

Информация. Кодирование. Системы счисления

Лекция 1

В настоящее время

- Компьютер стал средством, «орудием труда» сотрудников ОВД и т.п.
- Появились новые виды преступлений, связанные с незаконным хищением, копированием и распространением информации.

Квалификация современного специалиста в области юридической и правоохранительной деятельности включает в себя понимание устройства и основных принципов работы персонального компьютера (ПК), необходимые навыки алгоритмизации и программирования задач, знание современных принципов сбора, хранения и переработки информации, в том числе с использованием телекоммуникаций (компьютерных сетей) и т.п.

Оглавление

- Информатика и информация
- Виды информации
- Свойства информации
- Информационные процессы
- Кодирование информации
- Измерение информации
 - Единица измерения
 - Подходы к измерению информации
- Системы счисления

Информатика и информация

Французский язык:

informatique = *information* + *automatique*

информатика

информация

автоматика

Английский язык:

computer science

компьютер + наука

= наука о компьютерах

Информатика — это основанная на использовании компьютерной техники дисциплина, изучающая структуру и общие свойства информации, а также закономерности и методы ее создания, хранения, поиска, преобразования, передачи и применения в различных сферах человеческой деятельности.

Информатику обычно представляют состоящей из следующих частей:

- *технические средства* – это аппаратура компьютеров;
- *программные средства* – это совокупность всех программ, используемых компьютерами, а также вся область деятельности по их созданию и применению;
- *алгоритмические средства* – это разработка алгоритмов и изучение методов и приемов их построения.

Информатика изучает ...

- информацию и ее свойства
 - процессы
 - хранения...
 - обработки...
 - и передачи информации
- с помощью компьютеров.

Информация

любые сведения об окружающем мире, которые человек получает с помощью органов чувств:

- **глаза** (зрение, 90 процентов информации)
- **уши** (слух)
- **язык** (вкус)
- **нос** (обоняние)
- **кожа** (осязание)

Латинский язык:

informatio – разъяснение, сведения

Информация – это настолько общее и глубокое понятие, что его нельзя объяснить одной фразой. В это слово вкладывается различный смысл в технике, науке и в жизненных ситуациях.

Определения информации

- Информация – сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые воспринимают информационные системы (живые организмы, управляющие машины др.) в процессе жизнедеятельности и работы.
- Информация – это некоторая последовательность символических обозначений (букв, цифр, закодированных графических образов и звуков и т.п.), несущую смысловую нагрузку и представленную в понятном компьютеру виде. Каждый новый символ в такой последовательности символов увеличивает информационный объем сообщения.

Виды информации

- Символ (знак, жест)
- Текст (состоит из символов, важен их порядок)
КОТ ≠ ТОК
- Числовая информация
- Графическая информация (рисунки, картины, чертежи, фото, схемы, карты)
- Звук
- Тактильная информация (осязание)
- Вкус
- Запах

Свойства информации



Информация должна быть

- **объективной** (не зависящей от чьего-либо мнения)
«На улице тепло», «На улице 28°C».
- ~~понятной~~ (английский язык?)
- **полезной** (получатель решает свои задачи)
- **достоверной** (правильной)
дезинформация, помехи, слухи, байки
- ~~актуальной~~ — ~~должна быть важна в данный момент~~
(погода, землетрясение)
устаревшая, ненужная
- ~~полной~~ (~~достаточной~~ для принятия правильного решения)
«Концерт будет вечером», история

Информационные процессы

- **Получение** (через органы чувств)
- **Хранение**
 - мозг, бумага, камень, береста, ...
 - память ПК, дискеты, винчестеры, CD, DVD
- **Обработка**
 - **создание** новой информации
 - **кодирование** – изменение формы, запись в некоторой знаковой системе (в виде кода)
 - **поиск**
 - **сортировка** – расстановка элементов списка в заданном порядке
- **Передача**



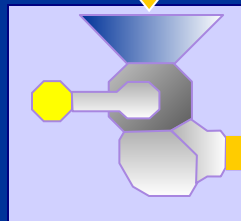
Кодирование информации

Кодирование – это запись информации с помощью некоторой знаковой системы (языка).



Зачем кодируют информацию?

кодирование



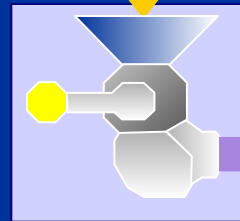
данные (код)

10101001010

передача

Информация передается, обрабатывается и хранится в виде кодов.

борьба с помехами
(специальные способы
кодирования)



данные (код)

1111100010

передача



обработка

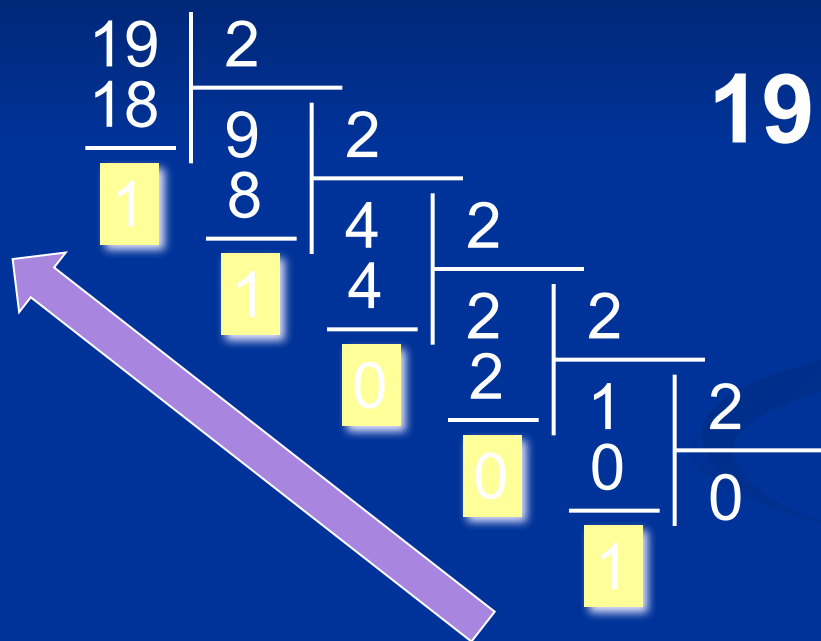
хранение

Кодирование чисел (двоичная система)

Алфавит: 0, 1

Основание (количество цифр): 2

10 → 2



$$19 = 10011_2$$

система
счисления

2 → 10

4 3 2 1 0 разряды

$$10011_2 = 1 \cdot 2^4 + \cancel{0 \cdot 2^3} + \cancel{0 \cdot 2^2} + 1 \cdot 2^1 + 1 \cdot 2^0$$
$$= 16 + 2 + 1 = 19$$

Кодирование символов

Текстовый файл

- на экране (символы)
- в памяти – двоичные коды



1000001_2

65

1000010_2

66

1000011_2

67

1000100_2

68

! В файле хранятся не изображения символов, а их числовые коды в двоичной системе!

А где же хранятся изображения?

Кодирование символов

1. Сколько символов надо использовать одновременно? **256** или 65536 (UNICODE)
2. Сколько места надо выделить на символ:

$$256 = 2^8 \implies 8 \text{ бит на символ}$$

3. Выбрать **256 любых символов** (или 65536) - алфавит.
4. Каждому символу – **уникальный код 0..255** (или 0..65535). Таблица символов:

коды

65

66

67

68

...

A

B

C

D

...

5. Коды – в **двоичную систему**.

Задача. Сколько места в памяти надо выделить для хранения предложения

Привет, Вася!

- считаем все символы, включая знаки препинания и пробелы (здесь **13** символов)
- если нет дополнительной информации, то считаем, что 1 символ занимает **1 байт**
- в кодировке UNICODE 1 символ занимает **2 байта**

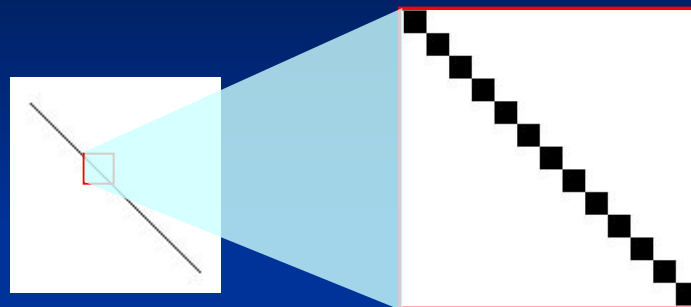
Ответ: 13 байт или 104 бита

(в UNICODE: 26 байт или 208 бит)

Два типа кодирования рисунков

- **растровое кодирование**

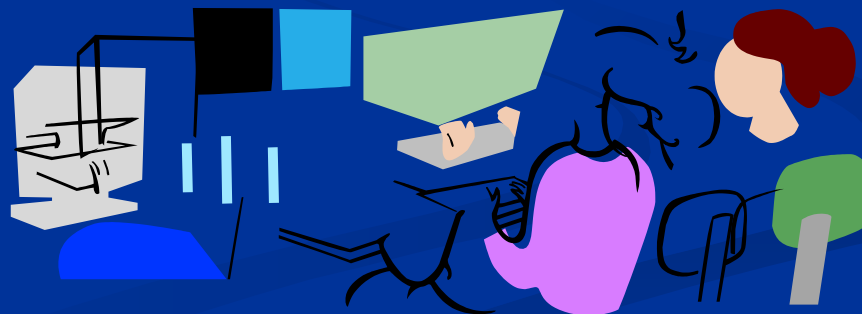
точечный рисунок, состоит из **пикселей**



фотографии, размытые изображения

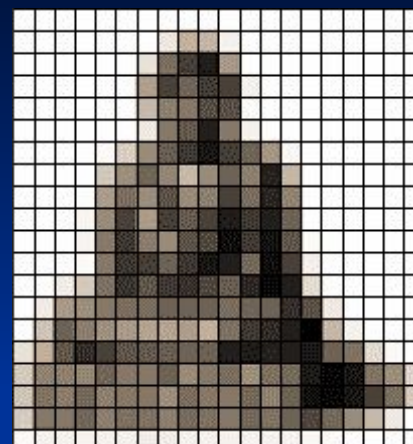
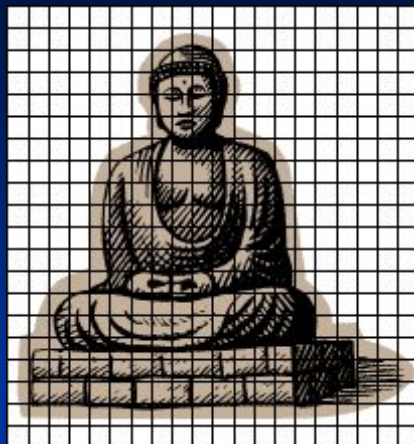
- **векторное кодирование**

рисунок, состоит из **отдельных геометрических фигур**



чертежи, схемы, карты

Растровое кодирование



Шаг 1. Дискретизация:
разбивка на *пиксели*.

Пиксель – это наименьший элемент рисунка, для которого можно независимо установить цвет.

Шаг 2. Для каждого пикселя
определяется **единый цвет**.



Есть потеря информации!

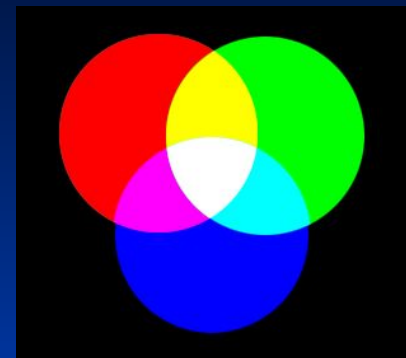
- почему?
- как ее уменьшить?

Разрешение: число пикселей на дюйм, *pixels per inch (ppi)*
экран 96 ppi, печать 300-600 ppi, типография 1200 ppi

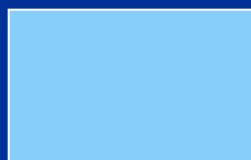
Растровое кодирование (*True Color*)

Шаг 3. От цвета – к числам: модель RGB

цвет = **R** + **G** + **B**
red *green* *blue*
красный зеленый синий
0..255 0..255 0..255



R = 218
G = 164
B = 32



R = 135
G = 206
B = 250

Шаг 4. Числа – в двоичную систему.



Сколько разных цветов можно кодировать?

$256 \cdot 256 \cdot 256 = 16\,777\,216$ (*True Color*)

Глубина
цвета



Сколько памяти нужно для хранения цвета 1 пикселя?

R: $256=2^8$ вариантов, нужно 8 бит = 1 байт

R G B: всего 3 байта

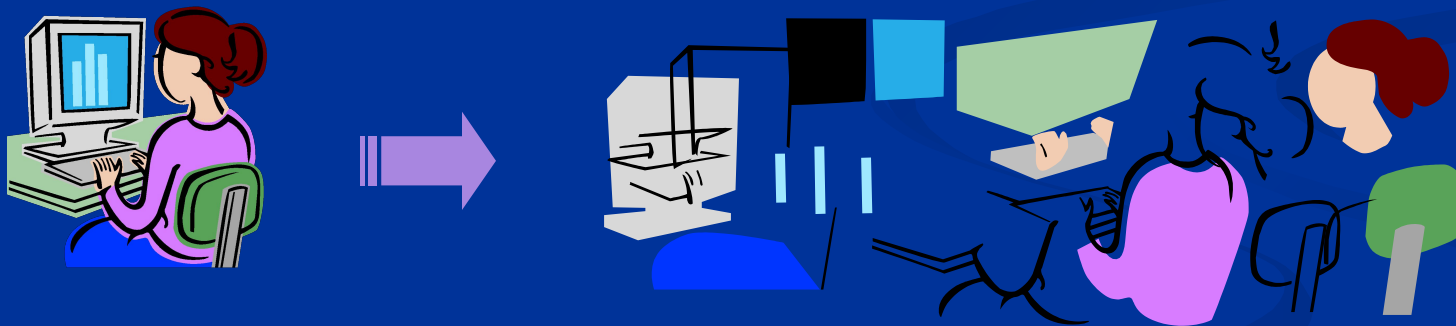
Векторные рисунки

Строятся из геометрических фигур:

- отрезки, ломаные, прямоугольники
- окружности, эллипсы, дуги
- сглаженные линии (кривые Безье)

Для каждой фигуры в памяти хранятся:

- размеры и координаты на рисунке
- цвет и стиль границы
- цвет и стиль заливки (для замкнутых фигур)



Форматы файлов:

- **WMF** (*Windows Metafile*)
- **AI** (*Adobe Illustrator*)
- **CDR** (*CorelDraw*)
- **FH** (*FreeHand*)

Форматы

■ Растровые рисунки:

- BMP
- JPG
- GIF
- PNG

■ Векторные рисунки:

- WMF
- CDR
- AI
- FH

Кодирование

Задача 1. Закодируйте свое имя с помощью азбуки Морзе.

А	• —	П	• — — — •	Ь	— • • —
Б	— • • •	Р	• — •	Ы	— • — — —
В	• — — —	С	• • •	Й	• — — — —
Г	— — — •	Т	—		
Д	— • •	У	• • —	1	• — — — — —
Е	•	Ф	• • — •	2	• • — — — —
Ж	• • • —	Х	• • • •	3	• • • — — —
З	— — — • •	Ц	— • — — •	4	• • • • —
И	• •	Ч	— — — — •	5	• • • • •
К	— • — —	Ш	— — — — —	6	— • • • •
Л	• — — • •	Щ	— — — • —	7	— — — • • •
М	— — —	Э	• • — — • •	8	— — — — • •
Н	— •	Ю	• • — — —	9	— — — — — •
О	— — — —	Я	• — — • —	0	— — — — — —

ВАСЯ

• — — — • — — • • • • — — — —



Код неравномерный, нужен разделитель!

Кодирование

Задача 2. Закодируйте свое имя с помощью кодовой таблицы (*Windows-1251*):

	0	1	2	3	4	5	6	7	8	9	А	В	С	Д	Е	Ф
С	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
Д	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я

В А С Я

ВАСЯ С2 С0 Д1 ДF



Код равномерный, разделитель **НЕ** нужен!

Как измерить информацию?

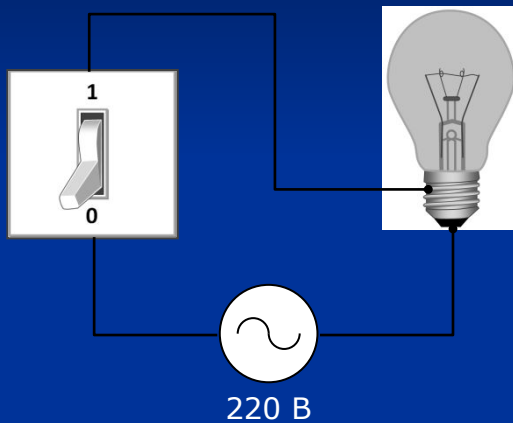
- ❑ Что такое «много информации» и «мало информации»?
- ❑ Как определить, в каком сообщении больше информации?

- количество информации определяется временем ее передачи
- количество информации – это длина сообщения, с помощью которого её можно закодировать.

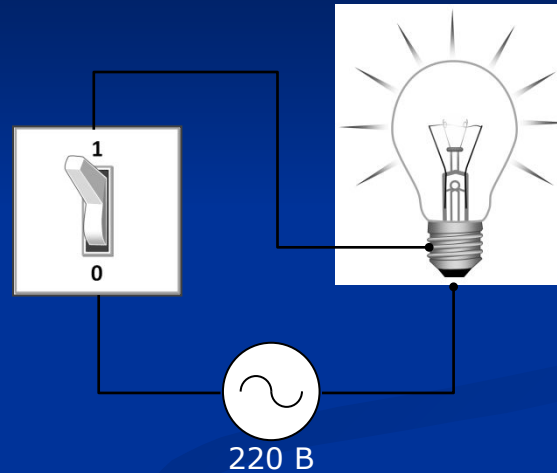
- ❑ От чего зависит длина сообщения?
от алфавита!
- ❑ Какой алфавит выбрать?
абвг...эюя?
abcd...хуz?

Какой код использовать?

использовать тот код, который применяется в компьютерной технике



«0»



«1»

ДВОИЧНЫЙ КОД

Код, в котором используются только два знака, называется **двоичным**. В компьютерах применяется двоичный код.

1 бит – это количество информации, которое можно передать с помощью одного знака в двоичном коде («0» или «1»).

bit = binary digit, двоичная цифра

Единицы измерения

1 бит – это количество информации, которое мы получаем при выборе одного из двух возможных вариантов (вопрос: «Да» или «Нет»?)

Примеры:

Эта стена – зеленая? Да.

Дверь открыта? Нет.

Сегодня выходной? Нет.

Это новый автомобиль? Новый.

Ты будешь чай или кофе? Кофе.

Сколько вариантов?

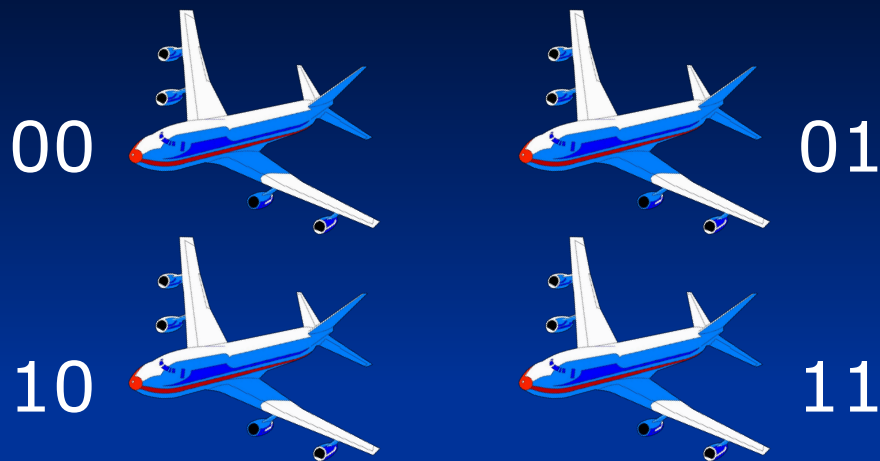
1 бит: **0 1**

2 бита: **00 01**
 10 11

3 бита: **000 001 100 101**
 010 011 110 111

4 бита: 16 вариантов!

Если вариантов больше...



4 варианта – **2** бита

? вариантов – **3** бита

? вариантов – **4** бита

? варианта – **5** бит

? варианта – **6** бит

? вариантов – **7** бит

Если вариантов больше...

Количество вариантов	2	4	8	16	32	64	128	256	512	1024
Количество бит информации	1	2	3	4	5	6	7	8	9	10

6 вариантов – между **4** (2 бита) и **8** (3 бита)

Ответ: количество информации между
2 и 3 битами

Единицы измерения

1 байт (*byte*) = 8 бит

1 Кбайт (килобайт) = 1024 байта

1 Мбайт (мегабайт) = 1024 Кбайт

1 Гбайт (гигабайт) = 1024 Мбайт

1 Тбайт (терабайт) = 1024 Гбайт

1 Пбайт (петабайт) = 1024 Тбайт

2¹⁰

Перевод в другие единицы

$$5 \text{ Кбайт} = 5 \cdot 1024 \text{ байт} = 5120 \text{ байт}$$

$$15 \text{ байт} = 20 \text{ бит} = 120 \text{ бит}$$

$$2048 \text{ Кбайт} = 2048 : 1024 \text{ Мбайт} = 2 \text{ Мбайта}$$

$$1024 \text{ Мбайт} = 1024 : 1024 \text{ Гбайт} = 1 \text{ Гбайт}$$

$$3 \text{ Мбайта} = 3 \cdot 1024 \text{ Кбайт} = 3072 \text{ Кбайта}$$





При вычислении количества информации
выделяют два подхода:

- Алфавитный (технический);
- Вероятностный (энтропийный).

Алфавитный подход

Алфавит – набор знаков, используемых при кодировании информации с помощью некоторого языка.

Примеры:

АБВГДЕЖЗИЙКЛМНОПРСТУФХЦЧШЩЪЫЬЭЮЯ	32
ABCDEFGHIJKLMNOPQRSTUVWXYZ	26
× 0	2
0123456789	10

Мощность алфавита – количество символов.



Все символы несут одинаковую информацию:

информационная
емкость символа

$$I = \log_2 N$$

мощность
алфавита

Задача. Определить объем информации в сообщении
ПРИВЕТВАСЯ
для кодирования которого используется русский алфавит (только заглавные буквы).

Решение:

- считаем все символы (здесь **10** символов)
- мощность алфавита – 32 символа ($32=2^5$)
- 1 символ несет **5 бит** информации

Ответ: $10 \cdot 5 \text{ бит} = 50 \text{ бит}$

Вероятностный подход

Вероятность события – число от 0 до 1, показывающее, как часто случается это событие в большой серии одинаковых опытов.

$p = 0$ событие **никогда** не происходит
(нет неопределенности)

$p = 0,5$ событие происходит в половине случаев (есть неопределенность)

$p = 1$ событие происходит **всегда**
(нет неопределенности)



Полная система событий: одно из N событий обязательно произойдет (и только одно!).

p_i – вероятность выбора i -ого варианта ($i=1, \dots, N$)

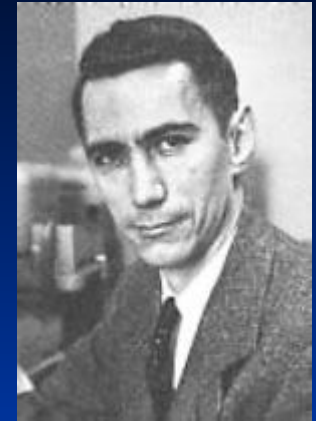
$$0 \leq p_i \leq 1, \quad p_1 + p_2 + \dots + p_N = 1$$

Вероятностный подход

Как посчитать информацию, если варианты не равновероятны?

Клод Шеннон (1916 — 2001)

американский математик и электротехник, один из создателей математической теории информации и криптографии.



если случается менее вероятное событие, мы получаем больше информации.

$0 \leq p_i \leq 1$ – вероятность выбора i -ого варианта ($i=1, \dots, N$)

Если произошло событие i , мы получаем информацию

$$I_i = \log_2 \frac{1}{p_i}$$

Задача 1. В пруду живут 100 рыб, из них 20 карасей, 30 пескарей, а остальные – окуни. *Сколько информации* несет сообщение о том, что рыбак поймал карася (пескаря, окуня), если все рыбы одинаково голодны?

Формула:

$$I_i = \log_2 \frac{1}{p_i}$$

Решение:

карась $p_1 = \frac{20}{100} = 0,2$ $I_1 = -\log_2 0,2 = \log_2 5 \approx 2,32$ бита

пескарь $p_2 = \frac{30}{100} = 0,3$ $I_2 = -\log_2 0,3 \approx \log_2 3,33 \approx 1,74$ бита

окунь $p_3 = \frac{50}{100} = 0,5$ $I_3 = -\log_2 0,5 = \log_2 2 = 1$ бит

Информация и знание



Неопределенность – недостаток знаний (незнание).

- при получении информации знания увеличиваются, неопределенность уменьшается
- чем больше получено информации, тем больше уменьшается неопределенность
- информация – мера уменьшения неопределенности



Как измерить неопределенность?

Формула Шеннона (1948)

Неопределенность (энтропия системы)

$$I = \sum_1^N p_i I_i = p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2} + \dots + p_N \log_2 \frac{1}{p_N}$$

Информация = снятая неопределенность!



Когда неопределенность наибольшая?

Задача. В коробке имеет 50 шаров, из них 40 белых и 10 черных. Очевидно, что при вытаскивании «не глядя» попадание белого или черного шара не равновероятны.

$P_1 = \frac{40}{50}$ - вероятность того, что будет вынут белый шар.

$P_2 = \frac{10}{50}$ - вероятность того, что будет вынут черный шар.

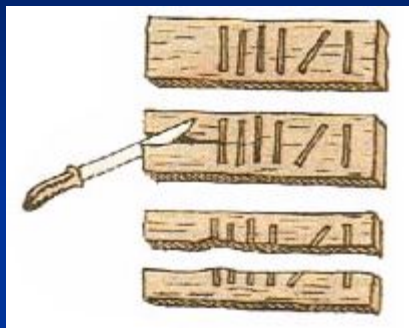
$$S = -(p_1 \log_2 p_1 + p_2 \log_2 p_2) = -\left(\frac{40}{50} \log_2 \frac{40}{50} + \frac{10}{50} \log_2 \frac{10}{50}\right) = -(-0,25754 - 0,46439) = 0,72193$$

Системы счисления

- Система счисления – это способ записи чисел с помощью специальных знаков – цифр.
- Числа:
123, 45678, 1010011, CXL
- Цифры:
0, 1, 2, ... I, V, X, L, ...
- Алфавит – это набор цифр. $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$
- Типы систем счисления:
 - непозиционные – значение цифры не зависит от ее места (*позиции*) в записи числа;
 - позиционные – зависит...

Непозиционные системы

Унарная – одна цифра обозначает единицу (1 день, 1 камень, 1 баран, ...)



Римская:

I – 1 (палец), V – 5 (раскрытая ладонь, 5 пальцев),
X – 10 (две ладони), L – 50,
C – 100 (*Centum*), D – 500 (*Demimille*),
M – 1000 (*Mille*)

Славянская система счисления

алфавитная система счисления (непозиционная)



Часы
Суздальского
Кремля

Позиционные системы

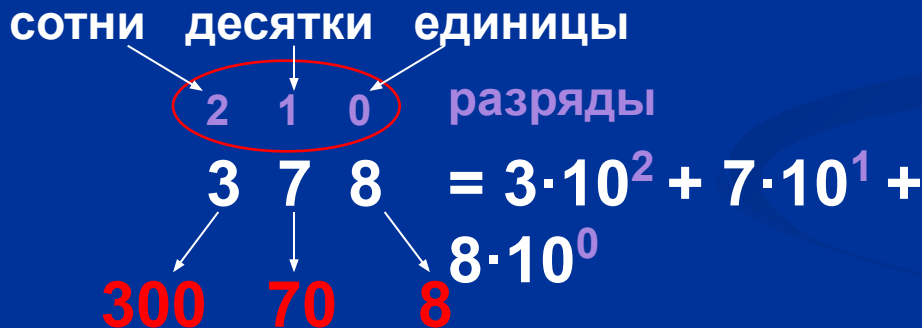
Позиционная система: значение цифры определяется ее позицией в записи числа.

Десятичная система:

первоначально – счет на пальцах
изобретена в Индии, заимствована арабами, завезена в Европу

Алфавит: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9

Основание (количество цифр): 10



Другие позиционные системы:

- двоичная, восьмеричная, шестнадцатеричная (информатика)
- двенадцатеричная (1 фут = 12 дюймов, 1 шиллинг = 12 пенсов)
- двадцатеричная (1 франк = 20 су)
- шестидесятеричная (1 минута = 60 секунд, 1 час = 60 минут)

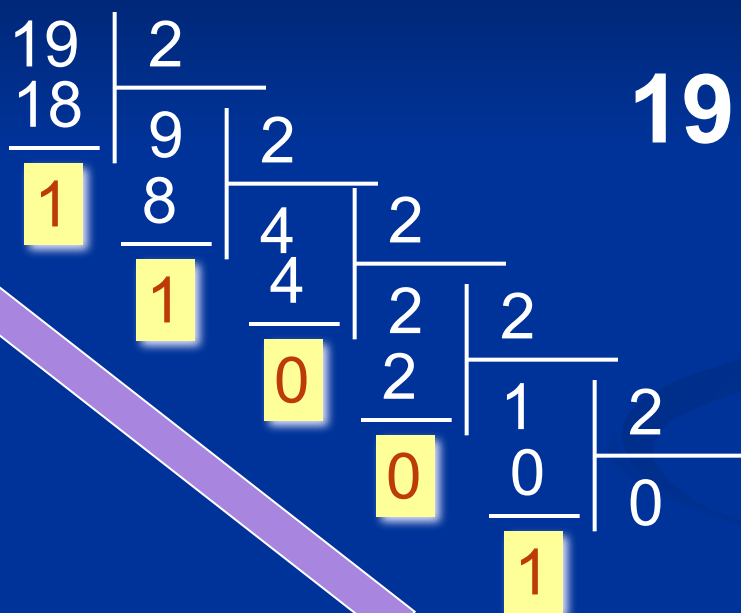
Перевод целых чисел

Двоичная система:

Алфавит: 0, 1

Основание (количество цифр): 2

10 → 2



$$19 = 10011_2$$

система
счисления

2 → 10

4 3 2 1 0 разряды

$$10011_2 = 1 \cdot 2^4 + \cancel{0 \cdot 2^3} + \cancel{0 \cdot 2^2} + 1 \cdot 2^1 + 1 \cdot 2^0$$
$$= 16 + 2 + 1 = 19$$

Перевод дробных чисел

10 → 2

$$0,375 = 0,011_2$$

$$\begin{array}{r} \times \quad 2 \\ \hline \end{array}$$

$$0,750$$

$$0,75$$

$$\begin{array}{r} \times \quad 2 \\ \hline \end{array}$$

$$1,50$$

$$0,5$$

$$\begin{array}{r} \times \quad 2 \\ \hline \end{array}$$

$$1,0$$



$$0,7 = ?$$

$$0,7 = 0,101100110\dots$$

$$= 0,1(0110)_2$$

Многие дробные числа нельзя представить в виде **конечных** двоичных дробей.

Для их точного хранения требуется **бесконечное** число разрядов.

Большинство дробных чисел хранится в памяти с ошибкой.

2 → 10

$$2^{-2} = \frac{1}{2^2} = 0,25$$

2 1 0 -1 -2 -3 разряды

$$101,011_2$$

$$= 1 \cdot 2^2 + 1 \cdot 2^0 + 1 \cdot 2^{-2} + 1 \cdot 2^{-3}$$

$$= 4 + 1 + 0,25 + 0,125 = 5,375$$

Арифметические операции

сложение

$$0+0=0 \quad 0+1=1$$

$$1+0=1 \quad 1+1=10_2$$

$$1 + 1 + 1 = 11_2$$

перенос

вычитание

$$0-0=0 \quad 1-1=0$$

$$1-0=1 \quad 10_2-1=1$$

заем

$$\begin{array}{r} \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ 10110_2 \\ + 111011_2 \\ \hline 1010001_2 \end{array}$$

$$\begin{array}{r} \cdot \quad \cdot \\ 01110_2 \quad 010_2 \\ \cancel{1000101}_2 \\ - \quad 11011_2 \\ \hline 0101010_2 \end{array}$$

Примеры:

$$\begin{array}{r} 101101_2 \\ + 11111_2 \\ \hline \end{array}$$

$$\begin{array}{r} 10111_2 \\ + 101110_2 \\ \hline \end{array}$$

$$\begin{array}{r} 111011_2 \\ + 11011_2 \\ \hline \end{array}$$

$$\begin{array}{r} 111011_2 \\ + 10011_2 \\ \hline \end{array}$$

Примеры:

$$\begin{array}{r} 101101_2 \\ - 11111_2 \\ \hline \end{array}$$

$$\begin{array}{r} 110011_2 \\ - 10101_2 \\ \hline \end{array}$$

$$\begin{array}{r} 11011_2 \\ - 110101_2 \\ \hline \end{array}$$



$$\begin{array}{r} 110101_2 \\ - 11011_2 \\ \hline \end{array}$$

Арифметические операции

умножение

$$\begin{array}{r} 10101_2 \\ \times 10 \\ \hline 1_210101_2 \\ + 10101_2 \\ \hline 1101001_2 \end{array}$$

деление

$$\begin{array}{r} 10101_2 \big| 111_2 \\ - 111_2 \\ \hline 111_2 \\ - 111_2 \\ \hline 0 \end{array}$$

11₂

Восьмеричная система

Основание (количество цифр): 8

Алфавит: 0, 1, 2, 3, 4, 5, 6, 7

10 → 8

$$\begin{array}{r|l} 100 & 8 \\ \hline 96 & 12 \\ \hline 4 & 8 \\ \hline & 10 \\ \hline & 0 \\ \hline & 1 \end{array}$$

$$100 = 144_8$$

система
счисления

8 → 10

2 1 0 разряды

$$\begin{aligned} 144_8 &= 1 \cdot 8^2 + 4 \cdot 8^1 + 4 \cdot 8^0 \\ &= 64 + 32 + 4 = 100 \end{aligned}$$

Шестнадцатеричная система

Основание (количество цифр): 16

Алфавит: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, **A**, **B**, **C**, **D**, **E**, **F**
10 11 12 13 14 15

10 → 16

$$\begin{array}{r|l} 107 & 16 \\ \hline 96 & \\ \hline & 6 \\ & \hline & 0 \\ & \hline & 0 \end{array}$$

$$107 = 6B_{16}$$

система
счисления

16 → 10

2 1 0 разряды

$$1C5_{16} = 1 \cdot 16^2 + 12 \cdot 16^1 + 5 \cdot 16^0$$

$$= 256 + 192 + 5 = 453$$