5. Контекстносвободни граматики и езици. Стекови автомати

1. Контекстносвободна граматика, дърво на синтактичен анализ, контекстносвободен език

Деф: Контекстносвободна граматика

Контекстносвободна граматика наричаме четворката $G = \langle V, \Sigma, R, S\rangle$, където:
- Σ - азбука на символите (терминали)
- V - променливи (нетерминали) ($V \cap \Sigma = \emptyset$)
- $R \subseteq V \times (V \cup \Sigma)^{+}$, $|R| < \infty$ (крайно), $\langle \alpha, \beta \rangle \in R$ означаваме с $\alpha \underset{G}{\rightarrow} \beta$
- $S \in V$ - начална променлива

Деф: Релация на преход (извод) $\Rightarrow_G \subseteq (V \cup \Sigma)^* \times (V \cup \Sigma)^*$

Имаме $G = \langle V, \Sigma, R, S\rangle$

Извод на дума $v \in (V \cup \Sigma)^*$ от дума $u \in (V \cup \Sigma)^*$

$u \Rightarrow_G v$ т.с.т.к. $u = xAz$, $v = x\alpha z$, където $(\mathrm{A}, \alpha) \in \mathrm{R}$, $x, y \in (V \cup \Sigma)^*$

Деф: Релация на преход $\overset{n}{\Rightarrow}, \overset{*}{\Rightarrow}$

Нека $G = \langle V, \Sigma, R, S\rangle$

Дължина на извод:
- $\forall u \in (V \cup \Sigma)^* : u \underset{G}{\overset{0}{\Rightarrow}} u$
- $\forall u, v, w \in (V \cup \Sigma)^* : u \underset{G}{\Rightarrow} v \;\&\; v \underset{G}{\overset{n}{\Rightarrow}} w \rightarrow u \underset{G}{\overset{n+1}{\Longrightarrow}} w$

Извод: $u \underset{G}{\overset{*}{\Rightarrow}} v \overset{def}{\iff} (\exists n \in \mathbb{N}) \left[u \underset{G}{\overset{n}{\Rightarrow}} v \right]$. Когато се подразбира граматиката, пишем без долен индекс.

$\overset{*}{\Rightarrow}$ е рефлексивно и транзитивно затваряне на $\Rightarrow$

Деф: Език дефиниран от граматика (контекстносвободен език - КСЕ)

Нека $G = \langle V, \Sigma, R, S\rangle$

$\mathcal{L}(G) = \left\{ \omega \in \Sigma^* \mid S \underset{G}{\overset{*}{\Rightarrow}} \omega \right\}$ - език генериран от контекстносвободна граматика G

Твърдение: За произволни естествени числа l_1 и l_2 е изпълнено, че:

$\alpha \overset{l_1}{\Rightarrow} xBy$, $B \overset{l_2}{\Rightarrow} \beta$, то $\alpha \overset{l_1 + l_2}{\Longrightarrow} x\beta y$ (доказва се с индукция по l_1)

Твърдение: Нека G е КСГ и $\alpha, \beta \in (V \cup \Sigma)^*$са непразни думи, т.ч. $\alpha\beta \overset{\mathrm{n}}{\Rightarrow} \omega$.

Тогава $\exists n_1, n_2 \in \mathbb{N}$, $\exists \alpha_1, \beta_1 \in (V \cup \Sigma)^*$, за които:

$\alpha \overset{\mathrm{n_1}}{\Rightarrow} \alpha_1, \quad \beta \overset{\mathrm{n_2}}{\Rightarrow} \beta_1, \quad \omega = \alpha_1\beta_1$ и $n = n_1 + n_2$

Д-во:

Индукция по дължината на извода n:
- Нека $n = 0$, тогава $\alpha = \alpha_1, \beta = \beta_1$
- Нека $n > 0$ и $(A, \rho) \in R$: $\alpha\beta = xAy \Rightarrow_G x\rho y \overset{\mathrm{n-1}}{\Longrightarrow} \omega$, т.е. $\underbrace{xAy}_{\alpha\beta} \overset{\mathrm{n}}{\Rightarrow} \omega$

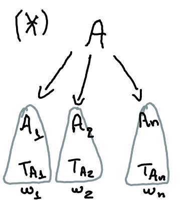
  БОО, нека A е част от α, т.е. $\alpha = uAv_1$ и $y = v_2\beta$. От ИП за $\overbrace{u\rho v_1}^{x_1} v_2\beta \overset{\mathrm{n}}{\Rightarrow} \omega$, то съществува представяне на $\omega = \alpha_1\beta_1$, т.ч. $x_1 \overset{\mathrm{n_1}}{\Rightarrow} \alpha_1$, $v_2\beta \overset{\mathrm{n_2}}{\Rightarrow} \beta_1$, $n - 1 = n_1 + n_2$. Следователно съществува представяне на $\omega = \alpha_1\beta_1$ със следния извод:

  $\alpha\beta = uAv_1v_2\beta \Rightarrow_G u\rho v_1 v_2 \beta \overset{\mathrm{n-1}}{\Longrightarrow} \alpha_1\beta_1 = \omega$, т.е. $\alpha\beta \overset{\mathrm{n}}{\Rightarrow} \alpha_1\beta_1$

Деф: Дърво на синтактичен анализ (дърво на извода) (индуктивно)

Нека $G = \langle V, \Sigma, R, S\rangle$ е КСГ. Дърво на синтактичен анализ с корен A и резултат W по G наричаме:
1. $\forall a \in \Sigma$, a е дърво с корен a и резултат a.
2. S е дърво с корен S и резултат ε,
3. Ако A е правило в G:

   $A \rightarrow \mathrm{A}_1 \dots \mathrm{A}_n \in R$ в G и за $\mathrm{A}_1, \dots, \mathrm{A}_n$ имаме дефинирани синтактични дървета $T_{\mathrm{A}_1}, \dots, T_{\mathrm{A}_n}$ с резултати съответно $w_1, \dots, w_n$, то (*) е дърво с корен A и резултат $w_1 \dots w_n$



Твърдение: Нека G е КСГ и нека $X_1 \dots X_k \overset{n}{\underset{G}{\Rightarrow}} \beta$, където $X_i \in V \cup \Sigma$ и $k \geq 2$. Тогава съществуват думи $\beta_1, \dots, \beta_k$ т.ч. за $i = 1, \dots, k$ е изпълнено $X_i \overset{n_i}{\Rightarrow} \beta_i$ и $n = \sum_{i=1}^{k} n_i$

Доказва се с пълна индукция по k като използваме предното твърдение.

Наблюдение: Следните са еквивалентни за $G = \langle V, \Sigma, R, S \rangle$ - КСГ, $\omega \in \Sigma^*$

1. $S \overset{*}{\underset{G}{\Rightarrow}} \omega$ $(\omega \in \mathcal{L}(G))$
2. Има синтактично дърво на извод с корен S и резултат ω по G.
3. $S \overset{*}{\underset{G}{\Rightarrow}} \omega$ има най-ляв извод на ω от S.

1. Доказателство на теоремите за затвореност на контекстносвободните езици

Твърдение: Безконтекстните езици са затворени относно операциите обединение, конкатенация и звезда на Клини

Д-во:

Нека $G_1 = \langle V_1, \Sigma, R_1, S_1 \rangle, G_2 = \langle V_2, \Sigma, R_2, S_2 \rangle$ са КСГ и $V_1 \cap V_2 = \emptyset$

1. Има КСГ G, т.ч. $\mathcal{L}(G) = \mathcal{L}(G_1) \cup \mathcal{L}(G_2)$
   $G = \langle V_1 \cup \mathrm{V}_2 \cup \mathrm{S}, \Sigma, R_1 \cup \mathrm{R}_2 \cup \{S \to S_1 | S_2\}, S \rangle$, където $S \notin (V_1 \cup V_2)$
   Д-во:
   - $\omega \in \mathcal{L}(G)$, $\quad S \overset{*}{\Rightarrow} \omega,\ S \in V, \omega \in \Sigma^* \Rightarrow$ или $\begin{matrix} S \Rightarrow S_1 \overset{*}{\Rightarrow} \omega,\ \omega \in \mathcal{L}(G_1) \\ S \Rightarrow S_2 \overset{*}{\Rightarrow} \omega,\ \omega \in \mathcal{L}(G_2) \end{matrix}$
   - $\omega \in \mathcal{L}(G_1) \cup \mathcal{L}(G_2)$, нека БОО $\omega \in \mathcal{L}(G_1)$, $S_1 \overset{*}{\Rightarrow} \omega$, значи $S \Rightarrow S_1 \overset{*}{\Rightarrow} \omega$ и $\omega \in \mathcal{L}(G)$

2. Има КСГ G, т.ч. $\mathcal{L}(G) = \mathcal{L}(G_1) \cdot \mathcal{L}(G_2)$
   $G = \langle V_1 \cup \mathrm{V}_2 \cup \mathrm{S}, \Sigma, R_1 \cup \mathrm{R}_2 \cup \{S \to S_1 S_2\}, S \rangle$, където $S \notin (V_1 \cup V_2)$
   Д-во:
   - $\omega \in \mathcal{L}(G)$, $\quad S \overset{*}{\Rightarrow} \omega$
     $S \Rightarrow S_1 S_2 \overset{*}{\Rightarrow} \omega$, от твърдението $\omega = uv$: $\begin{matrix} S_1 \overset{*}{\Rightarrow} u,\ \text{същият извод е в } G_1,\ u \in \mathcal{L}(G_1) \\ S_2 \overset{*}{\Rightarrow} v,\ \text{същият извод е в } G_2,\ u \in \mathcal{L}(G_2) \end{matrix}$
   - $\omega \in \mathcal{L}(G_1) \cdot \mathcal{L}(G_2), \omega = uv,\ u \in \mathcal{L}(G_1), v \in \mathcal{L}(G_2)$, $\quad \left. \begin{matrix} S_1 \overset{*}{\Rightarrow} u \\ S_2 \overset{*}{\Rightarrow} v \end{matrix} \right\} S \Rightarrow S_1 S_2 \overset{*}{\Rightarrow} u S_2 \overset{*}{\Rightarrow} uv$

3. Има КСГ G, т.ч. $\mathcal{L}(G) = \mathcal{L}(G_1)^*$
   $G = \langle V_1 \cup \{\mathrm{S}\}, \Sigma, R_1 \cup \{S \to \varepsilon \mid \mathrm{SS}_1\}, S \rangle$, където $S \notin V_1$
   Д-во:
   - $\omega \in \mathcal{L}(G)$.
     - $\omega = \varepsilon$, $\quad \omega = \varepsilon \in \mathcal{L}(G_1)^*$
     - $\omega \neq \varepsilon, S \overset{n}{\Rightarrow} \omega, n \geq 1$. Индукция по n
       - $n = 1$: $\omega = \varepsilon$, но $\omega \neq \varepsilon$ - абсурд
       - $n = 2$: $S \Rightarrow SS_1 \Rightarrow \omega$ - не може
       - $n = 3$: $S \Rightarrow SS_1 \Rightarrow S_1 \Rightarrow \omega$, значи $S_1 \Rightarrow \omega \in R_1$, следователно $\omega \in \mathcal{L}(G_1)$
       - Стъпка: $n \to n+1$: $S \Rightarrow SS_1 \overset{n}{\Rightarrow} \omega$

         По твърдение: $\omega = uv$, $\begin{matrix} S \overset{<n}{\Rightarrow} u,\ \text{от ИП имаме } u \in \mathcal{L}(G_1)^* \\ S_1 \overset{<n}{\Rightarrow} v \in R_1, \text{значи } v \in \mathcal{L}(G_1) \end{matrix} \Rightarrow \omega \in \mathcal{L}(G_1)^*$
   - $\omega \in \mathcal{L}(G_1)^*$
     - $\omega = \varepsilon$, $\quad S \to \varepsilon \in R_1 \cup \{S \to \varepsilon \mid \mathrm{SS}_1\}$, значи $\omega \in \mathcal{L}(G)$
     - $\omega = u_1 \dots u_k$, $\quad u_i \in \mathcal{L}(G_1), i \in \{1, \dots, k\}$

       $S_1 \overset{*}{\underset{G_1}{\Rightarrow}} u_i$, $\quad i \in \{1, \dots, k\}$

       $S \Rightarrow_G SS_1 \underset{2}{\Rightarrow} \dots \underset{k}{\Rightarrow} S \underbrace{S_1 \dots S_1}_{k} \Rightarrow S_1 \dots S_1 \overset{*}{\Rightarrow} u_1 S_1 \dots S_1 \overset{*}{\Rightarrow} \dots \overset{*}{\Rightarrow} u_1 \dots u_k$, т.е. $\omega \in \mathcal{L}(G)$

Освен това съществуват КСГ за езиците $\{a\}$ $(S \to a)$, $\emptyset$ $(S \to S)$, значи всвки регулярен език е КСЕ.

Деф: Нормална форма на Чомски

Една безконтекстна граматика е в НФЧ, ако всяко правило е от вида $A \to BC$ и $A \to a$

3. Недетерминистичен стеков автомат. Изпълнение в недетерминиран стеков автомат. Език, разпознаван от недетерминиран стеков автомат.

Деф: Недетерминиран стеков автомат (НСА) $P = \langle Q, \Sigma, \Gamma, \#, \Delta, S, F \rangle$

- Q - крайно множество от състояния
- Σ - крайна входна азбука
- Γ - крайна стекова азбука
- $\# \in \Gamma$ - символ за дъно на стека
- $\Delta: Q \times \Sigma_\varepsilon \times \Gamma \to \mathcal{P}(Q \times \Gamma^{\leq 2})$ е функция на преходите
- $S \in Q$ - начално състояние
- $F \in Q$ - заключително състояние

Деф: Конфигурация (моментно описание) на изчислението със стеков автомат е тройка от вида $(q, \alpha, \gamma) \in Q \times \Sigma^* \times \Gamma^*$, т.е. Автоматът:

- Е в *състояние* q
- Остава да прочете *думата* α
- Има съдържание на *стека* γ

Деф: $P = \langle Q, \Sigma, \Gamma, \#, \Delta, S, F \rangle$ - НСА, $b \in \Sigma$,

$(p, \beta) \in \Delta(q, \varepsilon, A)$, то бележим: $(q, \alpha, A\gamma) \vdash_P (p, \alpha, \beta\gamma)$

$(p, \beta) \in \Delta(q, b, A)$, то бележим: $(q, b\alpha, A\gamma) \vdash_P (p, \alpha, \beta\gamma)$

Деф: Нека k и k' са конфигурации. Дефинираме $\vdash_P^n$ над $Q \times \Sigma^* \times \Gamma^*$, която ни казва, че k се променя до k' след изчисление от n стъпки на стековия автомат:

- $k \vdash_P^0 k$ (рефлексивност)
- $k \vdash_P k'',\ k'' \vdash_P^n k'$, то $k \vdash_P^{n+1} k'$ (транзитивност)

$$k \vdash_P^* k' \overset{def}{\iff} (\exists n \in \mathbb{N})[k \vdash_P^n k']$$

Деф: Език разпознаван от стековия автомат $P = \langle Q, \Sigma, \Gamma, \#, \Delta, S, F \rangle$

$\mathcal{L}(P) \overset{\text{def}}{=} \{\omega \in \Sigma^* \mid (S, \omega, \#) \vdash_P^* (F, \varepsilon, \varepsilon)\}$

4. Свеждане на контекстносвободна граматика към еквивалентен недетерминиран стеков автомат

Лема: За всяка КСГ $G = \langle V, \Sigma, R, S \rangle$ съществува стеков автомат P, т.ч. $\mathcal{L}(G) = \mathcal{L}(P)$

Д-во:

Нека е дадена КСГ $G = \langle V, \Sigma, R, S \rangle$ в НФЧ. Дефинираме $P = \langle Q, \Sigma, \Gamma, \#, \Delta, q_S, F \rangle$:

- $Q = \{S, q, F\}$
- $\Gamma = \Sigma \cup V \cup \{\#\}$
- Δ:
  (1) $\Delta(q_S, \varepsilon, \#) \overset{\text{def}}{=} \{(q, q_S\#)\}$
  (2) $\Delta(q, \varepsilon, A) \overset{\text{def}}{=} \{(q, \alpha) \mid A \to_G \alpha$ е правило в $G\}$ за всяка променлива $A \in V$:
  (3) $\Delta(q, a, a) \overset{\text{def}}{=} \{(q, \varepsilon)\}$ за всяка буква $a \in \Sigma$:
  (4) $\Delta(q, \varepsilon, \#) \overset{\text{def}}{=} \{(F, \varepsilon)\}$

Ще докажем, че

$S \overset{*}{\Rightarrow} \omega\alpha$ (ляв извод) $\Leftrightarrow \langle q, \omega, S \rangle \vdash_P^* \langle q, \varepsilon, \alpha \rangle, \qquad \omega \in \Sigma^*, \alpha \in \{\varepsilon\} \cup V(\Sigma \cup V)^*$

Д-во:

Нека $S \overset{*}{\Rightarrow} \omega\alpha$. Нека $S \overset{n}{\Rightarrow} \omega\alpha$, $\omega \in \Sigma^*, \alpha \in \{\varepsilon\} \cup V(\Sigma \cup V)^*$. Индукция по n:

- $n = 0$: $S = w\alpha$, т.е. $\omega = \varepsilon, \alpha = S$, имаме $\langle q, \underbrace{\omega}_{\varepsilon}, S \rangle \vdash_P^* \langle q, \varepsilon, \underbrace{\alpha}_{S} \rangle$
- ИП: Нека твърдението е вярно за n
- Стъпка:
  $S \overset{n+1}{\Longrightarrow} \omega\alpha, \qquad S \overset{n}{\Rightarrow} xA\beta \Rightarrow \omega\alpha, \qquad x \in \Sigma^*, A \in V, \beta \in (\Sigma \cup V)^*$
  $A \to \gamma \in R \quad xA\beta \Rightarrow x\gamma\beta = \omega\alpha.$
  $\omega = xy, \qquad$ за някое $y \in \Sigma^*, \qquad x\gamma\beta = xy\alpha \Rightarrow \gamma\beta = y\alpha$
  За $S \overset{n}{\Rightarrow} xA\beta$ по ИП $\langle q, x, S \rangle \vdash_P^* \langle q, \varepsilon, A\beta \rangle$

$A \to \gamma \in R$, следователно

$$\langle q, xy, S\rangle \vdash_P^* \langle q, y, A\beta\rangle \vdash_P \Big\langle q, y, \underbrace{\gamma\beta}_{y\alpha}\Big\rangle \underbrace{\vdash_P^*}_{\substack{(3),\\ |y|\text{ пъти}}} \langle q, \varepsilon, \alpha\rangle$$

Обратно:

Нека $\langle q, \omega, S\rangle \vdash_P^* \langle q, \varepsilon, \alpha\rangle$

Индукция по броя n на преходите от тип (2):

- $n = 0$: S не е терминал, няма преходи, $\omega = \varepsilon, \alpha = S$
- $n \to n+1$: $\langle q, \omega, S\rangle \vdash_P^{\text{n тип (2)}} \langle q, y, A\beta\rangle \vdash_P \Big\langle q, y, \underbrace{\gamma\beta}_{y\alpha}\Big\rangle \vdash_P^* \langle q, \varepsilon, \alpha\rangle$,

  Където $\omega = xy,\ x, y \in \Sigma^*, A \to \gamma \in R, \gamma\beta = y\alpha$

  Значи $\langle q, x, S\rangle \vdash_P^{\text{n тип (2)}} \langle q, \varepsilon, A\beta\rangle$, то по ИП $S \overset{*}{\Rightarrow} xA\beta$

  $S \overset{*}{\Rightarrow} xA\beta \Rightarrow x\gamma\beta = xy\alpha = \omega\alpha$, т.е. $S \overset{*}{\Rightarrow} \omega\alpha$

Следствие: За $\alpha = \varepsilon, \forall\omega \in \Sigma^*$

$$\omega \in \mathcal{L}(G) \leftrightarrow S \overset{*}{\Rightarrow} \omega \leftrightarrow \langle q, \omega, S\rangle \vdash_P^* \langle q, \varepsilon, \varepsilon\rangle \xleftrightarrow{P-def} \langle q_s, \omega, \#\rangle \vdash_P^* \langle F, \varepsilon, \varepsilon\rangle \leftrightarrow \omega \in \mathcal{L}(\mathrm{P})$$

3. **Лема за разрастване на контекстносвободни езици**

Деф: *Дължина* на път се нарича броят на преходите в пътя

Височина на дърво е дължината на най-дългия път

Твърдение: Ако T е двоично дърво, в което всеки път е с дължина $< k$, то броят на листата е $< 2^k$

Лема (PL): Ако $L \subseteq \Sigma^*$, е КСЕ, то има число $n \in \mathbb{N}$, т.ч. за $\forall$ дума $z \in L,\ |z| \geq n$ има думи $z = uvwxy$, т.ч.

(1) $|vx| \geq 1$

(2) $|vwx| \leq n$

(3) $\forall i \in \{0,1,\ldots\}$: $uv^iwx^iy \in L$

Д-во:

Нека $G = \langle V, \Sigma, R, S\rangle$ е КСГ, т.ч. $\mathcal{L}(G) = L$ и БОО, нека G е в НФЧ.

Нека $|V| = k$. Избираме $n = 2^k$. Нека $z \in \Sigma^*, |z| = m \geq n = 2^k$

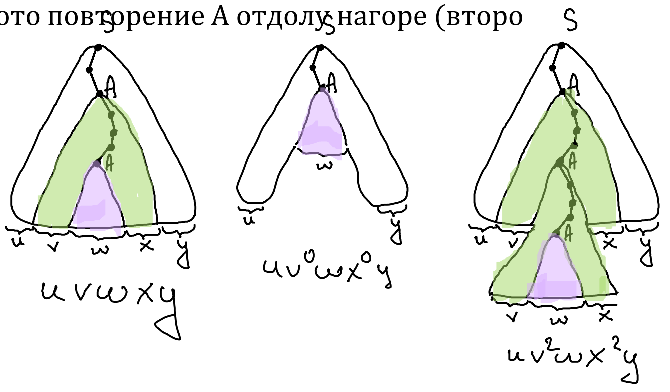
Ще разгледаме синтактичното дърво T на извод на z. То има макс.имален брой наследници=2, значи има $\geq 2^k$ листа.

$\Rightarrow$ от твърдението има път с дължина $k \Rightarrow$ променливите по пътя са поне $\mathrm{k}+1$, но $|V| = k$

$\Rightarrow$ поне 2 променливи са еднакви. Избираме първото повторение A отдолу нагоре (второ появяване отдолу нагоре). Имаме че:

- $|vx| \geq 1$: ако $vx = \varepsilon$, тонямa 2 срещания на A
- $|vwx| \leq n = 2^k$ (първо повторение, най-дългият път в поддървото на 2то A отдолу нагоре е дълъг $\leq k$, от твърдението значи резултатът (листата) $|vwx| \leq 2^k = n$
- $uv^iwx^iy \in L$ от:

  $(\mathrm{S} \Rightarrow^* uAy, \qquad A \Rightarrow^* w, \qquad A \Rightarrow^* vAx)$



4. **Примери за езици, които не са КСЕ**

Пример: $L = \{a^nb^nc^n \mid n \in \mathbb{N}\}$ не е КСЕ

Д-во:

Допускаме, че L е КСЕ, тогава PL е в сила, т.е. $\exists n: \forall z \in L,\ |z| \geq n,\ \exists u, v, w, x, y \in \Sigma^*(x = uvwxy)$ и $|vx| \geq 1, \qquad |vwx| \leq n$

Нека n_0 е такова число. Избираме $z = a^{n_0}b^{n_0}c^{n_0} \Rightarrow$ във vx не могат да участват и трите букви.

- 1сл. $vx = a^lb^k$, като $l + k \geq 1 \Rightarrow l \geq 0 \vee k \geq 0$

  $uwy = a^{n_0-l}b^{n_0-k}c^{n_0}$, или $n_0 - l < n_0$ или $n_0 - k < n_0$, значи $uwy \notin L$, но по PL $uwy \in L$.

  Противоречие
- 2сл $vx = b^lc^k$ - аналогично

5. Незатвореност на КСЕ относно допълнение и сечение

Пример: Сечението не запазва контекстносвободноста:

$L_1 = \{a^n b^n c^k \mid n, k \in \mathbb{N}\}$ – КСЕ

$L_2 = \{a^n b^k c^k \mid n, k \in \mathbb{N}\}$ – КСЕ

$L_1 \cap L_2 = \{a^n b^n c^n \mid n \in \mathbb{N}\}$ – не е КСЕ

Знаем, че $L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$, ако бяха затворени относно допълнението, щяха да са затворени и относно сечението, но дадохме контрапример, следователно не са затворени относно допълнението и сечението