Bienvenido: Ingresar
location: WebHome / TrabajosPracticos / PracticoFLOAT8

Trabajo Práctico Nro. 8 Punto Flotante

Ejercicios en el uso de Punto Flotante IEEE-754

web para pasar decimal a float http://babbage.cs.qc.cuny.edu/IEEE-754/Decimal.html

Ejercicio 1

Convertir en numero 234,625 en formato de float y double

Solución

En primer lugar calculamos

$$2^n = 234,625$$

$$n = \frac{\log 234,625}{\log 2} = 7,874212935$$

separamos el exponente en parte entera y decimal

$$2^{0,8742} \cdot 2^7 = 234,625$$

$$1,8330078125 \cdot 2^7 = 234,625$$

La mantisa calculada está comprendida entre 1 <= m < 2, como corresponde a un número normalizado, resta ahora codificar cada elemento del número.

Mantisa

La debemos pasar a binario

0,8330078125	x 2	1,666015625
0,666015625	x 2	1,33203125
0,33203125	x 2	0,6640625
0,6640625	x 2	1,328125
0,328125	x 2	0,65625
0,65625	x 2	1,3125
0,3125	x 2	0,625
0,625	x 2	1,25
0,25	x 2	0,5
0,5	x 2	1

el valor final será

1,1101010101

el valor en la mantisa a almacenar solo posee los valores a la derecha de la coma

mantisa para float	11010101010000000000000

mantisa para double	11010101010000000000000000000000000000000000

Como vemos es muy importante evitar el redondeo, el mismo puede producir un error en los bit menos significativos y en este caso al no caer a 0 deberíamos haber seguido calculando hasta el bit 23 para float o 52 para double

exponente

La codificación del exponente exige que le sumemos al valor calculado un corrimiento, este corrimiento es la mitad del máximo valor que se puede guardar según el tipo de numero

float

$$e_{float} = e + 2^{n-1} - 1 = 7 + 2^{7} - 1 = 134$$

$$e_{double} = e + 2^{n-1} - 1 = 7 + 2^{10} - 1 = 1030$$

exponente para float	134	10000110
exponente para double	1030	10000000110

signo

el signo en este caso es positivo = 0

resultado

float

31	30 23	22 0
0	10000110	11010101010000000000

double

63	62 52	51 0
0	10000000110	11010101010000000000000000000000000000000000000000

Ejercicio 2

Expresar en hexadecimal la representación de 58,75 y -58,75 en formato float

Solución: 0x426B0000 y 0xC26B0000

Ejercicio 3

Expresar en hexadecimal la representación de 4580 en formato float y double

Solución: 0x458F2000 y 0x40B1E40000000000

Ejercicio 4

Realizar un programa en assembler que reciba como argumento un número de 32 bits entero y con signo y devuelva el valor transformado a float.

efectuada por GuillermoSteiner)