

Модуль «Информация и информационные процессы»

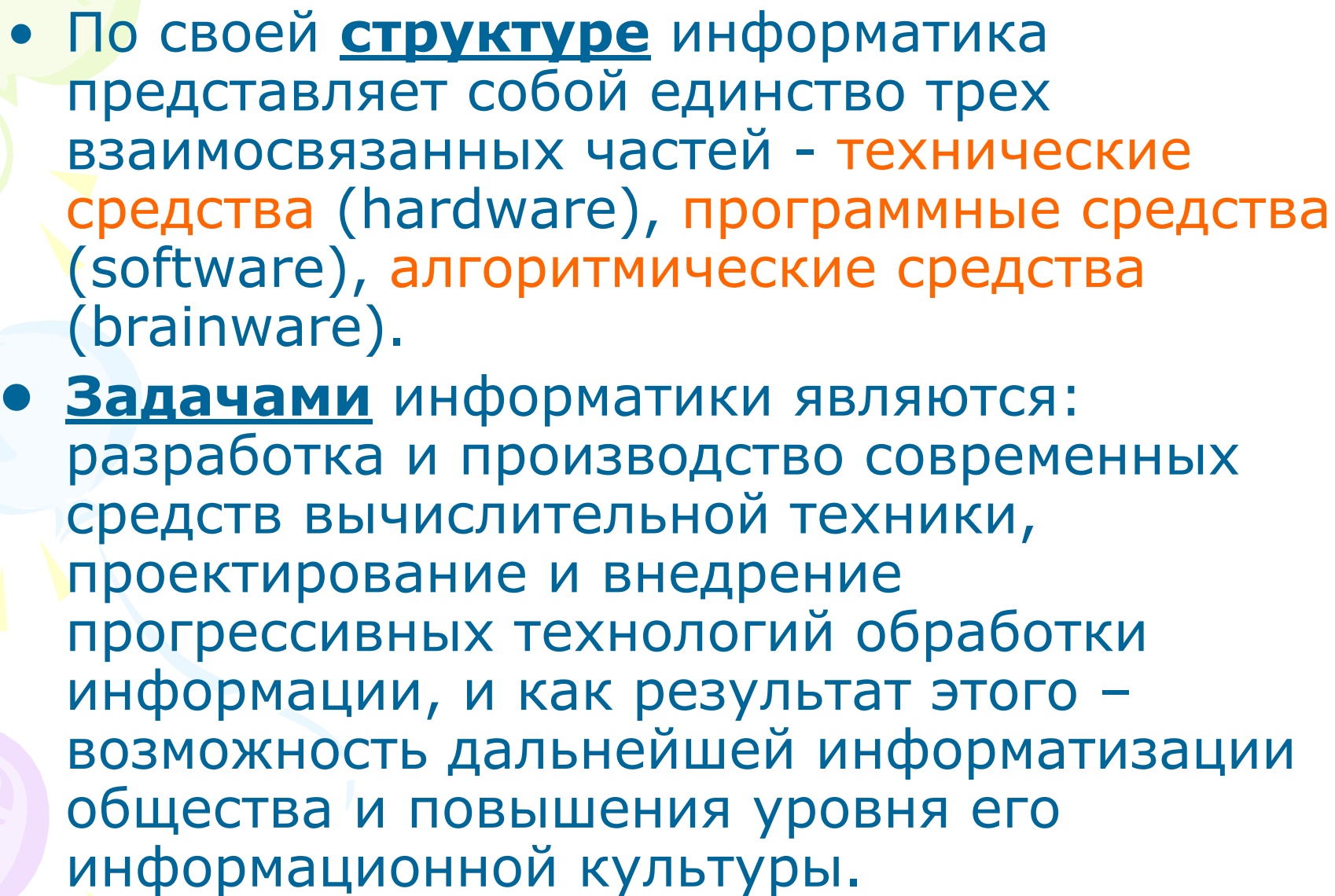
1. Подходы к понятию и измерению информации. Информационные объекты различных видов. Универсальность дискретного (цифрового) представления информации. Представление информации в двоичной системе счисления

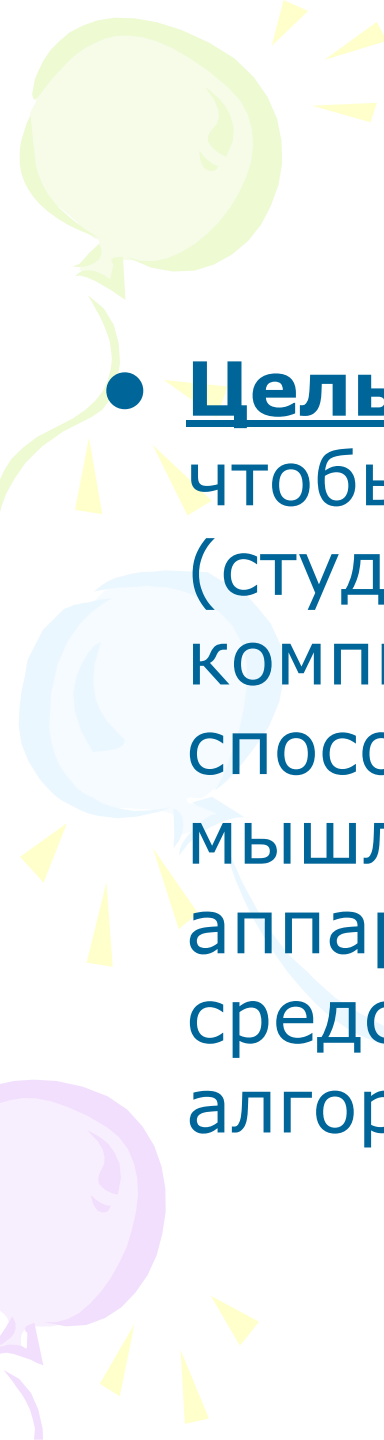
**Лектор –
Максимчук Ольга Владимировна,
к.т.н., доцент**



Понятие и задачи информатики

- **Информатика** – термин заимствован из французского языка в 1960х гг. и обозначает название области, связанной с автоматизированной обработкой информации с помощью электронных вычислительных машин.
- **Информатика** – это наука, занимающаяся исследованием форм и методов сбора, хранения, обработки и передачи информации.
- Это комплексная дисциплина, тесно связанная с другими науками, у которых есть общий объект исследования – **информация**.

- 
- По своей **структуре** информатика представляет собой единство трех взаимосвязанных частей - **технические средства (hardware), программные средства (software), алгоритмические средства (brainware)**.
 - **Задачами** информатики являются: разработка и производство современных средств вычислительной техники, проектирование и внедрение прогрессивных технологий обработки информации, и как результат этого – возможность дальнейшей информатизации общества и повышения уровня его информационной культуры.

- 
- **Цель курса** информатики состоит в том, чтобы помочь всем желающим (студентам) овладеть основами компьютерной грамотности, способствовать развитию их логического мышления, познакомиться с аппаратными и программными средствами компьютера, освоить основы алгоритмизации и программирования.

Понятие информации

- **Информация** – это сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые **уменьшают степень неопределенности**, имеющуюся о них в реальности.
- Наряду с информацией в информатике часто употребляется понятие «**данные**».
- **Данные** обычно рассматриваются как признаки или записанные наблюдения, которые не используются, а только хранятся.

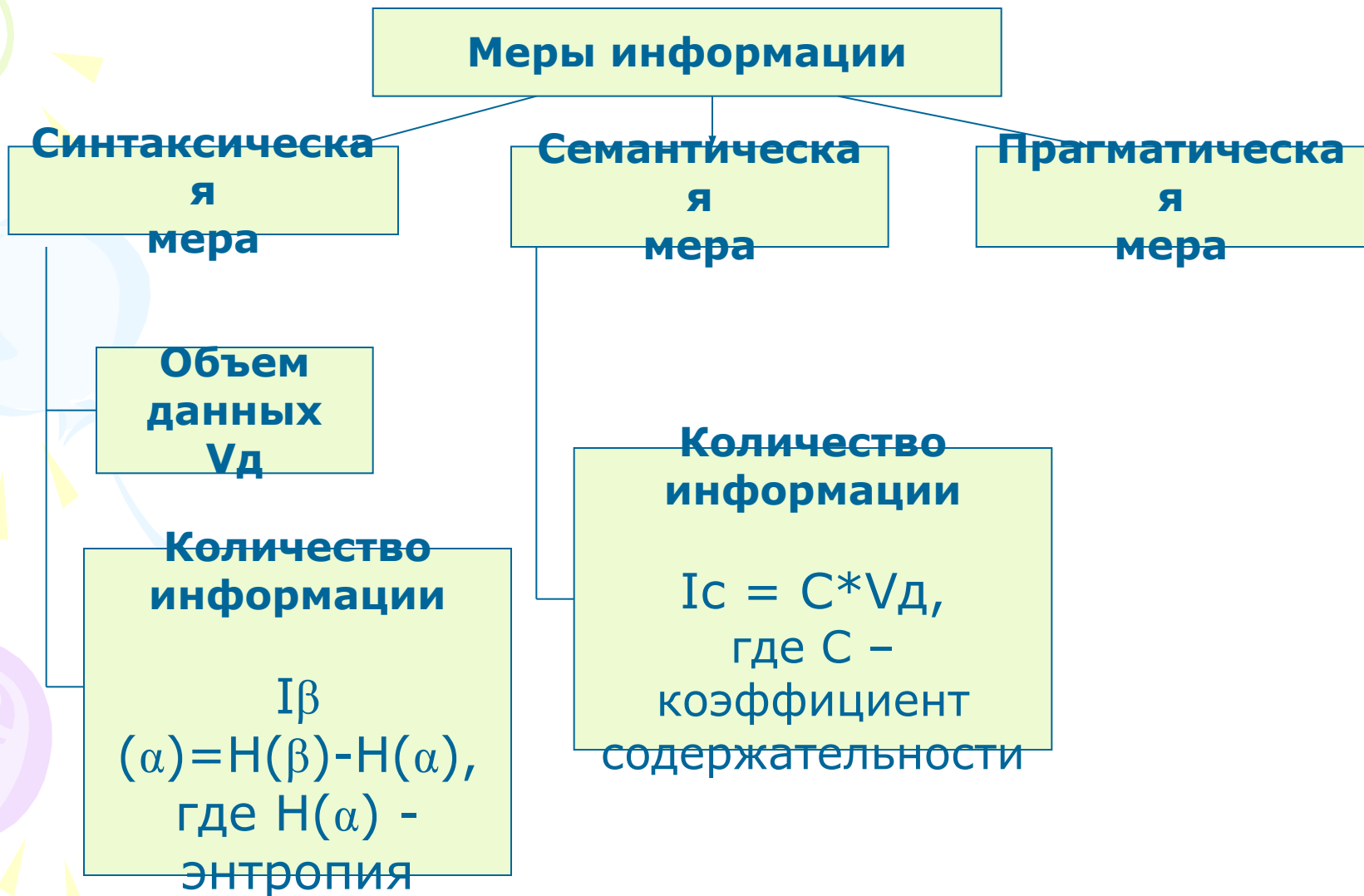
- Понятие «информация» обычно предполагает наличие двух объектов – «**источника**» информации и «**приемника**» (потребителя, адресата) информации.
- Информация передается от источника к приемнику **в материально-энергетической форме в виде сигналов** (например, электрических, световых, звуковых и т. д.), распространяющихся в определенной среде.
- **Сигнал** (от лат. signum – знак) – физический процесс (явление), несущий сообщение (информацию) о событии или состоянии объекта наблюдения.
- Информация может поступать в **аналоговом** (непрерывном) виде или **дискретно** (в виде последовательности отдельных сигналов). Соответственно различают **аналоговую и дискретную информацию.**

Информация и ее свойства

- **Адекватность информации** – важнейшее свойство информации, это определенный уровень соответствия образа, создаваемого с помощью полученной информации, реальному объекту, процессу, явлению ...
- **Формы адекватности информации:**
- **Синтаксическая** адекватность (отвечает за формально-структурные характеристики информации)
- **Семантическая** адекватность (отвечает за смысл)
- **Прагматическая** (потребительская) адекватность (ценность, полезность информации)

Меры информации

Классификация мер



Синтаксическая мера информации

- **В технике** часто используют способ определения количества информации называемый **объемным**. Он основан на подсчете числа символов в сообщении, т.е. связан с его длиной и не зависит от содержания.
- **В вычислительной технике** (ВТ) применяют две стандартные единицы измерения:
- **бит** (двоичный разряд) и **байт**. **1 байт = 8 бит**
- **Бит** – минимальная единица измерения информации, которая представляет собой двоичный знак двоичного алфавита {0;1}.
- **Байт** – единица количества информации в СИ, представляющая собой восьмиразрядный двоичный код, с помощью которого можно представить один символ.
- **Информационный объем сообщения V_d** - количество информации в сообщении, измеренное в стандартных единицах или производных от них
 - 1 байт = 8 бит
 - 1 Кбайт = 2^{10} байт = 1024 байт
 - 1 Мбайт = 2^{10} Кбайт = 2^{20} байт
 - 1 Гбайт = 2^{10} Мбайт = 2^{20} Кбайт = 2^{30} байт
 - 1 Терабайт = 2^{40} байт, 1 Петабайт = 2^{50} байт

Количество информации. Синтаксический подход

- В теории информации **количеством информации** называют числовую характеристику сигнала, не зависящую от его формы и содержания, и характеризующую **неопределенность (Энтропия)**, которая исчезнет после получения сообщения в виде данного сигнала.
- В этом случае **количество информации зависит от вероятности** получения сообщения о том или ином событии.
- Для абсолютно достоверного события (вероятность равна 1) количество информации в сообщении о нем равно 0. Чем неожиданнее событие, тем больше информации он несет.

Формула Хартли

- Научный подход к оценке количества информации был предложен в 1928 году **Р. Хартли**.
- Расчетная формула для равновероятных событий имеет вид
- **$I = \log_2 K$** ,
где K - количество равновероятных событий;
 I - количество бит в сообщении.
- Тогда можно записать **$K = 2^I$** .
Иногда формулу Хартли записывают так:
- **$I = \log_2 K = \log_2 (1 / p)$**
т. к. каждое из K событий имеет равновероятный исход **$p = 1 / K$** , то **$K = 1 / p$** .
- **Задача:** Шарик находится в одной из трех урн: А, В, С. Определить сколько бит информации содержит сообщение о том, что он находится в урне В

Формула Шеннона

- Более общий подход к вычислению количества информации в сообщении об одном из N , но уже неравновероятных событий, был предложен **Клодом Шенноном** в 1948 году
- Потребитель имеет заранее некоторые сведения о системе α . Мера его неосведомленности (**энтропия системы α**) – $H(\alpha)$.
- После получения сообщения β потребитель приобретает **дополнительную информацию $I_\beta(\alpha)$** , которая уменьшает его неосведомленность до $H_\beta(\alpha)$.
- **Количество информации $I_\beta(\alpha)$** можно определить так:
- **$I_\beta(\alpha) = H(\alpha) - H_\beta(\alpha)$**

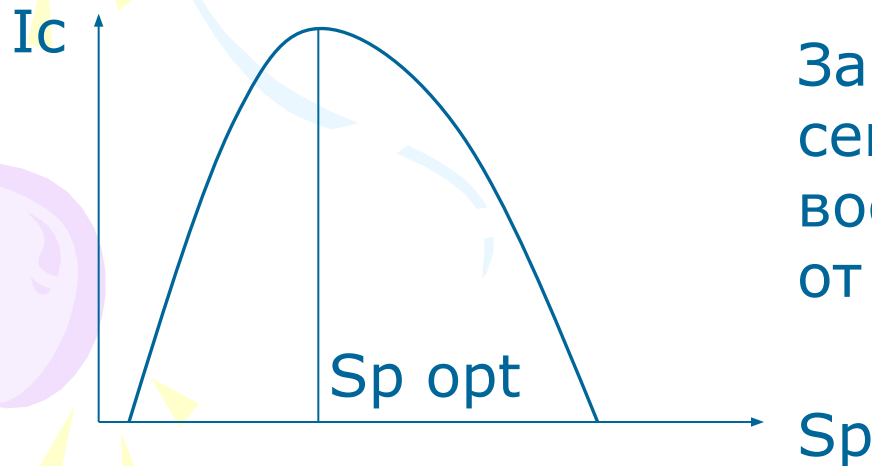
- Энтропия системы $H(\alpha)$, имеющая N возможных состояний, определяется по формуле Шеннона

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

- где P_i – вероятность того, что система находится в i -м состоянии.

Семантическая мера информации (тезаурусная мера информации)

- **Тезаурус** – совокупность сведений, которыми располагает пользователь или система.
- S – смысловое содержание информации
- Sp – тезаурус пользователя



Зависимость количества семантической информации, воспринимаемой пользователем, от его тезауруса



Прагматическая мера информации

- Прагматическая мера информации определяет полезность (ценность) информации для пользователя.
- Это величина относительная, обусловленная особенностями использования информации в определенной сфере деятельности.



Качество информации

- Показатели качества информации:
- Репрезентативность
- Содержательность
- Достаточность (полнота)
- Доступность
- Актуальность
- Своевременность
- Точность
- Достоверность
- Устойчивость

ПРЕДСТАВЛЕНИЕ И КОДИРОВАНИЕ ИНФОРМАЦИИ

- Для обмена информацией с другими людьми человек использует **естественные языки** (русский, английский, китайский и др.). В основе языка лежит алфавит, т.е. набор символов (знаков).
- Наряду с естественными языками были разработаны **формальные языки** (системы счисления, язык алгебры, символы, языки программирования и др.).
- В процессе обмена информацией часто приходится производить операции **кодирования** и **декодирования** информации.
- Преобразование информации из одной формы представления в другую называют **кодированием**.
- Средством кодирования служит **таблица соответствия знаковых систем**, которая устанавливает взаимнооднозначное соответствие между знаками или группой знаков двух различных знаковых систем.

Двоичное кодирование

- В компьютере для представления информации используется **двоичное кодирование**, т.к. используются технические устройства, которые могут сохранять и распознавать не более двух различных состояний (цифр):
 - Электромагнитные реле (замкнуто/разомкнуто), широко использовались в конструкциях первых ЭВМ.
 - Участок поверхности магнитного носителя информации (намагничен/размагничен).
 - Участок поверхности лазерного диска (отражает/не отражает).
 - И т.д.
- Все виды информации в компьютере кодируются на машинном языке, в виде логических последовательностей нулей и единиц.

Системы счисления

- **Система счисления** – это знаковая система, в которой числа записываются по определенным правилам с помощью символов некоторого алфавита, называемых цифрами.
- Все системы счисления делятся на две большие группы: **позиционные** и **непозиционные**.
- В позиционных системах счисления значение цифры зависит от ее положения (позиции) в числе.
- **Позиция** цифры в числе называется **разрядом**.
Например, десятичная: 343, 222 и т.д.
- В непозиционных значение цифры не зависит от ее положения в числе, например, Римская непозиционная система счисления: I(1), V(5), X(10), L(50), C(100), D(500), M(1000): XXX (тридцать) = X+X+X.



Системы счисления

- Наиболее распространенными в настоящее время позиционными системами являются **десятичная, двоичная, восьмеричная, шестнадцатеричная.**
- Каждая позиционная система имеет определенный **алфавит цифр** и **основание.** (см. таблицу)
- В **позиционных системах счисления** основание системы равно количеству цифр (знаков в ее алфавите) и определяет, во сколько раз различаются значения одинаковых цифр, стоящих в соседних позициях.

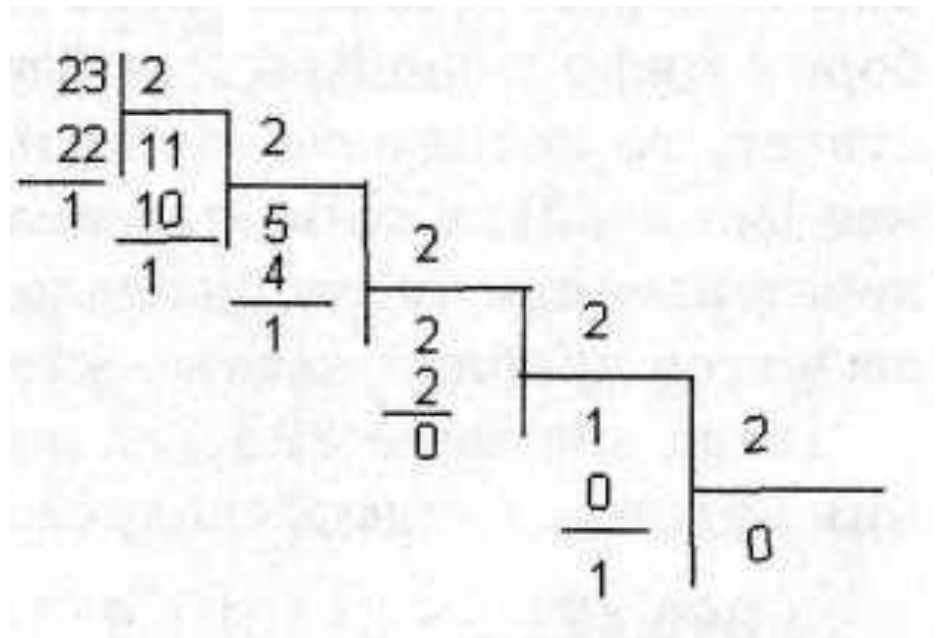
Системы счисления

Система счисления	Алфавит
Десятичная ($p=10$)	0,1,2,3,4,5,6,7,8,9
Двоичная ($p=2$)	0,1
Восьмеричная ($p=8$)	0,1,2,3,4,5,6,7
Шестнадцатеричная ($p=16$)	0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F

Алгоритм перевода десятичных чисел в двоичные. Целая часть

1. Разделить число на 2. Зафиксировать остаток (0 или 1) и частное.
2. Если частное не равно 0, то разделить его на 2, и так далее, пока частное не станет равно 0. Если частное равно 0, то записать все полученные остатки, начиная с первого, справа налево.

В примере ответ 10111.



Алгоритм перевода десятичных чисел в двоичные. Дробная часть

1. Умножить дробную часть на 2.

Зафиксировать целую часть результата (0 или 1). $0,75 * 2 = \underline{1},5$

2. Отбросить целую часть результата и продолжить умножение на 2.

$0,5 * 2 = \underline{1},0$

• Получить требуемое количество знаков после запятой. $0,11$

Перевод числа в десятичную систему счисления

- $1011,01_2 = 1*2^3 + 0*2^2 + 1*2^1 + 1*2^0 + 0*2^{-1} + 1*2^{-2} = 11,25_{(10)}$

- $1011,01$

↓ ↓ ↓ ↓ ↓ ↓
3 2 1 0 -1 -2

Местоположение цифр в числе (разряд)

- $F5C,F6_{16} = F*16^2 + 5*16^1 + C*16^0 + F*16^{-1} + 6*16^{-2}$

Перевод числа из восьмеричной системы в шестнадцатеричную и наоборот

8 cc	0	1	2	3	4	5	6	7
2 cc	000	001	010	011	100	101	110	111

16 cc	0	1	2	3	4	5	6	7
2 cc	0000	0001	0010	0011	0100	0101	0110	0111
16 cc	8	9	A	B	C	D	E	F
2 cc	1000	1001	1010	1011	1100	1101	1110	1111

Перевод числа из восьмеричной системы в шестнадцатеричную и наоборот

• $546,12_8 =$

$= 101100110,001010_2 =$

$= 166,28_{16}$

• $F5C,F6_{16} =$

$= 111101011100,11110110_2 =$

$= 7534,754_8$