PRIMER ZADATKA SA RANIJIH ISPITNIH ROKOVA [1]

(Februar 2006.) *Unicode* standard propisuje jedinstvene kodove za karaktere iz većine poznatih jezika, pri čemu su karakterima iz najviše korišćenih jezika dodijeljeni kodovi u opsegu $[0, 2^{16}-1]$ i taj skup karaktera se označava terminom *BMP* (*Basic Multilingual Plane*). Jedan često korišćen način zapisa *Unicode* kodova je tzv. *UTF-8*, u kome se karakteri iz *BMP*-a zapisuju sa 1, 2 ili 3 bajta, prema sledećoj tabeli:

Unicode kod	*UTF-8* zapis
`0x0000-0x007f`	`0xxxxxxx`
`0x0080-0x07ff`	`110xxxxx 10xxxxxx`
`0x0800-0xffff`	`1110xxxx 10xxxxxx 10xxxxxx`

U gornjoj tabeli, `x` su bitovi koda karaktera, pri čemu se krajnje desni `x` odnosi na bit najmanje težine i onda redom ulijevo prema bitovima veće težine. Napisati asemblersku proceduru `bmp2utf8()` koja određuje *UTF-8* zapis za dati karakter iz *BMP*-a. Procedura prihvata kao argumente *Unicode* kod karaktera kao cio broj, te pointere na lokacije za smještanje dužine odn. elemenata odgovarajućeg *UTF-8* zapisa. Napisati potom i *C* program koji sa standardnog ulaza učitava *Unicode* kod karaktera kao cio broj u heksadecimalnom formatu, poziva proceduru `bmp2utf8` da izvrši konverziju i potom štampa na standardni izlaz elemente koji sačinjavaju *UTF-8* zapis datog karaktera, opet u heksadecimalnom formatu. Primjer - za unos oblika (koji inače odgovara *Unicode* kodu ćiriličnog slova `A`):

```
0x0410
```

ispis treba da bude:

```
0xd0 0x90
```

[1] Tekst zadatka, kao i rešenje pripremio Aleksandar Samardžić