UFSC / CTC / INE
Disciplina: Introdução à Informática

Curso de Sistemas de Informação: INE5602-138

Prof. Dr. João Dovicchi*

Solução dos exercícios em sala 3

1. Converta os números decimais abaixo para a representação IEEE de precisão simples (32 bits):

a. 12.35

Solução: Primeiro encontramos o número na base 2

$$\begin{aligned} 12_{10} &= 1100_2 \\ 0.35_{10} &= 0.0101100110011\ldots \\ 12.35_{10} &= 1100.010110011001100\ldots \end{aligned}$$

o ponto flutuante deve ser deslocado até o próximo 1 mais significativo ($1.10001011001100110011001 \times 2^3$) resultando na mantissa e um expoente que deve ser equalizado com 127 ($2^{n-1} - 1$).

$$3 + 127 = 130_{10} = 10000010_2$$

Resposta: `0 10000010 10001011001100110011001`

b. -0.525

Solução:

*`http://www.inf.ufsc.br/~dovicchi --- dovicchi@inf.ufsc.br`

$$
\begin{aligned}
0.525_{10} &= 0.1000011001100\ldots \\
&= 1.00001100110011001100110 \times 2^{-1} \\
\\
-1 + 127 &= 126_{10} \\
&= 01111110_2
\end{aligned}
$$

Resposta: `1 01111110 00001100110011001100110`

c. 3.25×10^{-4}

Solução:

$$
\begin{aligned}
3.25 \times 10^{-4} &= 0.000325 \\
&= 0.0000000000010101010011001001100001 0_2 \\
&= 1.01010100110010011000010 \times 2^{-12} \\
\\
-12 + 127 &= 115_{10} \\
&= 01110011_2
\end{aligned}
$$

Resposta: `0 01110011 01010100110010011000010`

Obs: Note que como o ponto flutuante teve que ser deslocado de 12 casas, é necessário calcular o número com 35 casas $(12 + 23)$.

d. $\arctan(1)$ (considere $\pi = 3.14159$)

Solução: Uma vez que $\arctan(1) = \tan^{-1}(1)$ e, portanto corresponde ao arco delimitado pelo ângulo de 45^o. Uma vez que o arco correspondente a π rad, corresponde ao arco formado pelo ângulo de 180^o, então $\arctan(1) = \pi/4 = 0.785398$.

$$
\begin{aligned}
0.785398_{10} &= 0.110010010000111111011000_2 \\
&= 1.10010010000111111011000 \times 2^{-1} \\
\\
-1 + 127 = 126_{10} &= 01111110_2
\end{aligned}
$$

Resposta: `0 01111110 10010010000111111011000`

e. $\sqrt{3}/2$

Solução:

$$
\begin{aligned}
\frac{\sqrt{3}}{2} &= \frac{1.732050}{2} \\
&= 0.866025 \\
&= 0.110111011011001111010000_2 \\
&= 1.10111011011001111010000 \times 2^{-1}
\end{aligned}
$$

$$-1 + 127 = 126_{10} = 01111110_2$$

Resposta: `0 01111110 10111011011001111010000`

2. Encontre os números decimais representados pelos números abaixo, considerando que os mesmos estejam no formato IEEE de 32 bits:

a. `0xB00CAA00`

Solução:

$$B00CAA00_{16} = 10110000000011001010101000000000_2$$

O bit de sinal é 1, portanto o número é negativo. O expoente é $01100000_2 = 96_{10}$ portanto, o expoente de 2 é $96 - 127 = 31$. Isto significa que a mantissa do número foi obtida com o ponto flutuante deslocado 31 casas para a direita. Assim o número deve ser:

$$
\begin{array}{llll}
 & & & 1.00011001010101000000000 \\
 & & \leftarrow & 31 \text{ casas para a esquerda} \\
0. & 000000000000000000000000000000 & & 100011001010101000000000
\end{array}
$$

O número:

`0.00000000000000000000000000000001000110010101010000000000`

equivale à soma decimal das potências negativas de 2:

$$2^{-31} + 2^{-35} + 2^{-36} + 2^{-39} + 2^{-41} + 2^{-43} + 2^{-45} = -5.1173288 \times 10^{-10}$$

Obs.: Provavelmente, muitas das calculadoras serão de pouca ajuda nestes cálculos. Uma dica para resolver o problema é utilizar a escala logarítmica.

Lembrando que, por exemplo, se $x = 2^{-32}$, então $\log_2 x = -32$ e $\log_2 x = \log_2 e \cdot \ln x$. Oras, se

$$\log_2 e = \frac{\ln e}{\ln 2}$$

Pode-se calcular x por meio de seu logaritmo natural.

b. `0xC06051EC`

$$C06051EC_{16} = 11000000011000000101000111101100_2$$

Resposta: `-3.5050001`

c. `0x3F7FFFFF`

Resposta: Aproximadamente 1.0 ou `0.999999999...`

d. `0xFF800000`

Resposta: $-\infty$

e. `0x7F80A0A0`

Resposta: `NaN`

3. Converta o número 1.097373175×10^5 (constante de Rydberg/cm) em sua representação IEEE de precisão dupla (64 bits).

Resposta: `0x40FACA95147AE148`

4. Quais os números binários máximo e mínimo que podem ser representados com 64 bits (IEEE de precisão dupla)?

Resposta: O maior valor que pode ser representado com 64 bits (IEEE 754) é:

$$\pm(1 - (1/2)^{53}) \times 2^{1024} \approx \pm 1.7976931348623157 \times 10^{308}$$

O menor número (mais próximo de zero pelo lado positivo ou negativo) que pode ser representado de 2 formas: uma com o campo de expoente em zeros e um único bit contendo 1 na casa menos significativa da mantissa. Seu valor é de:

$$\pm 2^{-1074} \approx \pm 5 \times 10^{-324}$$

outra forma é a normalizada, ou seja, com o bit menos significativo do expoente contendo 1 e a mantissa com zeros. Seu valor é:

$$\pm 2^{-1022} \approx \pm 2.2250738585072020 \times 10^{-308}$$