

СИНТАКСА ПРОГРАМСКИХ ЈЕЗИКА

Мирјана Ивановић, Нови Сад

Језик као средство комуникације, једноставно се може посматрати као скуп симбола који имају нека одређена својства. У основи можемо разликовати природне и вештачке језике. Посматрајмо један вештачки језик тј. језик који чине сви синтаксно исправни програми у програмском језику Pascal. Ако програм посматрамо као дугачак низ знакова тада наведени језик можемо дефинисати као скуп:

$$\text{PascalProgrami} = \{\omega \mid \omega \in \Sigma^* \text{ и } \omega \text{ је синтаксно исправан Pascal програм}\},$$

где је Σ абеледа језика а Σ^* је скуп свих низова знакова произвољне дужине који се састоје од знакова из абеледе Σ .

Приметимо да су сви програми низови знакова над задатом абеледом, али да не важи и обрнуто тј. да су сви низови знакова над абеледом програми. Pascal компајлер је програм, који између осталих активности, утврђује да ли је дати низ знакова синтаксно исправан Pascal програм.

Претходна дефиниција језика, помоћу скупа, није довољно јасна да би се на основу ње конструисао одговарајући компајлер. Прецизније речено, она не даје правила помоћу којих би се могло утврдити да ли је посматрани низ знакова из скупа PascalProgrami или није тј. не даје значење тврђења: синтаксно исправан Pascal програм.

Грамматика представља један од начина да се установе тј. запишу правила језика. Како је представљање грамматике за скуп исправних Pascal програма комплексан посао у наставку ће бити дато неколико примера једноставнијих језика.

ОСНОВНИ ПОЈМОВИ ВЕЗАНИ ЗА ГРАМАТИКЕ

Грамматика представља погодан начин за дефинисање језика. Она често представља прецизан и јасан начин да се напише рекурзиван алгоритам за одређивање да ли дати низ знакова припада језику. Такав алгоритам се назива алгоритам за препознавање језика. За запис грамматике користићемо неколико специјалних ознака:

- 1) $x \mid y$ - значи x или y ,
- 2) xy - значи да y следи x (ако је потребно ово јасније истаћи пише се $x \cdot y$),
- 3) $\langle \text{геџ} \rangle$ - значи да је геџ потребно даље дефинисати.

Посматрајмо употребу граматике на примеру једног једноставног језика.

$$\text{PascalIdent} = \{\omega \mid \omega \text{ је исправан Pascal идентификатор}\}.$$

Pascal идентификатор мора почети словом иза кога може да следи произвољан број (укључујући и 0) слова или цифара. Ради једноставности може се поставити да не постоји ограничење на дужину идентификатора.

Одговарајућа граматика овог језика би се могла записати на следећи начин.

$$\begin{aligned}\langle \text{PascalIdent} \rangle &= \langle \text{slovo} \rangle \mid \langle \text{PascalIdent} \rangle \langle \text{slovo} \rangle \mid \langle \text{PascalIdent} \rangle \langle \text{cifra} \rangle. \\ \langle \text{slovo} \rangle &= a \mid b \mid \dots \mid z \mid A \mid \dots \mid Z. \\ \langle \text{cifra} \rangle &= 0 \mid 1 \mid \dots \mid 9.\end{aligned}$$

Ова једноставна дефиниција се може прочитати као: Pascal идентификатор је слово или Pascal идентификатор иза кога следи слово или цифра. Као што се може видети у дефиницији Pascal идентификатора се користио исти тај појам тј. симбол. Такав симбол се назива рекурзиван симбол, а граматика у којој се јављају рекурзивни симболи назива се рекурзивна граматика.

Да би се одредило да ли је неки низ знакова Pascal идентификатор користи се дефиниција помоћу граматике: Ако је низ знакова ω дужине 1 и ако је то слово онда је ω из језика. Ако је ω низ знакова чија је дужина већа од 1, он је из језика ако важи један од следећих услова:

- 1) последњи знак у низу ω је слово, а низ ω без последњег знака је Pascal идентификатор
- 2) последњи знак у низу ω је цифра, а низ ω без последњег знака је Pascal идентификатор.

Користећи овај поступак одговарајућа процедура за одређивање да ли је задати низ знакова Pascal идентификатор би се у псеудо језику могла записати на следећи начин:

```
FUNCTION Id(w): BOOLEAN;
BEGIN
  IF (w је duzine 1) THEN
    Id := (w је slovo)
  ELSE
    IF (poslednji znak od w је slovo) THEN
      Id := Id(w - poslednji znak)
    ELSE
      IF (poslednji znak od w је cifra) THEN
        Id := Id(w - poslednji znak)
      ELSE
        Id := FALSE
END;
```

ПАЛИНДРОМИ

Палиндроми су низови знакова који се читају исто и слева у десно и здесна у лево. Један од опште познатих палиндрома је *анаболимилована*. Језик палиндрома се може дефинисати следећим скупом:

$$\text{Palindromi} = \{\omega \mid \omega \text{ се исто чита и слева у десно и здесна у лево}\}.$$

Да бисмо дефинисали одговарајућу граматiku за овај језик треба да одредимо правила помоћу којих се утврђује да ли је неки низ знакова палиндром или не. Уочавамо да постоји релација између следећих реченица:

- 1) ω је палиндром,
- 2) ω без првог и последњег знака је палиндром.

Другачије речено: низ знакова ω је палиндром ако и само ако су његов први и последњи знак исти и ако је низ знакова ω , без првог и последњег знака, такође палиндром. Потребан нам је још једноставан случај до кога долазимо низом уклањања првог и последњег знака. Уочавамо два случаја.

- 1) Ако је дужина низа ω паран број, при низу узастопних уклањања првог и последњег знака долазимо до случаја када у низу остају два знака или ниједан знак. Низ знакова дужине 0 је празан низ и он је палиндром.
- 2) Ако је дужина низа ω непаран број, при низу узастопних уклањања првог и последњег знака долазимо до случаја када у низу остаје један знак и када даље не можемо наставити поступак и у том случају закључујемо да је низ знакова дужине 1 палиндром.

Сумирајући дате анализе долазимо до граматике за језик палиндрома која је дата у наставку.

$$\begin{aligned} \langle \text{Pali} \rangle &= \langle \text{PrazanNiz} \rangle \mid \langle \text{Znak} \rangle \mid a \langle \text{Pali} \rangle a \mid b \langle \text{Pali} \rangle b \mid \dots \mid Z \langle \text{Pali} \rangle Z. \\ \langle \text{Znak} \rangle &= a \mid b \mid \dots \mid z \mid A \mid \dots \mid Z. \end{aligned}$$

На основу ове граматике можемо написати одговарајућу рекурзивну процедуру за препознавање палиндрома.

```
FUNCTION Pali(w): BOOLEAN;
BEGIN
  IF (w је prazan niz) OR (w је niz duzine 1) THEN
    Pali := TRUE
  ELSE
    IF prvi znak od w = poslednji znak od w THEN
      Pali := Pali(w без prvog i poslednjeg znaka)
    ELSE
      Pali := FALSE
END;
```

НИЗОВИ ОБЛИКА $a^n b^n$

Ознаком $a^n b^n$ су представљени низови који се састоје од n знакова a иза чега следи n знакова b . Овај језик се може представити скупом облика:

$$L = \{\omega \mid \omega \text{ је облика } a^n b^n \text{ за } n \geq 0\}.$$

Граматика овог језика је врло слична граматичи палиндрома. Из полазног низа знакова уклањамо први и последњи знак и проверавамо да ли је у преосталом низу први знак a , а последњи знак b . Одговарајућа граматика је облика:

$$\langle \text{IspravnaRec} \rangle = \langle \text{PrazanNiz} \rangle \mid a \langle \text{IspravnaRec} \rangle b.$$

Одговарајућа процедура, за проверу да ли је дати низ знакова из овог језика, се заснива на претходном правилу.

```

FUNCTION Prepoznaj(w): BOOLEAN;
BEGIN
  IF (duzina w је 0) THEN
    Prepoznaj := TRUE
  ELSE
    IF (prvi znak niza w је a) AND (poslednji znak niza w је b) THEN
      Prepoznaj := Prepoznaj(w без prvog i poslednjeg znaka)
    ELSE
      Prepoznaj := FALSE
END.

```

ЗАДАЦИ

1. За сваки од датих низова знакова дефинисати скуп који представља одговарајући језик, написати одговарајућу граматичу и процедуру која за дати низ знакова проверава да ли припада језику:

$$\begin{array}{ll}
 \text{a) } a^n b^{2n}, n \geq 0 & \text{в) } a^n b^m a^m b^{2n}, n \geq 0, m \geq 1 \\
 \text{б) } a^n c^m b^{2n}, n \geq 0, m \geq 0 & \text{г) } a^n b a^m, m \geq 1, n \geq 1.
 \end{array}$$

2. Дата је граматика облика: $\langle S \rangle = a \langle A \rangle$ $\langle A \rangle = a \langle B \rangle$ $\langle C \rangle = c \langle B \rangle$
 $\langle A \rangle = a \langle A \rangle$ $\langle B \rangle = b \langle C \rangle$ $\langle C \rangle = c.$

Одредити језик који генерише дата граматика и написати процедуру за препознавање низова знакова који припадају том језику (полазни симбол је $\langle S \rangle$).

**Статијата прв пат е објавена во списанието Тангента на ДМ
на Србија во 1997/98 година**