

Информация и информационные процессы Кодирование информации Системы счисления



Предмет информатики

Информатика – фундаментальная естественная наука, изучающая общие свойства информации, процессы, методы и средства ее обработки (сбор, хранение, преобразование, передача)

Теоретическая

Выявляет общие законы и принципы

Базируется на теории вероятностей и дискретной математике

Прикладная

Обеспечивает создание и использование информационных систем для решения практических задач

Понятие информации

Информация — сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые уменьшают степень неопределенности, неполноты знаний

Информация – общенаучное понятие, включающее в себя обмен сведениями между людьми, обмен сигналами в природе и в технике

Данные — представление фактов и идей в формализованном виде, пригодном для использования в некотором информационном процессе, результат фиксации информации на каком-либо материальном носителе

Данные могут рассматриваться как записанные наблюдения, которые не используются, а только хранятся. Если появляется возможность использовать эти данные для уменьшения неопределенности, они превращаются в информацию

Информационный процесс — процесс, связанный операциями над информацией, в ходе которых может измениться *содержание информации* или ее *форма*.

Основные процессы — **получение, передача, хранение, обработка**

Сообщение. Сигнал

Абстрактная информация всегда представляется в виде конкретного сообщения.

Сообщение – материальная оболочка для представления и передачи информации

Носитель информации - материальный объект или среда, которые служат для представления или передачи информации

Изменение характеристики носителя, которое используется для представления информации, называется **сигналом**, а значение этой характеристики – **параметром сигнала**

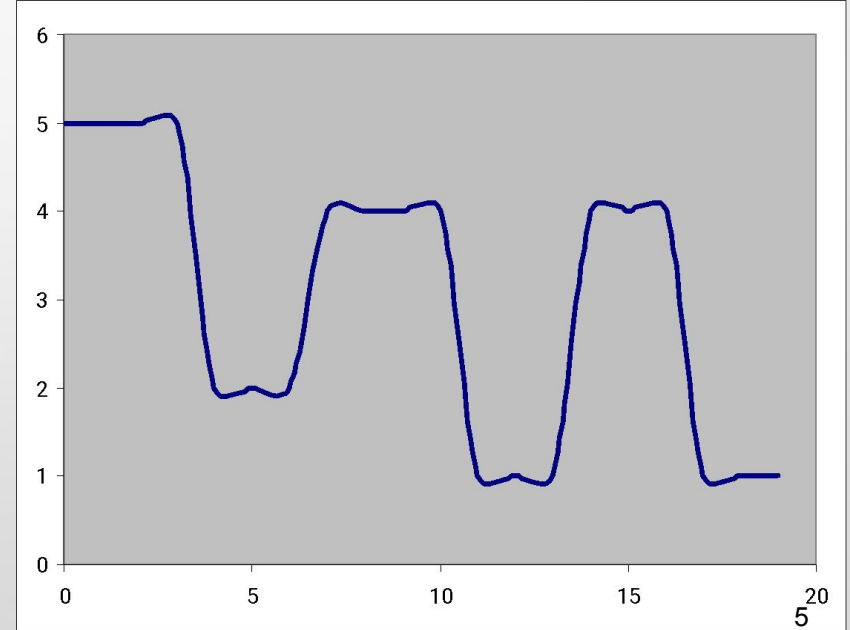
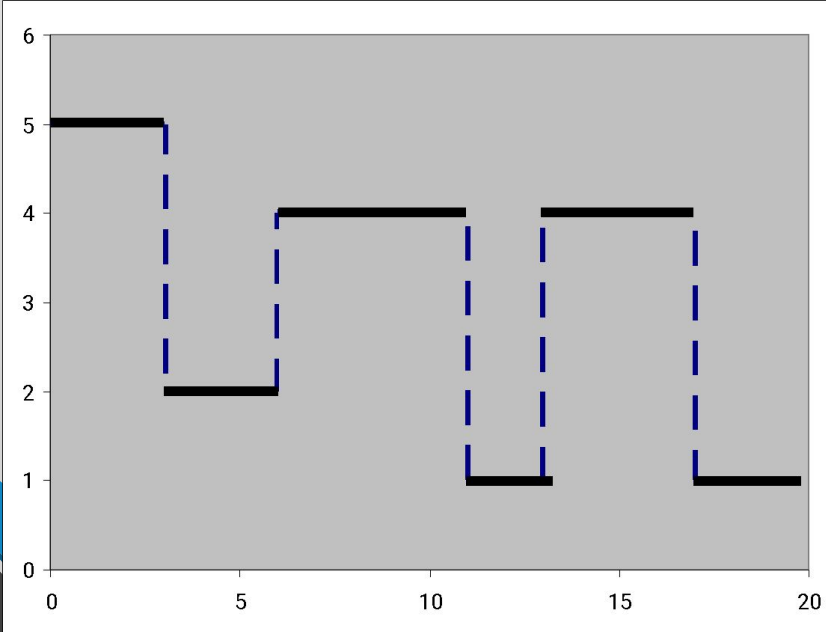
Последовательность сигналов – **сообщение**

Соответствие между сообщением и содержащейся в нем информацией называется **правилом интерпретации сообщения**

Виды сигналов

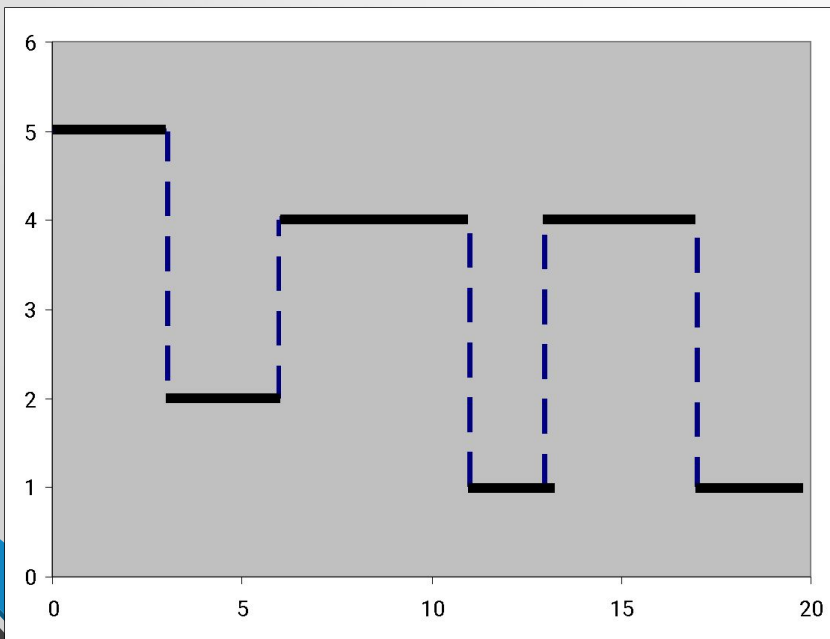
Сигнал называется **дискретным** (или **цифровым**), если его параметр может принимать конечное число значений

Сигнал называется **непрерывным** (или **аналоговым**), если его параметр может принимать любое значение в пределах некоторого интервала



Виды сигналов

Сигнал называется **дискретным** (или **цифровым**), если его параметр может принимать конечное число значений



Знак – элемент некоторого конечного множества отличных друг от друга сущностей

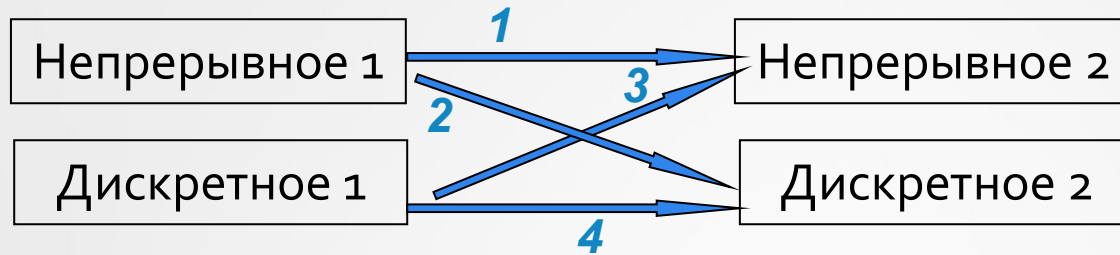
Алфавит – упорядоченная совокупность знаков

Порядок следования знаков в алфавите называется **лексикографическим**

Минимальный алфавит содержит два знака и называется **двоичным**

Эти знаки принято обозначать **0** и **1**

Преобразование сообщений



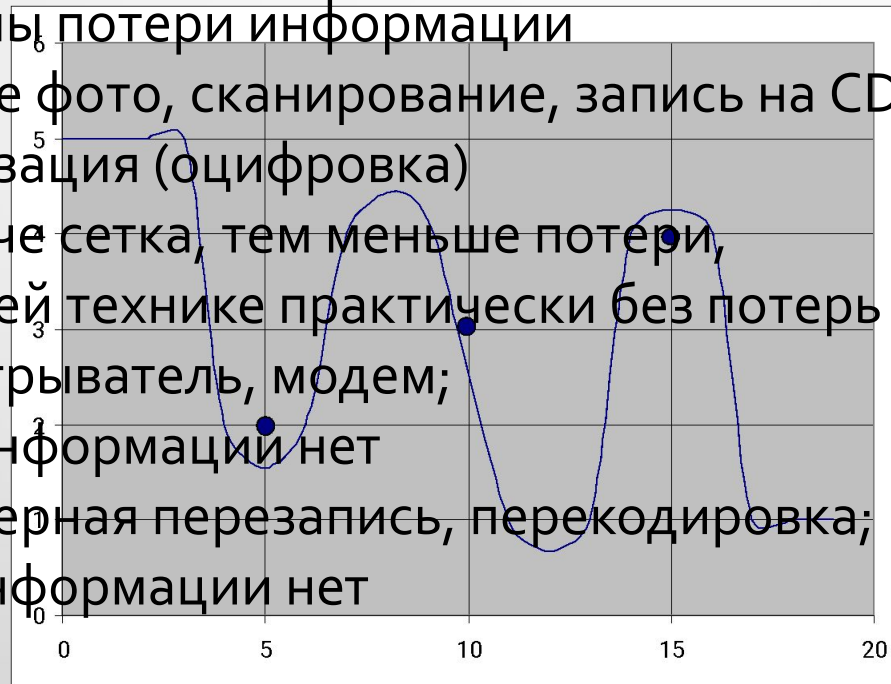
1 – микрофон, магнитофон, радиоприемник;
неизбежны потери информации

2 – цифровое фото, сканирование, запись на CD;
дискретизация (оцифровка)

чем мельче сетка, тем меньше потери,
на хорошей технике практически без потерь

3 – CD-проигрыватель, модем;
потерь информации нет

4 – компьютерная перезапись, перекодировка;
потерь информации нет



Преимущества дискретной формы представления информации

- Возможность передачи без потерь и искажений
- Простота, надежность и относительная дешевизна устройств по обработке информации
- Универсальность устройств, обусловленная возможностью приведения любых дискретных сообщений к единому алфавиту.

В качестве такого единого алфавита принят двоичный алфавит, на котором базируется вся компьютерная техника

Количество информации

В информатике измерению подвергается информация, представленная дискретным сигналом. При этом различают следующие подходы

Структурный – измеряет количество информации простым подсчетом информационных элементов, составляющих сообщение. Применяется для оценки возможностей запоминающих устройств, объемов передаваемых сообщений, инструментов кодирования без учета статистических характеристик их эксплуатации

Статистический – учитывает вероятность появления сообщений: более информативным считается сообщение, которое менее вероятно, т.е. менее всего ожидалось. Применяется при оценке значимости получаемой информации

Семантический – учитывает целесообразность и полезность информации. Применяется при оценке эффективности получаемой информации и ее соответствия реальности

Статистический (вероятностный) подход к измерению количества информации

Информация – это снятие неопределенности, связанной с наступлением некоторого события (уменьшение энтропии)
Чем меньше вероятность события, тем больше информации несет сообщение о его наступлении

1 бит – количество информации о наступлении одного из двух равновероятных событий

Количество информации, которое вмещает один символ **N**-элементного алфавита, определяется по **формуле Хартли (1928 г):**

$$k = \log_2 N$$



Ральф Хартли

Статистический (вероятностный) подход к измерению количества информации

Пусть мы имеем алфавит, состоящий из N символов, с частотной характеристикой P_1, P_2, \dots, P_N , где P_i – вероятность появления i -го символа.

Все вероятности неотрицательны и их сумма равна 1.

Тогда средний информационный вес символа (количество информации, содержащееся в символе)

такого алфавита выражается **формулой Шеннона (1948 г):**

$$H = -(P_1 \log_2 (1/P_1) + P_2 \log_2 (1/P_2) + \dots + P_N \log_2 (1/P_N))$$



Клод Шеннон

Структурный (объемный, технический) подход к измерению количества информации

В качестве меры количества информации принимается длина сообщения, записанного с использованием двоичного алфавита

В качестве основной единицы измерения принимается **1 бит** – сообщение длиной 1 символ двоичного алфавита

Производные единицы

1 байт = 8 бит – 4 варианта (00, 01, 10, 11)

1 Кбайт = 1024 байт – 8 бит – 68 вариантов (2^3)

1 Мбайт = 1024 Кбайт – 4 бита – Кбайт вариантов

1 Гбайт = 1024 Мбайт – 5 битов – Мбайт вариантов (телеграфный код)

1 Тбайт = 1024 Гбайт – 6 битов – Гбайт вариантов

1 Пбайт = 1024 Тбайт – 7 битов – Тбайт вариантов (достаточно для одного алфавита и специальных символов)

8 битов – 256 вариантов (достаточно для двух алфавитов – латинского и национального)

Формы представления информации

- текстовая
- графическая
- звуковая
- видео

Иногда отдельно выделяется

- числовая
- табличная
- музыкальная
- комбинированная
- ...

Выделение форм имеет условный характер, т.к. одна и та же информация может быть представлена посредством различных сообщений, в том числе и отличающихся характером сигналов

Кодирование текстовой информации

В персональных компьютерах и телекоммуникационных системах принят международный стандарт кодирования **ASCII** (American Standard Code for Information Interchange) или **ANSI** (American National Standards Institute)

В этом стандарте 1 символ = 1 байт (8 бит)

Он охватывает первую часть кодовой таблицы (коды от 0 до 127, первый бит 0). Это латинский алфавит, цифры, знаки, специальные символы и др.

Вторая часть кодовой таблицы (коды от 128 до 255) используется для национальных алфавитов и имеет свой стандарт. Например, КОИ-8 или ISO для кириллицы

В настоящее время все чаще используется стандарт **Unicode**:
1 символ = 2 байта (16 бит)

Это позволяет закодировать $2^{16} = 65536$ различных символов и использовать единую кодовую таблицу для большинства популярных алфавитов

Существует несколько стандартов Юникода.
1 символ может потребовать до 4 байтов

Знак	Код
пробел	00100000
A	01000001
Z	01011010
0	00110000
9	00111001
[Esc]	00011011
[Enter]	00001101

Кодирование числовой информации. Системы счисления

Система счисления – способ записи чисел с помощью заданного набора знаков (цифр)

Позиционными называются системы счисления, в которых значение каждой цифры в записи числа определяется ее положением (позицией) в ряду других цифр

Привычной для нас является десятичная система счисления, в которой для записи чисел используется 10 цифр (от 0 до 9)

$$12325 = 5 * 10^0 + 2 * 10^1 + 3 * 10^2 + 2 * 10^3 + 1 * 10^4$$

Компьютерное представление чисел основано на двоичном алфавите, поэтому использует двоичную систему счисления с цифрами 0 и 1

$$10101 = 1 * 2^0 + 0 * 2^1 + 1 * 2^2 + 0 * 2^3 + 1 * 2^4$$

Возможны системы счисления с любыми основаниями, но на практике кроме десятичной используется восьмеричная и шестнадцатеричная системы

Перевод из одной системы счисления в другую



perevod2_10.swf



10

Perevod.swf

Шестнадцатеричная система счисления

Набор цифр: 0, 1, ..., 9, A, B, C, D, E, F

Соответствие между двоичными, десятичными и шестнадцатеричными числами

Каждый разряд умножается на степень числа 16, например

$$\begin{aligned} A3B_{16} &= 11 \cdot 16^0 + 3 \cdot 16^1 + 10 \cdot 16^2 = \\ &= 11 + 48 + 2560 = 2619_{10} \end{aligned}$$

Перевод из шестнадцатеричной системы в двоичную выполняется поразрядно:

$$\text{Например, } A1B_{16} = 1010\ 0001\ 1011_2$$

Удобна для сокращения записи двоичных чисел

В стандартном калькуляторе в режиме ПРОГРАММИСТ есть возможность работы с двоичными, шестнадцатеричными и восьмеричными числами

Десят	Двоич	Шестн
0	0000	0
1	0001	1
2	0010	2
3	0011	3
4	0100	4
5	0101	5
6	0110	6
7	0111	7
8	1000	8
9	1001	9
10	1010	A
11	1011	B
12	1100	C
13	1101	D
14	1110	E
15	1111	F

HEX	0
DEC	0
OCT	0
BIN	0

Компьютерная арифметика

Для хранения числа отводится ограниченная память, поэтому диапазон чисел конечен. Выход за пределы этого диапазона вызывает сообщения об ошибке

Целые и вещественные числа представляются и обрабатываются в компьютере по-разному



Для хранения целого числа отводится, как правило, 4 байта (32 бита) памяти

Сложение неотрицательных целых чисел выполняется по правилам двоичной арифметики, например

Отрицательные числа хранятся в дополнительном коде: выполняется инверсия (0 заменяется на 1, 1 на 0), потом к результату прибавляется 1



Вычитание выполняется как сложение с отрицательным числом с отбрасыванием старшего разряда, например

Представление вещественных чисел

Вещественное число представляется в нормализованной форме:
 $X = \pm M * 10^{\pm k}$

M называется мантиссой,
 $0,1 \leq M < 1$,
k называется порядком

Число	Нормализ. форма	Мантисса	Порядок
123,45	$0,12345 * 10^3$	0,12345	3
-12000	$-0,12 * 10^5$	-0,12	5
0,00015	$0,15 * 10^{-3}$	0,15	-3

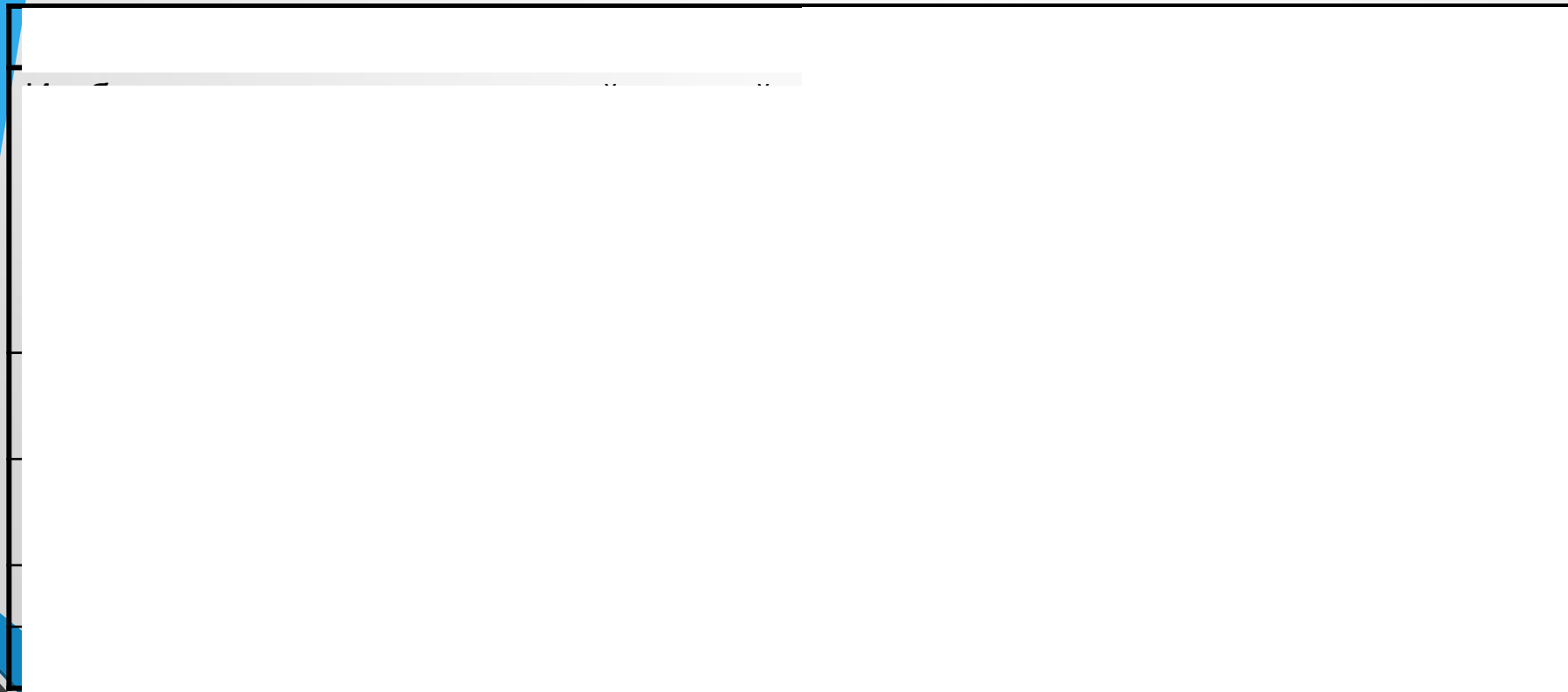
В памяти компьютера мантисса и порядок хранятся как два целых двоичных числа. На мантиссу обычно отводится 4 байта, на порядок 2 байта, т.е. 6 байтов на число

Вещественные вычисления всегда производятся с округлением, целые – точно

Вычисления с вещественными числами выполняются по сложным алгоритмам, поэтому требуют больше времени, чем целочисленные.

Представление графической информации

Для представления графической информации существует два принципа – растровый и векторный



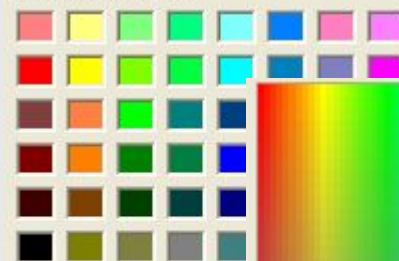
đãñò_ââèò.swf

Цветовые палитры. Глубина цвета

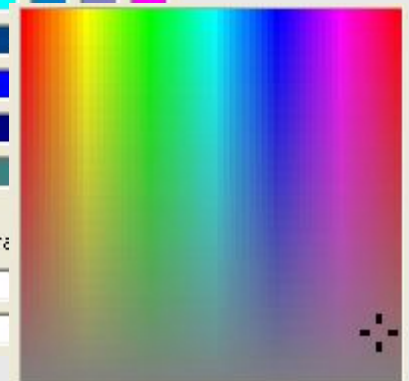
От цветовой палитры зависит качество и объем растрового изображения.

Количество бит, используемых для указания цвета одного пикселя в растровом изображении называется глубиной цвета

Основные цвета:

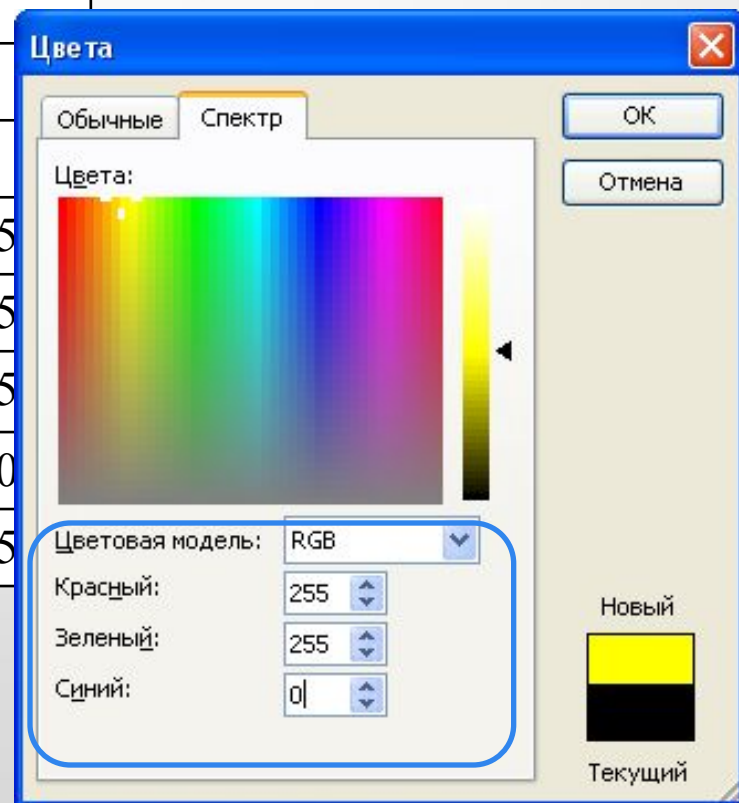


Дополнительные цвета:



Некоторые цвета палитры с 24-битной глубиной цвета

Цвет	Интенсивность		
	R	G	B
Черный	0	0	0
Красный	255	0	0
Зеленый	0	255	0
Синий	0	0	255
Желтый	0	255	255
Фиолетовый	255	0	255
Серый	100	100	100
	255	255	255



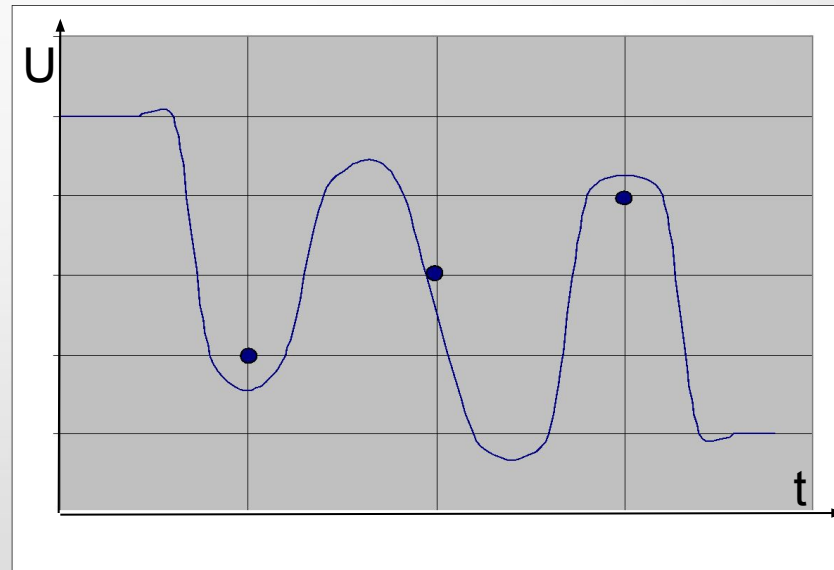
Кодирование звука

Звук – это воспринимаемые человеческим ухом колебания воздуха с частотами от 16 Гц до 20 кГц

Звук является сложной смесью колебаний разной частоты, интенсивности и фазы и представляет собой аналоговый сигнал, который может быть преобразован в изменения напряжения или тока (электромагнитные колебания). Амплитуда колебаний определяет силу звука (громкость), а частота колебаний – высоту звука

Полученный аналоговый сигнал преобразуется в дискретный. Частота дискретизации стандартизирована и составляет 44,1 кГц

Каждое измеренное значение аналогового сигнала представляется 16, 24 или 32 битами



Хранение информации. Типы файлов

- Текстовая («плоский» текст) .txt, .cs, .htm
- Текстовая («размеченный» текст) .docx, .rtf, .pdf
- Графическая .bmp, .jpg, .gif, .tiff, .png, .cdr
- Звуковая .wav, .mp3
- Мультимедийная .avi, .mp4, .pptx
- ...



Примеры

Задание N 15.

Сообщение из 50 символов было записано в 8-битной кодировке Windows-1251. После вставки в текстовый редактор сообщение было перекодировано в 16-битный код Unicode. Количество памяти, занимаемое сообщением, увеличилось на ...

Варианты ответа:

- 50 бит
- 50 байт
- 100 бит
- 400 байт

Задачи

Черно-белое (без градаций серого) растровое графическое изображение имеет размер 20×20 точек. Какой объем памяти займет это изображение?

Решение

Каждая точка монохромной палитры занимает 1 бит.

$$20 \times 20 = 400 \text{ битов}$$

$$400 / 8 = 50 \text{ байтов}$$

Цветное (с палитрой из 256 цветов) растровое графическое изображение имеет размер 20×20 точек. Какой объем памяти займет это изображение?

Решение

$$256 = 2^8$$

Каждая точка монохромной палитры занимает 8 битов или 1 байт.

$$20 \times 20 = 400 \text{ байтов}$$

В процессе преобразования растрового графического изображения количество цветов уменьшилось с 65536 до 16. Во сколько раз уменьшится объем занимаемой им памяти?

Решение

Для 16 цветов на пиксель требуется 4 бита, а для 65536 – 16, значит, объем уменьшится в 4 раза.

256-цветный рисунок содержит 120 байт информации. Из скольких точек он состоит?

Решение

На пиксель требуется 1 байт, значит 120 байтов содержит 120 точек.

Для хранения изображения размером 640 × 320 точек выделено 64 Кбайт памяти. Определите, какое максимальное число цветов допустимо использовать в этом случае.

Решение

$$\frac{64 \times 1024 \times 8}{640 \times 320} = 2,56 \text{ т.е. } 2 \text{ бита на } 1 \text{ пиксель или } 2^2 = 4 \text{ цвета}$$

■ Основные понятия и методы теории информации и кодирования. Сигналы, данные, ... / Позиционные системы счисления

Помощь ?

■ Задание N 16.

Значение выражения $10_{16} - \frac{10_8}{10_2}$ в двоичной системе счисления равно ...

■ Варианты ответа:

- 1010₂
- 0011₂
- 0101₂
- 1100₂



Текст задания:

При сложении восьмеричных чисел $2...7$ и 565 получается восьмеричное число 1064 . Это означает, что в первом слагаемом пропущена цифра...

Варианты ответов:

Должен быть выбран один правильный вариант ответа

7

5

4

6

Текст задания:

При вычитании из восьмеричного числа 601 восьмеричного числа $4\dots 4$, получаем восьмеричное число 125. Это означает, что в вычитаемом пропущена цифра...

Варианты ответов:

Должен быть выбран один правильный вариант ответа

5

7

6

4

Текст задания:

Самое большое число среди перечисленных: 001_2 ; 001_8 ; 001_{10} ; 001_{16} , равно...

Варианты ответов:

Должен быть выбран один правильный вариант ответа

Никакое

001_2

001_{16}

001_8

Текст задания:

Число 123_{10} соответствует:

Варианты ответов:

Должен быть выбран один правильный вариант

$7B_{16}$

1001011_2

$1100.1010.0101_{2-10}$

165_8

Текст задания:

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

для вычисления энтропии системы носит имя

Варианты ответов:

Должен быть выбран один правильный вариант ответа

Шеннона

Ланье

Шнейдера

Харгли

■ **Текст задания:**

Формула Шеннона, учитывающая вероятность p_i наступления i – го события из набора N событий

$$I = - (p_1 \log_2 p_1 + p_2 \log_2 p_2 + \dots + p_N \log_2 p_N),$$

используется для определения ...

■ **Варианты ответов:**

Должен быть выбран один правильный вариант ответа

количества информации

символов в сообщении

количества событий в сообщении

количества наборов информации

1. В студенческой группе 32 человека. Сколько информации содержит сообщение о том, что данный студент назначен старостой группы?

Решение: применим формулу Хартли.

$$N = 32$$

$$H = \log_2 N = 5$$

Ответ: 5 битов

2. В составе поезда 16 вагонов, среди которых К (купейные), П (плацкартные), СВ(спальные). Известно, что сообщение о том, что ваш босс приезжает в вагоне СВ, содержит 3 бита информации. Сколько вагонов СВ в составе поезда?

Решение

1 бит информации дает снятие неопределенности в 2 раза, 3 бита – в $2^3=8$ раз

$$16 / 8 = 2$$

Ответ: 2 вагона СВ

■ Основные понятия и методы теории информации и кодирования. Сигналы, данные, ... / Меры и единицы количества и объема информации

Помощь ?

■ Задание N 19.

В зрительном зале две прямоугольные области зрительских кресел: одна – 6 на 12, а другая – 8 на 4. Минимальное количество бит, которое потребуется для кодирования каждого места в автоматизированной системе, равно ...

■ Варианты ответа:

- 2
- 128
- 104
- 7

Количество мест в зале = $6 \cdot 12 + 8 \cdot 4 = 104$
Ближайшая степень двойки, большая этой величины,
 $128 = 2^7$
7 битов позволяет закодировать 128 мест (достаточно),
6 битов – только $2^6 = 64$ (недостаточно)



■ **Текст задания:**

Выберите вариант, в котором объемы памяти расположены в порядке **возрастания**.

■ **Варианты ответов:**

Должен быть выбран один правильный вариант ответа

15 бит, 2 байта, 20 бит, 1010 байт, 1 Кбайт

15 бит, 20 бит, 2 байта, 1 Кбайт, 1010 байт

15 бит, 20 бит, 2 байта, 1010 байт, 1 Кбайт

15 бит, 2 байта, 20 бит, 1 Кбайт, 1010 байт

■ Блок 1. Тема: Файловая система и файловая структура ОС. Операции с файлами

■ Задание № 4

Общим признаком объединена группа расширений имен файлов ...

■ Варианты ответа

Укажите один вариант ответа

.bmp, jpeg, .cdr, .png

.zip, .com, .ppt, .mp3

.txt, .doc, .rtf, .bat

.bmp, jpeg, .mpeg, .wav

■ Задание № 4

← развернуть

Общим признаком объединена группа расширений имен файлов ...

■ Варианты ответа

Укажите один вариант ответа

.bmp, .jpeg, .cdr, .png

.zip, .com, .ppt, .mp3

.txt, .doc, .rtf, .bat

.bmp, .jpeg, .mpeg, .wav

Лекция окончена
Спасибо за внимание

