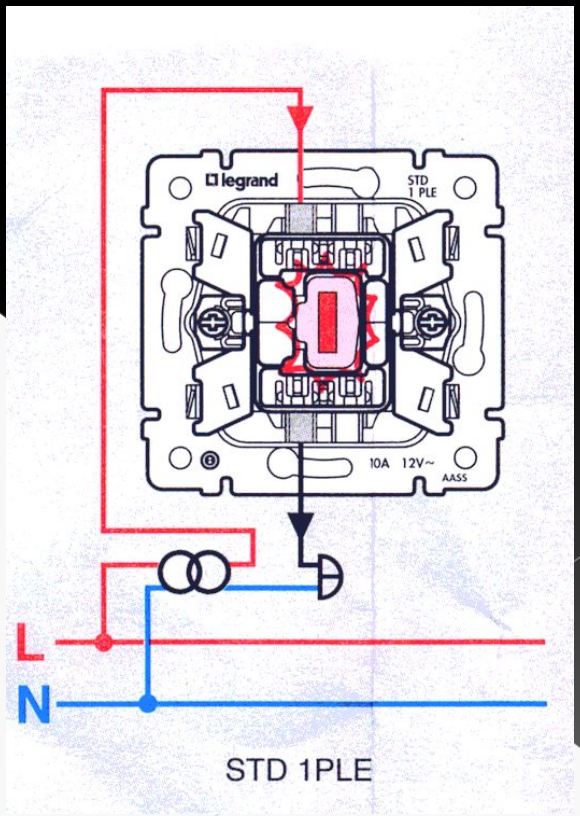
1 – наличие электрического сигнала.

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 находится 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.

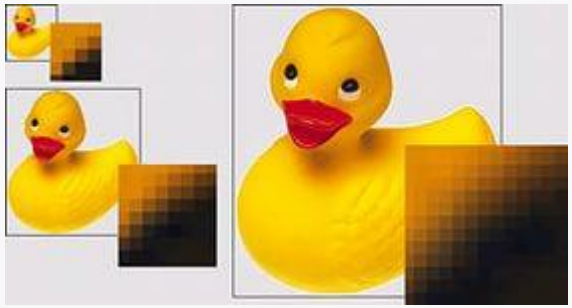
В результате после ряда неудачных попыток разработчики пришли к выводу о невозможности построения компьютера на основе десятичной системы счисления. И в основу представления чисел в компьютере была положена именно двоичная система счисления.

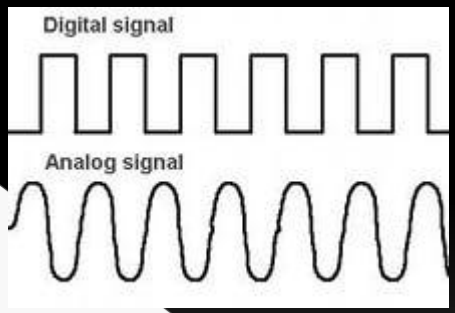


Способы кодирования и декодирования информации в компьютере, в первую очередь, зависит от вида информации, а именно, что должно кодироваться: числа, текст, графические изображения или звук.

Аналоговый и дискретный способ кодирования

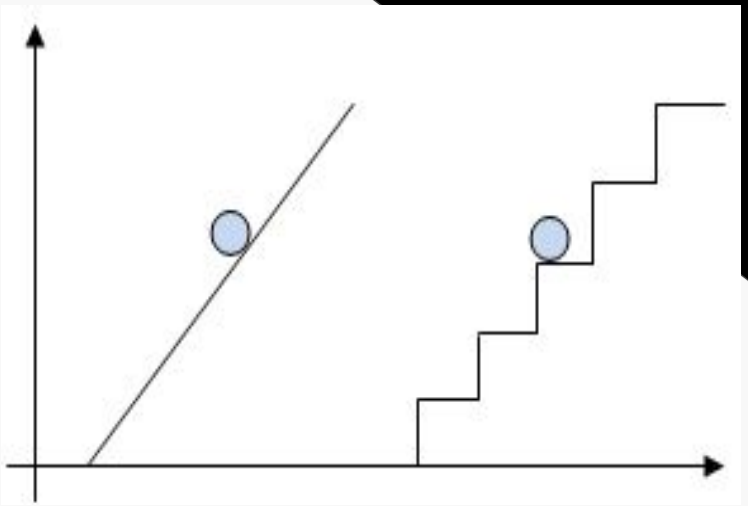
Информация, в том числе графическая и звуковая, может быть представлена в **аналоговой** или **дискретной** форме. При ***аналоговом*** представлении физическая величина принимает ***бесконечное множество значений***, причем её значения изменяются непрерывно. При ***дискретном*** представлении физическая величина принимает ***конечное множество значений***, причем её величина изменяется скачкообразно.





Человек способен воспринимать и хранить информацию в форме образов (зрительных, звуковых, осязательных, вкусовых и обонятельных). Зрительные образы могут быть сохранены в виде изображений (рисунков, фотографий и так далее), а звуковые — зафиксированы на пластинках, магнитных лентах, лазерных дисках и так далее.

Примеры



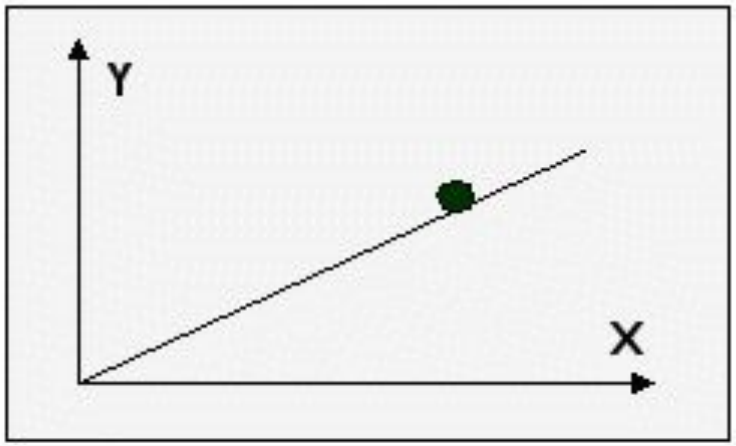
Приведем пример аналогового и дискретного представления информации. Положение тела на наклонной плоскости и на лестнице задается значениями координат X и У. При движении тела по наклонной плоскости его координаты могут принимать бесконечное множество непрерывно изменяющихся значений из определенного диапазона, а при движении по лестнице — только определенный набор значений, причем меняющихся скачкообразно.

Примером аналогового представления графической информации может служить, например, живописное полотно, цвет которого изменяется непрерывно, а дискретного — изображение, напечатанное с помощью струйного принтера и состоящее из отдельных точек разного цвета. Примером аналогового хранения звуковой информации является виниловая пластинка (звуковая дорожка изменяет свою форму непрерывно), а дискретного — аудиокомпакт-диск (звуковая дорожка которого содержит участки с различной отражающей способностью).

Способ представления

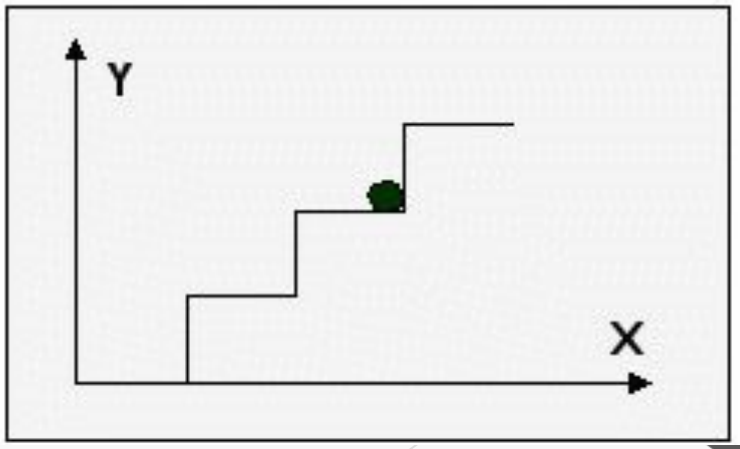
Аналоговый

Физическая величина может принимать **бесконечное** множество значений в пределах заданного интервала.



Дискретный

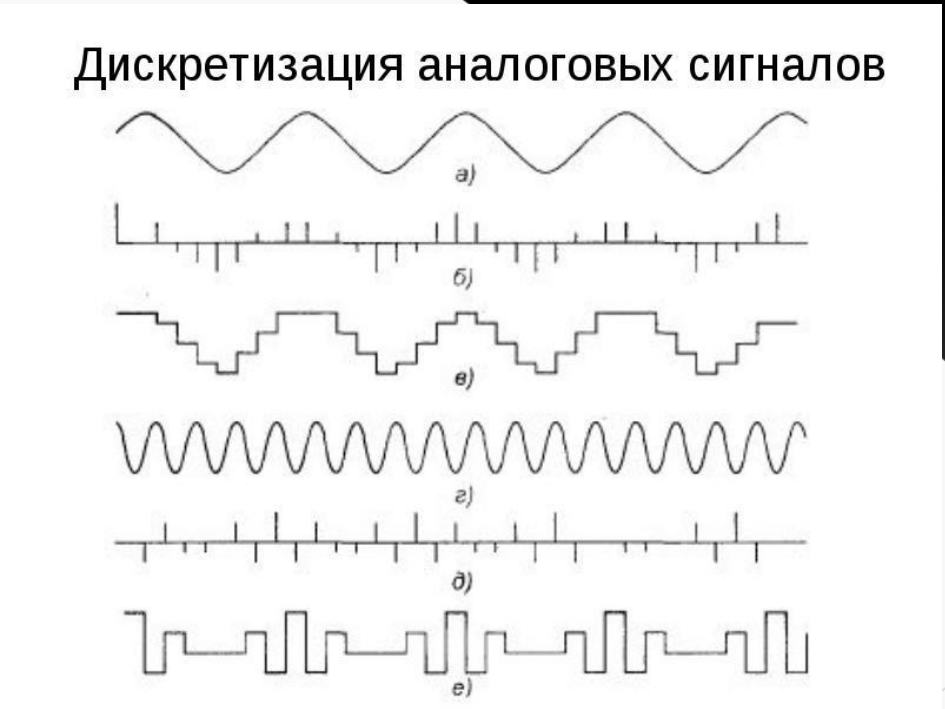
Физическая величина может принимать **конечное** множество значений в пределах заданного интервала. Эта величина изменяется скачкообразно.



Дискретизация

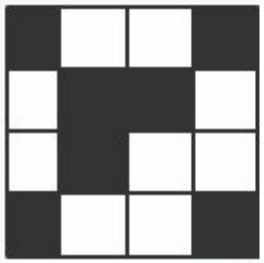
Преобразование графической и звуковой информации из аналоговой формы в дискретную производится путем дискретизации, то есть разбиения непрерывного графического изображения и непрерывного (аналогового) звукового сигнала на отдельные элементы. В процессе дискретизации производится кодирование, то есть присвоение каждому элементу конкретного значения в форме кода.

Дискретизация – это преобразование непрерывных изображений и звука в набор дискретных значений в форме кодов.

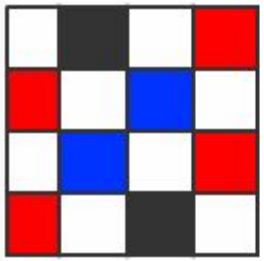


Кодирование изображений

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



1 0 0 1
0 1 1 0
0 1 0 0
1 0 0 1



00 11 00 01
01 00 10 00
00 10 00 01
01 00 11 00

Растровое изображение представляет собой совокупность точек (пикселей) разных цветов.

Пиксель - минимальный участок изображения, цвет которого можно задать независимым образом.

Кодирование изображений

В процессе
дискретизация
сравнить с постр
маленьких разноцветных стекол).
Изображение разбивается на отдельные маленькие фрагменты (точки), причем каждому фрагменту присваивается значение его цвета, то есть код цвета (красный, зеленый, синий и так далее).

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).

Качество изображения зависит от количества точек (чем меньше размер точки и, соответственно, больше их количество, тем лучше качество) и количества используемых цветов (чем больше цветов, тем качественнее кодируется изображение).

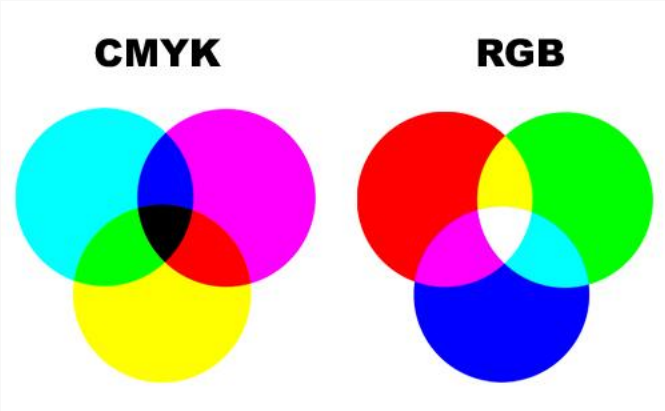
Кодирование изображений

изображения зависит от количества точек (чем меньше размер точки и, соответственно, больше их количество, тем лучше качество) и количества используемых цветов (чем больше цветов, тем качественнее кодируется изображение).

Для представления цвета в виде числового кода используются две обратных друг другу цветовые модели: RGB или CMYK. Модель **RGB** используется в телевизорах, мониторах, проекторах, сканерах, цифровых фотоаппаратах… Основные цвета в этой модели: красный (**R**ed), зеленый (**G**reen), синий (**B**lue).Цветовая модель *CMYK* используется в полиграфии при формировании изображений, предназначенных для печати на бумаге.

Цветные изображения могут иметь различную глубину цвета, которая задается количеством битов, используемых для кодирования цвета точки.

Если кодировать цвет одной точки изображения тремя битами (по одному биту на каждый цвет RGB), то мы получим все восемь различных цветов.



Кодирование изображений

На практике же, для сохранения информации о цвете каждой точки цветного изображения в модели RGB обычно отводится 3 байта (т.е. 24 бита) - по 1 байту (т.е. по 8 бит) под значение цвета каждой составляющей. Таким образом, каждая RGB-составляющая может принимать значение в диапазоне от 0 до 255 (всего 2^8=256 значений), а каждая точка изображения, при такой системе кодирования может быть окрашена в один из 16 777 216 цветов. Такой набор цветов принято называть True Color (правдивые цвета), потому что человеческий глаз все равно не в состоянии различить большего разнообразия.

Для того чтобы на экране монитора формировалось изображение, информация о каждой точке (код цвета точки) должна храниться в видеопамяти компьютера.

Рассчитаем необходимый объем видеопамяти для одного из графических режимов. В современных компьютерах разрешение экрана обычно составляет 1280x1024 точек. Т.е. всего 1280 * 1024 = 1310720 точек. При глубине цвета 32 бита на точку необходимый объем видеопамяти: 32 * 1310720 = 41943040 бит = 5242880 байт = 5120 Кб = 5 Мб.

Кодирование изображений

Растровые изображения очень чувствительны к масштабированию (увеличению или уменьшению). При уменьшении растрового изображения несколько соседних точек преобразуются в одну, поэтому теряется различимость мелких деталей изображения. При увеличении изображения увеличивается размер каждой точки и появляется ступенчатый эффект, который можно увидеть невооруженным глазом.

Векторное изображение представляет собой совокупность графических примитивов (точка, отрезок, эллипс…). Каждый примитив описывается математическими формулами. Кодирование зависит от прикладной среды.
Достоинством векторной графики является то, что файлы, хранящие векторные графические изображения, имеют сравнительно небольшой объем.
Важно также, что векторные графические изображения могут быть увеличены или уменьшены без потери качества.

Графические форматы файлов

Форматы графических файлов определяют способ хранения информации в файле (растровый или векторный), а также форму хранения информации (используемый алгоритм сжатия).

Наиболее популярные растровые форматы:

BMP
GIF
JPEG
TIFF
PNG

- **Bit MaP image (BMP)** — универсальный формат растровых графических файлов, используется в операционной системе Windows. Этот формат поддерживается многими графическими редакторами, в том числе редактором Paint. Рекомендуется для хранения и обмена данными с другими приложениями.

- **Tagged Image File Format (TIFF)** — формат растровых графических файлов, поддерживается всеми основными графическими редакторами и компьютерными платформами. Включает в себя алгоритм сжатия без потерь информации. Используется для обмена документами между различными программами. Рекомендуется для использования при работе с издательскими системами.

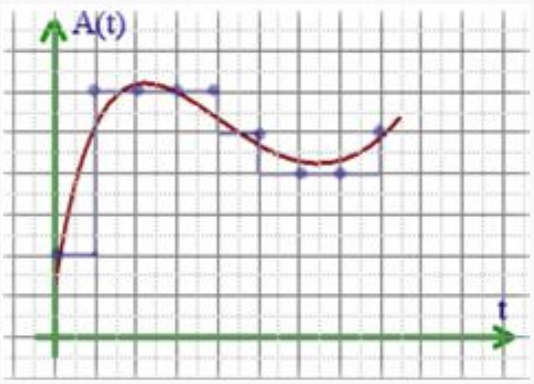
- **Graphics Interchange Format (GIF)** — формат растровых графических файлов, поддерживается приложениями для различных операционных систем. Включает алгоритм сжатия без потерь информации, позволяющий уменьшить объем файла в несколько раз. Рекомендуется для хранения изображений, создаваемых программным путем (диаграмм, графиков и так далее) и рисунков (типа аппликации) с ограниченным количеством цветов (до 256). Используется для размещения графических изображений на Web-страницах в Интернете.

- **Portable Network Graphic (PNG)** — формат растровых графических файлов, аналогичный формату GIF. Рекомендуется для размещения графических изображений на Web-страницах в Интернете.

- **Joint Photographic Expert Group (JPEG)** — формат растровых графических файлов, который реализует эффективный алгоритм сжатия (метод JPEG) для отсканированных фотографий и иллюстраций. Алгоритм сжатия позволяет уменьшить объем файла в десятки раз, однако приводит к необратимой потере части информации. Поддерживается приложениями для различных операционных систем. Используется для размещения графических изображений на Web-страницах в Интернете.

Использование компьютера для обработки звука началось позднее, нежели чисел, текстов и графики.

Звук – волна с непрерывно изменяющейся амплитудой и частотой. Чем больше амплитуда, тем он громче для человека, чем больше частота, тем выше тон.



Звуковые сигналы в окружающем нас мире необычайно разнообразны. Сложные непрерывные сигналы можно с достаточной точностью представлять в виде суммы некоторого числа простейших синусоидальных колебаний. Причем каждое слагаемое, то есть каждая синусоида, может быть точно задана некоторым набором числовых параметров – амплитуды, фазы и частоты, которые можно рассматривать как код звука в некоторый момент времени.

В процессе кодирования звукового сигнала производится его временная дискретизация – непрерывная волна разбивается на отдельные маленькие временные участки и для каждого такого участка устанавливается определенная величина амплитуды.

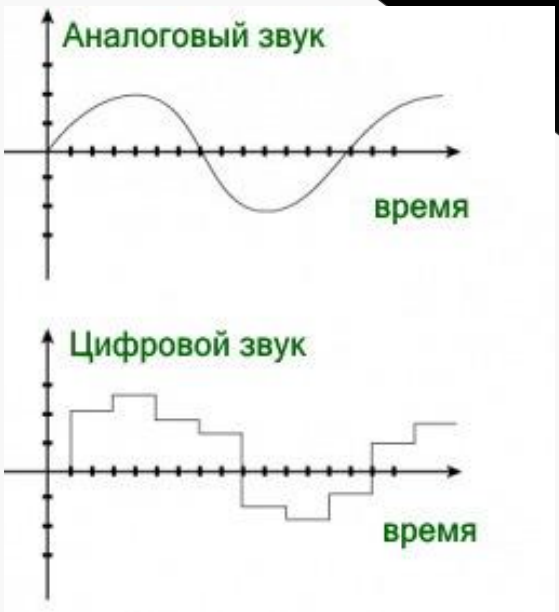
Таким образом непрерывная зависимость амплитуды сигнала от времени заменяется на дискретную последовательность уровней громкости (см. рисунок).

Каждому уровню громкости присваивается его код. Чем большее количество уровней громкости будет выделено в процессе кодирования, тем большее количество информации будет нести значение каждого уровня и тем более качественным будет звучание.

Качество двоичного кодирования звука определяется глубиной кодирования и частотой дискретизации.

Частота дискретизации – количество измерений уровня сигнала в единицу времени.

Количество уровней громкости определяет глубину кодирования. Современные звуковые карты обеспечивают 16-битную глубину кодирования звука. При этом количество уровней громкости равно $N = 2^I = 2^{16} = 65536$.



Задача (самостоятельно)

Определить информационный объем цифрового аудио файла длительностью звучания которого составляет 10 секунда при частоте дискретизации 22,05 кГц и разрешении 8 битов.

Запись условия

T=10 сек

I=8 бит

H= 22,05 кГц

Моно- ×1

V=?

Решение

$V= T \times I \times H \times 2$

$V=10 \times 8 \times 22\,050 \times 1=$

$10 \times 8 \times 22\,050$ бит/8

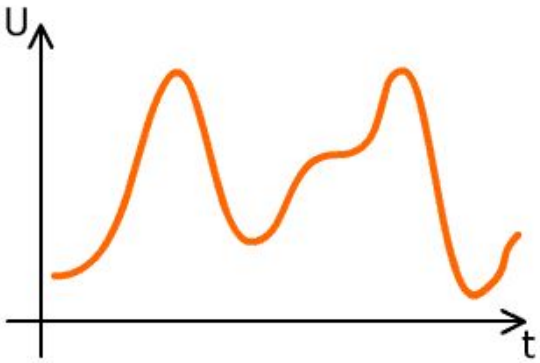
= 220500 байт/1024

= 215,332/1024

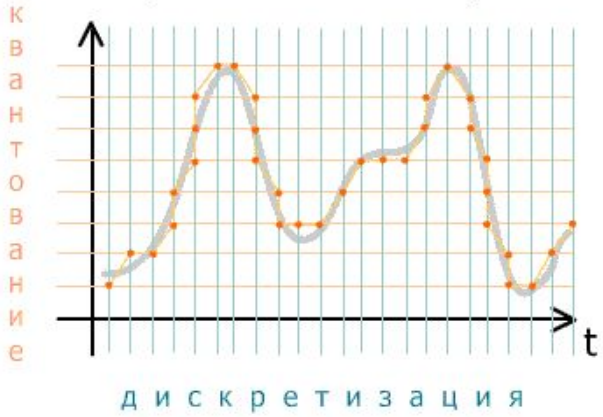
Кбайт = 0,21 Мбайт

Представление аналогового сигнала в цифровой форме

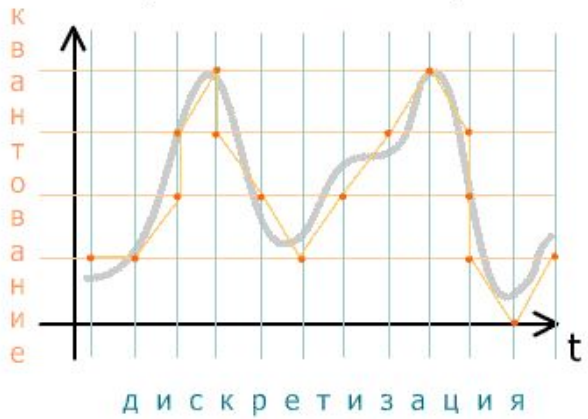
Пример аналогового сигнала



Кодирование сигнала в цифровой вид (высокое качество)



Кодирование сигнала в цифровой вид (низкое качество)



В компьютере хранится информация только о точках.

В последнее время компьютер все чаще используется для работы с видеоинформацией. Простейшей такой работой является просмотр кинофильмов и видеоклипов. Следует четко представлять, что обработка видеоинформации требует очень высокого быстродействия компьютерной системы.

Что представляет собой фильм с точки зрения информатики? Прежде всего, это сочетание звуковой и графической информации. Кроме того, для создания на экране эффекта движения используется дискретная по своей сути технология быстрой смены статических картинок. Исследования показали, что если за одну секунду сменяется более 10-12 кадров, то человеческий глаз воспринимает изменения на них как непрерывные.

Казалось бы, если проблемы кодирования статической графики и звука решены, то сохранить видеоизображение уже не составит труда. Но это только на первый взгляд, поскольку, как показывает разобранный выше пример, при использовании традиционных методов сохранения информации электронная версия фильма получится слишком большой. Достаточно очевидное усовершенствование состоит в том, чтобы первый кадр запомнить целиком (в литературе его принято называть ключевым), а в следующих сохранять лишь отличия от начального кадра (разностные кадры).

Существует множество различных форматов представления видеоданных.
В среде Windows, например, уже более 10 лет (начиная с версии 3.1) применяется формат Video for Windows, базирующийся на универсальных файлах с расширением AVI (Audio Video Interleave – чередование аудио и видео).
Более универсальным является мультимедийный формат Quick Time, первоначально возникший на компьютерах Apple.
Все большее распространение в последнее время получают системы сжатия видеоизображений, допускающие некоторые незаметные для глаза искажения изображения с целью повышения степени сжатия. Наиболее известным стандартом подобного класса служит MPEG (Motion Picture Expert Group), который разработан и постоянно развивается созданным в 1988 году Комитетом (группой экспертов) международной организации ISO/IEC (International Standards Organization/International Electrotechnical Commission) по стандартам высококачественного сжатия движущихся изображений. Методы, применяемые в MPEG, непросты для понимания и опираются на достаточно сложную математику.
Большее распространение получила технология под названием DivX (происходит от сокращения слов Digital Video Express). Благодаря DivX удалось достигнуть степени сжатия, позволившей вмесить качественную запись полнометражного фильма на один компакт-диск – сжать 4,7 Гб DVD-фильма до 650 Мб.

Мультимедиа (multimedia, от англ. multi - много и media - носитель, среда) - совокупность компьютерных технологий, одновременно использующих несколько информационных сред: текст, графику, видео, фотографию, анимацию, звуковые эффекты, высококачественное звуковое сопровождение.

Под словом «мультимедиа» понимают воздействие на пользователя по нескольким информационным каналам одновременно. Можно еще сказать так: **мультимедиа** – это объединение изображения на экране компьютера (в том числе и графической анимации и видеокадров) с текстом и звуковым сопровождением.

Наибольшее распространение системы мультимедиа получили в области обучения, рекламы, развлечений.

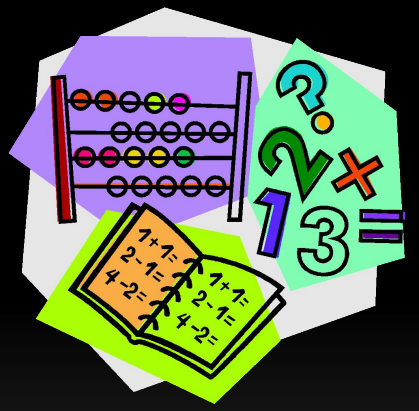
Измерение информации

1. Содержательный подход

Информация – это знания человека.

Сообщение информативно, если:

- содержит новые сведения
- понятно для человека

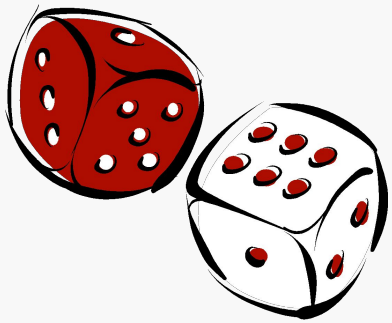


$$\int_b^a f_x = F(a) - F(b)$$

Неопределенность знаний и единица информации

Сообщение, уменьшающее неопределенность знаний в два раза, несет **1 бит** информации.

Неопределенность знаний о некотором событии – это количество возможных результатов события (бросания монеты, кубика; вытаскивание жребия)



Алфавитный подход к измерению информации

«Достаточный» алфавит содержит 256 символов

$$2^i = 256$$

$$I = 8 \text{ бит}$$

При алфавитном подходе количество информации зависит не от содержания, а от объема текста (то есть от числа знаков)

Единицы измерения информации

1 байт = 8 бит

1 Кб = 2^{10} байт = 1024 байта

1 Мб = 2^{10} Кб = 1024 Кб

1Гб = 2^{10} Мб = 1024 Мб

1Тб = 2^{10} Гб = 1024 Гб

1Пб = 2^{10} Тб = 1024 Тб

Примеры объемов информации

Страница книги	2,5 Кбайт
Учебник	0,5 Мбайт
Большая Советская Энциклопедия	120 Мбайт
Газета	150 Кбайт
Черно-белый телевизионный кадр	300 Кбайт
Цветной кадр из 3-х цветов	1 Мбайт
1,5-часовой цветной художественный фильм	135 Гбайт

В 100 Мб можно уместить:

Страниц текста	50 000
Цветных слайдов высочайшего качества	150
Аудиозапись	1,5 часа
Музыкальный фрагмент качества CD - стерео	10 минут
Фильм высокого качества записи	15 секунд
Протоколы операций по банковским счетам	За 1000 лет

Представление числовой информации с помощью систем счисления

Система счисления – это совокупность правил записи чисел с помощью определенного набора символов.

- Для записи чисел могут использоваться не только цифры, но и буквы (например, запись римскими цифрами – XXI, IV)
- Одно и то же число может быть по-разному представлено в различных системах счисления.

Системы счисления бывают *позиционными и непозиционными*.

- В позиционной системе счисления значение каждой цифры числа зависит от того, в каком месте (позиции или разряде) она записана.
  - Например, меняя позицию цифры 2 в десятичной системе счисления можно записать разные числа: 2, 20, 2000, 0,02.
- В непозиционной системе счисления цифры не изменяют своего значения при изменении их расположения в числе.
  - Примером непозиционной системы счисления может служить римская система, в которой независимо от местоположения одинаковый символ имеет неизменное значение (например, символ X в числе XXV)

Количество **(p)** различных символов, используемых для записи числа в позиционных системах счисления, называется основанием системы счисления.

В десятичной системе счисления **p=10**, для записи любого числа используются 10 различных цифр 0,1,…9.

- В компьютере наиболее подходящей и надежной системой счисления оказалась двоичная **(p=2)**, в которой для представления любого числа используются две цифры 0 и 1.
- Кроме того, удобно использовать еще 8-ричную и 16-ричную систему счисления.

Системы счисления			
Десятичная	Двоичная	Восьмеричная	Шестнадцатеричная
0	0	0	0
1	1	1	1
2	10	2	2
3	11	3	3
4	100	4	4
5	101	5	5
6	110	6	6
7	111	7	7
8	1000	10	8
9	1001	11	9
10	1010	12	A
11	1011	13	B
12	1100	14	C
13	1101	15	D
14	1110	16	E
15	1111	17	F

Соотношение 10-тичной, двоичной, 8-ричной, 16-ричной

Числа в десятичной		В 8-ричной		В 16-ричной	
0	**0**	0	**000**	0	**0000**
1	**1**	1	**001**	1	**0001**
2	**10**	2	**010**	2	**0010**
3	**11**	3	**011**	3	**0011**
4	**100**	4	**100**	4	**0100**
5	**101**	5	**101**	5	**0101**
6	**110**	6	**110**	6	**0110**
7	**111**	7	**111**	7	**0111**
8	**1000**			8	**1000**
9	**1001**			9	**1001**
10				A	**1010**
11				B	**1011**
12				C	**1100**
13				D	**1101**
14				E	**1110**

представим двоичное число

1010110101111_2

в 8-ричной системе счисления,

1 010 110 101 111

1 2 6 5 7

$100\ 011_2 = 43_8$

$1000\ 1010_2 = 8A_{16}$

- Набор символов, используемый для обозначения цифр, называется *алфавитом.*

Позиционная система счисления

Система счисления	*Основание*	*Алфавит*
Десятичная	**10**	0;1;2;3;4;5;6;7;8;9
Восьмеричная	**8**	0;1;2;3;4;5;6;7
Двоичная	**2**	0; 1
Шестнадцате-ричная	**16**	0;1;2;3;4;5;6;7;8;9; A;B;C;D;E;F

Перевод чисел из позиционной СС с основанием q в десятичную систему счисления

Правило перевода:

- представить число в развернутой форме;

555 – свернутая форма записи числа

$555_{10} = 5*10^2+5*10^1+5*10^0$ - развернутая форма

$$a_{n-1}q^{n-1} + a_{n-2}q^{n-2} + \ldots + a_1q^1 + a_0q^0 + a_{-1}q^{-1} + \ldots + a_{-m}q^{-m}$$

- вычислить сумму ряда.

Полученный результат является значением числа в 10-ой СС.

Пример: число 3201_5 перевести в 10-ую СС

$$3201_5 = \overset{3\ 2\ 1\ 0}{3 \cdot 5^3 + 2 \cdot 5^2 + 0 \cdot 5^1 + 1 \cdot 5^0} =$$

$$= 3 \cdot 125 + 2 \cdot 25 + 1 = 426 \qquad 3201_5 = 426$$

Перевод чисел из позиционной СС с основанием q в десятичную систему счисления

Пример. Перевести число 1011,1 из двоичной системы счисления в десятичную.

разряды 3 2 1 0 -1

число $1\ 0\ 1\ 1,1_2 = 1\cdot2^3 + 0\cdot2^2 + 1\cdot2^1 + 1\cdot2^0 + 1\cdot2^{-1} = 11,5_{10}$

Пример. Перевести число 276,8 из восьмеричной системы счисления в десятичную.

разряды 2 1 0 -1

число $2\ 7\ 6,5_8 = 2\cdot8^2 + 7\cdot8^1 + 6\cdot8^0 + 5\cdot8^{-1} = 190,625_{10}$

Пример. Перевести число 1F3 из шестнадцатеричной системы счисления в десятичную.

разряды 2 1 0

число $1\ F\ 3_{16} = 1\cdot16^2 + 15\cdot16^1 + 3\cdot16^0 = 499_{10}$

В меню

Перевод целых чисел из десятичной системы счисления

Алгоритм перевода:

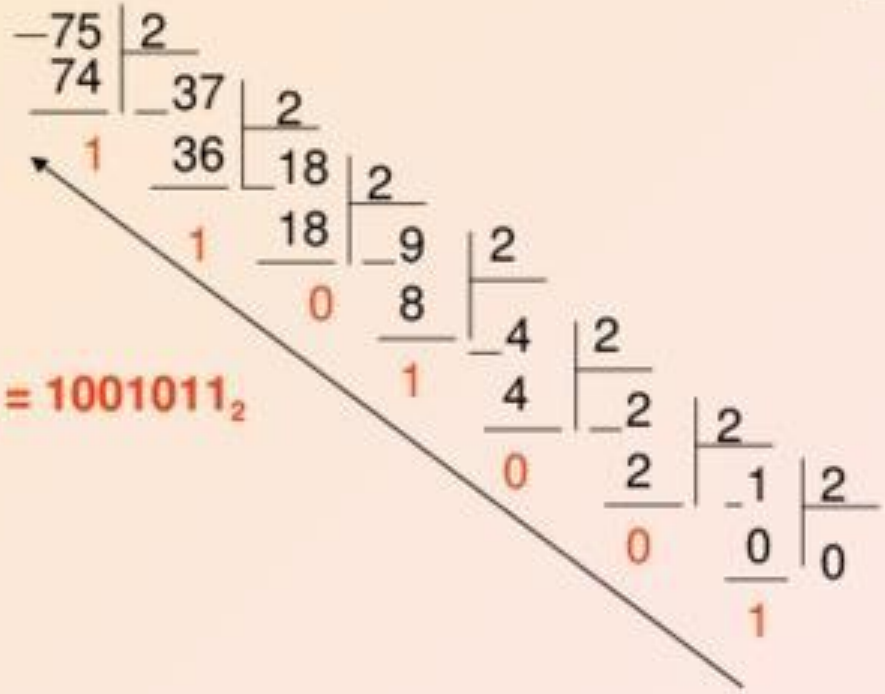
1. Последовательно делить с остатком данное число и получаемые целые частные на основание новой системы счисления до тех пор, пока частное не станет равно нулю.

2. Полученные остатки выразить цифрами алфавита новой системы счисления

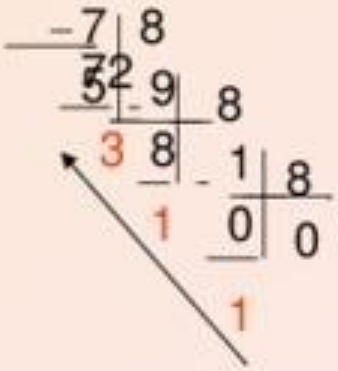
3. Записать число в новой системе счисления из полученных остатков, начиная с последнего.

Перевод целых чисел из десятичной системы счисления

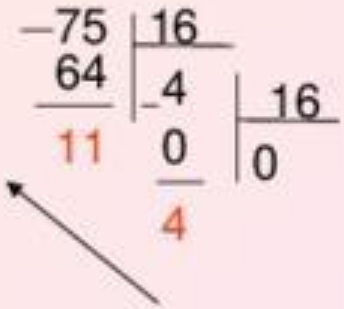
Пример. Перевести число 75 из десятичной системы счисления в двоичную, восьмеричную и шестнадцатеричную.



$75_{10} = 1001011_2$



$75_{10} = 113_8$



$75_{10} = 4B_{16}$

Перевести следующие числа в десятичную систему счисления:

XXIV=

MCDXXII=

XIV=

$123,245_8$=

$2222,1_3$=

$6548,4D_{16}$=

$10111,1_2$=

$1101,001_2$=

$1010,11_2$=

$D1,F_{16}$=

$333,43_6$=

$222,12_4$=