

Zad 9.

niedziela, 19 marca 2023 10:33

Zadanie 9. Uzupełnij ciało poniższej procedury konwertującej binarną reprezentację liczby typu «float» umieszczoną w zmiennej «f» do wartości typu «int32_t». Wynik należy zaokrąglić w kierunku zera. Jeśli konwersja spowoduje nadmiar lub f ma wartość NaN, zwróć wartość 0x80000000 (tj. MIN_INT). Kod procedury zaczyna się wyodrębnieniem poszczególnych elementów liczby zmiennopozycyjnej:

```
1 int32_t float2int(int32_t f) {
2     int32_t s = f >> 31;          /* -1 jeśli znak był ujemny */
3     int32_t e = (f >> 23 & 255) - 127; /* wykładnik po odjęciu bias */
4     uint32_t m = f << 8 | 0x80000000; /* mantysa 1.xxx... dosunięta do lewej */
5     /* TODO */
6 }
```

Wskazówka: Wzorcówka ma dodatkowe cztery linie kodu i używa jednej instrukcji warunkowej!

```
30 int32_t float2int (int32_t f) {
31     int32_t s = f >> 31;
32     uint32_t e = f >> 23 & 0xFF;
33     uint32_t m = f & 0x7FFFFFFF;
34
35     // special cases
36     if (e == 0xFF) // +/-inf and NaN
37         return 0x80000000;
38     else if (e == 0) // denormalized or +/-0
39         return 0;
40
41     // overflows and numbers rounded to 0
42     int32_t ex = e - 127;
43     if (ex >= 31) // overflow
44         return 0x80000000;
45     if (ex < 0) // f \in (-1, 1)
46         return 0;
47
48     // ex \in [0, 30]
49     int32_t x = 1 << ex;
50     if (ex < 23) {
51         x |= m >> (23 - ex); // truncate
52     } else {
53         x |= m << (ex - 23); // add trailing 0's
54     }
55     if (s == -1) // return -x if s negative
56         return ~x+1;
57     else
58         return x;
59
60 }
```