



*Российский государственный
университет нефти и газа
им. И.М. Губкина
Кафедра «Информатика»*

Лекция

Понятие информации Измерение информации

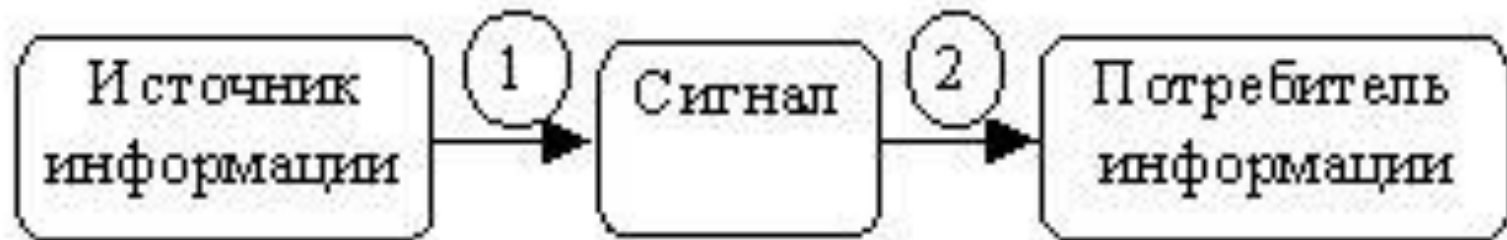
Информация – сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые уменьшают имеющуюся о них степень неопределённости, неполноты знаний.

Данные - признаки или записанные наблюдения, которые по каким – то причинам не используются, а только хранятся.

Информация - используемые данные

Информационные коммуникации

- пути и процессы, обеспечивающие передачу сообщений от источника информации к её потребителю



Адекватность информации

– это определённый уровень соответствия создаваемого с помощью полученной информации образа реальному объекту, процессу, явлению, и т.д.

Адекватность информации может выражаться в трех формах:

- ✓ семантической,
- ✓ синтаксической,
- ✓ прагматической.

Синтаксическая адекватность.

– отображает формально-структурные характеристики информации и не затрагивает ее смыслового содержания.

Информацию, рассматриваемую только с синтаксических позиций, обычно называют **данными, так как при этом не имеет значения смысловая сторона**

На синтаксическом уровне учитываются:

- **ТИП НОСИТЕЛЯ**
- **способ представления информации**
- **скорость передачи и обработки**
- **размеры кодов представления информации**
- **надежность и точность преобразования этих кодов**
- **и т.п.**

Семантическая (смысловая) адекватность

- определяет степень соответствия образа объекта и самого объекта
- предполагает учет смыслового содержания информации

Прагматическая адекватность

- ✓ отражает отношение информации и ее потребителя
- ✓ связана с практическим использованием информации

Меры информации

Синтаксическая
мера

Объём данных V_d

Количество информации
 $I_\beta(\alpha)$

уменьшение неопределённости

Прагматическая
мера

Семантическая
мера

Количество
информации
 $I_c = C V_d$,

C-коэффициент
содержательности

Синтаксическая мера

Объём данных (V_D) - информационный объём сообщения или объём памяти, необходимый для хранения сообщения без каких-либо изменений (измеряется в **битах**).

бит - количество информации, которое содержит сообщение , уменьшающее неопределенность в два раза.

Количество информации

$$I_{\beta}(\alpha) = H(\alpha) - H_{\beta}(\alpha)$$

$H(\alpha)$

энтропия системы - мера неопределенности состояния системы

$H_{\beta}(\alpha)$

неопределенность состояния системы после получения сообщения β

$I_{\beta}(\alpha)$

количество информации о системе, полученной в сообщении β

формула Шеннона

$$H(\alpha) = - \sum_{i=1}^N p_i \log_2 p_i$$

где p_i — вероятность того, что система находится в i -м состоянии из N возможных состояний

Для случая, когда все состояния системы
равновероятны, т.е. их вероятности

равны

$$p_i = \frac{1}{N}$$

$$H(\alpha) = - \sum_{i=1}^N \frac{1}{N} \log_2 \frac{1}{N}$$

При кодировании информации
числовыми кодами в той или иной
системе счисления

$$N = m^n$$

N — число всевозможных отображаемых состояний;

m — основание системы счисления;

n — число разрядов (символов) в сообщении.

Формула Хартли

- Р. Хартли (1928 г.) процесс получения информации рассматривал как выбор одного сообщения из конечного наперёд заданного множества из N равновероятных сообщений,
- количество информации I , содержащееся в выбранном сообщении:

$$I = \log_2 N$$

Задача 1

Шарик находится в одной из трех урн: А, В или С. Определить сколько бит информации содержит сообщение о том, что он находится в урне В.

Решение.

$$*I = \log_2 3 = 1,585 \text{ бита}*$$

информации.

Каждая буква русского алфавита (если считать, что е=ё) несет информацию

5 бит (32 = 2⁵).

Семантическая мера информации

Тезаурусная мера связывает семантические свойства информации (смысловое содержание) со способностью пользователя принимать поступившее сообщение.

тезаурус пользователя - это совокупность сведений, которыми располагает пользователь или система.

количество семантической информации I_c

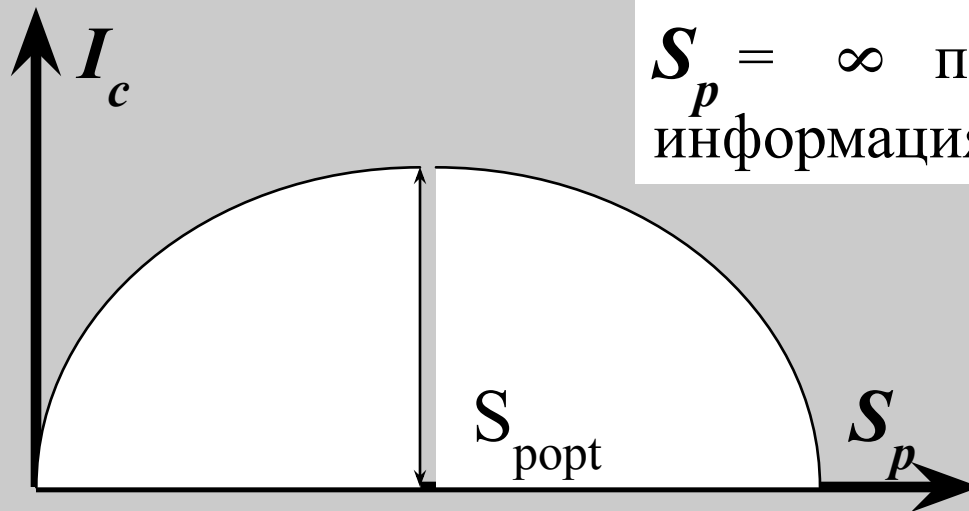
$S_p = 0$ - пользователь
не понимает
информацию

$$C = \frac{I_c}{V_\partial}$$

- коэффициент
содержательности

S-смысловое содержание

S_p -тезаурус пользователя



$S_p = \infty$ пользователь всё знает
информация не нужна

Прагматическая мера информации

Определяет полезность информации (ценность) для достижения пользователем поставленной цели.

Величина относительная, обусловленная особенностями использования этой информации в той или иной системе.

Мера информации	Единицы измерения	Примеры (для компьютерной области)
Синтаксическая: шенноновский подход	Степень уменьшения неопределенности	Вероятность события
компьютерный подход	Единицы представления информации	Бит, байт, Кбайт и т.д.

<p align="center">Мера информации</p>	<p align="center">Единицы измерения</p>	<p align="center">Примеры (для компьютерной области)</p>
<p>Семантическая</p>	<p>Тезаурус</p> <p>Экономические показатели</p>	<p>Пакет прикладных программ, персональный компьютер, компьютерные сети и т. д.</p> <p>Рентабельность, производительность, коэффициент амортизации и т.д.</p>

Мера информации	Единицы измерения	Примеры (для компьютерной области)
Прагматическая	Ценность использования	Емкость памяти, производительность компьютера, скорость передачи данных и т.д. Денежное выражение Время обработки информации и принятия решений

Представление информации в компьютере.

- электромагнитные реле (замкнуто/разомкнуто), широко использовались в конструкциях первых ЭВМ;
- участок поверхности магнитного носителя информации (намагничен/размагничен);
- участок поверхности лазерного диска (отражает/не отражает);
- триггер может устойчиво находиться в одном из двух состояний, широко используется в оперативной памяти компьютера.

Кодирование данных двоичным кодом

- ❑ бит (bit) - от английского словосочетания **Binary digit** (двоичная цифра).
- ❑ Один бит кодирует два понятия:
0 или **1** (да или нет, черное или белое, истина или ложь и т.п.)
- ❑ два бита - четыре различных значения: **00**
01 10 11
- ❑ Три бита можно закодировать восемь различных значений: **000 001 010 011 100**
101 110 111

общая формула имеет вид: $N = 2^m$

где N - количество
независимых
кодируемых значений;
m - разрядность двоичного
кодирования

m	N
1	2
2	4
3	8
4	16
...	...
8	256

Единицы измерения информации:

1 байт = $2^3 = 8$ бит

1 Килобайт = 1024 байт.

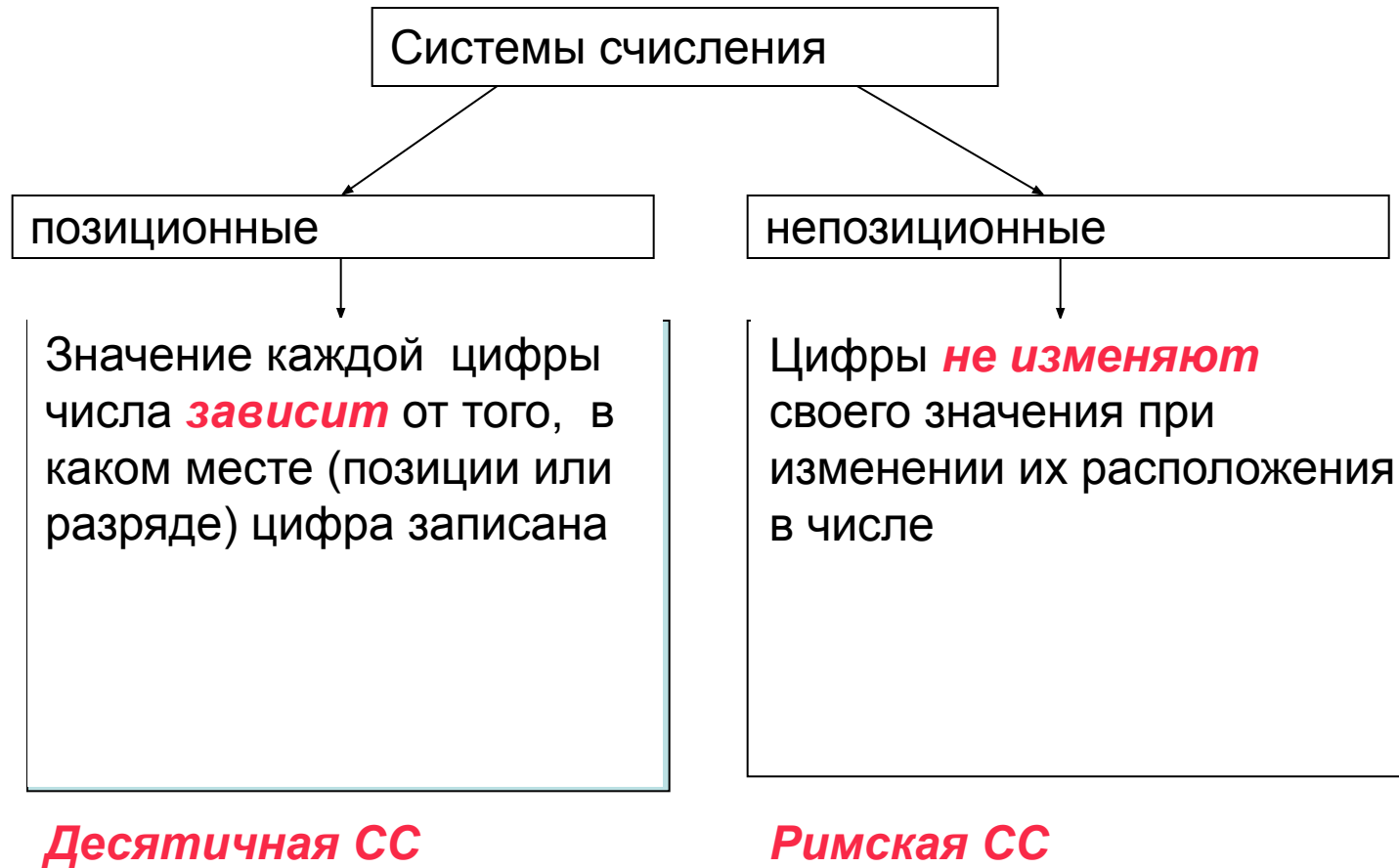
Килобайт	Кбайт = 2^{10} байт
Мегабайт	Мбайт = 2^{10} Кбайт
Гигабайт	Гбайт = 2^{10} Мбайт
Терабайт	Тбайт = 2^{10} Гбайт
Петабайт	Пбайт = 2^{10} Тбайт
Эксобайт	Эбайт = 2^{10} Пбайт

Что такое система счисления?

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

Для записи чисел могут использоваться не только цифры, но и буквы.

Что такое система счисления?



Не позиционные системы счисления

Римская система счисления

- Является непозиционной, т.к. каждый символ обозначает всегда одно и тоже число;
- Цифры обозначаются латинскими буквами:

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

Например: XXX – 30; XLI - 41

Позиционные системы счисления

Основание ПСС – это количество цифр, используемое для представления чисел;

Алфавит – набор символов, используемый для обозначения цифр.

- Значение цифры зависит от ее позиции, т.е. одна и та же цифра соответствует разным значениям в зависимости от того, в какой позиции числа она стоит;

Например: 888: 800; 80; 8

- Любое позиционное число можно представить в виде суммы степеней основания системы.

Позиционные системы счисления

Десятичная СС

- Основание системы – число 10;
- Алфавит (10 цифр): 0, 1, 2, 3, 4, 5, 6, 7, 8, 9;
- Любое десятичное число можно представить в виде суммы степеней числа 10 – основания системы;

$$2345_{10} = 2 \cdot 10^3 + 3 \cdot 10^2 + 4 \cdot 10^1 + 5 \cdot 10^0$$

Позиционные системы счисления

Двоичная СС

- Основание системы – 2;
- Алфавит (2 цифры): 0; 1;
- Любое двоичное число можно представить в виде суммы степеней числа 2 – основания системы;

$$10101_2 = 1 \cdot 2^4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0$$

Кодирование Графической информации

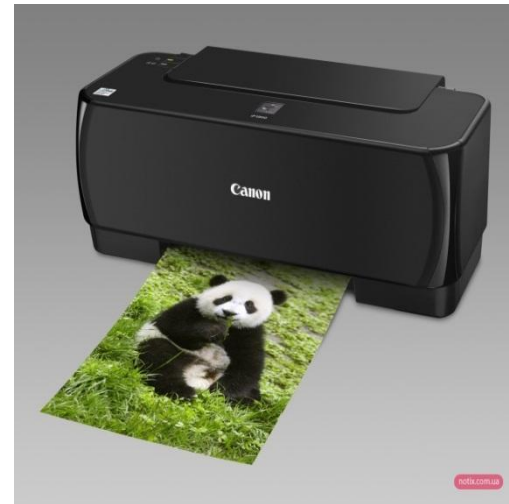
формы представления графической информации

Аналоговая



живописное полотно, цвет
которого изменяется непрерывно

Дискретная

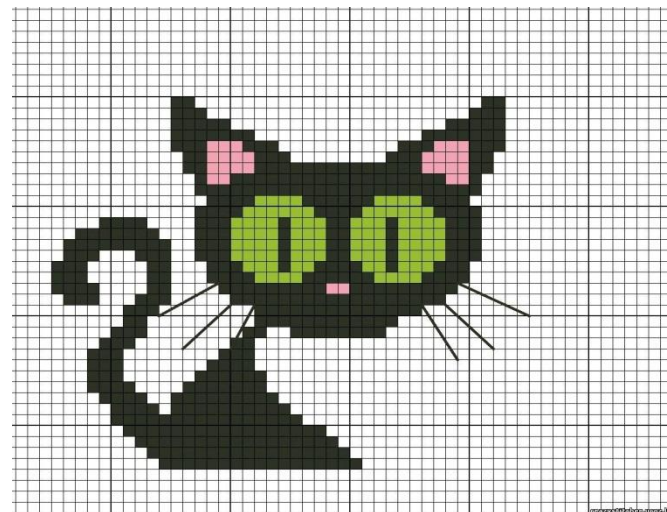


изображение, напечатанное с помощью
струйного принтера и состоящее из

отдельных точек разного цвета

Графические изображения из аналоговой (непрерывной) формы в цифровую (дискретную) преобразуются путем пространственной дискретизации.

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



Все компьютерные изображения разделяют на два основных типа: растровые и векторные

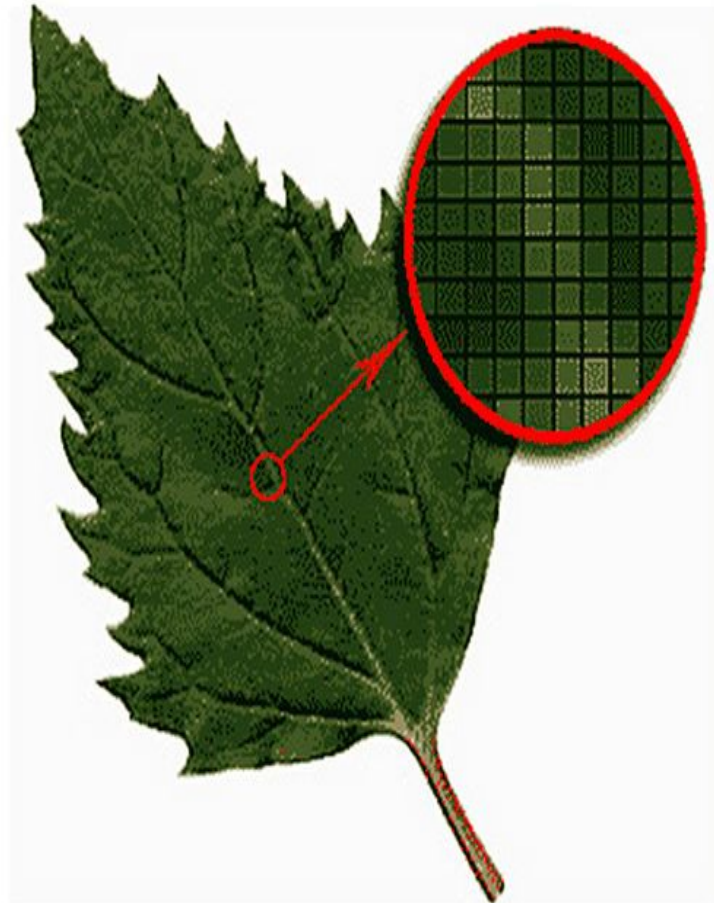
Виды компьютерных изображений



Векторное изображение представляет собой совокупность графических примитивов (точка, отрезок, эллипс и т.д.). Каждый примитив описывается математическими формулами.



растровое изображение – это набор пикселей, расположенных на прямоугольной сетке.



Качество изображения зависит от **разрешающей способности**. Разрешающая способность растрового изображения определяется количеством точек по горизонтали и количеством точек по вертикали на единицу длины изображения.

- Чем меньше размер точки, тем больше разрешающая способность (больше строк растра и точек в строке) и, соответственно, выше качество изображения.

Величина разрешающей способности выражается в dpi (dot per inch - точек на дюйм), т.е. в количестве точек в полоске изображения длиной в 1 дюйм (1 дюйм = 2,54 см).

Количество информации, которое используется для кодирования цвета точки изображения, называется *глубиной* цвета.

Каждый цвет можно рассматривать как возможное состояние точки. Количество цветов N в палитре и количество информации для кодирования цвета каждой точки связаны между собой формулой: $N=2^I$.

Глубина цвета I	Количество отображаемых цветов N
4	$2^4 = 16$
8	$2^8 = 256$
16 (hige color)	$2^{16} = 65\ 536$
24 (true color)	$2^{24} = 16\ 777\ 216$
32 (true color)	$2^{32} = 4\ 294\ 967\ 296$

Цветное изображение на экране монитора формируется за счет смешивания трех базовых цветов: красного, зеленого и синего. Такая цветовая модель называется RGB-моделью по первым буквам английских названий цветов (Red, Green, Blue).



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.

Система кодировки RGB

R	G	B	Цвет
1	1	1	белый
1	1	0	желтый
1	0	1	пурпурный
1	0	0	красный
0	1	1	голубой
0	1	0	зеленый
0	0	1	синий
0	0	0	черный

Для получения богатой палитры цветов базовым цветам могут быть заданы различные интенсивности. Например, при глубине цвета в 24 бита на каждый из цветов выделяется по 8 бит, то есть для каждого из цветов возможны

$N = 2^8 = 256$ уровней интенсивности, заданные двоичными кодами (от минимальной - 00000000 до максимальной - 11111111)

Название цвета	Интенсивность		
	Красный	Зеленый	Синий
Черный	00000000	00000000	00000000
Красный	11111111	00000000	00000000
Зеленый	00000000	11111111	00000000
Синий	00000000	00000000	11111111
Голубой	00000000	11111111	11111111
Желтый	11111111	11111111	00000000
Белый	11111111	11111111	11111111

Графические режимы

МОНИТОРОВ

Графический режим вывода изображения на экран монитора определяется величиной разрешающей способности и глубиной цвета. Для того чтобы на экране монитора формировалось изображение, информация о каждой его точке (код цвета точки) должна храниться в видеопамати компьютера.

видеопамати

можно рассчитать по формуле:

$$\text{Объем видеопамати } I_{\text{п}} = I \times X \times Y,$$

где $I_{\text{п}}$ - информационный объем видеопамати памяти в битах;

$X \times Y$ - количество точек изображения (X - количество точек по горизонтали, Y - по вертикали);

I - глубина цвета в битах на точку.

Пример

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

Всего точек на экране: $800 \times 600 = 480\,000$.

Необходимый объем видеопамяти:

$24 \text{ бит} \times 480\,000 = 11\,520\,000 \text{ бит} = 1\,440\,000 \text{ байт} = 1406,25 \text{ Кбайт} = 1,37 \text{ Мбайт}$.

Аналогично рассчитывается необходимый объем видеопамяти для других графических режимов.

Кодирование звуковой информации

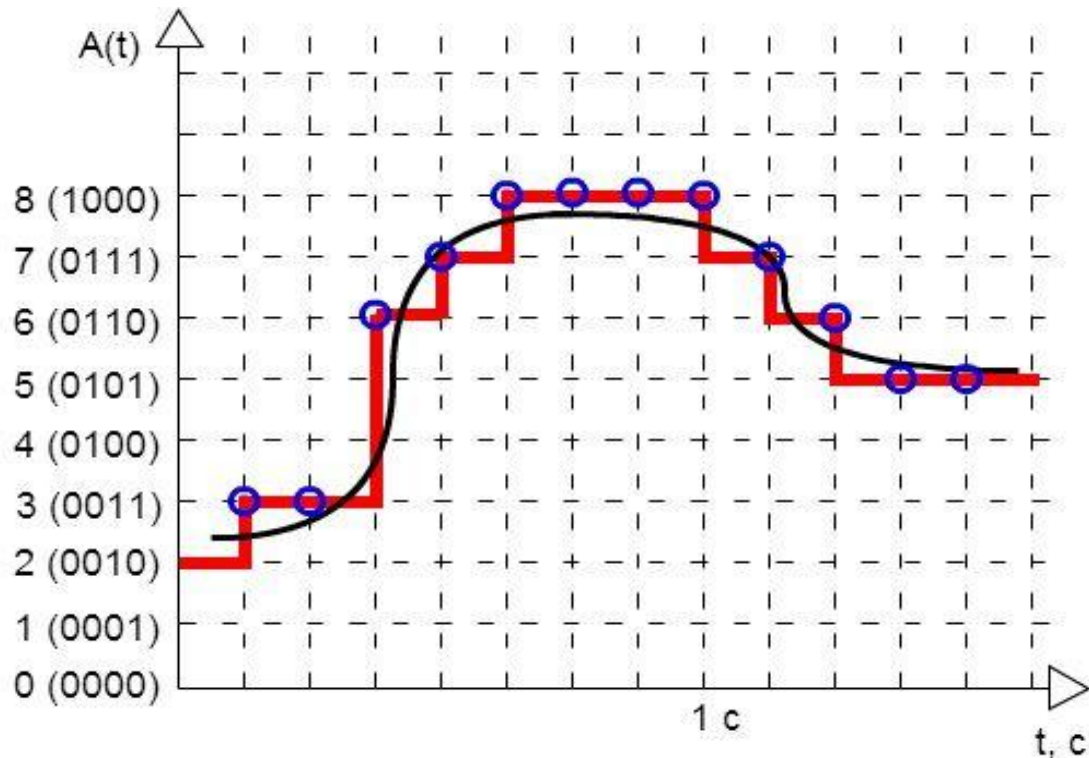
Звук — физическое явление, представляющее собой распространение механических колебаний в упругой среде в виде упругих волн .



Цифровое представление звука

Для того чтобы компьютер мог обрабатывать звук, непрерывный звуковой сигнал должен быть превращен в последовательность электрических импульсов (двоичных нулей и единиц).

Каждой «ступеньке» присваивается значение уровня громкости звука, его код (1, 2, 3 и так далее).



Информационный объем оцифрованного звука

$$2^b = N$$

$$I = H \times t \times b$$

N – количество уровней громкости

b – глубина кодирования звука, разрядность квантования

H – частота дискретизации

I – информационный объем звукового файла

Задача

Оценить информационный объем
цифрового стерео звукового
файла длительностью звучания 1
минута при среднем качестве
звука
(16 битов, 24 кГц).

Вывести ответ в мегабайтах.

Решение:

Дано:

Сtereo

$t = 1$ мин

$H = 24$ кГц

$b = 16$ бит

$$I = (H * t * b) * \text{Количество каналов}$$

$$I = (24.000 \text{ Гц} * 60 \text{ с} * 16 \text{ бит}) * 2$$

$$I = 46.080.000 \text{ бит}$$

$$I = 46.080.000 / 8 \text{ байт} = 5.760.000 \text{ байт}$$

$$I = 5.760.000 / 1024 \text{ байт} = 5.625 \text{ кбайт}$$

$$I = 5.625 / 1024 \text{ байт} \approx 5,5 \text{ Мбайт}$$

Ответ: 5,5 Мбайт.

Найти:

I (мб) - ?

Кодирование текстовых данных

СИМВОЛЫ	Число кодов
Цифры 0-9	10
Знаки препинания, математические символы, специальные символы (\$, # и т.д.)	34
Прописные и строчные латинские буквы	26+26=52
Управляющие символы	32
Итого:	128

$$(2^7 = 128)$$

каждый текстовый символ

(из 128 + нац. алфавит и псевдографические символы) кодируется 8-ми битовым двоичным числом.

$$2^8 = 256 \text{ символов.}$$

Стандарты

Институт стандартизации США (ANSI - *American National Standard Institute*) ввел в действие систему кодирования **ASCII** (*American Standard Code for Information Interchange* - стандартный код информационного обмена США).

В системе ASCII закреплены **две таблицы кодирования** - базовая и расширенная. **Базовая** таблица закрепляет значения кодов от 0 до 127, а **расширенная** относится к символам с номерами от 128 до 255.

Первые 32 кода базовой таблицы, начиная с нулевого, отданы **производителям аппаратных средств**. В этой области размещаются так называемые **управляющие коды**, которым не соответствуют никакие символы языков, и, соответственно, эти коды не выводятся ни на экран, ни на устройства печати.

Начиная с кода **32 по код 127** размещены коды символов английского алфавита, знаков препинания, цифр, арифметических действий и некоторых вспомогательных символов.

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

Отечественной версией кода ASCII является код **КОИ – 7** (код обмена информацией, семизначный).

Только в России можно указать три действующих стандарта, кодировки:

Windows – 1251 (кодировка введена фирмой Microsoft для символов русского языка);

КОИ -8 (код обмена информацией, восьмизначный);

ISO (International Standard Organization – международный институт стандартизации) – международный стандарт, в котором предусмотрена кодировка символов русского алфавита.

Увеличение числа разрядов для кодирования символов в два раза позволит увеличить число кодируемых символов до $2^{16} = 65536$ – этого вполне достаточно для размещения в одной таблице символов большинства языков планеты. Такая система, основанная на 16-ти разрядном кодировании, получила название *универсальной* – **UNICODE**.

примеры

