

Приложение. МАШИННЫЕ ЧИСЛА

Число в компьютере представлено своим двоичным кодом - конечным набором двоичных цифр (нулей и единиц). Длина этого набора определяется конструкцией компьютера. Так, например, IBM PC использует для размещения числа либо 16 двоичных разрядов (бит), либо 32, либо 64, либо, наконец, 80.

Если длина кода, представляющего в компьютере число, равна n бит, то количество различных чисел, представимых в этом компьютере не может быть больше, чем 2^n .

Таким образом, множество *машинных чисел* - чисел, представимых в компьютере - конечно.

В зависимости от способа кодирования машинные числа разделяются на два типа - INTEGER и REAL

П.1. Числа типа INTEGER

Взаимно однозначное соответствие между машинными числами типа INTEGER и их двоичными кодами устанавливается с помощью правила (рис. П.1).

$$X = (-1)^{X_{n-1}} \cdot (2^{n-2}X_{n-2} + 2^{n-3}X_{n-3} + \dots + 2^1X_1 + 2^0X_0).$$

X_{n-1}	X_{n-2}	X_{n-3}	\dots	X_1	X_0
-----------	-----------	-----------	---------	-------	-------

Рис.П.1

Здесь X_0, \dots, X_{n-1} - двоичные цифры (единицы или нули). Цифра X_{n-1} кодирует знак числа ($X_{n-1} = 0$ при $X > 0$, $X_{n-1} = 1$ при $X < 0$). Отметим, что коды

$$X_0 = \dots = X_{n-2} = 0, \quad X_{n-1} = 0 \quad \text{и} \quad X_0 = \dots = X_{n-2} = 0, \quad X_{n-1} = 1$$

не различаются и обозначают число нуль.

Таким образом, при длине кода, равной n , существует $2^n - 1$ машинных чисел типа INTEGER. Наибольшее среди них -

$$X_{I_{max}} = (-1)^0 \cdot (2^{n-2} + \dots + 2^0) = \frac{2^{n-1} - 1}{2 - 1} = 2^{n-1} - 1,$$

наименьшее -

$$X_{I_{min}} = (-1)^1 \cdot (2^{n-2} + \dots + 2^0) = \frac{2^{n-1} - 1}{2 - 1} = -(2^{n-1} - 1),$$

Серьезное предупреждение. В некоторых операционных системах результат арифметической операции над числами типа INTEGER, выходящий за границы диапазона машинных чисел, искажается без аварийного останова. Незнание этого факта может привести к трудно диагностируемым ошибкам.

П.2. Числа типа REAL

Число 0 типа REAL кодируется набором нулей.

Отличное от нуля число типа REAL записывается в виде

$$X = M \cdot 2^P.$$

Здесь число M называется *мантиссой* числа X ($\frac{1}{2} \leq M < 1$), а целое число P - его *порядком*.

Из n бит, предназначенных для размещения кода числа X типа REAL, m бит выделяется под код мантиссы, а p - под код порядка ($m + p = n$).

Поскольку порядок P - целое число, он кодируется как число типа INTEGER¹, т.е. существует $2^p - 1$ различных значений порядка (+0 и -0 не различаются). При этом

$$P_{max} = 2^{p-1} - 1, \quad P_{min} = -(2^{p-1} - 1).$$

Кодирование мантиссы представлено на рис П.2:

$$M = (-1)^{X_0} \cdot (2^{-1}X_{-1} + 2^{-2}X_{-2} + \dots + 2^{-(m-1)}X_{-(m-1)}).$$

¹Применяются различные способы кодирования порядка, но мы не рассматриваем технические подробности

$$\boxed{X_0 \parallel X_{-1} \mid X_{-2} \mid \dots \mid X_{-(m-1)}}$$

Рис. П.2

В силу условия $\frac{1}{2} \leq |M| < 1$ для любого числа, кроме нуля, $X_{-1} = 1$. Наибольшее значение модуля мантиссы (все цифры равны единице)

$$|M|_{max} = 2^{-1} + \dots + 2^{-(m-1)} = 1 - 2^{-(m-1)}.$$

Таким образом, наибольшее число типа REAL –

$$X_{Rmax} = (1 - 2^{-(m-1)}) \cdot 2^{p-1} - 1,$$

а диапазон этих чисел –

$$\Delta_R = 2X_{Rmax} \approx 2^{Pmax+1}.$$

Существенно, что, в отличие от чисел типа INTEGER, числа типа REAL распределены по своему диапазону неравномерно. В каждом положительном и в каждом отрицательном поддиапазоне, соответствующем фиксированному значению порядка P , существует одинаковое количество -2^{m-2} – различных чисел типа REAL. Длина же поддиапазона, равная 2^{P-1} , увеличивается по мере удаления от нуля, и расстояние между парой соседних чисел типа равно

$$2^{-(m-2)} \cdot 2^{P-1} = 2^{P-m+1}.$$

Наиболее плотно расположены числа типа REAL в ближайшем к нулю поддиапазоне. Например, в формате REAL(4) для IBM PC наименьшее расстояние между соседними числами равно $2^{-149} \approx 1.4 \cdot 10^{-45}$. В следующем поддиапазоне оно уже вдвое больше, а в самом дальнем поддиапазоне составляет $2^{104} \approx 2 \cdot 10^{31}$ (как видно, ближайшие соседи не так уж и близки!).

Особо отметим, что среди чисел типа кроме наибольшего положительного существует и *наименьшее положительное*

$$X_{Rmin}^+ = 2^{-(Pmax+1)}.$$

Пример. Пусть $n = 7$, $p = 3$, $m = 4$. Тогда порядок числа может принимать следующие $2^3 - 1 = 7$ значений

$$P = -3, -2, -1, 0, +1, +2, +3;$$

ненулевая мантисса может принимать следующие $2^{(4-1)} = 8$ значений

$$M = -7/8, -6/8, -5/8, -4/8, +4/8, +5/8, +6/8, +7/8.$$

Всего же в нашем компьютере существует $7 \cdot 8 = 56$ ненулевых чисел типа REAL:

-7	-6	-5	-4	+4	+5	+6	+7
-7/2	-6/2	-5/2	-4/2	+4/2	+5/2	+6/2	+7/2
-7/4	-6/4	-5/4	-4/4	+4/4	+5/4	+6/4	+7/4
-7/8	-6/8	-5/8	-4/8	+4/8	+5/8	+6/8	+7/8
-7/16	-6/16	-5/16	-4/16	+4/16	+5/16	+6/16	+7/16
-7/32	-6/32	-5/32	-4/32	+4/32	+5/32	+6/32	+7/32
-7/64	-6/64	-5/64	-4/64	+4/64	+5/64	+6/64	+7/64

На рис. П.3 представлены три ближайших к нулю положительных поддиапазона.

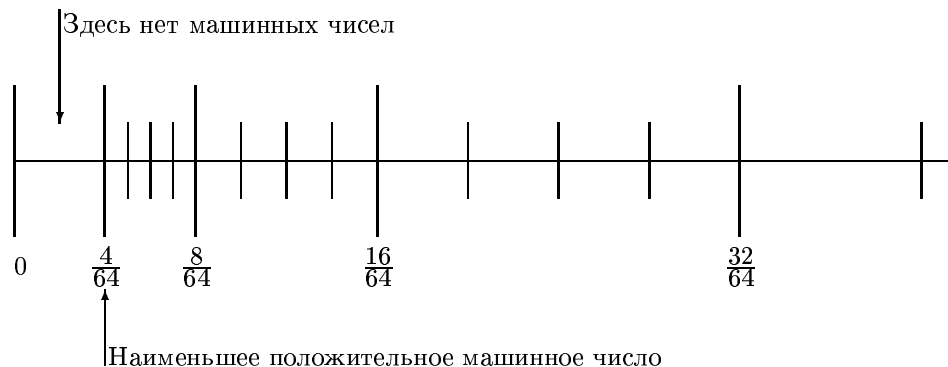


Рис. П.3

Серьезное предупреждение. Если "точный" результат арифметической операции не является машинным числом, но лежит в диапазоне машинных чисел, то он искажается (либо усекается, либо округляется). Если же он оказывается вне этого диапазона, то поведение компьютера определяется установленной на нем операционной системой.

Это одна из причин, по которым пользователю не следует самому писать даже простейшие вычислительные программы. Созданные и в течение многих лет оттестированные мировым сообществом библиотеки Фортран-программ, как правило, позволяют избежать неприятностей, связанных с дискретностью и конечностью сетки машинных чисел.

П3. Машинный эpsilon

Термином *машинный epsilon* называют наименьшее положительное машинное число, удовлетворяющее неравенству

$$1_M + \varepsilon_M > 1_M.$$

Здесь 1_M – машинная единица, совпадающая с общеизвестной, а ε_M – машинный epsilon.

Вследствие минимальности машинного epsilon его мантисса содержит единственную единицу. Поэтому представим машинный epsilon в виде

$$\varepsilon_M = 2^{-1} \cdot 2^{P_\varepsilon},$$

где P_ε – искомый порядок.

Тогда

$$1_M + \varepsilon_M = 2^{-1} \cdot 2^1 + 2^{-1} \cdot 2^{P_\varepsilon} = (2^{-1} + 2^{P_\varepsilon-2}) \cdot 2^1.$$

Если $P_\varepsilon - 2 = -(m - 1)$, то

$$1_M + \varepsilon_M > 1_M \quad (\text{рис.П4}).$$

0	1	0	...	0	1
---	---	---	-----	---	---

Рис.П.4

Если же $P_\varepsilon - 2 < -(m - 1)$, то будет иметь место равенство

$$1_M + \varepsilon_M = 1_M$$

и, следовательно, порядок машинного epsilon находится из равенства

$$P_\varepsilon - 2 = -(m - 1), \text{ т.е. } P_\varepsilon = -(m - 3) \text{ или } \varepsilon_M = 2^{-1} \cdot 2^{-(m-3)} = 2^{-(m-2)}.$$

Машинный epsilon характеризует относительную погрешность выполнения компьютером арифметических операций. Для IBM PC

$$\varepsilon_M = \begin{cases} 2^{-23} \approx 1.19 \cdot 10^{-7} & \text{в формате REAL(4),} \\ 2^{-52} \approx 2.22 \cdot 10^{-16} & \text{в формате REAL(8).} \end{cases}$$