9. Безконтекстни Граматики (КСГ)

Def Безконтекстна (Контекстно-свободна) граматика наричаме всяка наредена четворка $G = (\Gamma, \Sigma, S, R)$, където:

Γ, Σ - крайни азбуки, $\Sigma \subseteq \Gamma$

$S \in \Gamma$

$R \subseteq \Gamma \times (\Gamma \cup \Sigma)^*$ е крайно мн-во

S - начален символ на граматиката.

Σ - терминални символи на граматиката

Γ - нетерминални символи на граматиката

R - правила на граматиката

Пример: $G = (\{a, b, S\}, \{a, b\}, S,$
$\{(S, \varepsilon), (S, aSb)\})$ е безконт. граматика.

За кратко пишем:

$G: \quad S ::= aSb \mid \varepsilon$

Def $(A, \alpha) \in R$ пишем като $A \to_G \alpha$ или $A \to \alpha \in R$ и казваме, че в G можем да заменим A със α или да изведем α от A.

Def За вс. $w_1, w_2 \in \Gamma^*$ и вс. $A \to \alpha \in R$ деф. релацията

$$\Rightarrow_G \subseteq (\Gamma^*)^2 \quad \text{и} \quad w_1 A w_2 \Rightarrow_G w_1 \alpha w_2$$

Def $\Rightarrow_G^* \subseteq (\Gamma^*)^2$ е рефлексивното и транзитивно затваряне на $\Rightarrow_G$

Def Език на граматиката $G = (\Gamma, \Sigma, S, R)$ наричаме

$$L(G) = \{ w \in \Sigma^* \mid S \Rightarrow_G^* w \}$$

Def $L \subseteq \Sigma^*$ наричаме безконтекстен ако има безконтекстна граматика G над азбуката Σ с език $L(G) = L$.

Примери:

G_1: $S ::= Hello\_World$

G_2: $S ::= aS \mid \varepsilon$

$G_{\mathbb{Z}}$: $S ::= N \mid -N$

$N ::= 1N' \mid 2N' \mid \dots \mid 9N' \mid 0$

$N' ::= 1N' \mid 2N' \mid \dots \mid 9N' \mid \varepsilon \mid 0N'$

$G_{()}$: $S ::= (B)$

$B ::= S \mid SB$

$\{a^n b^n \mid n \in \mathbb{N}\}$ не е регулярен. Безконтекстен ли е?

как се извежда $(()(()())())$

как се извежда $aaabbb$

граматика за аритметични изрази?

конкатенация?

обединение?

звезда?

сечение??

Def граматиката $G = (\Gamma, \Sigma, S, R)$ наричаме регулярна / дясно линейна ако правилата в G са от вида

$$A \rightarrow wB, \quad w \in \Sigma^*, \quad B \in \Gamma \setminus \Sigma$$

$$A \rightarrow w, \quad w \in \Sigma^*$$

Твърдение рег езици $\subseteq$ езици на рег. граматики

Д-во Ще докажем с индукция по регулярните изрази.

Първо $G_\emptyset = (\{\emptyset\}, \Sigma, \emptyset, \emptyset)$ очевидно $L(G_\emptyset) = \emptyset$

$G_\varepsilon = (\{S\}, \Sigma, S, \{S \rightarrow \varepsilon\})$ очевидно $L(G_\varepsilon) = \{\varepsilon\}$

за вс. $\sigma \in \Sigma$, $G_\sigma = (\{S\}, \Sigma, S, \{S \rightarrow \sigma\})$ очевидно $L(G_\sigma) = \{\sigma\}$

Готови сме с базата, сега трябва за произволни L_1, L_2 - рег. езици, които са езици на рег. граматики да докажем, че $L_1 \circ L_2$, $L_1 \cup L_2$, L_1^* са също езици на рег. граматики.

иначе казано, трябва за две произволни регулярни граматики G_1, G_2 т.че $L(G_1) = L_1$ и $L(G_2) = L_2$ да конструираме регулярни граматики $G_\cup . G_\circ , G^*$

т.че $L(G_\cup) = L(G_1) \cup L(G_2)$

$L(G_\circ) = L(G_1) \circ L(G_2)$

$L(G^*) = L(G_1)^*$

Ще използваме следните конструкции:

Нека $G_1 = (N_1, \Sigma, S_1, R_1)$

$G_2 = (N_2, \Sigma, S_2, R_2)$ и Б.О.О. $N_1 \cap N_2 = \emptyset$

$G_\cup = (N_1 \cup N_2 \cup \{S\}, \Sigma, S, R_\cup)$,

където $S \notin N_1 \cup N_2$ и

$R_\cup = R_1 \cup R_2 \cup \{S \to S_1, S \to S_2\}$

Ясно е, че правилата на $G_\cup$ са от искания вид, т.е. $G_\cup$ е регулярна граматика.

Ще докажем, че $L(G_\cup) = L(G_1) \cup L(G_2)$

$\supseteq$ Нека $w \in L(G_1) \cup L(G_2)$

сл. $w \in L(G_1)$ или $w \in L(G_2)$

сл. $S_1 \overset{*}{\underset{G_1}{\Rightarrow}} w$ или $S_2 \overset{*}{\underset{G_2}{\Rightarrow}} w$

но $R_\cup \supseteq R_1$ и $R_\cup \supseteq R_2$ и следователно

$S_1 \overset{*}{\underset{G_\cup}{\Rightarrow}} w$ или $S_2 \overset{*}{\underset{G_\cup}{\Rightarrow}} w$

но $S \underset{G_\cup}{\Rightarrow} S_1$ и $S \underset{G_\cup}{\Rightarrow} S_2$, сл. е вярно

или $S \Rightarrow S_1 \overset{*}{\Rightarrow} w$ или $S \Rightarrow S_2 \overset{*}{\Rightarrow} w$,

т.е. винаги $S \overset{*}{\Rightarrow} w$, т.е. $w \in L(G_\cup)$

(c) Нека $w \in L(G_U)$. Тогава $S \Rightarrow_{G_U} w$

Но S участва само в две правила $S \to S_1$ и $S \to S_2$

и тогава изводът на $\underline{\underline{w}}$ в $\underline{\underline{G_U}}$ може да изглежда

само така:

$$S \Rightarrow_{G_U} S_1 \Rightarrow^*_{G_U} w$$

или

$$S \Rightarrow_{G_U} S_2 \Rightarrow^*_{G_U} w$$

И в двата случая, тъй като $N_1 \cap N_2 = \emptyset$, то правилата след първото в извода са или само от R_1, или само от R_2, т.е.

$$\text{или } S \Rightarrow_{G_U} S_1 \Rightarrow^*_{G_1} w,$$
$$\text{или } S \Rightarrow_{G_U} S_2 \Rightarrow^*_{G_2} w$$

, т.е. $w \in L(G_1)$ или $w \in L(G_2)$

иначе казано, $w \in L(G_1) \cup L(G_2)$

$\square$.

Сега продължаваме с конкатенацията:

$G_o = (N_1 \cup N_2, \Sigma, S_1, R_o)$, където

$$R_o = \{ A \to wB \mid A \to wB \in R_1,\ w \in \Sigma^*,\ B \in N \} \cup$$
$$\{ A \to w S_2 \mid A \to w \in R_1,\ w \in \Sigma^* \} \cup$$
$$R_2$$

Ще докажем, че $L(G_o) = L(G_1) \circ L(G_2)$

Пример: G_1: $S_1 ::= a S_1$ $\quad G_2$ $S_2 ::= c a D$

$S_1 ::= b$ $\quad D ::= a d$

G_0: $S_1 := aS_1$

$S_1 := bS_2$

$S_2 := caD$

$D := ad$

Д-во

(⊆) Нека $w \in L(G_0)$, т.е. $S_1 \Rightarrow^*_{G_0} w$

Всички правила в G_0 съдържат 0 или 1 нетерминали, т.е. във всяка изведена дума има 0 или 1 нетерминали. Освен това, във всеки извод може да се прилага правило, несъдържащо нетерминали най-много веднъж.

Щом $S_1 \Rightarrow^*_{G_0} w$ и в $\underline{w}$ няма нетерминали, то правило без нетерминали е приложено точно веднъж. Такива правила в G_0 са породени само от G_2, и тъй като $N_1 \cap N_2 = \emptyset$, то някъде в извода има S_2. Иначе казано, има думи $w', w'' \in \Sigma^*$ т.че

$$S_1 \Rightarrow^*_{G_0} w' S_2 \Rightarrow^* w'w'' = w$$

Тук $w'' \in L(G_2)$ тъй като $S_2 \Rightarrow^* w''$ и приложените правила са само от R_2.

Щом $S_1 \Rightarrow^*_{G_0} w' S_2$, значи е приложено правило $A \to v S_2$ от G_0. То е породено от правилото $A \to v \in R_1$ за някои $A \in N_1$ и $v \in \Sigma^*$. Тогава има дума $u \in \Sigma^*$ т.че

$$S_1 \Rightarrow^*_{G_0} uA \Rightarrow_{G_0} uvS_2 = w'S_2$$

В извода, съответстващ на $S_1 \Rightarrow^*_{G_0} uA$ може да са използвани само правила от R_1 и сл.

$$S_1 \Rightarrow^*_{G_1} uA \Rightarrow_{G_1} uv = w', \text{ сл. } w' \in L(G_1)$$

Така $w'w'' = w$ и $w' \in L(G_1)$ и $w'' \in L(G_2)$,

т.е. $w \in L(G_1) \circ L(G_2)$

$\supseteq$ Нека $w \in L(G_1) \circ L(G_2)$. Тогава $\exists w', w'' \in \Sigma^*$ т.че.

$w = w'w''$, $w' \in L(G_1)$ $w'' \in L(G_2)$, т.е

$$S_1 \Rightarrow^*_{G_1} w' \text{ и } S_2 \Rightarrow^*_{G_2} w''.$$

Последното приложено правило в извода на $\underline{\underline{w_1}}$ от $\underline{\underline{S_1}}$ в $\underline{\underline{G_1}}$ е от вида $A \longrightarrow v$, където $v \in \Sigma^*$.

Тогава има дума $u \in \Sigma^*$ т.че $uv = w'$ и

$$S_1 \Rightarrow^*_{G_1} uA \Rightarrow uv = w'$$

Тогава в G_0 имаме следния извод:

$$S_1 \Rightarrow^*_{G_0} uA \Rightarrow_{G_0} \underbrace{uv}_{w'} S_2 \Rightarrow^*_{G_0} w'w'' = w$$

сл. $w \in L(G_0)$

□

Звезда на Клини

$G^* = (N_1 \cup \{S\}, \Sigma, S, R^*)$, където

$S \notin N_1$ и

$R^* = R \cup \{S \to \varepsilon, S \to S_1\} \cup$

$\{A \to wB \mid A \to wB \in R_1, w \in \Sigma^*, B \in N\} \cup$

$\{A \to wS \mid A \to w \in R_1, w \in \Sigma^*\} \cup$

Пример:

G_1: $S_1 ::= abcA$
$A ::= aA$
$A ::= \varepsilon$

G^*: $S ::= S_1$
$S ::= \varepsilon$
$S_1 ::= abcA$
$A ::= aA$
$A ::= S$

Ще докажем, че $L(G^*) = L(G_1)^*$

$\subseteq$ Нека $w \in L(G^*)$, т.е. $S \overset{*}{\Rightarrow}_{G^*} w$. Всички правила в R^* освен $S \to \varepsilon$ съдържат поне един нетерминал, следователно в извода, съответстващ на $S \Rightarrow^*_{G^*} w$ или има само едно правило $S \to \varepsilon$ и $w = \varepsilon$ (и очевидно $\in L(G_1)^*$), или някъде се среща $\underline{S}$ за втори път.

Индукция по броя срещания на $\underline{S}$ в извода $S \Rightarrow^*_{G^*} w$.

База: 1: Тогава $w = \varepsilon$ ok

И.Х Нека за всяка дума $w \in \Sigma^*$, т.че $S \Rightarrow^*_{G^*} w$ се извежда с $\underline{n}$ срещания на $\underline{S}$ в извода следва, че $w \in L(G_1)^*$

И.С. Нека $S \underset{G^*}{\Rightarrow^*} w$ и в съответстващия извод

$\underline{\underline{S}}$ се среща $\underline{\underline{n+1}}$ пъти.

Тогава $S \Rightarrow^* w'S \Rightarrow^* w'w'' = w.$

От И.Х. $w'' \in L(G_1)^*$ тъй като в остатъка от

извода се среща $\underline{\underline{S}}$ точно $\underline{\underline{n}}$ пъти.

Сега тъй като $S \Rightarrow^* w'S$, то точно веднъж е било

приложено правило $A \rightarrow w_2 S$, т.е.

$$S \Rightarrow^* w_1 A \Rightarrow w_1 w_2 S \quad и \quad w_1 w_2 = w'$$

Тогава в G_1 може да се направи следния извод:

$$S \Rightarrow^* w_1 A \Rightarrow w_1 w_2 = w'$$

Оттук $w' \in L(G_1)$, сл. $w'w'' \in L(G_1) \cdot L(G_1)^* \subseteq L(G_1)^*$

сл. $w \in L(G_1)^*$

$\square$

А сега ще докажем и обратната посока, т.е. че

ако L е език на рег. граматика, сл. L е регулярен.

За целта ще конструираме автомат от граматиката.

Нека $G = (N, \Sigma, S, R)$ е регулярна граматика.

Първо ще конструираме граматиката G' по следния начин:

- Всяко правило $A \to wB$ заменяме с правилата:

$$A \to w_1 A_1$$
$$A_1 \to w_2 A_2$$
$$\vdots$$
$$A_{|w|-1} \to w_{|w|} B \quad ,$$

където $A_1 \dots A_{|w|-1}$ са нови нетерминали

- Всяко правило $A \to w$ заменяме с правилата:

$$A \to w_1 A_1$$
$$A_1 \to w_2 A_2$$
$$\vdots$$
$$A_{|w|-1} \to w_{|w|} A_{|w|}$$
$$A_{|w|} \to \varepsilon$$

където $A_1 \dots A_{|w|}$ са нови нетерминали.

Така е очевидно, че получаваме $L(G') = L(G)$.

Пример:

$S := abc$
$S := aA$
$A := aA$
$A := \varepsilon$

$\longrightarrow$

$S := aS_1 \mid aA$
$S_1 := bS_2$
$S_2 := cS_3$
$S_3 := \varepsilon$
$A := aA$
$A := \varepsilon$

Нека $G' = (N', \Sigma, S', R')$.

Def $A_G = (\Sigma, N', \{S'\}, \Delta, F)$

$F = \{ A \in N' \mid A \to \varepsilon \in R' \}$

$\Delta = \{ (A, \sigma, B) \mid A \to \sigma B \in R' \}$

Тв $L(A_G) = L(G') = L(G)$

ясно

Д-во

($\supseteq$) Нека $w \in L(G')$.

Аналогично на предни разсъждения, точно веднъж в извода $S' \Rightarrow^* w$ се прилага правило $A \to \varepsilon$ от R'.

т.е. извода изглежда така:

$$S' \Rightarrow w_1 A_1 \Rightarrow w_1 w_2 A_2 \Rightarrow w_1 w_2 w_3 A_3 \Rightarrow \dots \Rightarrow w_1 \dots w_n A_n \Rightarrow w \varepsilon.$$

Тогава от построението на Δ, има път π в A_G:

$\pi : S' \xrightarrow{w_1} A_1 \xrightarrow{w_2} A_2 \xrightarrow{w_3} \dots \xrightarrow{w_{|w|}} A_n$, където S' е начално и $A_n \to \varepsilon \in R'$,

т.е. $A_n \in F$. Тогава $w \in L(A_G)$

($\subseteq$) Нека $w \in L(A_G)$. Тогава има път π:

$\pi : S' \xrightarrow{w_1} A_1 \xrightarrow{w_2} A_2 \xrightarrow{w_3} \dots \xrightarrow{w_{|w|}} A_n$, където $A_n \in F$.

Тогава

$S \to w_1 A_1$, $A_1 \to w_2 A_2 \dots A_{n-1} \to w_n A_n \in R'$ и $A_n \to \varepsilon \in R'$

Оттук, можем да видим, че в G' има извод:

$$S' \Rightarrow w_1 A_1 \Rightarrow w_1 w_2 A_2 \Rightarrow \dots \Rightarrow w_1 \dots w_n A_n \Rightarrow w_1 \dots w_n = w$$

Тогава $S' \overset{*}{\underset{G'}{\Rightarrow}} w$, т.е. $w \in L(G')$.

$\square$