

Jedanaesti čas

1. Format datoteke sa opisom DKA je takav da se u posebnim linijama (u baš tom redosledu) zadaju azbuka (samo karakteri engleske abecede razdvojeni belinama uz korišćenje specijalnog znaka -), stanja (pozitivni celi brojevi razdvojeni belinama uz korišćenje specijalnog znaka -), početna stanja, završna stanja, a potom prelazi, svaki u pojedinačnoj liniji. Između ovih linija mogu se eventualno naći jednolinijski komentari, koji počinju simbolom #.
 - (a) Napisati *Python* skript koji proverava da li je datoteka koja se zadaje kao prvi argument komandne linije u ispravnom formatu. Ispravna datoteka je ona koja ima datu strukturu i u kojoj se nalazi strogo deterministički KA.
 - (b) Program treba potom da čita sa ulaza nisku, za koju proverava da li pripada jeziku učitano automata. (domaći)

Leksička analiza - program flex

Instalacija:

```
$ sudo apt-get install flex
```

flex - program za generisanje leksičkog analizatora (info *flex* dokumentacija).

- pišemo skripte sa ekstenzijom .lex
- pozivanje i pokretanje programa:

```
$ flex naziv_fajla.lex
$ gcc -Wall -Wextra lex.yy.c
$ ./a.out
```
- program koji pišemo se sastoji iz tri dela razdvojena %%:
 - sekcija za definicije (sadrži podsekciju %{...%} u okviru koje pišemo C kôd)
 - sekcija za pravila
 - sekcija za korisnički kôd

Važne promenljive u *lex.yy.c* koji se automatski generiše pozivanjem programa *flex*:

- **char*** *yytext* - tekuća poklopljena leksema,
- **FILE*** *yyin* - ulaz sa koga se čitaju lekseme,
- **FILE*** *yyout* - izlaz na koji se piše.

Funkcija **extern int yylex (void)** je zadužena za leksičku analizu. Čita slovo po slovo sa ulaza, gradi lekseme i njima pridružuje odgovarajuće tokene. Ključna reč **extern** označava da se definicija nalazi u nekoj drugoj datoteci.

U drugoj sekciji pravila se čitaju redom kojim su napisana. **VAŽNO:**

- Ako dva pravila imaju jednaka poklapanja (npr. ključna reč i identifikator) izvršavaju se akcije prvog pravila koje je navedeno - zato je važno pravilo za ključne reči definisati pre pravila za identifikatore.
- Funkcija *yylex* uvek poklapa najviše teksta što može. Ako su dva pravila poklopljena izvršavaju se akcije onog pravila kod kojeg je duže poklapanje.

Program *flex* postoji odavno tako da nema podršku za neka proširenja regularnih izraza kao što su: \s, \w...

U drugoj sekciji se specijalno značenje karakterima može oduzeti korišćenjem navodnika, karaktera \ ili karakterske klase.

2. Napisati program koji broji nove redove i karaktere sa standardnog ulaza. Rezultat analize ispisuje na standardni izlaz.
3. Napisati program koji proverava da li su zagrade ispravno uparene.
4. Napisati program koji čisti *C* datoteku od komentara. Pretpostaviti da neće biti višelinijjskih komentara koji se završavaju sa ****/**.

5. Napisati leksički analizator koji u tekstu pronalazi celobrojne i realne konstante i vraća različite tokene. Nakon što se prepozna neki token, potrebno je saopštiti glavnom programu o kom tokenu se radi. Glavni program prijavljuje korisniku koji token je prepoznat i šta je prepoznato.