

Napomene:

- Savjetuje se navedene zadatke riješiti ubrzo nakon predavanja
- Savjetuje se ne gledati rješenja prije nego se pokuša samostalno riješiti zadatke

6. vježbe uz predavanja

1. Odredite najveću moguću relativnu i najveću moguću apsolutnu pogrešku koja se može očekivati pri pohrani broja $2 \cdot 10^{22}$ u IEEE 754 formatu jednostruke preciznosti.
2. Gdje se (i zašto) u sljedećem odsječku programa nalaze sintaktičke pogreške:

```
int thin, tall, short;  
float which, while, when, why, who;  
char single, double, triple;  
signed long a777, 7b, _19;
```
3. Pronađite koje su konstante ispravno, a koje neispravno napisane. Za ispravno napisane konstante odredite kojeg su tipa i koliko okteta zauzimaju u memoriji:

2	4u	7f	9.1	14.5U	0101u	12.1L	12.1e+22F
12.1e22	12.1Fe-22			12.1E11L	12.1E11u	0x22L	0xABC
0x2f	2F	0x2F.1F		021.1f			
4. Napisati sadržaj registra u kojem je, prema IEEE 754 standardu za prikaz brojeva u **dvostrukoj preciznosti** pohranjen broj -0.25_{10} . Sadržaj registra napisati u heksadekadskom obliku.
5. U registru od 64 bita upisan je broj $C0\ 3D\ 80\ 00\ 00\ 00\ 00\ 00_{16}$. Napisati koji je broj predstavljen u tom registru, ako registar služi za pohranu varijable `double x`. Rezultat napisati u dekadskom brojevnom sustavu.
6. Napisati sadržaj registra u kojem je, prema IEEE 754 standardu za prikaz brojeva u **dvostrukoj preciznosti** pohranjen broj $-\infty$. Sadržaj registra napisati u heksadekadskom obliku.
7. Napisati sadržaj registra u kojem je, prema IEEE 754 standardu za prikaz brojeva u **dvostrukoj preciznosti** pohranjena vrijednost NaN. Sadržaj registra napisati u heksadekadskom obliku.
8. Odredite najveću moguću relativnu i najveću moguću apsolutnu pogrešku koja se može očekivati pri pohrani broja $2 \cdot 10^{22}$ u IEEE 754 formatu dvostruke preciznosti.

Rješenja

1. Najveća moguće relativna pogreška ovisi isključivo o broju bitova mantise m . Vodite računa o tome da parametar m uključuje i skriveni bit. Kod prikaza prema IEEE 754 standardu jednostruke preciznosti $m = 24$.

Najveća moguća relativna pogreška iznosi $2^{-24} \approx 6 \cdot 10^{-8}$

Najveća moguća apsolutna pogreška ovisi o parametru m i konkretnom broju x koji se prikazuje:

Najveća moguća apsolutna pogreška iznosi $x \cdot 2^{-24} \approx 2 \cdot 10^{22} \cdot 6 \cdot 10^{-8} = 1.2 \cdot 10^{15}$

2. `int thin, tall, short;`
`float which, while, when, why, who;`
`char single, double, triple;`
`signed long a777, 7b, _19;`

U prvom retku se za ime varijable koristi ključna riječ `short`;

U drugom retku se za ime varijable koristi ključna riječ `while`;

U trećem retku se za ime varijable koristi ključna riječ `double`;

U četvrtom retku ime varijable `7b` započinje znamenkom (nije dopušteno)

3.

<code>2</code>	signed int - 4 okteta
<code>4u</code>	unsigned int - 4 okteta
<code>7f</code>	pogreška: nedostaje točka
<code>9.1</code>	double - 8 okteta
<code>14.5U</code>	pogreška: ne postoji tip unsigned double
<code>0101u</code>	unsigned int u oktalnom obliku - 4 okteta
<code>12.1L</code>	long double - 8 okteta
<code>12.1e+22F</code>	float - 4 okteta
<code>12.1e22</code>	double - 8 okteta
<code>12.1Fe-22</code>	pogreška: F na pogrešnom mjestu
<code>12.1E11L</code>	long double - 8 okteta
<code>12.1E11u</code>	pogreška: ne postoji tip unsigned double
<code>0x22L</code>	long int u heksadekadskom obliku - 4 okteta
<code>0xABC</code>	int u heksadekadskom obliku - 4 okteta
<code>0x2f</code>	int u heksadekadskom obliku - 4 okteta
<code>2F</code>	pogreška: nedostaje točka
<code>0x2F.1F</code>	pogreška: ne može se realni broj zapisati u heksadekadskom obliku
<code>021.1F</code>	float, 0 na početku nema nikakvo značenje, ali nije pogreška - 4 okteta

4. `BFD00000000000000`

5. `-29.5`

6. FFF0000000000000

7. Prikazano je jedno od mogućih rješenja. Bitno je da su svi bitovi karakteristike postavljeni na 1, te da je barem jedan bit mantise postavljen na 1

FFF8000000000000

8. Najveća moguća relativna pogreška iznosi $2^{-53} \approx 1.1 \cdot 10^{-16}$

Najveća moguća apsolutna pogreška iznosi $x \cdot 2^{-53} \approx 2.2 \cdot 10^6$