

Prevodioci i interpretatori - septembar 2003. - praktični deo

1. Napisati program koji HTML datoteku čije se ime prosledjuje kao argument komandne linije "oslobadja" etiketa, tj. generiše datoteku sa istim imenom i *txt* ekstenzijom na sledeći način:
 - Svaki red sadrži najviše 80 karaktera, pri čemu je poželjno da se reči ne "lome" na kraju reda.
 - Svi višestruki separatori se zamenjuju jednim razmakom.
 - Zaglavlje koje je navedeno unutar <HEAD> etiketa se obrađuje tako što se u datoteku sa imenom *html* datoteke i ekstenzijom *header*, prepíše njegov celokupni sadržaj (zajedno sa svim etiketama koje sadrži).
 - Dalje je potrebno obraditi samo tekst koji se nalazi unutar <BODY> etiketa.
 - Etiketa <P> se zamenjuje jednim praznim redom.
 - Etiketa
 se zamenjuje početkom novog reda bez ostavljanja praznog reda.
 - Tekst ogradjen etiketama <Hn> i <\Hn> kao i <CENTER> i <\CENTER> se centrira. Obratiti pažnju na to da dužina teksta može da predje 80 karaktera i da se u tom slučaju centriranje prenosi i u sledeće redove.
 - Ukoliko se naidje na nabranja , svaka etiketa se zamenjuje odgovarajućim rednim brojem. Voditi računa da nabranja mogu da budu "ugnježdjena".
 - Tabele koje se zadaju u okviru <TABLE> etiketa koristeći etikete <TR> i <TD> treba prepisati u tzv. *tab-delimited* format, tj. prepisati ih kao niz redova, pri čemu su pojedinačne ćelije svakog reda odvojene tabovima. Možete pretpostaviti da sve ćelije sadrže samo običan tekst.
 - etikete se zamenjuju imenom slike tj. vrednošću svog SRC atributa navedenog između uglastih zagrada.
 - Preostali tekst se jednostavno prepisuje, dok se sve ostale etikete preskaču.