

Представление и кодирование информации в компьютере

Понятие информации

Что такое информация?

Нет однозначного ответа.
Смысл зависит от
контекста



Человек

Информация – это сведения, которые человек получает из окружающего мира через свои органы чувств

Смысл – это значение, которое придаёт информации человек

Компьютер



Компьютер - это универсальный программно-управляемый автомат для работы с информацией

Информацию, циркулирующую в устройствах компьютера, правильнее называть **данными**

Понятие информации



В Толковом словаре В. И. Даля нет слова «информация»



Термин «информация» начал широко употребляться с середины XX века

Теория информации

Результат развития теории связи (К. Шеннон)

Информация – содержание, заложенное в знаковые (сигнальные) последовательности

Кибернетика

Исследует информационные процессы в системах управления (Н. Винер)

Информация – содержание сигналов, передаваемых по каналам связи в системах управления

Нейрофизиология

Изучает информационные процессы в механизмах нервной деятельности человека

Информация – содержание сигналов электрохимической природы, передающихся по нервным волокнам организма

Генетика

Изучает механизмы наследственности, пользуется понятием «наследственная информация»

Информация – содержание генетического кода – структуры молекул ДНК, входящих в состав клетки живого организма



Философские концепции

Атрибутивная

Информация в мире возникла вместе со Вселенной

Информация – всеобщее свойство (атрибут) материи

Функциональная

Информация появилась лишь с возникновением жизни

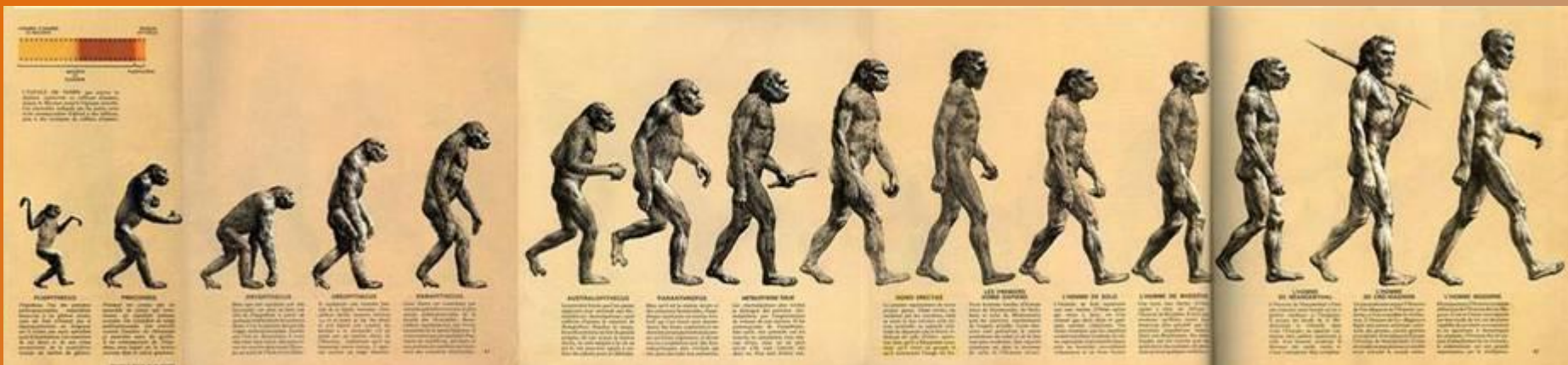
Информация и информационные процессы присущи только живой природе, являются ее функцией

Антропоцентрическая

Информация существует лишь в человеческом сознании

Информация и информационные процессы присущи только человеку

Представление информации, языки, кодирование



Историческое развитие человека, формирование человеческого общества связано с развитием речи, с появлением и распространением языков.



Язык – это знаковая система для представления и передачи информации



Люди сохраняют свои знания в записях на различных носителях. Благодаря этому знания передаются не только в пространстве, но и во времени.



Языки кодирования информации

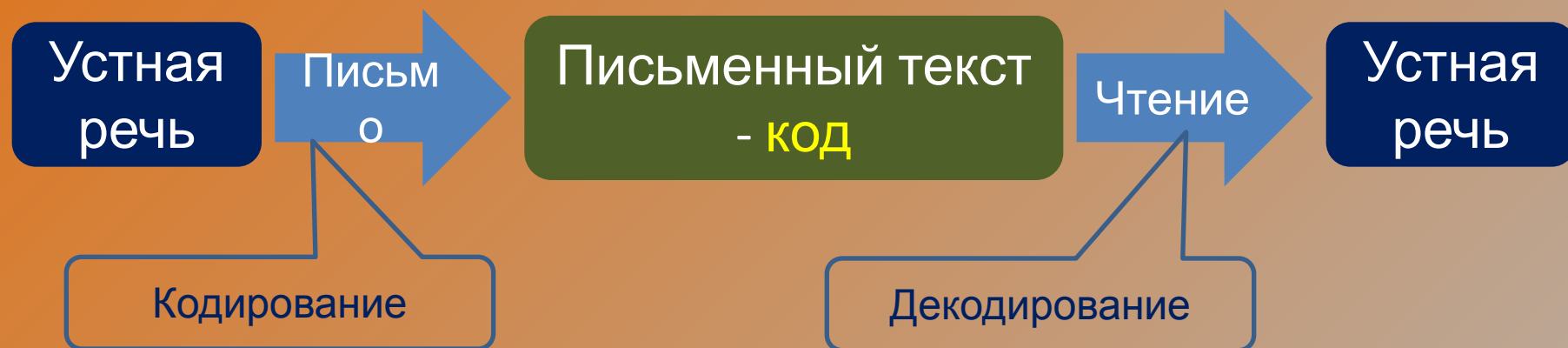
Естественные

русский,
китайский,
английский
и др.

Формальные

язык математики,
нотная грамота,
языки
программирования
и др.

СХЕМА ПЕРЕДАЧИ ИНФОРМАЦИИ ЧЕРЕЗ ПИСЬМЕННОСТЬ



Кодирование

Цели кодирования

Засекречивание

Шифрование

Алгоритмы криптографии

Быстрый способ записи

Стенография

Один знак – слово или сочетание букв

Передача по техническим каналам связи

Телеграфный код

Код Морзе:
неравномерный, троичный код

Код Бодо:
равномерный, двоичный код

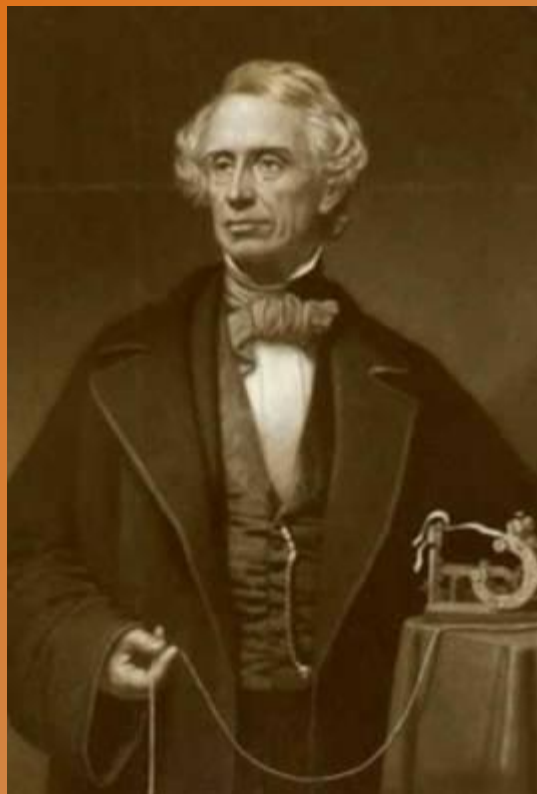
Выполнение математических вычислений

Системы счисления

Для человека:
десятичная система счисления

Для компьютера:
двоичная система счисления

ИСТОРИЯ ТЕХНИЧЕСКИХ СПОСОБОВ КОДИРОВАНИЯ ИНФОРМАЦИИ



Сэмюэль Морзе
(1791-1892), США

А	● —	Р	● — ●
Б	— ● ●	С	● ●
В	● — ●	Т	—
Г	— ● ●	У	● ● —
Д	— ● ●	Ф	● ● ●
Е	●	Х	● ● ●
Ж	● ● ● —	Ц	— ● ● ●
З	— ● ● ●	Ч	— ● ● ●
И	● ●	Ш	— ● ● ●
Й	● — ● —	Щ	— ● ● ●
К	— ● ●	Ъ	● ● ● ●
Л	— ● ●	Ы	— ● ● ●
М	— ● ●	Ь	— ● ● ●
Н	— ● ●	Э	● ● ● ● ●
О	— ● ●	Ю	● ● ● ● ●
П	— ● ●	Я	● ● ● ● ●

Кодовая таблица азбуки Морзе



Первый телеграф Морзе,
изобретённый в 1837 году



ИСТОРИЯ ТЕХНИЧЕСКИХ СПОСОБОВ КОДИРОВАНИЯ ИНФОРМАЦИИ



Жан Морис Бодо
(1845-1903),
Франция

Оригинальный код Бодо			
Управляющие символы			
о.	...	пробел,	перейти к таблице букв
.о	...	пробел,	перейти к таблице цифр
оо	...	удалить	последний знак
таблица букв		таблица цифр	
.. о..	A оо о..	K	.. о.. 1 о. о..
.. оо.	É оо оо.	L	.. .о. 2 о. .о.
.. .о.	E оо .о.	Mо 3 о. ..о
.. .оо	I оо .оо	N	.. о.о 4 о. о.о
.. ооо	O оо ооо	P	.. ооо 5 о. ооо
.. о.о	U оо о.о	Q	.. оо. 1/ о. оо.
.. ..о	Y оо ..о	R	.. .оо 3/ о. .оо
.о .оо	V о. .оо	S	.о о.. 6 оо о..
.о о.о	C о. о.о	T	.о .о. 7 оо .о.)
.о ооо	D о. ооо	V	.о ..о 8 оо ..о
.о .оо	F о. .оо	W	.о о.о 9 оо о.о
.о .о.	G о. .о.	X	.о ооо 0 оо ооо
.о оо.	H о. оо.	Z	.о оо. 4/ оо оо. =
.о о..	J о. о..	—	.о .оо 5/ оо .оо £

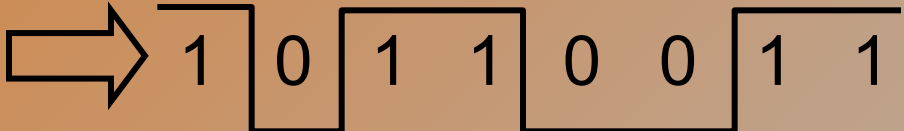


Аппарат Бодо-дуплекс.
Для печати на пяти клавишах
используются два пальца левой руки и
три – правой



Перфорированная лента

Кодирование информации

Вид информации	Двоичный код
Числовая	
Текстовая	
Графическая	
Звуковая	
Видео	

Кодирование информации

- 1 и 0 – два равновероятных события.
- Воспользовавшись формулой $2^I = N$, где I – количество информации в сообщении о событии, N – количество равновероятных событий, получим, что
- одна двоичная цифра несет количество информации равное 1 бит.

Кодирование числовой информации

- Форматы
- представления чисел
 - целочисленный
 - целые
 - положительные числа
 - целые числа
 - со знаком
 - с плавающей точкой

Кодирование числовой информации

- Целые числа без знака обычно занимают в памяти компьютера один или два байта. В однобайтовом формате принимают значения от 00000000_2 до 11111111_2 . В двухбайтовом формате - от $00000000\ 00000000_2$ до $11111111\ 11111111_2$.

Формат числа в байтах	Диапазон	
	Запись с порядком	Обычная запись
1	$0 \dots 2^8 - 1$	0 ... 255
2	$0 \dots 2^{16} - 1$	0 ... 65535

Правило для представления целого положительного числа

1. Перевести число в двоичную систему.
2. Результат дополнить слева незначащими нулями в пределах выбранного формата.

Пример

Запишем число $+156_{10}$ в 1-байтовом и 2-байтовом форматах.

формат 1 байт: 10011100.

формат 2 байта: 00000000 10011100.

Кодирование числовой информации

- Целые числа со знаком обычно занимают в памяти компьютера один, два или четыре байта, при этом самый левый (старший) разряд содержит информацию о знаке числа.

Формат числа в байтах	Диапазон	
	Запись с порядком	Обычная запись
1	$-2^7 \dots 2^7 - 1$	-128 ... 127
2	$-2^{15} \dots 2^{15} - 1$	-32768 ... 32767
4	$-2^{31} \dots 2^{31} - 1$	-2147483648 ... 2147483647

Правило для представления целого отрицательного числа

1. Модуль числа перевести в двоичную систему.
2. Результат дополнить слева незначащими нулями в пределах выбранного формата.
3. Полученное число перевести в обратный код – нули заменить единицами, а единицы – нулями.
4. К полученному коду прибавить 1.
 - **Примечание:** дополнительный код положительного числа равен его прямому коду.

Пример

Запишем число -156_{10} в 2-байтовом формате.

1. 10011100.
2. 00000000 10011100.
3. 11111111 01100011.
4. 11111111 01100100.

Задания для самостоятельной работы

1. Представьте числа в 2-байтовом формате:

a) 126_{10} ;

b) -345_{10} .

Проверка:

c) 00000000 01111110.

d) 11111110 10100111.

Задания для самостоятельной работы

2. Найдите десятичное представление целого числа со знаком:
- a) с дополнительным кодом 11111001;
 - b) с прямым кодом 10111111.

Решение:

- c) 1. 11111000.
2. 00000111.
3. -7.
- b) -63.

Кодирование числовой информации

- При записи вещественных чисел в программах вместо привычной запятой принято ставить точку.
- Для отображения вещественных чисел, которые могут быть как очень маленькими, так и очень большими, используется форма записи чисел с порядком основания системы счисления.
- Например, десятичное число 1.25 в этой форме можно представить так:
$$1.25 \cdot 10^0 = 0.125 \cdot 10^1 = 0.0125 \cdot 10^2 = \dots$$

Запись вещественных чисел

- Любое число N в системе счисления с основанием q можно записать в виде

$$N = M \cdot q^p$$

- где M - множитель, содержащий все цифры числа (**мантисса**), а p - целое число, называемое **порядком**.
- Такой способ записи чисел называется **представлением числа с плавающей точкой** или **записью числа в экспоненциальной форме**.

Запись вещественных чисел

- Мантисса должна быть правильной дробью, у которой первая цифра после точки (запятой в обычной записи) отлична от нуля:

$$0.1 \leq M < 1$$

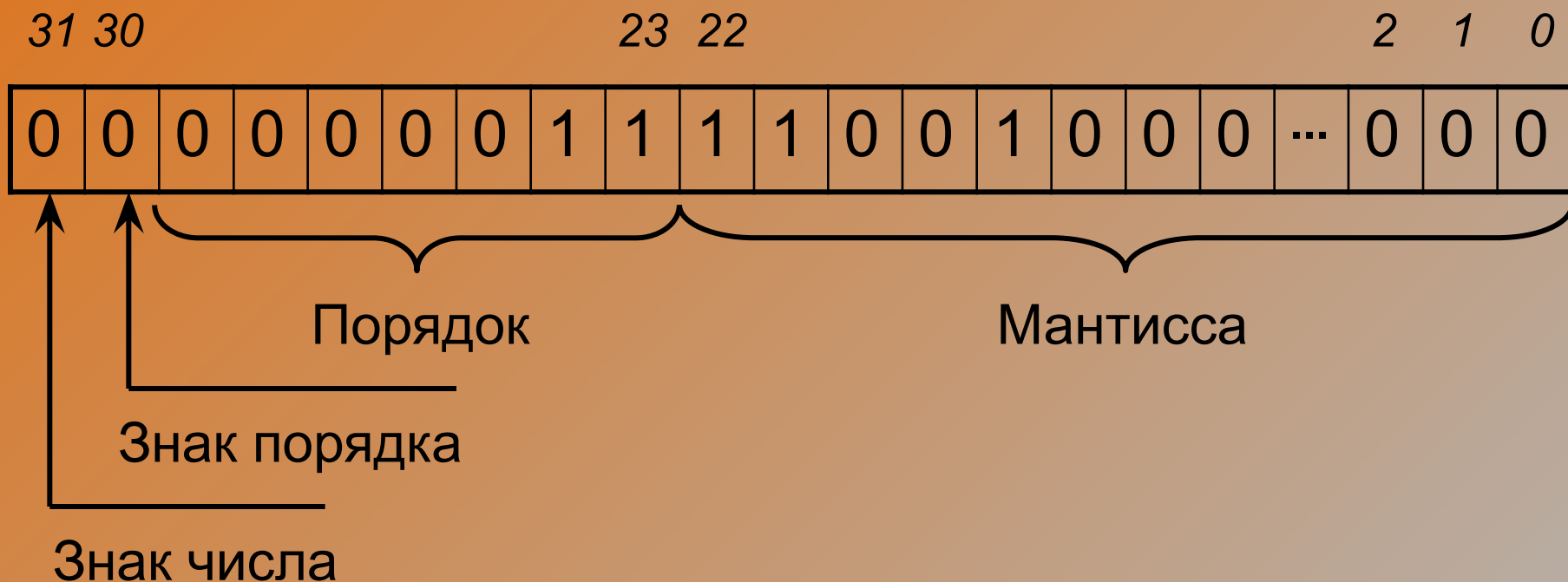
- Если это требование выполнено, то число называется нормализованным.
- Мантиссу и порядок q -ичного числа принято записывать в системе с основанием q , а само основание - в десятичной системе.

Примеры

Десятичная система	Двоичная система
$753.15 = 0.75315 \cdot 10^3$	$-101.01 = -0.10101 \cdot 2^{11}$ (порядок $11_2 = 3_{10}$)
$-0.000034 = -0.34 \cdot 10^{-4}$	$0.000011 = 0.11 \cdot 2^{-100}$ (порядок $-100_2 = -4_{10}$)

Пример

$$6,25_{10} = 110,01_2 = 0,11001 \cdot 2^{11}.$$



Кодирование текстовой информации

- Для кодирования одного символа требуется 1 байт информации.
- Можно закодировать 256 различных символов:
$$N = 2^l \rightarrow 2^8 = 256.$$
- 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.

символ	10-й код	2-й код	символ	10-й код	2-й код	символ	10-й код	2-й код	символ	10-й код	2-й код
Ъ	128	10000000		160	10100000	А	192	11000000	а	224	11100000
Г	129	10000001	Ѹ	161	10100001	Б	193	11000001	б	225	11100001
,	130	10000010	ѹ	162	10100010	В	194	11000010	в	226	11100010
г	131	10000011	Ј	163	10100011	Г	195	11000011	г	227	11100011
„	132	10000100	о	164	10100100	Д	196	11000100	д	228	11100100
...	133	10000101	Г	165	10100101	Е	197	11000101	е	229	11100101
†	134	10000110	ј	166	10100110	Ж	198	11000110	ж	230	11100110
‡	135	10000111	§	167	10100111	З	199	11000111	з	231	11100111
€	136	10001000	Ё	168	10101000	И	200	11001000	и	232	11101000
‰	137	10001001	©	169	10101001	Й	201	11001001	й	233	11101001
Љ	138	10001010	€	170	10101010	К	202	11001010	к	234	11101010
<	139	10001011	«	171	10101011	Л	203	11001011	л	235	11101011
Њ	140	10001100	¬	172	10101100	М	204	11001100	м	236	11101100
Ќ	141	10001101	-	173	10101101	Н	205	11001101	н	237	11101101
Ѝ	142	10001110	@	174	10101110	О	206	11001110	о	238	11101110
Њ	143	10001111	Ї	175	10101111	П	207	11001111	п	239	11101111
ђ	144	10010000	°	176	10110000	Р	208	11010000	р	240	11110000
‘	145	10010001	±	177	10110001	С	209	11010001	с	241	11110001
’	146	10010010	І	178	10110010	Т	210	11010010	т	242	11110010
“	147	10010011	і	179	10110011	У	211	11010011	у	243	11110011
”	148	10010100	г	180	10110100	Ф	212	11010100	ф	244	11110100
•	149	10010101	и	181	10110101	Х	213	11010101	х	245	11110101
–	150	10010110	¶	182	10110110	Ц	214	11010110	ц	246	11110110
—	151	10010111	·	183	10110111	Ч	215	11010111	ч	247	11110111
□	152	10011000	ë	184	10111000	Ш	216	11011000	ш	248	11111000
™	153	10011001	№	185	10111001	Щ	217	11011001	щ	249	11111001
љ	154	10011010	ε	186	10111010	Ъ	218	11011010	ъ	250	11111010
›	155	10011011	»	187	10111011	Ы	219	11011011	ы	251	11111011
њ	156	10011100	ј	188	10111100	Ь	220	11011100	ь	252	11111100
ќ	157	10011101	ѕ	189	10111101	Э	221	11011101	э	253	11111101
ћ	158	10011110	ѕ	190	10111110	Ю	222	11011110	ю	254	11111110
џ	159	10011111	ї	191	10111111	Я	223	11011111	я	255	11111111

Кодировка русских букв

- Существует пять различных кодовых таблиц для русских букв КОИ-8, CP1251, CP866, Mac, ISO.
- Поэтому тексты, созданные в одной кодировке, не будут правильно отображаться в другой.

- Например, код

221 194 204

- соответствует

Кодировка	Слово
CP1251	ЭВМ
КОИ-8	щбл
ISO	нТЬ

Международный стандарт

- Unicode отводит на каждый символ не 1 байт, а 2, и поэтому с его помощью можно закодировать 65536 различных СИМВОЛОВ.
- Эту кодировку поддерживают последние версии платформы Microsoft Windows&Office (начиная с 1997 года).

Форматы текстовых файлов

- TXT
- DOC
- RTF
- PDF
- HTML



Задания для самостоятельной работы

1. Средняя скорость чтения учащихся 9-11 классов составляет 160 слов в минуту (одно слово в среднем – 6 символов). Сколько Кбайт успеет переработать ученик за четыре часа непрерывного чтения?

Решение:

$$I = (160 \cdot 6 \cdot 60 \cdot 4) / 1024 = 225 \text{ (Кб)}.$$

Задания для самостоятельной работы

2. Сведения о сотруднике хранятся в виде строки из 2048 символов. На каком минимальном числе дискет емкостью 1,2 Мб можно разместить сведения обо всех 8192 сотрудниках?

Решение:

$$K = (2048 \cdot 8192) / (1024 \cdot 1024 \cdot 1,2) = 14 \text{ (дискет).}$$

Кодирование графической информации

- В процессе кодирования изображения производится его пространственная дискретизация.

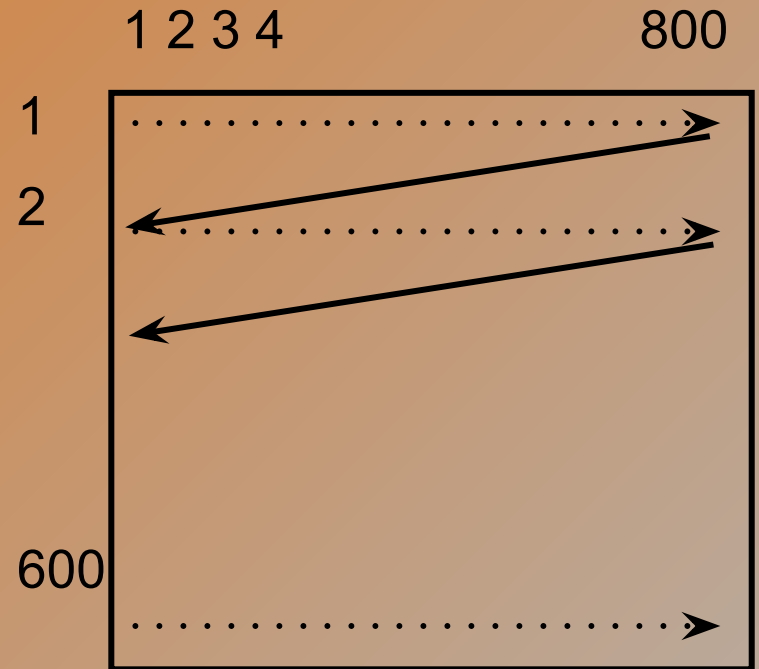
- Качество кодирования изображения зависит от двух параметров:

1. качество тем выше, чем меньше размер точки и соответственно большее количество точек составляет изображение.
2. чем большее количество цветов, используется, тем более качественно кодируется изображение.



Формирование изображения

- Разрешающая способность монитора – это размер экрана монитора по ширине и высоте в пикселях.
- Самыми популярными являются разрешающие способности: 640x480, 800x600, 1024x768, 1280x1024.



Кодирование цвета

Для кодирования черно-белого изображения достаточно одного бита памяти:

1 – белый

0 – черный.

Для кодирования 4-цветного изображения требуется два бита на пиксель, поскольку два бита могут принимать 4 различных состояния. Может использоваться, например, такой вариант кодировки цветов:

00 – черный

01 – красный

10 – зеленый

11 – коричневый.

На цветном экране все разнообразие красок получается из сочетаний трех базовых цветов: красного, зеленого, синего. Такая цветовая модель называется **RGB моделью**, по первым буквам английских названий цветов (Red, Green, Blue):

К	З	С	Цвет
0	0	0	Черный
0	0	1	Синий
0	1	0	Зеленый
0	1	1	Циан
1	0	0	Красный
1	0	1	Фиолетовый
1	1	0	Желтый
1	1	1	Белый

Шестнадцатичервная палитра получается при использовании четырехразрядной кодировки пикселя: к трем битам базовых цветов добавляется один бит интенсивности. Этот бит управляет яркостью всех трех цветов одновременно:

И	К	З	С	Цвет
0	0	0	0	черный
0	0	0	1	синий
0	0	1	0	зеленый
0	0	1	1	голубой
0	1	0	0	красный
0	1	0	1	розовый
0	1	1	0	коричневый
0	1	1	1	серый
1	0	0	0	темно-серый
1	0	0	1	ярко-синий
1	0	1	0	ярко-зеленый
1	0	1	1	ярко-голубой
1	1	0	0	ярко-красный
1	1	0	1	ярко-розовый
1	1	1	0	ярко-желтый
1	1	1	1	белый

Количество бит, которое используется для кодирования цвета точки, называется **глубиной цвета**. Тогда количество цветов, отображаемых на экране монитора, может быть вычислено по формуле:

$$N=2^I,$$

где N – количество цветов, I – глубина цвета.

Наиболее распространенными значениями глубины цвета являются 4, 8, 16 или 24 бита на точку.

Глубина цвета	Кол-во цветов
4	16
8	256
16	65536
24	16777216

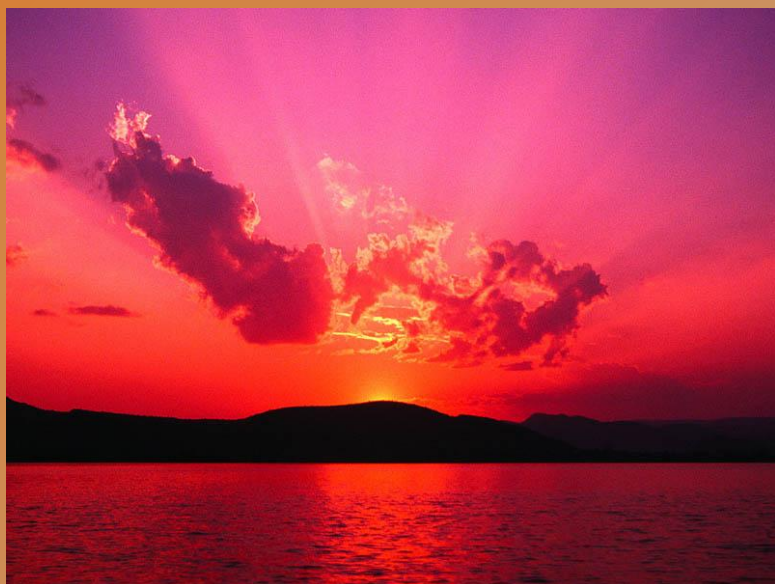
Пример

Рассчитаем необходимый объем видеопамяти для одного из графических режимов, например, с разрешением 800 на 600 точек и глубиной цвета 24 бит на точку.

$$V = 800 \cdot 600 \cdot 24 = 11520000 \text{ (бит)} = 1440000 \text{ (байт)} = 1406,25 \text{ (Кб)} = 1,37 \text{ (Мб)}.$$

Виды изображений

- **Растровое** – совокупность точек (пикселей), используемых для его отображения на экране монитора.
- **Векторное** – совокупность графических примитивов (линий, прямоугольников, эллипсов и т.д.).



Форматы графических файлов

Растровые:

- BMP
- GIF
- JPEG
- TIFF
- PCX

Векторные:

- WMF
- CDR



Задания для самостоятельной работы

1. Графическое изображение занимает на экране дисплея область размером 100×100 пикселей. Цветовая палитра – 256 цветов. Найти объем оперативной памяти, занимаемой этой иллюстрацией.

Решение:

1. $N = 2^l, 256 = 2^l \rightarrow l = 8$ (бит).
2. $V = 100 \cdot 100 \cdot 8 = 80000 = 10000$ (байт).

Задания для самостоятельной работы

2. В процессе преобразования растрового графического файла количество цветов уменьшилось с 65 536 до 256. Во сколько раз уменьшится информационный объем файла?

Решение:

1. $N = 2^l, 65536 = 2^{l_1} \rightarrow l_1 = 16$ (бит).
2. $256 = 2^{l_2} \rightarrow l_2 = 8$ (бит).
3. $l_1 = 2l_2 \rightarrow$ Информационный объем уменьшится в два раза.

Кодирование звуковой информации

- **Звук** – непрерывный сигнал, звуковая волна 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).

Кодирование звуковой информации

- В процессе кодирования звуковой информации непрерывный сигнал заменяется дискретным, то есть превращается в последовательность электрических импульсов.
- Качество звукового сигнала определяется глубиной и частотой дискретизации.
- Глубина дискретизации – количество бит, отводимое на один звуковой сигнал.
- Например, 16-битные звуковые карты могут обеспечить

$$N = 2^l \rightarrow 2^{16} = 65536$$

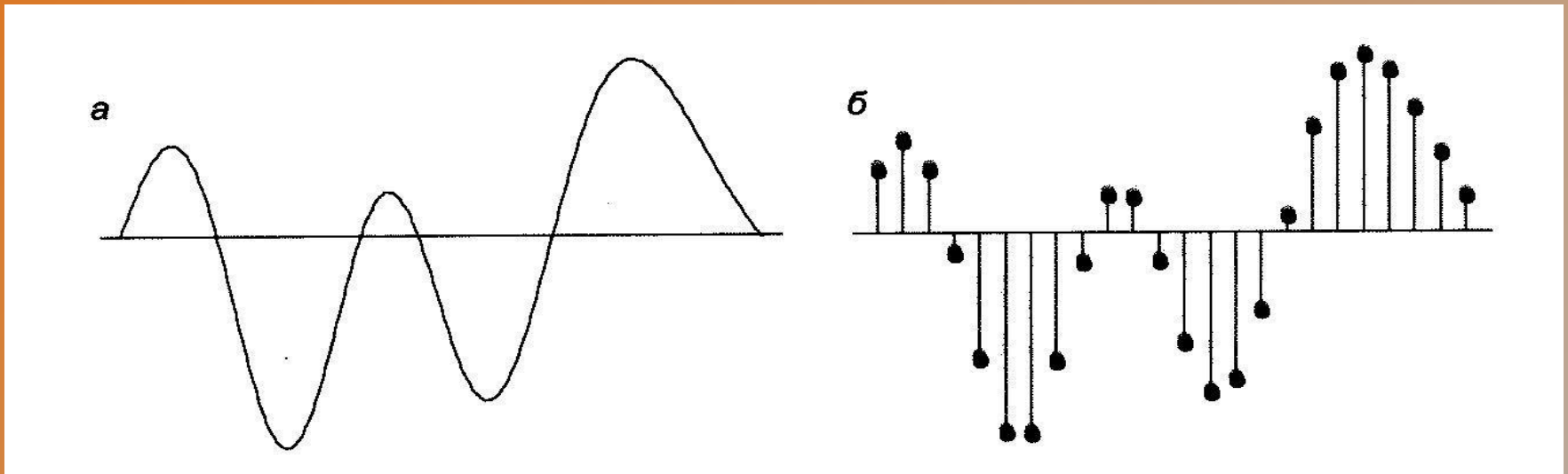
- уровней сигнала.

Кодирование звуковой информации

- **Частота дискретизации** – количество измерений уровня сигнала за 1 секунду.
- Количество измерений в секунду может лежать в диапазоне от 8000 до 48 000, то есть частота дискретизации звукового сигнала может принимать значения от 8 до 48 кГц.
- При частоте 8 кГц качество дискретизации соответствует качеству радиотрансляции, а при частоте 48 кГц – качеству звучания аудио-CD.

Методы кодирования звуковой информации

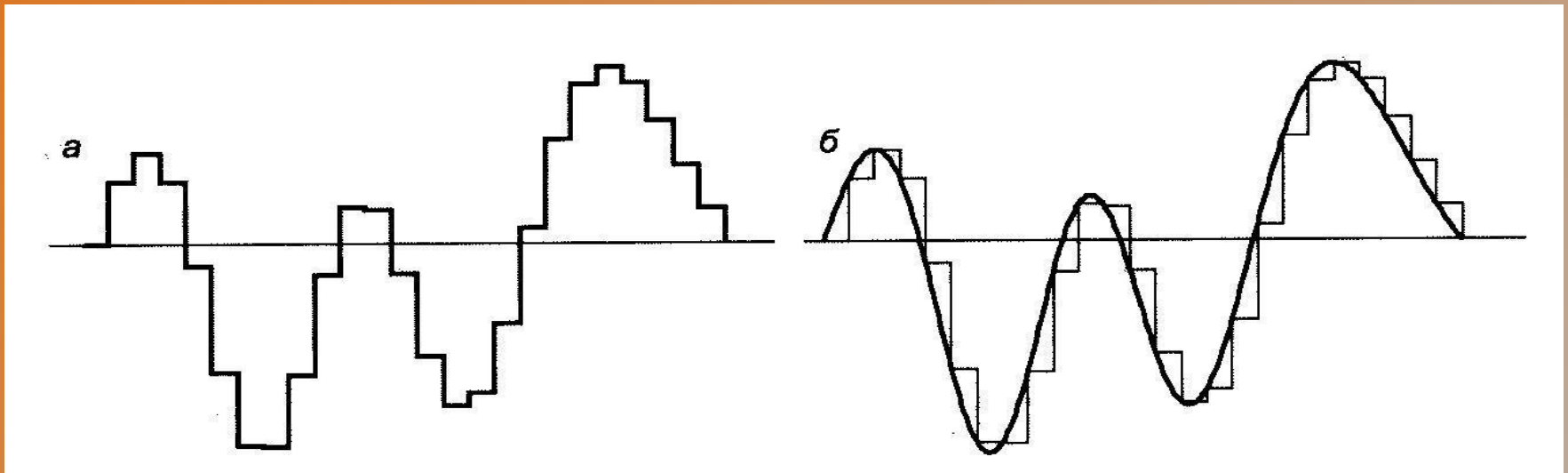
- **Метод FM (Frequency Modulation).**



Преобразование звукового сигнала в дискретный сигнал: а – звуковой сигнал на входе АЦП (аналогово-цифровые преобразователи); б – дискретный сигнал на выходе АЦП

Методы кодирования звуковой информации

- **Метод FM (Frequency Modulation).**



Преобразование дискретного сигнала в звуковой сигнал:
а – дискретный сигнал на входе ЦАП (цифро-аналоговые преобразователи); б – звуковой сигнал на выходе ЦАП

Методы кодирования звуковой информации

- **Таблично-волновой метод (Wave-Table)**
основан на том, что в заранее подготовленных таблицах хранятся образцы звуков окружающего мира, музыкальных инструментов и т.д.
- Числовые коды выражают высоту тона, продолжительность и интенсивность звука и прочие параметры.
- Качество звука приближается к качеству звучания реальных музыкальных инструментов.

Пример

Найдем информационный объем моноаудиофайла, длительность звучания которого 1 секунда, качество звука среднее (16 бит, 24 кГц):

$$V = 16 \cdot 24000 = 384000 \text{ (бит)} = 48000 \text{ (байт)} = 47 \text{ (Кбайт)}.$$

Форматы звуковых файлов

- MIDI
- WAV
- MP3



Задания для самостоятельной работы

1. Объем свободной памяти на диске – 5,25 Мб, разрядность звуковой платы – 16 бит. Какова длительность звучания цифрового аудиофайла, записанного с частотой дискретизации 22,05 кГц?

Решение:

$$X = (5,25 \cdot 1024 \cdot 1024 \cdot 8) / (16 \cdot 22050) = 124,8 \text{ (с)}.$$

Задания для самостоятельной работы

2. Две минуты записи цифрового аудиофайла занимают на диске 5,1 Мб. Частота дискретизации - 22050 Гц. Какова разрядность аудиоадаптера?

Решение:

$$X = (5,1 \cdot 1024 \cdot 1024 \cdot 8) / (2 \cdot 60 \cdot 22050) = 16 \text{ (бит)}$$

Кодирование видеоинформации

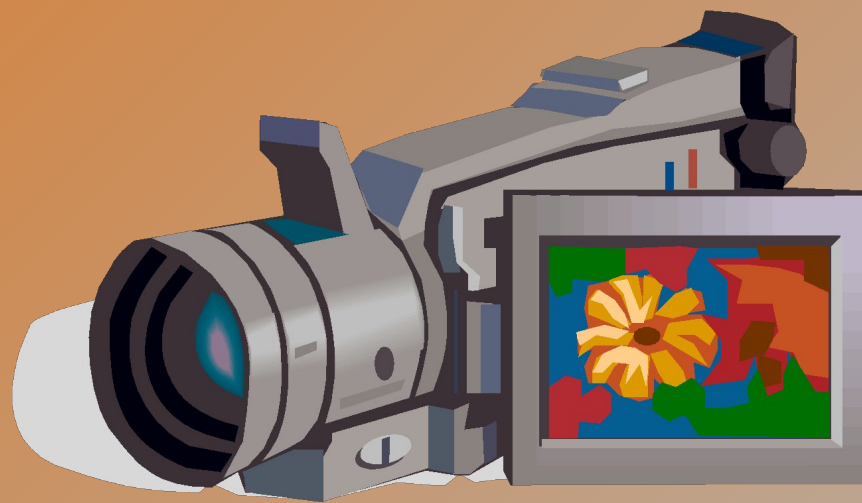
- Преобразование оптического изображения в последовательность электрических сигналов осуществляется видеокамерой.
- Сигналы несут информацию о яркости и цветности отдельных участков изображения.
- Сигналы сохраняются на носителе в виде изменения намагниченности видеоленты (аналоговая форма) или в виде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

Аналого-цифровое преобразование видеоинформации

- **Дискретизация** – непрерывный сигнал заменяется последовательностью мгновенных значений через равные промежутки времени.
- **Квантование** – величина каждого отсчета заменяется округленным значением ближайшего уровня.
- **Кодирование** – каждому значению уровней квантования сопоставляются их порядковые номера в двоичном виде.

Форматы видео файлов

- AVI
- MPEG
- DIVX



Источники информации

- Информатика и ИКТ. Учебник. 10 класс. Базовый уровень / Под ред. проф. Н.В. Макаровой. – СПб.: Питер, 2007.
- Угринович Н.Д. Информатика и информационные технологии. Учебное пособие для 10-11 классов. Углубленный курс. – М.: Лаборатория Базовых Знаний, 2000.
- Шауцукова Л.З. Информатика. 10-11 классы. – М.: Просвещение, 2000.