

5. Контекстно-свободни граматики и езици. Стекови автомати.

Дефиниция: Четворката $G = \langle V, \Sigma, P, S \rangle$, където

V – крайно множество от променливи (нетерминали)

Σ – крайно множество от букви (терминали), $V \cap \Sigma = \emptyset$

$P \subseteq V \times (V \cup \Sigma)^*$ – крайно множество от правила

S – начална променлива,

наричаме контекстно-свободна граматика (КСГ).

Дефиниция: Дефинираме релацията за преход $\Rightarrow_G$ (или само $\Rightarrow$)

$$u \Rightarrow_G v \Leftrightarrow u = xyz \in (V \cup \Sigma)^* \text{ \& } v = xy'z \in (V \cup \Sigma)^* \text{ \& } \exists y \rightarrow y' \in P$$

(u отива директно във v)

Дефиниция: Дефинираме извод на w_n : редицата от думи

$$(w_1, w_2, \dots, w_n) : w_i \Rightarrow w_{i+1}, \forall i = 1 \dots n$$

Дефиниция: Дефинираме индуктивно дължина на извод:

$$\forall u \in (V \cup \Sigma)^* : u \overset{0}{\Rightarrow} u$$

$$\forall u, v, w \in (V \cup \Sigma)^* : u \Rightarrow v \text{ \& } v \overset{n}{\Rightarrow} w \Rightarrow u \overset{n+1}{\Rightarrow} w$$

Означаваме рефлексивно и транзитивно затваряне на $\Rightarrow$ с $\Rightarrow^*$ т.е. $u \Rightarrow^* v \Leftrightarrow \exists n \in \mathbb{N}_0 : u \overset{n}{\Rightarrow} v$.

Дефиниция: Езикът, който се генерира от G , е $L(G) = \{ w \in \Sigma^* : S \Rightarrow^* w \}$. Език, разпознаван от КСГ, наричаме контекстно-свободен език.

Дефиниция: Нека $G = \langle V, \Sigma, P, S \rangle$ е КСГ. Дървото $D(W, E)$ (корен на D е $r \in W$) наричаме синтактично дърво (дърво на извод) на граматиката G , ако то е с наредба на синовете и за функцията $f : W \rightarrow V \cup \Sigma$ е изпълнено :

1. $\forall$ нелисто $v \in W$, $f(v)$ е нетерминал, като $f(r) = S$.
2. $\forall$ листо $v \in W$, $f(v)$ е терминал или ε .
3. За всеки нелист $v \in W$ с наредена последователност на синовете $v_{i_0} \dots v_{i_e}$ е изпълнено, че $\exists f(v) \rightarrow f(v_{i_0}) \dots f(v_{i_e}) \in P$.

Дефиниция: Граматиката $G(V, \Sigma, P, S)$ е в **нормална форма на Чомски**, ако правилата P са от вида :

$$\begin{aligned} A &\rightarrow a, & a &\in \Sigma \\ A &\rightarrow BC, & B, C &\in V \end{aligned}$$

Теорема: За всяка КСГ $G(V, \Sigma, P, S)$ съществува КСГ $G'(V', \Sigma, P', S')$ в нормална форма на Чомски, така че $L(G') = L(G) \setminus \{ \varepsilon \}$.

Доказателство: Ще илюстрираме верността на твърдението, като покажем алгоритъм за свеждане.

- 1) Елиминирание на правила, свеждащи до ϵ

```

 $V_\epsilon = \{ A \in V \mid \exists A \rightarrow \epsilon \in P \}$ 
while (  $\exists p = A \rightarrow \alpha \in P \ \& \ A \notin V_\epsilon \ \& \ \alpha \in V_\epsilon^*$  )
{
    insert A into  $V_\epsilon$ 
}
while (  $\exists X \rightarrow \alpha\beta \in P \ \& \ A \in V_\epsilon \ \& \ X \rightarrow \alpha\beta \notin P$  ){
    insert  $X \rightarrow \alpha\beta$  into P
}
 $P = P \setminus ( V \times \{ \epsilon \} )$ 

```

- 2) Премахване на преименуващите правила (от вида $A \rightarrow B$ ($A, B \in V$))

- премахване на циклични преименуващи правила

```

while (  $\exists$  цикъл  $(Y_1 \dots Y_k)$  ){
    замества  $Y_2 \dots Y_k$  с  $Y_1$  във всички правила
     $P = P \setminus \{ Y_1 \rightarrow Y_1 \}$ 
}

```

Избираме една от променливите, участваща в цикъла, и заместваме всички останали, участващи с нея.

- премахване на нецикличните преименуващи правила

```

while (  $\exists X \rightarrow Y \in P$  и  $\nexists Y \rightarrow Z \in P, X, Y, Z \in V$  ){
     $P = P \cup \{ X \rightarrow x \mid \exists Y \rightarrow x \in P, \exists X \rightarrow Y \in P \}$ 
     $P = P \setminus V \times \{ Y \}$ 
}

```

Т.е. ако имаме преименуващи правила : $X_1 \rightarrow X_2 \dots X_{n-1} \rightarrow X_n$ – избираме нетерминала X_n (той не се преименува от никой нетерминал), след това добавяме всички правила (които той генерира) към променливите, които биват преименувани на него. След това премахваме правилата, в които той е от дясната страна.

- 3) Елиминирание на дългите правила :

$A \rightarrow \alpha, |\alpha| \geq 3$

Ако съществува терминал $a \in \alpha$, то създаваме фиктивно правило $X_a \rightarrow a$ (ако такова вече съществува – използваме него) и заместваме срещанията на a в α с X_a . Така α може да се представи посредством само нетерминали.

(*) $A \rightarrow A_1 A_2 \dots A_n, |n| \geq 3$

Разбиването на това правило до правила от вида $A \rightarrow BC$ става чрез въвеждане на нови нетерминални символи и премахвайки (*) от P :

$A \rightarrow A_1 B_1$

$$B_1 \rightarrow A_2 B_2$$

...

$$B_{n-3} \rightarrow A_{n-2} B_{n-2}$$

$$B_{n-2} \rightarrow A_{n-1} A_n$$

Дефиниция: Недетерминиран стеков автомат (НСА) наричаме шесторката

$K = \langle Q, \Sigma, \Gamma, \delta, s, \# \rangle$, където :

Q – крайно множество от състояния

Σ – крайно множество от символи (азбука)

Γ – крайно множество от символи (азбука на стека), $\Sigma \subseteq \Gamma$ и $\# \in \Gamma$

$\delta : Q \times (\Sigma \cup \{ \epsilon \}) \times \Gamma \rightarrow 2^{Q \times \Gamma^*}$ – функция на прехода

s – начално състояние, $s \in Q$

$\#$ – начален стеков символ, $\# \notin \Sigma$

Дефиниция: Текуща конфигурация $(q, w, x) \in Q \times \Sigma^* \times \Gamma^*$

q – текущо състояние

w – остатък от думата

x – съдържание на стека

Дефиниция: Дефинираме релацията $\vdash$ между конфигурации :

$(p, aw, bx) \vdash (q, w, b'x)$, ако $(q, b') \in \delta(p, a, b)$

$(p, w, bx) \vdash (q', w, b'x)$, ако $(q, b') \in \delta(p, \epsilon, b)$

Рефлексивно и транзитивно затваряне на $\vdash$ означаваме с $\vdash^*$.

Дефиниция: Ако K е НСА, то $L(K) = \{ w \in \Sigma^* \mid (s, w, \#) \vdash^* (q, \epsilon, \epsilon) \}$

Теорема: L – контекстно-свободен език $\Leftrightarrow \exists$ НСА $M : L(M) = L$.

Доказателство:

$\Rightarrow$) Нека L е КСЕ. Следователно съществува КСГ $G = \langle V, \Sigma, P, S \rangle$, $L(G) = L$. Построяваме НСА $K = \langle \{ q \}, \Sigma, V \cup \Sigma, \delta, q, S \rangle$, където

(1) $\forall A \rightarrow \alpha \in P : (q, \alpha) \in \delta(q, \epsilon, A)$

(2) $\forall a \in \Sigma : (q, \epsilon) \in \delta(q, a, a)$

Лема: Нека $w\alpha \in \Sigma^*(V(V \cup \Sigma)^* \cup \{ \epsilon \})$. Тогава $S \vdash_G^* w\alpha \Leftrightarrow (q, w, S) \vdash_K^* (q, \epsilon, \alpha)$.

Доказателство:

$\Rightarrow$) Нека $S \vdash_G^* w\alpha$. Тогава съществува извод $u_0 = S \Rightarrow u_1 \Rightarrow u_2 \Rightarrow \dots \Rightarrow u_{n-1} \Rightarrow u_n = w\alpha$.

Индукция по дължината на извода :

а) $n = 0 : u_0 = S = w\alpha \Rightarrow w = \epsilon \ \& \ \alpha = S$

Очевидно $(q, w, S) = (q, \epsilon, S) \vdash^* (q, \epsilon, \alpha)$ (рефлексивно затваряне)

б) $n \Rightarrow n + 1 : u_0 = S \Rightarrow u_1 \Rightarrow u_2 \Rightarrow \dots \Rightarrow u_n \Rightarrow u_{n+1} (u_n = xVy; u_{n+1} = x\beta y)$

Съществува правило $B \rightarrow \beta \in P$, $x \in \Sigma^*$, $V \in V$. От $B \rightarrow \beta \in P \Rightarrow (q, \beta) \in \delta(q, \epsilon, V)$.

От ИП :

$(q, x, S) \vdash^* (q, \epsilon, Vy)$,

$$w\alpha = x\beta y, w \in \Sigma^*, \alpha \in V(V \cup \Sigma)^* \cup \{\varepsilon\},$$

$$w = xz, z \in \Sigma^* \Rightarrow \beta y = z\alpha$$

Следователно : $(q, w, S) \vdash^* (q, z, \beta y) = (q, z, z\alpha) \vdash^* (q, \varepsilon, \alpha)$ ($w = xz$ и $\beta y = z\alpha$ и прилагаме $|z|$ пъти правилата от тип (1))

$\Leftarrow$) Нека $(q, w, S) \vdash^* (q, \varepsilon, \alpha)$. Индукция по броя приложения преходи от тип (1)

а) $n = 0$

$(q, w, S) \vdash^* (q, \varepsilon, \alpha) \Rightarrow w = \varepsilon, \alpha = S, S \Rightarrow^* S = w\alpha$ (рефлексивно затваряне)

б) $n \Rightarrow n + 1$

$(q, w, S) \vdash^* [n \text{ пъти тип (1)}] (q, z, A\beta) \vdash (q, z, \gamma\beta); w = xz, z \in \Sigma^*; \gamma\beta = z\alpha; \alpha \in V(V \cup \Sigma)^* \cup \{\varepsilon\}$

Съществува $A \rightarrow \gamma \in P$. Следователно $(q, w, S) \vdash^* (q, z, z\alpha) \vdash^* [\text{прилагаме тип (1)}] (q, \varepsilon, \alpha)$.

По ИП : $S \Rightarrow^* xA\beta \Rightarrow x\gamma\beta = xz\alpha = w\alpha$, т.е. $S \Rightarrow^* w\alpha$.

Окончателно, при $\alpha = \varepsilon$ в лемата имаме, че $S \Rightarrow^* w\alpha \Leftrightarrow (q, w, S) \vdash^* (q, \varepsilon, \varepsilon)$.

$\Leftarrow$) Нека M е НСА с едно състояние $M = \langle \{q\}, \Sigma, \Gamma, \delta, q, \# \rangle$. Дефинираме граматиката $G = \langle \Gamma \setminus \Sigma, \Sigma, P, \# \rangle$, като $P = \{A \rightarrow a\alpha \mid (q, \alpha) \in \delta(q, a, A)\}$ (a може да е ε)

1) $L(G) \subseteq L(M)$

Нека $x \in L(G) \Rightarrow$ съществува ляв извод $\# \Rightarrow^* x$. Ще покажем, че за всяко начало w на x ($x = wu$), съществува редица от конфигурации на M .

$(q, x, \#) \vdash \dots \vdash (q, y, \alpha)$, т.е. $\# \Rightarrow^* w\alpha \Leftrightarrow (q, x, \#) \vdash^* (q, y, \alpha)$.

Индукция по $|w|$:

а) Нека $w = \varepsilon$. От $\# \Rightarrow^* \alpha \Rightarrow \{\alpha = \# \text{ и } x = y = \varepsilon\}$.

$(q, x = \varepsilon, \#) \vdash^* (q, y = \varepsilon, \#)$

б) $n \Rightarrow n + 1$

$\# \Rightarrow^* wA\beta \Rightarrow w\alpha\beta, a \in \{\varepsilon\} \cup \Sigma$

$x = w\alpha y \in \Sigma^*$

От ИП : $(q, x, \#) \vdash^* (q, ay, A\beta)$. Т.к. $A \rightarrow a\alpha \in P$, то $(q, \alpha) \in \delta(q, a, A) \Rightarrow (q, ay, A\beta) \vdash (q, y, \alpha\beta)$.

Така $(q, x, \#) \vdash^* (q, y, \alpha\beta)$.

Следователно, ако $\# \Rightarrow^* x$, то $(q, x, \#) \vdash^* (q, \varepsilon, \varepsilon)$ ($w = x, y = \varepsilon, \alpha = \varepsilon$)

1) $L(M) \subseteq L(G)$

Нека $x \in L(M)$. Тогава съществува редицата от конфигурации : $(q, x, \#) \vdash \dots \vdash (q, \varepsilon, \varepsilon)$. Трябва да покажем, че на това съответства ляв извод $\# \Rightarrow^* x$, т.е. за всяко начало w на x ($x = wu$), ако $(q, x, \#) \vdash^* (q, y, \alpha) \Rightarrow \exists$ извод $\# \Rightarrow^* w\alpha$. Доказва се аналогично. Тогава за $w = x, \alpha = \varepsilon, y = \varepsilon$ от $(q, x, \#) \vdash^* (q, \varepsilon, \varepsilon) \Rightarrow \# \Rightarrow^* x$.

Теорема: (Pumping lemma, лема за покачването) Нека L е КСЕ. Тогава $\exists n \in \mathbb{N} : \forall z \in L,$

$|z| \geq n, \exists u, v, w, x, y : z = uvwxu$:

1. $|vx| \geq 1$ (поне едната от v и $x \neq \varepsilon$)

2. $|vwx| \leq n$

3. $uv^iwx^iy \in L, \forall i \in \mathbb{N}_0$

Доказателство:

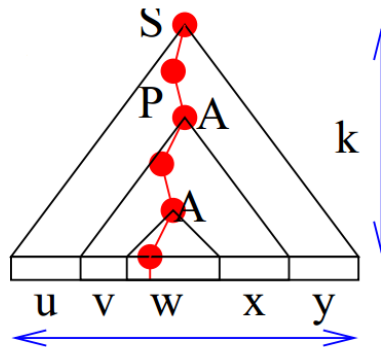
Лема: Всяко двоично дърво с височина k има най-много 2^k листа.

Доказателство: Индукция по k :

1) $k = 0 \Rightarrow$ дървото има само 1 връх (корена); той е единственото листо (2^0 на брой)

- 2) $k \rightarrow k + 1$. Разглеждаме поддървото, получено чрез премахване на листата $\rightarrow$ то е с височина k . От ИП $\Rightarrow$ поддървото има най-много 2^k листа. Следователно дървото има най-много $2 \cdot 2^k = 2^{k+1}$ листа. Следователно всяко дърво с поне 2^k листа има височина поне k .

Т.к. L е КСЕ, то съществува КСГ $G = \langle V, \Sigma, P, S \rangle$ (в НФЧ), такава че $L(G) = L$. Избираме $n = 2^{|V|+1}$, нека $k = |V|$. Нека $w' \in L$, $|w'| \geq n = 2^{k+1}$. Построяваме синтактично дърво T за w' . Т.к. $|w'| \geq 2^{k+1}$, то T има височина $\geq k+1$. Тогава съществува път P в T , който има дължина поне $(k+1)$ – той съдържа $(k+2)$ символа, единият от които е терминал, а другите – нетерминали. Следователно, от принципа на Дирихле, съществува $A \in V$, A се среща 2 пъти в пътя P . Нека изберем последната двойка, повтарящи се променливи в този път P . Фиксираме u, v, w, x, y , както е показано на чертежа.



Нека T_1 е поддървото с корен, първото срещане на A , а T_2 – поддърво с корен, второто срещане.

- 1) $|vx| \geq 1$ – ако допуснем, че $|v| = |w| = 0$, то съществува правило $A \rightarrow A$ – противоречие с нормалната форма на Чомски
- 2) $|vwx| \leq n$. Ако допуснем, че височината на T_1 е $> k+1$, то следователно съществува път с дължина $k+2 \Rightarrow$ ще намерим друга съвпадаща двойка по пътя P след първото срещане на A – противоречие с избирането на двойката. Следователно $h_{T_1} \leq k+1 \Rightarrow T_1$ има най-много 2^{k+1} листа. Следователно $|vwx| \leq 2^{k+1} = n$.
- 3) Можем многократно да добавяме дървото T_1 във второто срещане на променливата A (изключвайки T_2), включително 0 пъти. Така синтактичното дърво продължава да е коректно. Следователно $uv^iwx^iy \in L, \forall i \in \mathbb{N}_0$.

Пример: Езикът $L = \{ a^n b^n c^n \mid n \geq 0 \}$ не е КСЕ.

Доказателство:

Допускаме, че L е КСЕ. Тогава от pumping лемата $\exists n \in \mathbb{N}, \forall w' \in L$

$|w'| \geq n : w' = uvwx^iy : uv^iwx^iy \in L, \forall i$

Нека $w' = a^n b^n c^n$. Фиксираме u, v, w, x, y . Тогава от PL $\Rightarrow |vwx| \leq n$. Но тогава, както и да ги изберем uvw ще съдържа небалансирани a, b, c .

Свойства на затвореност на КСЕ:

- 1) Затворени относно обединение :

Нека L_1 и L_2 са КСЕ : $\exists G_1, G_2 : L(G_1) = L_1, L(G_2) = L_2, G_1, G_2$ – КСГи. Нека $G_1 = \langle V_1, \Sigma, P_1, S_1 \rangle$ и $G_2 = \langle V_2, \Sigma, P_2, S_2 \rangle$. Тогава $G = \langle \{S\} \cup V_1 \cup V_2, \Sigma, P_1 \cup P_2 \cup \{S \rightarrow S_1, S \rightarrow S_2\}, S \rangle : L(G) = L_1 \cup L_2$.

2) Затворени относно конкатенация :

Нека L_1 и L_2 са КСЕ : $\exists G_1, G_2 : L(G_1) = L_1, L(G_2) = L_2, G_1, G_2$ – КСГи. $G_1 = \langle V_1, \Sigma, P_1, S_1 \rangle$ и $G_2 = \langle V_2, \Sigma, P_2, S_2 \rangle$. Тогава $G = \langle \{S\} \cup V_1 \cup V_2, \Sigma, P_1 \cup P_2 \cup \{S \rightarrow S_1 S_2\}, S \rangle : L(G) = L_1 L_2$.

3) Затворени относно звезда на Клини :

Нека L_1 е КСЕ : $\exists G_1 = \langle V_1, \Sigma, P_1, S_1 \rangle$ – КСГ и $L(G_1) = L$. Тогава $G = \langle \{S\} \cup V_1, \Sigma, P_1 \cup \{S \rightarrow \epsilon, S \rightarrow S_1, S \rightarrow S_1 S_1\} \setminus \{S_1 \rightarrow \epsilon\}, S \rangle : L(G) = L_1^*$.

4) Незатвореност относно сечение :

$L_1 = \{a^i b^j c^j : i > 0, j > 0\}, L_2 = \{a^i b^i c^j : i > 0, j > 0\}$ – КСЕ.

Да допуснем, че $L_1 \cap L_2$ е КСЕ $\Rightarrow L_1 \cap L_2 = \{a^i b^i c^i : i > 0, j > 0\}$ е КСЕ – противоречие.

5) Незатвореност относно допълнение :

Допускаме, че има затвореност относно допълнение. Тогава $L_1 \cap L_2 = \overline{\overline{L_1} \cap \overline{L_2}} = \overline{\overline{L_1} \cup \overline{L_2}}$ е КСЕ – противоречие с 4).