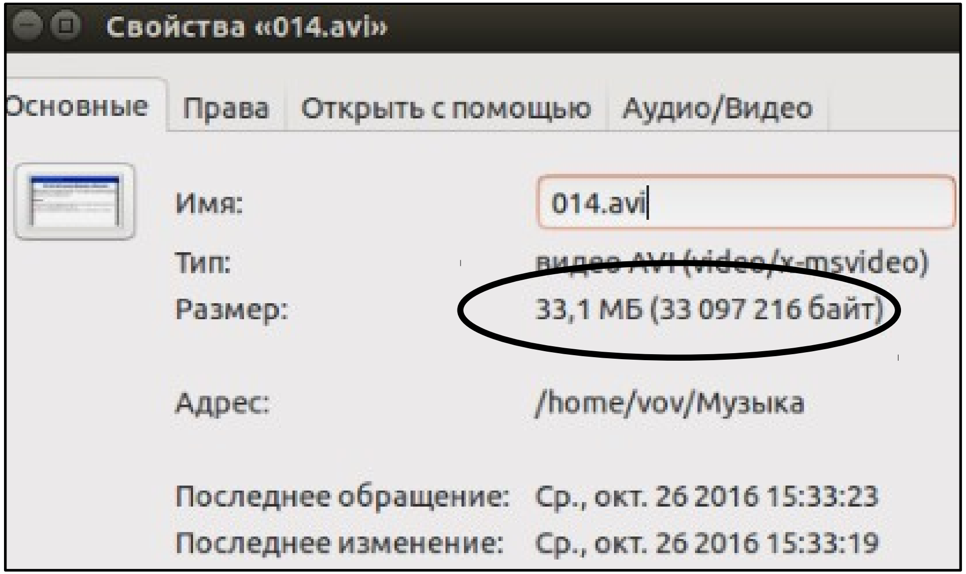
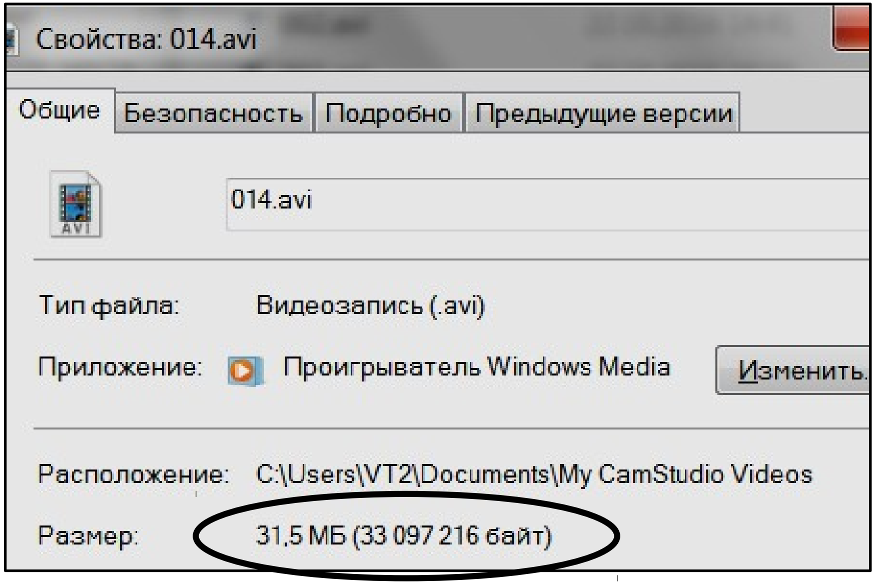
Приставки для единиц измерения количества информации/данных: проблема

Linux Ubuntu 14



Microsoft Windows 7



33 097 216 байт – это 33,1 МБ или 31,5 МБ?

Приставки для единиц измерения количества информации/данных: решение

1. IEEE 1541-2002 – Институт инженеров по электротехнике и радиоэлектронике.
2. ISO/IEC 80000-13:2008 – Международная организация по стандартизации.
3. ГОСТ IEC 60027-2-2015 – Международная электротехническая комиссия.

Приставки единиц СИ	Новые двоичные префиксы	Δ,%
килобайт (kB) = 10^3 байт	кибибайт (KiB, КиБ) = 2^{10} байт	2
мегабайт (MB) = 10^6 байт	мебибайт (MiB, МиБ) = 2^{20} байт	5
гигабайт (GB) = 10^9 байт	гибибайт (GiB, ГиБ) = 2^{30} байт	7
терабайт (TB) = 10^{12} байт	тебибайт (TiB, ТиБ) = 2^{40} байт	10

Краткое обозначение битов и байтов: b = bit = бит, B = Б = байт

1024 B = 1024 Б = 8192 b = 8192 бит = 8 Кибит = 1 КиБ = 1 KiB

Приставки для единиц измерения количества информации/данных: детали

Полное произношение названий приставок

3 КиБ = «три кибибайта» = «три килобинарных (kilobinary) байта».

7 Гибит = «семь гибибитов» = «семь гигабинарных (gigabinary) битов».

Сложившаяся практика использования приставок

Объем памяти (HDD, RAM, Cache): 512 KiB = 524 288 bytes.

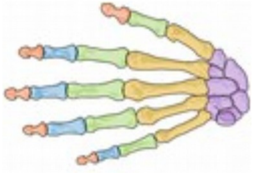
Скорость передачи данных: 512 kbps = 512 000 bps = 512 000 бит/с.

Типовая задача

Сколько мегабит содержится в двух гигабинарных байтах?

$$2\,\text{ГиБ} \;=\; 2\cdot 2^{30}\,\text{Б} \;=\; 16\cdot 2^{30}\,\text{бит} \;=\; \frac{16\cdot 2^{30}}{1000000}\,\text{Мбит} \;\approx\; 17180\,\text{Мбит (округл.)}$$

Системы счисления: историческая справка

Основание	Кто и как использовал	
нет	Австралийские племена	3=два-один, 4=два-два, 5=два-два-один, 6=два-два-два, 7=много
5	Африканские племена	
12	Тибетцы, нигерийцы	
20	Индейцы Майя, кельты	
60	Вавилоняне, шумеры	
10	5 век (Индия) 16 век (Европа) 17 век (Россия)	

Позиционные системы счисления (СС)

$X = 2017{,}042 = 2*1000 + 0*100 + 7*1 + 4/100 + 2/1000$

$$\boxed{X_{(q)} = x_{n-1}x_{n-2}x_1x_0{,}\,x_{-1}x_{-2}x_{-m}}$$

$X_{(q)}$ — запись числа в системе счисления с основанием q;
x_i — натуральные числа меньше q, т.е. цифры;
n — число разрядов целой части;
m — число разрядов дробной части.

$$\boxed{X_{(q)}=x_{n-1}q^{n-1} + x_{n-2}q^{n-2} + \ldots + x_1q^1 + x_0q^0 + x_{-1}q^{-1} + x_{-2}q^{-2} + \ldots + x_{-m}q^{-m}}$$

$$\boxed{X_{(q)} = \sum_{i=-m}^{n-1} x_i \cdot q^i}$$

ПРИМЕРЫ: $123_{(4)} = 1*4^2 + 2*4 + 3$ (если основание СС не указано => 10-ричная СС)

$456{,}78_{(10)} = 4*10^2 + 5*10^1 + 6*10^0 + 7*10^{-1} + 8*10^{-2}$

Перевод из одной СС в другую. Пример 1

$231_{(10)}$ = $ABC_{(10)}$ = $\ldots HGFE_{(8)}$ = $\ldots + H*8^3 + G*8^2 + F*8 + E$, при натуральных H, G, F, E < 8.

Как найти E, F, G, H?

Решение: $(\ldots + H*8^3 + G*8^2 + F*8 + E)/8 = \ldots + H*8^2 + G*8^1 + F$ (плюс остаток E)

=> $(\ldots HGFE_{(8)})/8 = \ldots HGF_{(8)}$ (с остатком E)

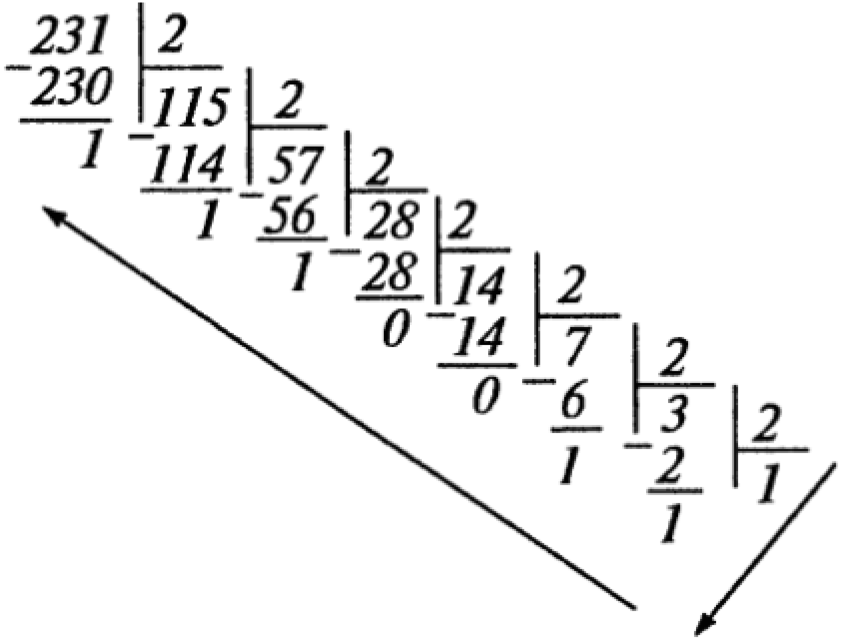
Номер шага (*i*)	0	1	2	3	4	…
Частное от деления на 8	231	28	3	0	0	0
Остаток от деления на 8	0	7	4	3	0	0

Ответ: E=7, F=4, G=3, H=0.

$231_{(10)} = 347_{(8)}$

Перевод из одной СС в другую. Пример 2

Задача: $231_{(10)} = ?_{(2)}$



Ход решения →

Ответ: $231_{(10)} = 11100111_{(2)}$

Перевод из одной СС в другую. Пример 3

Задача: $0{,}15_{(10)} = \ ?_{(3)} = 0{,}ABCD\ldots_{(3)} = A/3^1 + B/3^2 + C/3^3 + D/3^4 + \ldots$

Решение: $(A/3^1 + B/3^2 + C/3^3 + D/3^4 + \ldots)*3 = A*3^0 + (B/3^1 + C/3^2 + D/3^3 + \ldots)$

$$=> 3*0{,}ABCD\ldots_{(3)} = A{,}BCD\ldots_{(3)}$$

Номер шага (*i*)	0	1	2	3	4	5	…
Целая часть после умножения дробной части на 3	0	0	1	1	1	0	…
Дробная часть после умножения на 3	0,15	0,45	0,35	0,05	0,15	0,45	…

Ответ: $0{,}15_{(10)} = 0{,}011101110\ldots_{(3)} = 0{,}(0111)_{(3)}$

Перевод из одной СС в другую. Пример 4

Задача: $0{,}8125_{(10)} = ?_{(2)}$

Ход решения →

0	,8125 2
1	,625 2
1	,250 2
0	,5 2
1	0

Ответ: $0{,}8125_{(10)} = 1 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-4} = 0{,}1101_{(2)}$

Перевод из одной СС в другую. Пример 5

$$231_{(10)} = 11100111_{(2)}$$
$$0{,}8125_{(10)} = 0{,}1101_{(2)}$$

$$231{,}8125_{(10)} = 11100111{,}1101_{(2)}$$

Перевод из СС с основанием 2 в СС с основанием 4

Сложный путь: 1) СС-2 -> СС-10: $10100_{(2)} = 20_{(10)}$

2) СС-10 -> СС-4: $20_{(10)} = 110_{(4)} \Rightarrow 10100_{(2)} = 110_{(4)}$

Примечание: «СС-*N*» означает «система счисления с основанием *N*»

Простой путь:

$$x_{i+1}2^{i+1} + x_i 2^i + \cdots + x_3 2^3 + x_2 2^2 + x_1 2^1 + x_0 2^0$$

$$\Downarrow$$

$$x_{2k+1}2^{2k+1} + x_{2k} 2^{2k} + \cdots + x_3 2^{2*1+1} + x_2 2^{2*1} + x_1 2^1 + x_0 2^0$$

$$\Downarrow$$

$$2^{2k}(x_{2k+1}2^1 + x_{2k}) + \cdots + 2^2(x_3 2^1 + x_2) + 2^0(x_1 2^1 + x_0)$$

$$\Downarrow$$

$$4^k(x_{2k+1}2^1 + x_{2k}) + \cdots + 4^1(x_3 2^1 + x_2) + 4^0(x_1 2^1 + x_0)$$

Преобразование из СС-2 в СС-2^k и обратно

Двоичная <-> Четверичная	Двоичная <-> Восьмеричная	Двоичная <-> Шестнадцатеричная
00 <-> 0	000 <-> 0	0000 <-> 0
01 <-> 1	001 <-> 1	0001 <-> 1
10 <-> 2	010 <-> 2	0010 <-> 2
11 <-> 3	011 <-> 3	0011 <-> 3
	100 <-> 4	…
	101 <-> 5	1101 <-> D
	110 <-> 6	1110 <-> E
	111 <-> 7	1111 <-> F

Пример: $1111110001{,}1110001_{(2)} = 0011\ 1111\ 0001{,}1110\ 0010_{(2)} = 3F1{,}E2_{(16)}$

Преобразование из СС-N в СС-N^k и обратно

Из СС-N в СС-N^k

- дополнить число, записанное в СС с основанием N, незначащими нулями так, чтобы количество цифр было кратно k;
- разбить полученное число на группы по k цифр, начиная от нуля;
- заменить каждую такую группу эквивалентным числом, записанным в СС с основанием N^k.

Задача: $1020101_{(3)} = ?_{(27)}$

Решение: $1020101_{(3)} = 001\ 020\ 101_{(3)} = 16A?_{(27)}$

Из СС-N^k в СС-N

- заменить каждую цифру числа, записанного в СС с основанием N^k, эквивалентным набором из k цифр СС с основанием N.

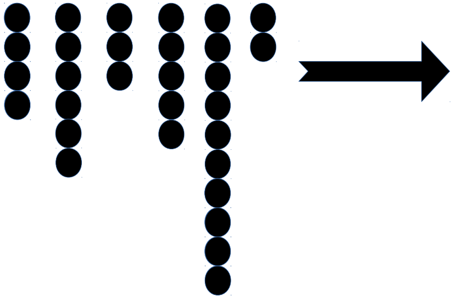
Задача: $2345_{(125)} = ?_{(5)}$

Решение: $2345_{(125)} = 002\ 003\ 004\ 010_{(5)} = 2003004010_{(5)}$

Оптимальная система счисления

Задача. Робинзон Крузо нашёл на острове 60 камней. Сколько прошедших дней можно ими закодировать в разных СС?

Пример СС-10:



463502-й день из 999999 возможных, где $999999 = 10^6 - 1$

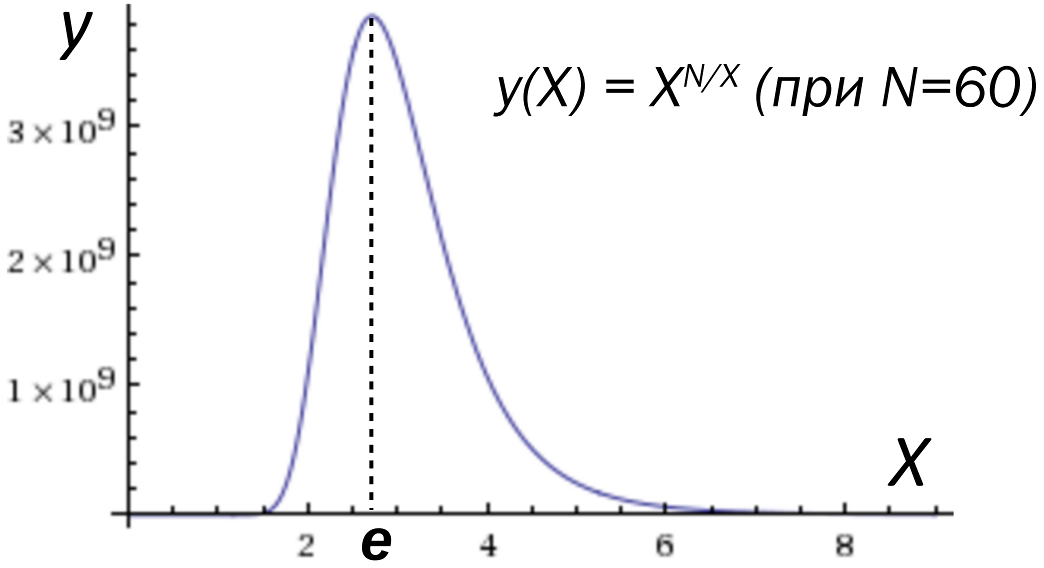
Возможные варианты в других СС:

2^{30}, 3^{20}, 4^{15}, 5^{12}, 6^{10}, 7^8, 8^7, 9^6, 10^6, 11^5, 12^5, ..., 20^3, ..., 30^2, ..., 60^1

В какой СС количество кодируемых дней наибольшее?

Оптимальная система счисления (продолжение)

Если взять *N* камней, а за основание СС принять число *X*, то получится *N/X* разрядов, которыми можно закодировать $y=X^{N/X}$ дней (для простоты полагаем, что количество разрядов может быть нецелым).



Вывод: оптимальная система счисления имеет основание *e*=2,7183....

Каким может быть основание позиционной СС?

$$\boxed{X_{(q)} = \sum_{k=-m}^{n-1} d_k \cdot q^k}$$

m — количество цифр справа от запятой,

n — количество цифр слева от запятой,

d_k — цифра числа, стоящая на *k*-й позиции,

q — основание системы счисления.

Пример: $\mathbf{789{,}13}_{\mathbf{10}} = \mathbf{7}{*}10^2 + \mathbf{8}{*}10^1 + \mathbf{9}{*}10^0 + \mathbf{1}{*}10^{-1} + \mathbf{3}{*}10^{-2}$

Что если *q* отрицательно? иррационально? переменно?

Система счисления Бергмана

Джорж Бергман (р. 1943)

Любое действительное число можно представить в виде

$$x = \sum_{k=-m}^{n-1} d_k \cdot z^k, \quad \text{где}\, d_k \in \{0,1\},\, z = \frac{1+\sqrt{5}}{2}$$

m — количество цифр справа от запятой, n — количество цифр слева от запятой, d_k — цифра числа, стоящая на k-й позиции, z — число золотой пропорции. Запись числа x в системе Бергмана имеет вид : $x_{(Б)} = d_{n-1} \ldots d_2 d_1 d_0, d_{-1} d_{-2} d_{-3} \ldots d_{-m\ (Б)}$

$2_{(10)} = 10{,}01_{(Б)} = z^1+z^{-2}$

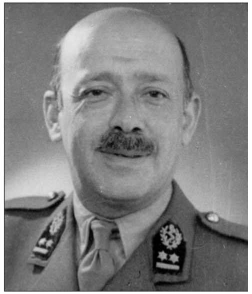
$3_{(10)} = 11{,}01_{(Б)} = z^1+z^0+z^{-2}$

$3_{(10)} = 100{,}01_{(Б)} = z^2+z^{-2}$

Чтобы исключить неоднозначность, используют запись с наибольшим количеством разрядов, т. е. $3_{(10)} = 100{,}01_{(Б)}$

Применение: запись иррациональных чисел конечным числом цифр: $10_{(Б)} = 2{,}61803399\ldots$, контроль арифметических операций, коррекция ошибок, самосинхронизация кодовых последовательностей при передаче по каналу связи.

Система счисления Цекендорфа (фибоначчиева СС)



Эдуард Цекендорф (1901–1983)

Любое целое число можно представить в виде

$$x=\sum_{k=1}^{n} d_k F_k, \text{ где } d_k \in \{0,1\}, \text{ а } F_k - \text{числа Фибоначчи (ЧФ)}$$

n — количество цифр в записи числа, d_k — цифра числа, стоящая на k-й позиции,

каждое ЧФ есть сумма двух предыдущих ЧФ: F_i = {1, 1, 2, 3, 5, 8, 13, ...} , где i = 0, 1,... .

Запись числа x в системе Цекендорфа будет иметь вид $x_{(Ц)}=d_n d_{n-1} \ldots d_{1(Ц)}$

Проблема неуникальности: 16 = 8+5+2+1 = 13+3, т.е. 16 = $11011_{(Ц)}$ = $100100_{(Ц)}$.

Чтобы исключить неоднозначность, введён запрет на использование двух единиц подряд: т. е. $16_{(10)}$ = $100100_{(Ц)}$, а запись $11011_{(Ц)}$ считается ошибочной!

Применение: минимизация числа зёрен маиса в счётах у инков,
кодирование данных с маркером завершения «11».

Факториальная СС: определение

Любое целое число можно представить в виде

$$x=\sum_{k=1}^{n} d_k k!,\quad \text{где}\quad 0\leqslant d_k\leqslant k,\quad k!=1\cdot 2\cdot 3\cdot \ldots \cdot k.$$

n — количество цифр в записи числа,

d_k — цифра числа, стоящая на k-й позиции,

Запись числа x в факториальной системе счисления будет иметь вид:

$$x_{(\Phi)}=d_n d_{n-1}\ldots d_{1(\Phi)}.$$

Примеры: $310_{(\Phi)} = 3*3! + 1*2! + 0*1! = 201_{(10)}$

$106_{(10)} = d_5*5! + d_4*4! + d_3*3! + d_2*2! + d_1*1! =$ ...подбор d_1, d_2, d_3, d_4, d_5 ...=

$= 0 *5! + 4 *4! + 1 *3! + 2 *2! + 0 *1! = 4120_{(\Phi)}$

Перевод чисел из СС-10 в факториальную СС

$$\text{Дано:}\quad x = d_4 d_3 d_2 d_{1(\Phi)} = (2\cdot 3\cdot 4)\,d_4 + (2\cdot 3)\,d_3 + (2)\,d_2 + (1)\,d_1 .$$

1) $(x\,\mathrm{div}\,2) \;=\; (3\cdot 4)d_4 + (3)d_3 + d_2$ (и остаток, равный d_1).

2) $(x\,\mathrm{div}\,2)\,\mathrm{div}\,3 \;=\; (4)d_4 + d_3$ (и остаток, равный d_2).

3) $((x\,\mathrm{div}\,2)\,\mathrm{div}\,3)\,\mathrm{div}\,4 \;=\; d_4$ (и остаток, равный d_3).

4) $(((x\,\mathrm{div}\,2)\,\mathrm{div}\,3)\,\mathrm{div}\,4)\,\mathrm{div}\,5 \;=\; 0$ (и остаток, равный d_4).

Примечание: «*A* div *B*» означает целочисленное деление *A* на *B*.

«*A* mod *B*» означает остаток от деления *A* на *B*.

Пример: $106_{(10)} = ?_{(\Phi)}$

1) 106 mod 2 = 53, d_1 = 106 mod 2 = 0
2) 53 mod 3 = 17, d_2 = 53 mod 3 = 2
3) 17 mod 4 = 4, d_3 = 17 mod 4 = 1
4) 4 mod 5 = 0, d_4 = 4 mod 5 = 4

$$x_{(\Phi)} = d_4 d_3 d_2 d_{1(\Phi)} = 4120_{(\Phi)}$$

Факториальная СС: применение

Проблема: как упорядочить перестановки букв АБВ: 1-АБВ, 2-АВБ, 3-ВБА, 4-ВАБ, 5-БАВ, 6-БВА.

Пример. Пусть имеется n=5 чисел (1,2,3,4,5) и нужно найти все их перестановки. Известно, что всего существует n! = 5! =120 таких перестановок. Как найти перестановку, если задан её номер *k*?

Решение. Найдём 21-ю перестановку (*k* = 21). Переведём *k* в факториальную систему:

$21 = 3*3! + 1*2! + 1*1! = 311_{(Ф)}$. Дополним его до (*n*-1) разрядов: $311_{(Ф)} \rightarrow 0311_{(Ф)}$.

Расставим символы по местам:

1) справа от «5» есть 0 меньших цифр (_ _ _ _ 5)
2) справа от «4» есть 3 меньшие цифры (4 _ _ _ 5)
3) справа от «3» есть 1 меньшая цифра (4 _ 3 _ 5)
4) справа от «2» есть 1 меньшая цифра (4 2 3 _ 5)

ОТВЕТ: 42315

Значение k	0	1	2	3	...	21	...	119
k-я перестановка	12345	21345	13245	23145	...	42315	...	54321

СС с отрицательным основанием или цифрами

1. Нега-позиционные (с отрицательным основанием). Примеры в нега-десятичной СС:

- $123_{(-10)} = 1{\cdot}(-10)^2 + 2{\cdot}(-10)^1 + 3{\cdot}(-10)^0 = 100 - 20 + 3 = 83_{(10)}$
- $58_{(-10)} = 5{\cdot}(-10)^1 + 8{\cdot}(-10)^0 = -50 + 8 = -42_{(10)}$

Числа с чётным количеством цифр — отрицательные.

2. Симметричные (с отрицательными цифрами). Например, в симметричной пятеричной СС вместо привычных цифр {0, 1, 2, 3, 4} используются {-2, -1, 0, 1, 2}:

- $20\bar{2}10_{(5С)} = (2){\cdot}5^4 + (0){\cdot}5^3 + (-2){\cdot}5^2 + (1){\cdot}5^1 + (0){\cdot}5^0 = 1250 - 50 + 5 = 1205_{(10)}$
- $\bar{2}02\bar{1}0_{(5С)} = (-2){\cdot}5^4 + (0){\cdot}5^3 + (2){\cdot}5^2 + (-1){\cdot}5^1 + (0){\cdot}5^0 = -1250 + 50 - 5 = -1205_{(10)}$

Симметричные СС определены только для нечётных оснований!

Применение. В негапозиционных и симметричных СС не требуется специального знака для обозначения отрицательных чисел. Это позволяет использовать их для представления отрицательных чисел в компьютерах.

Проблемы округления чисел

Классическое правило округления – к ближайшему целому:

	Число	Округл.
	1,1	1,0
	2,9	3,0
	5,0	5,0
	3,4	3,0
	8,6	9,0
Сумма	21,0	21,0

	Число	Округл.
	1,5	2,0
	2,5	3,0
	5,5	6,0
	3,5	4,0
	8,5	9,0
Сумма	21,5	24,0

Пример накопленной ошибки округления

Копеечная часть зарплаты	Округление до целых рублей
0,33	0
0,51	1
0,89	1
0,49	0
…	…
0,50	1
0,73	1
0,20	0

10000 строк

Сумма1 vs Сумма2

В первом столбце из 100 возможных значений только одно приводит к накоплению ошибки в 50 коп., поэтому

В среднем (Сумма2 – Сумма1) =
= (10000/100) * 50 коп. =
= 50 руб. переплаты!

Проблемы округления чисел в различных СС

В системах счисления с **чётным** основанием накапливается ошибка округления:

<u>Основание 10</u>: **1, 2, 3, 4,** ← округление в меньшую сторону
5, 6, 7, 8, 9, ← округление в бóльшую сторону
0 ← нет ошибки округления

В системах счисления с **нечётным** основанием этой проблемы нет:

<u>Основание 7</u>: **1, 2, 3** ← округление в меньшую сторону
4, 5, 6 ← округление в бóльшую сторону
0 ← нет ошибки округления

Актуальна ли проблема накопления ошибки округления для симметричных СС?

Решение проблемы с округлением в СС с чётным основанием

Суть решения — использовать неклассические правила округления:

- **Случайное округление**: используется датчик случайных чисел при принятии решения о том, в бо́льшую или меньшую сторону следует округлять.
- **Банковское округление** (к ближайшему чётному): $3{,}5 \approx 4$, но $2{,}5 \approx 2$.
- **К ближайшему нечётному**: $3{,}5 \approx 3$, но $2{,}5 \approx 3$. Аналогично: $4{,}3_{(6)} \approx 5_{(6)}$.
- **Чередующееся**: направление округления меняется на противоположное при каждой операции округления (необходимо «помнить» о предыдущем округлении).

Примечание. Каждое из правил можно применять как полностью универсально, так и комбинировано с классическим правилом округления, дополняя его лишь при округлении пограничных значений.

Непозиционные системы счисления

Славянская кириллическая нумерация:
аддитивная запись

А = 1, **В** = 2, **Г** = 3, **Д** = 4…
444 = **УМД**, = 400 (**У**) + 40 (**М**) + 4 (**Д**).

Китайская нумерация:
аддитивно-мультипликативная запись

444 = 四百四十四 =
= 4*100 + 4*10 + 4

Римская (латинская) нумерация:
аддитивно-субтрактивная запись

CX = 50 + 10 = 60
XC = 50 – 10 = 40

Недостатки непозиционных систем счисления по сравнению с позиционными:

- Сложно выполнять арифметические операции с большими числами.
- Длина записи числа (т. е. количество цифр) немонотонно зависит от его величины.