

Лекция 1

Весь курс дисциплины «Информатика» состоит из 2-х частей.

1 семестр

22 ч. лекции +

14 ч. практические занятия +

28 ч. лабораторные работы

Форма контроля – зачет

2 семестр

22 ч. лекции +

28 ч. лабораторные работы +

Форма контроля – экзамен

0.1. Список литературы

1. Информатика. Базовый курс. Учебник для ВУЗов. Под ред. Симонович С.В. СПб.: «Питер», 2014 (2013), 649 с.
2. Дьяконов В. П. MATLAB и SIMULINK для радиоинженеров. М.: ДМК, 2011, 976 с.
3. Стефанова И.А. Применение системы MATLAB+Simulink в технике связи. Задачи и методические указания к лабораторным работам по информатике, Самара, ПГУТИ, 2015, 72 с.
4. Стефанова И.А. Моделирование устройств телекоммуникаций в системе MATLAB+Simulink. Учебное пособие. Самара, ПГУТИ, 2016, 150 с.
5. Стефанова И.А. Обработка данных и моделирование в математических пакетах. (Учебно-методическое пособие). Самара: ИУНП ПГУТИ, 2016 г., 64 с.
6. Стефанова И.А. Методы обработки данных в системе Mathcad. Методическое пособие к лабораторным работам. Самара, ПГУТИ, 2015, 51 с.
7. Стефанова И.А. Программирование в системе Mathcad. Методическое пособие к лабораторным работам. Самара, ПГУТИ, 2015, 51 с.
8. Стефанова И.А. Офисный пакет LibreOffice. Часть 1. ПГУТИ, Самара, 2023 г. 95 с.

0.2 Информатика – предмет и задачи

Информатика изучает методы и средства обработки информации.

Методами обработки информации занимаются соответствующие разделы математики, а современным **средством** ее обработки является ЭВМ.

Информатика – наука, изучающая структуру и свойства информации, а также вопросы, связанные с ее сбором, хранением, обработкой, передачей и использованием в различных сферах человеческой деятельности.

Современная информатика коренным образом меняет не только сферу материального производства, но и сферу духовной жизни. Происходит бурный рост объема человеческих знаний, который называют «информационным взрывом».

Предметная область информатики изменяется чрезвычайно **динамично** по сравнению с другими техническими дисциплинами.

Например, мировой объем изданий книг по информатике составляет порядка 10 000 томов в год и полностью обновляется не реже, чем раз в два года.

Информатика включает в себя следующие основные разделы:

- теоретическую информатику,
- вычислительную технику,
- программирование и программное обеспечение,
- информационные системы,
- искусственный интеллект.

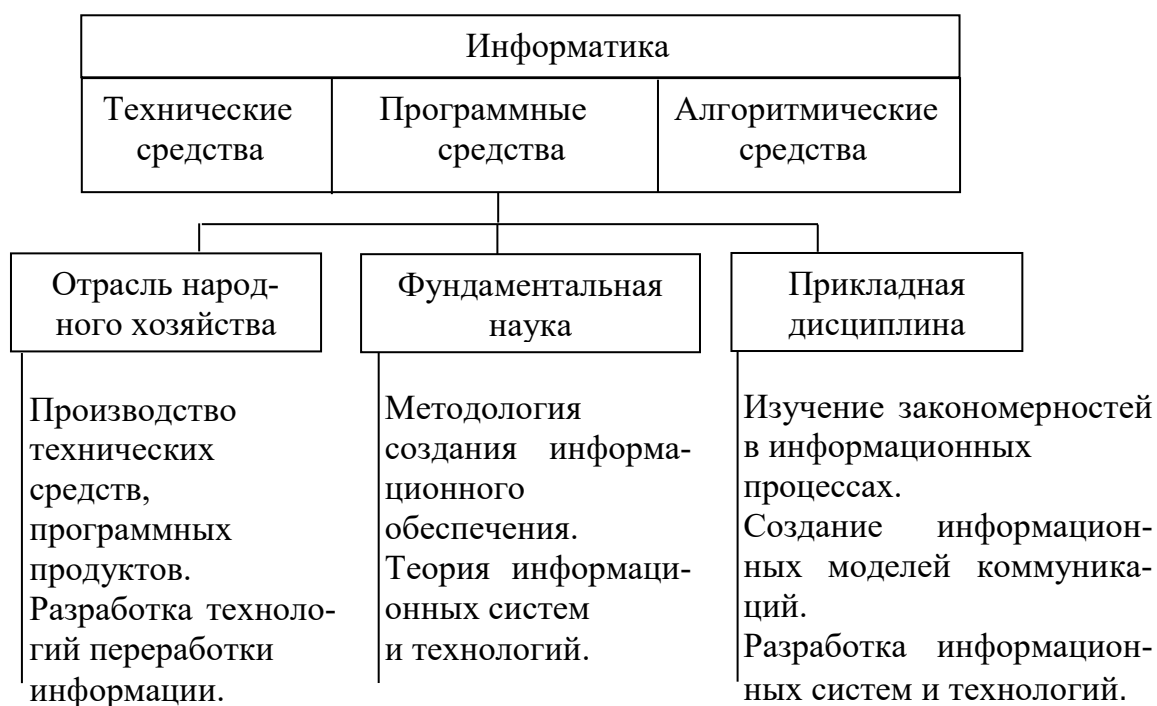
Ядро информатики – **информационная технология** – совокупность конкретных технических и программных средств, с помощью которых выполняются операции по обработке информации во всех сферах жизни и деятельности.

Центральное место в информатике занимает компьютер (*compute* – вычислять) – техническое устройство, предназначенное для обработки информации.

Задачи информатики:

- Исследование информационных процессов любой природы;
- Разработка информационной техники и создание новейшей технологии переработки информации;
- Решение научных и инженерных проблем создания, внедрения и обеспечения использования компьютерной техники и технологии во всех сферах общественной жизни

0.3 Структура информатики



0.4 Информация и ее свойства

Информация – это любые сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые уменьшают имеющуюся о них *степень неопределенности* (энтропию), неполноты знаний.

Однако сведения становятся информацией только тогда, когда они новы для воспринимающего их объекта. Информация динамически меняется.

Материальным носителем информации является *сообщение*.

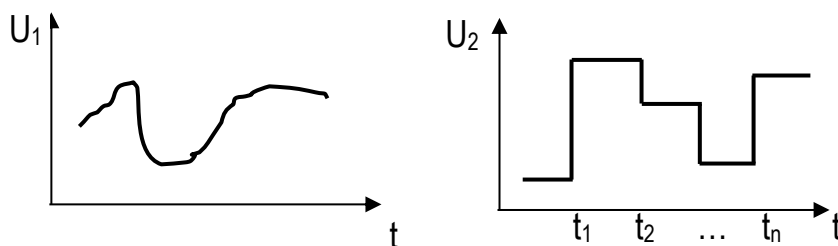
Под обработкой и передачей информации понимается обработка и передача сообщений.

Сообщение – это форма представления информации в виде речи, текста, изображения, графиков и т.п.

Сообщения строятся из элементов, образующих определенный алфавит.

Алфавитом речи являются фонемы, текста – буквы, музыки – тоны, рисунка – точки.

Во времени сообщения могут быть непрерывными аналоговыми и дискретными.



Аналоговые сообщения (U_1) имеют бесконечный алфавит.

Их примером являются речь, музыка, изображение.

Элементы дискретных сообщений (U_2) могут меняться только через определенные равные промежутки времени. Такие сообщения также могут иметь бесконечный алфавит. Их примером являются текст, данные, таблицы.

Важной характеристикой информации является её адекватность.

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

Данные – это зарегистрированные сигналы.

Данные могут рассматриваться как признаки некоторых физических объектов, которые по каким-то причинам не взаимодействуют с другими физическими объектами, а лишь хранятся.

Например, номера телефонов в справочнике.

Однако, как только эти **данные** начинают использоваться, возникают изменения свойств объектов и данные **превращаются в информацию**.

Информацией является используемые данные.

0.5 Меры информации.

Любые сообщения могут быть обработаны на ЭВМ, если они представлены в цифровом виде.

В качестве **алфавита** цифрового сообщения можно принять десять **цифр** привычной для нас десятичной системы счисления (0, 1 . . . 9).

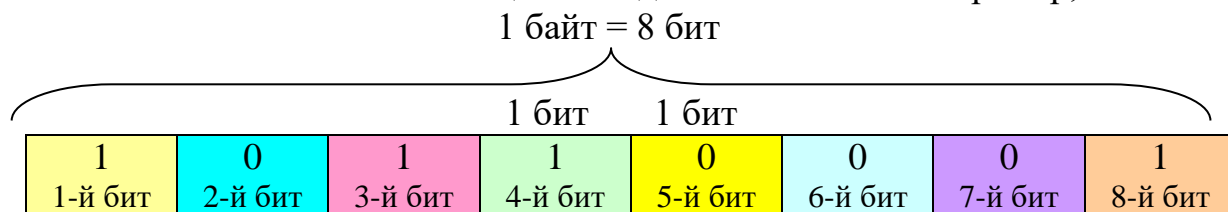
В качестве алфавита всех цифровых систем обработки информации приняты две цифры – 0 и 1 и сочетание нулей и единиц.

Например, десятичное число 12 в двоичной СС имеет вид 1100.

Отдельная позиция в двоичном числе получила название **бит**.

Т.о., бит имеет одно из двух значений – 0 или 1.

Байтом называется 8-позиционное двоичное число. Например, 10110001.



Используют еще более крупные единицы измерения:

$$1 \text{ Кбит} = 2^{10} \text{ бит},$$

$$1 \text{ Кбайт} = 2^{10} \text{ байт} = 8 \cdot 2^{10} \text{ бит},$$

$$1 \text{ Мбайт} = 2^{20} \text{ байт} = 8 \cdot 2^{20} \text{ бит},$$

$$1 \text{ Гбайт} = 2^{30} \text{ байт} = 8 \cdot 2^{30} \text{ бит}.$$

Для измерения информации вводятся два параметра:
количество информации I и объём данных V_D .

Объём данных V_D в сообщении измеряется количеством символом (разрядов) в этом сообщении; [измеряется в битах или байтах].

Количество информации I в сообщении определяется количеством новых знаний, получаемых пользователем и является величиной относительной; [бит; байт]. Количество информации измеряется изменением (уменьшением) неопределенности состояния исследуемой величины.

Коэффициент информативности (содержательность) Y определяется отношением: I / V_D .

Пусть мы имеем некоторые предварительные (априорные) сведения о системе α . Мерой неосведомленности о системе является функция $H(\alpha)$, которая в то же время служит и мерой неопределенности состояния системы.

После получения некоторого сообщения β мы получили некоторую дополнительную информацию $I_\beta(\alpha)$, уменьшившую нашу априорную неосведомленность так, что неопределенность состояния системы после получения сообщения β стала $H_\beta(\alpha)$.

Тогда количество информации $I_\beta(\alpha)$ в системе, полученной в сообщении β , определится как

$$I_\beta(\alpha) = H(\alpha) - H_\beta(\alpha).$$

Следовательно, *энтропия* (степень неопределенности) системы $H(\alpha)$ может рассматриваться как мера недостающей информации.

Один бит информации получает человек, когда он узнает, опаздывает с прибытием нужный ему поезд или нет (Да – 1, нет – 0), был ночью мороз или нет, присутствует на лекции студент Иванов или нет и т. д.

Вероятностный способ измерения количества информации впервые предложил в 1948 г. К. **Шеннон**. В соответствии с идеями К. Шеннона, информация – это сведения, уменьшающие неопределенность (энтропию), существовавшую до их получения.

Под неопределенностью состояния системы связи понимается тот факт, что на приемной стороне получатель информации не знает, какое сообщение пошлет отправитель информации, который находится на передающей стороне системы связи. Лишь после приема сообщения (букв, цифр, символов, звуков, изображения и т.д.) у получателя уменьшается неосведомленность относительно содержания передаваемого сообщения. Иначе говорят: полученная информация уменьшает энтропию системы.

Энтропия системы $H(\alpha)$, имеющая N возможных состояний определяется формулой:

$$H(\alpha) = - \sum_{i=1}^N P_i \cdot \log_2 P_i, \quad // \text{формула Шеннона}$$

где P_i – вероятность того, что система находится в i -м состоянии (вероятность появления i -го символа), N – число состояний (число допустимых символов). Для случая, когда **все состояния системы равновероятны**, т.е. их вероятности равны $P_i = 1/N$, ее энтропия определяется соотношением:

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

Знак минус в формуле поставлен для того, чтобы измерять информацию положительными числами. В противном случае логарифм числа, меньшего единицы, даст отрицательное значение. Заметим, что

$$\sum_{i=1}^N p_i = 1.$$

Установлено, что текстовые (осмысленные) сообщения содержат много избыточной информации. Анализ принятого сообщения позволяет восстановить пропущенный или недостающий фрагменты осмысленного текста.

Например, по принятой фразе по каналу связи «Спартак – чемпи», то легко «догадаться», что, вероятно, полное сообщение было «Спартак – чемпион».

Клод Шеннон установил, что в английских текстах **избыточная** информация достигает 80% от общего объема информации, заключенной в письменном сообщении. Оставшиеся 20% текста являются источником снятия неопределенности, непредсказуемости полученного текста. Именно эти 20% текста несут новую информацию, снимают неопределенность, дают знание.

В русском языке (так же, как и в языках других народов) существует статистическая устойчивость вероятности появления каждой буквы алфавита. Так буква «о» в русских текстах появляется значительно чаще, чем буква «ъ». Кроме того, существует зависимость вероятности появления буквы от того, какая буква в осмысленном тексте ей предшествовала. Очевидно, что вероятность появления сочетания букв «ма» значительно выше, чем у сочетания букв «мъ».

Знание правил построения слов сообщений позволяет предсказывать появление букв раньше, чем они будут переданы по каналу связи. Избыточность позволяет правильно понять содержание осмысленного текста даже при наличии в нем ошибок или пропуска символов.

Избыточность сообщения можно определить по формуле:

$$R = \left(1 - \frac{H}{\log_2 N} \right) \cdot 100\%$$

Информация в компьютере кодируется числовыми кодами. Одно и то же количество разрядов в разных системах счисления может передать разное число состояний отображаемого объекта, что можно представить в виде соотношения

$$N = m^n,$$

где N – число всевозможных отображаемых состояний; m – основание системы счисления (разнообразие символов, применяемых в алфавите); n – число разрядов (символов) в сообщении.

Например, в двоичной системе счисления при 8-разрядном кодировании:

$$N = 2^8 = 256.$$

Если по каналу связи передается n -разрядное сообщение, использующее m различных символов, то при равновероятном появлении любой из комбинаций, **количество информации**, будет равно:

$$I = \log_2(N) = n \cdot \log_2(m)$$

// формула Хартли является частным случаем формулы Шеннона

В двоичной системе счисления ($m = 2$) при 8-разрядном кодировании:

$$I = \log_2(256) = 8 \cdot \log_2(2) = 8 \text{ бит информации}$$

Пример расчета основных показателей информации для фразы:

Враги найдутся сами, а друзей надо искать

Mathcad - [Lec1.xmcd]

Файл Правка Вид Вставка Формат Сервис Символьные операции Окно Справка

Normal Arial 10 B I U

$S := \text{"Враги найдутся сами, а друзей надо искать"}$

Расчет количества информации на 1 двоичный символ

$m := 2$ Основание системы счисления

$n := 8$ Число разрядов на 1 символ

$M := m^n = 256$ Число возможных двоичных кодов

$I_c := \frac{\ln(M)}{\ln(2)} = 8$ бит, количество информации на 1 символ $I_c = \log_2(M)$

Расчет объема данных заданной строки s

$N := \text{strlen}(S) = 41$ Количество символов в строке S

$V := N \cdot n = 328$ бит, Объем данных

$V_{\text{byt}} := \frac{V}{8} = 41$ байт

Расчет количества информации в сообщении строки s

$a := 20$ Количество символов известной информации части строки "Враги найдутся сами,"

$b := N - a = 21$ Количество символов не известной информации строки " а друзей надо искать"

$I := I_c \cdot b = 168$ бит, количество информации на 21 символ новой информации

Расчет коэффициента информативности

$Y := \frac{I}{V} = 0.512$ Коэффициент информативности сообщения S= 51,2 %

Нажмите F1, чтобы открыть справку. АВТО NUM Страница 3

$$H = -\sum_{i=1}^N P_i \cdot \log_2 P_i \quad R = \left(1 - \frac{H}{\log_2 N}\right) \cdot 100\%$$

Mathcad - [Lec1.xmcd]

Файл Правка Вид Вставка Формат Сервис Символьные операции Окно Справка

Расчет энтропии сообщения заданной строки и избыточности сообщения для фразы S "Враги найдутся сами, а друзей надо искать"

Количество повторений символов в строке s

Вероятность появления символов в строке s

Сим :=

k :=

$\sum k = 41$

$p := \frac{k}{N} =$

$\sum p = 1$

Энтропия сообщения на символ

$H := -\sum h = 4.009$ бит / символ

Избыточность сообщения

$R := \left(1 - \frac{H}{\text{ceil}\left(\frac{\ln(N)}{\ln(2)}\right)}\right) \cdot 100 = 33.181 \%$

Нажмите F1, чтобы открыть справку.

АВТО NUM Страт

0.6 Свойства информации.

Объективность – относительное свойство, характеризующее повышение достоверности данных.

Полнота – характеризует качество информации, определяет достаточность данных для принятия решений или создания новых данных на основе имеющихся.

Достоверность – возникает в момент регистрации сигналов.

Адекватность – степень соответствия реальному объективному состоянию дела.

Доступность – мера возможности получить ту или иную информацию.

Актуальность – степень соответствия информации текущему моменту времени.

Своевременность – поступление не позже заранее назначенного времени.

0.7 Понятие о кодировании информации.

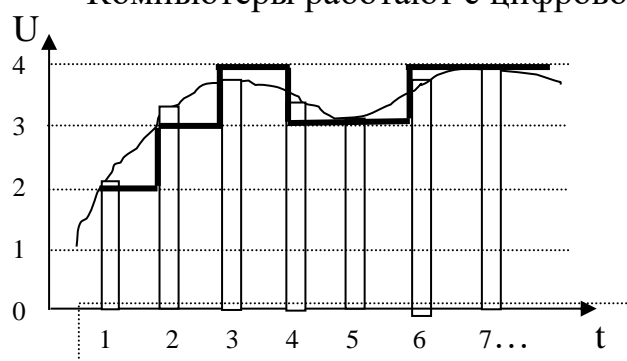
Для хранения информации её кодируют.

Любая информация всегда хранится в виде **кодов**.

Азбука (буквы) – коды письменности, ноты – коды музыкальных звуков.

В виде кодов хранятся и изображения.

Компьютеры работают с цифровой информацией, а не с аналоговой.



4 ... 100

3 ... 011

2 ... 010

1 ... 001

0 ... 000

Двоичный код

$$U_1(t) = 2_{10} = 010_2;$$

$$U_2(t) = U_4(t) = U_5(t) = 3_{10} = 011_2;$$

$$U_3(t) = U_6(t) = U_7(t) = 4_{10} = 100_2;$$

Кодирование чисел

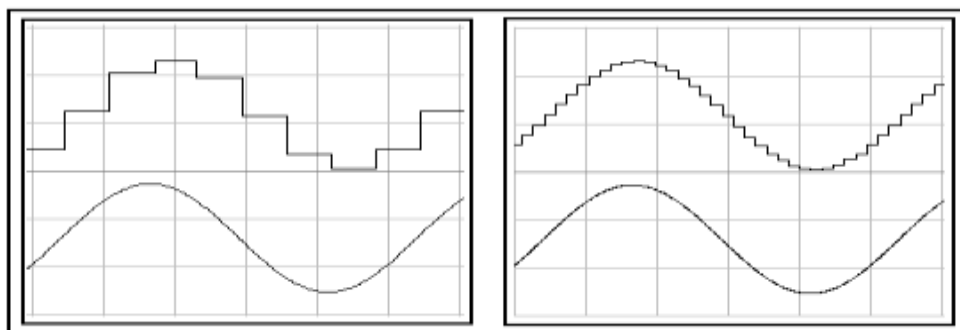
2	3	4	3	3	4	4	...	десятичные числа
010	011	100	011	011	100	100	...	двоичные числа

Для перевода используют **дискретизацию** во времени (в случае аналогового сообщения), **квантование** по значению (уровню) и последующее **кодирование** номера уровня.

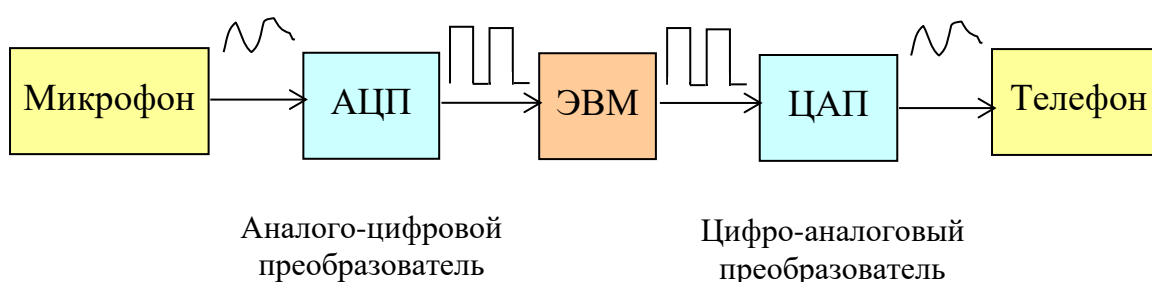
Дискретизация – замена аналогового сигнала совокупностью его отсчётов. От количества отсчетов (частоты) зависит качество преобразования.

Квантование – округление дискретного сигнала до ближайшего разрешенного уровня. От количества уровней квантования зависит качество преобразования.

Кодирование – замена квантованных значений последовательностью кодовых групп. На качество цифрового сигнала сильно влияет также его разрядность



Кодирование музыки и звука. Цифровая аудиозапись



Музыка является звуковыми колебаниям. На качество воспроизведения влияют частота дискретизации и разрешение /количество уровней квантования.

Разрешение – это количество уровней квантования, используемых для замены непрерывного аналогового сигнала цифровым сигналом. Восьмиразрядная выборка (8) позволяет получить только 256 различных уровней квантования цифрового сигнала, а шестнадцатиразрядная выборка (16) – 65 536 уровней.

Например, при записи музыки на компакт-диски используются 16 - разрядные значения и частота дискретизации 44,1 кГц. Музыка содержит множество оттенков, поэтому ее трудно обрабатывать и для ее хранения нужно иметь большую емкость диска.

Кодирование текста

Заключается в установлении соответствия между буквами, цифрами, знаками препинания, заранее установленным кодам.

Для хранения одного символа чаще всего используется 8-разрядная ячейка – один байт, иногда два байта (например, иероглифы, дополнительные символы). В байт можно записать 256 различных чисел, значит, это позволит закодировать 256 различных символов. Соответствие символов и их кодов задаётся в специальной таблице. Коды можно записывать в шестнадцатеричной системе, так как для записи числа из восьми разрядов нужно всего две шестнадцатеричных цифры.

А, Б, Я – 33 заглавных букв Кириллицы,

а, б,я – 33 строчных букв Кириллицы,

А, В, Z – 26 заглавных букв Латиницы,

а, b,z – 26 строчных букв Латиницы,

0, 1,9 – 10 арабских цифр,

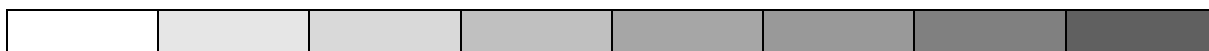
., + : ; ()– знаки препинания, и другие символы.

Всего порядка 150 символов. $2^7=128$, $2^8=256$ (с запасом).

Кодирование изображений

Картинку разбивают вертикальными и горизонтальными линиями на маленькие мозаичные квадратики, получим так называемый растр – двумерный массив квадратиков. Сами квадратики – элементы растра или пиксели (picture's element) – элементы картинки. Цвет каждого пикселя кодируется числом, тогда, задав по порядку номера цветов (слева направо или сверху вниз), можно описать любую картинку. Часть информации неизбежно потеряется, но чем больше растр (мельче пиксели), тем точнее воспроизводится картинка.

Для описания **черно-белых изображений** используются оттенки серого цвета, то есть при кодировании учитывается только яркость. Она описывается одним числом, поэтому для кодирования одного пикселя требуется от 1 до 8 бит: чёрный цвет – 0, белый цвет – $N=2^k-1$, где k – число разрядов, которые отводятся для кодирования цвета. Например, при длине ячейки в 8 бит это $256-1=255$. Человеческий глаз в состоянии различить от 100 до 200 оттенков серого цвета, поэтому **восьми разрядов** вполне хватает.



Цветные изображения воспринимаются нами как сумма трёх основных цветов – красного, зелёного и синего.

Например,

сиреневый = красный + синий;

жёлтый = красный + зелёный;

оранжевый = красный + зелёный, но в другой пропорции.

Поэтому достаточно закодировать цвет тремя числами – яркостью его красной, зелёной и синей составляющих. Этот способ кодирования называется RGB (Red—Green—Blue). Его используют в устройствах, способных излучать свет (мониторы). При рисовании на бумаге действуют другие правила, так как краски сами по себе не испускают свет, а только поглощают некоторые цвета спектра.

Если смешать красную и зелёную краски, то получится коричневый, а не жёлтый цвет. Поэтому при печати цветных изображений используют метод CMY (Cyan—Magenta—Yellow) — голубой, сиреневый, жёлтый цвета.

При таком кодировании:
 красный = сиреневый + жёлтый;
 зелёный = голубой + жёлтый.

Кодирование фильмов

Фильм представляет собой *последовательность* быстро сменяющих друг друга *кадров*, на которых изображены последовательные фазы движения.

Принципы кодирования отдельных кадров известны, и закодировать фильм как последовательность таких кадров несложно.

Звук записывают независимо от изображения. При демонстрации фильма важно только добиться синхронизации звука и изображения (в кино для этого используют хлопушку — по щелчку хлопушки совмещаются звук и изображение).

Закодированный фильм несёт в себе следующую информацию:

- о размере кадра в пикселях и количество используемых цветов;
- частоте и разрешении для звука;
- способе записи звука (покадровый или непрерывный — для всего фильма).

После этого следует последовательность закодированных картинок и звуковых фрагментов.

Технология упаковки информации

Упаковка нужна для сокращения длины сообщения и уменьшения времени для его передачи по каналам связи. Все используемые методы упаковки информации можно разделить на два класса:

- упаковка без потери информации и
- упаковка с потерей информации.

Упаковка без потери информации — исходное сообщение можно точно восстановить по упакованному сообщению. Первый способ такой упаковки представлен в 1952 году Хаффманом. По методу Хаффмана часто встречающиеся символы кодируются короткими последовательностями битов (то есть меньше 8), а редко встречающиеся — длинными (больше 8). В общей сложности, при такой кодировке в среднем требуется меньше 8 бит на символ.

Упаковка с потерей информации — используется для упаковки графических изображений. Для человеческого глаза яркость более существенна, чем информация о цветовом фоне или насыщенности. Поэтому при упаковке можно частично выбросить данные о цвете точки изображения (сохранив только её яркость), а при распаковке брать вместо выброшенного — цвет соседней точки. Распакованная картинка будет отличаться от исходной, но отличие будет практически незаметно. За высокое качество упаковки приходится платить большими затратами времени на распаковку.

Тест для контроля

Информатика изучает...

конструкцию компьютера, способы его включения и выключения
 методы и средства обработки информации
 совокупность программных средств, используемых для работы на ЭВМ
 дисциплины, чтобы использовать их для обработки информации
 свойства информации

Сообщение -

материальный носитель
 свойство информации
 вид данных
 обработанная информация
 форма представления информации

Установите соответствие

речь	фонемы
текст	буквы
музыка	тоны
рисунок	точки
график	линии

Данные -

буквы
 цифры
 линии
 точки
 зарегистрированные сигналы

Бит представляет

логический элемент
 минимальную единицу информации
 константу языка программирования
 элемент алгоритма
 число

Байт равен

1024 бит
 0 бит
 1бит
 8 бит
 10 бит

В 2 Гбайтах содержится

$2 \cdot 2^{30}$ байт

$2 \cdot 2^3$ байт
 $2 \cdot 2^{20}$ байт
 $2 \cdot 2^{31}$ байт
 2^{30} байт

В восьмибитной кодировке слово ИНФОРМАТИКА содержит

88 байт
11 бит
1 байт
11 байт
176 байт

Минимальной единицей измерения количества информации является:

1 бод
1 пиксель
1 байт
1 бит
8 бит