Chomského normálny tvar, CYK, FIRST, FOLLOW

Ing. Viliam Hromada, PhD.

C-510
Ústav informatiky a matematiky
FEI STU

viliam.hromada@stuba.sk

Konverzia na Chomského normálny tvar č. 1

Preveďte gramatiku $G = (N, T, P, S)$ do Chomského normálneho tvaru.
$N = (S, A, B)$, $T = \{a, b, c\}$. Pravidlá P:

- $S \rightarrow aSb \mid A$
- $A \rightarrow cB \mid cc$
- $B \rightarrow aSBc \mid \varepsilon$

Odstránenie ε-pravidiel

Prvý krok úpravy do CHNT je odstránenie ε-pravidiel. Pre túto gramatiku:
$N_\varepsilon = \{B\}$ a výsledná gramatika bez ε-pravidiel:

- $S \to aSb \mid A$
- $A \to cB \mid cc \mid c$
- $B \to aSBc \mid aSc$

Počiatočný neterminál

Ďalší krok je úprava, ktorá v prípade, že počiatočný neterminál S je na pravej strane niektorého z pravidiel pridá do gramatiky nový počiatočný neterminál $\acute{S}$ a pravidlo $\acute{S} \rightarrow S$.

Keďže v tejto gramatike **platí**, že S sa nachádza na pravej strane nejakého pravidla, konkrétne až troch:

- $S \rightarrow aSb$
- $B \rightarrow aSBc$
- $B \rightarrow aSc$

Tak úpravu vykonáme. Dostávame teda gramatiku s novým počiatočným neterminálom $\acute{S}$ a pravidlami:

- $\acute{S} \rightarrow S$
- $S \rightarrow aSb \mid A$
- $A \rightarrow cB \mid cc \mid c$
- $B \rightarrow aSBc \mid aSc$

Odstránenie jednoduchých pravidiel

V gramatike sú 2 jednoduché pravidlá: $\acute{S} \to S$ a $S \to A$. Musíme ich odstrániť. V tejto úprave: $N_{\acute{S}} = \{\acute{S}, S, A\}, N_S = \{S, A\}, N_A = \{A\}, N_B = \{B\}$.

Upravená gramatika bez jednoduchých pravidiel:

- $\acute{S} \to aSb \mid cB \mid cc \mid c$
- $S \to aSb \mid cB \mid cc \mid c$
- $A \to cB \mid cc \mid c$
- $B \to aSBc \mid aSc$

Odstránenie nadbytočných symbolov

Následne z gramatiky odstránime nadbytočné symboly. Najprv nadbytočné neterminály - pre ne $N_T = \{\acute{S}, S, A, C\}$, čiže v tomto kroku by sme neodstránili nič.

Dostupné symboly: $V_D = \{\acute{S}, a, S, b, c, B\}$. Zistili sme, že neterminál A nie je dostupný symbol, môžeme ho odstrániť. Dostávame gramatiku:

- $\acute{S} \rightarrow aSb \mid cB \mid cc \mid c$
- $S \rightarrow aSb \mid cB \mid cc \mid c$
- $B \rightarrow aSBc \mid aSc$

Skracovanie pravých strán

V tomto kroku generujeme ekvivalentnú gramatiku, v ktorej majú pravé strany dĺžku najviac 2 symboly. Pre tieto účely vytvárame nové pravidlá a neterminály pre každé pravidlo dĺžky viac ako 2. Napríklad pravidlo:

- $\acute{S} \rightarrow aSb$

Nahradíme dvomi pravidlami a jedným novým neterminálom:

1. $\acute{S} \rightarrow aA_1$
2. $A_1 \rightarrow Sb$

Teda:

- $\acute{S} \rightarrow aA_1 \mid cB \mid cc \mid c$
- $S \rightarrow aSb \mid cB \mid cc \mid c$
- $B \rightarrow aSBc \mid aSc$
- $A_1 \rightarrow Sb$

Skracovanie pravých strán

Pravidlo:

- $S \to aSb$

má rovnakú pravú stranu ako pred chvíľou spracované $\acute{S} \to aSb$, teda môžeme prepoužiť neterminál A_1:

- $\acute{S} \to aA_1 \mid cB \mid cc \mid c$
- $S \to aA_1 \mid cB \mid cc \mid c$
- $B \to aSBc \mid aSc$
- $A_1 \to Sb$

Skracovanie pravých strán

Pravidlo:

- $B \to aSBc$

rozpíšeme pomocou 2 nových neterminálov A_2, A_3 do pravidiel: $B \to aA_2$, $A_2 \to SA_3$, $A_3 \to Bc$:

- $\acute{S} \to aA_1 \mid cB \mid cc \mid c$
- $S \to aA_1 \mid cB \mid cc \mid c$
- $B \to aA_2 \mid aSc$
- $A_1 \to Sb$
- $A_2 \to SA_3$
- $A_3 \to Bc$

Skracovanie pravých strán

Pravidlo:

- $B \to aSc$

rozpíšeme pomocou nového neterminálu A_4 do pravidiel: $B \to aA_4$, $A_4 \to Sc$:

- $\acute{S} \to aA_1 \mid cB \mid cc \mid c$
- $S \to aA_1 \mid cB \mid cc \mid c$
- $B \to aA_2 \mid aA_4$
- $A_1 \to Sb$
- $A_2 \to SA_3$
- $A_3 \to Bc$
- $A_4 \to Sc$

Úprava pravých strán na $N \to NN$, $N \to T$

Dostávame teda gramatiku, v ktorej **každá pravá strana** je dĺžky max. 2.

- $\acute{S} \to aA_1 \mid cB \mid cc \mid c$
- $S \to aA_1 \mid cB \mid cc \mid c$
- $B \to aA_2 \mid aA_4$
- $A_1 \to Sb$
- $A_2 \to SA_3$
- $A_3 \to Bc$
- $A_4 \to Sc$

V poslednom kroku pravidlá upravíme, aby boli na pravej strane vždy 2 neterminály alebo jeden terminál.

- Každý **terminál** gramatiky (napr. a_1), ktorý sa vyskytuje v **niektorej pravej strane dĺžky 2**, nahradíme novým neterminálom (napr. V_1) v týchto pravidlách dĺžky 2.
- Zároveň do gramatiky pridáme pravidlo v tvare $V_1 \rightarrow a_1$.

V gramatike:

- $\acute{S} \rightarrow aA_1 \mid cB \mid cc \mid c$
- $S \rightarrow aA_1 \mid cB \mid cc \mid c$
- $B \rightarrow aA_2 \mid aA_4$
- $A_1 \rightarrow Sb$
- $A_2 \rightarrow SA_3$
- $A_3 \rightarrow Bc$
- $A_4 \rightarrow Sc$

máme terminál a v niekoľkých pravých stranách dĺžky 2. Vo všetkých týchto pravidlách ho nahradíme neterminálom V_1 a do gramatiky doplníme pravidlo $V_1 \rightarrow a$.

Dostávame:

- $\acute{S} \rightarrow V_1 A_1 \mid cB \mid cc \mid c$
- $S \rightarrow V_1 A_1 \mid cB \mid cc \mid c$
- $B \rightarrow V_1 A_2 \mid V_1 A_4$
- $A_1 \rightarrow Sb$
- $A_2 \rightarrow SA_3$
- $A_3 \rightarrow Bc$
- $A_4 \rightarrow Sc$
- $V_1 \rightarrow a$

Ďalší terminál b máme tiež v niekoľkých pravých stranách dĺžky 2. Vo všetkých týchto pravidlách ho nahradíme neterminálom V_2 a do gramatiky doplníme pravidlo $V_2 \rightarrow b$.

Dostávame:

- $\acute{S} \to V_1 A_1 \mid cB \mid cc \mid c$
- $S \to V_1 A_1 \mid cB \mid cc \mid c$
- $B \to V_1 A_2 \mid V_1 A_4$
- $A_1 \to SV_2$
- $A_2 \to SA_3$
- $A_3 \to Bc$
- $A_4 \to Sc$
- $V_1 \to a$
- $V_2 \to b$

Terminál c máme tiež v niekoľkých pravých stranách dĺžky 2. Vo všetkých týchto pravidlách ho nahradíme neterminálom V_3 a do gramatiky doplníme pravidlo $V_3 \to c$.

Dostávame:

- $\acute{S} \rightarrow V_1 A_1 \mid V_3 B \mid V_3 V_3 \mid c$
- $S \rightarrow V_1 A_1 \mid V_3 B \mid V_3 V_3 \mid c$
- $B \rightarrow V_1 A_2 \mid V_1 A_4$
- $A_1 \rightarrow SV_2$
- $A_2 \rightarrow SA_3$
- $A_3 \rightarrow BV_3$
- $A_4 \rightarrow SV_3$
- $V_1 \rightarrow a$
- $V_2 \rightarrow b$
- $V_3 \rightarrow c$

Čím sme dostali požadovaný výsledok - verziu pôvodnej gramatiky v Chomského normálnom tvare.

UPOZORNENIA

1. **Prvé varovanie!** Terminály nahrádzame neterminálmi **len v pravidlách dĺžky 2.** Všimnite si, že aj po nahradení c neterminálom V_3 nám zostali pravidlá ako: $S \to c$. Pretože ak by sme v tomto pravidle c nahradili V_3, dostali by sme **jednoduché pravidlo** $S \to V_3$, **čo v Chomského normálnom tvare nesmie byť!**
2. Pri nahrádzaní terminálov neterminálmi **musíme vždy vyrobiť nový neterminál.** Napr. pre c sme vyrobili nový neterminál V_3, hoci v gramatike už existoval iný neterminál $S \to c$. Avšak **nesmieme** prepísať všetky výskyty c neterminálom S, lebo by to **zmenilo jazyk** - viď ďalší slajd!

TOTO JE ZLE!!!

TOTO JE ZLE!!!

- $\acute{S} \rightarrow V_1A_1 \mid SB \mid SS \mid c$
- $S \rightarrow V_1A_1 \mid SB \mid SS \mid c$
- $B \rightarrow V_1A_2 \mid V_1A_4$
- $A_1 \rightarrow SV_2$
- $A_2 \rightarrow SA_3$
- $A_3 \rightarrow BS$
- $A_4 \rightarrow SS$
- $V_1 \rightarrow a$
- $V_2 \rightarrow b$

TOTO JE ZLE!!!

Konverzia na CHNT č. 2

Príklad: Je daná gramatika $G = (\{S, A, B, C\}, \{a, b, c\}, P, S)$, kde pravidlá P:

$S \to ABC \mid aSb$

$A \to C \mid aAb \mid c$

$B \to BabB \mid AA$

$C \to \varepsilon \mid baCab$

Nájdite k nej ekvivalentnú gramatiku v Chomského normálnom tvare!

Odstránenie ε-pravidiel

Pôvodná gramatika:

$$S \to ABC \mid aSb$$
$$A \to C \mid aAb \mid c$$
$$B \to BabB \mid AA$$
$$C \to \varepsilon \mid baCab$$

Množina N_ε:

$$N_\varepsilon = \{C, A, B, S\}$$

Nová gramatika (pridáme $\acute{S}$ pretože $\varepsilon \in L(G)$).

$$\acute{S} \to \varepsilon \mid S$$
$$S \to ABC \mid aSb \mid AB \mid AC \mid BC \mid A \mid B \mid C \mid ab$$
$$A \to C \mid aAb \mid c \mid ab$$
$$B \to BabB \mid AA \mid ab \mid Bab \mid abB \mid A$$
$$C \to baCab \mid baab$$

Počiatočný neterminál

V gramatike, ktorú sme dostali po odstránení ε-pravidiel:

$$\acute{S} \to \varepsilon \mid S$$
$$S \to ABC \mid aSb \mid AB \mid AC \mid BC \mid A \mid B \mid C \mid ab$$
$$A \to C \mid aAb \mid c \mid ab$$
$$B \to BabB \mid AA \mid ab \mid Bab \mid abB \mid A$$
$$C \to baCab \mid baab$$

určite platí, že jej počiatočný neterminál $\acute{S}$ na **nenachádza** na pravej strane žiadneho pravidla, čiže v tomto kroku nemusíme pridávať nový počiatočný neterminál.

Odstránenie jednoduchých pravidiel

Odstráňme jednoduché pravidlá z gramatiky:

$$\acute{S} \rightarrow \varepsilon \mid S$$
$$S \rightarrow ABC \mid aSb \mid AB \mid AC \mid BC \mid A \mid B \mid C \mid ab$$
$$A \rightarrow C \mid aAb \mid c \mid ab$$
$$B \rightarrow BabB \mid AA \mid ab \mid Bab \mid abB \mid A$$
$$C \rightarrow baCab \mid baab$$

Množiny N_X pre všetky neterminály X:

$$N_{\acute{S}} = \{\acute{S}, S, A, B, C\}$$
$$N_S = \{S, A, B, C\}$$
$$N_A = \{A, C\}$$
$$N_B = \{B, A, C\}$$
$$N_C = \{C\}$$

Odstránenie jednoduchých pravidiel

Po odstránení jednoduchých pravidiel:

$$
\begin{aligned}
\acute{S} &\to \varepsilon \mid ABC \mid aSb \mid AB \mid AC \mid BC \mid ab \\
\acute{S} &\to BabB \mid AA \mid Bab \mid abB \\
\acute{S} &\to aAb \mid c \mid baCab \mid baab \\
S &\to ABC \mid aSb \mid AB \mid AC \mid BC \mid ab \\
S &\to BabB \mid AA \mid Bab \mid abB \\
S &\to aAb \mid c \mid baCab \mid baab \\
A &\to aAb \mid c \mid ab \mid baCab \mid baab \\
B &\to BabB \mid AA \mid ab \mid Bab \mid abB \\
B &\to aAb \mid c \mid baCab \mid baab \\
C &\to baCab \mid baab
\end{aligned}
$$

Táto gramatika je redukovaná, neobsahuje nadbytočné ani nedostupné symboly.

Pred skracovaním pravidiel

$\acute{S} \rightarrow \varepsilon \mid ABC \mid aSb \mid AB \mid AC \mid BC \mid ab$

$\acute{S} \rightarrow BabB \mid AA \mid Bab \mid abB$

$\acute{S} \rightarrow aAb \mid c \mid baCab \mid baab$

$S \rightarrow ABC \mid aSb \mid AB \mid AC \mid BC \mid ab$

$S \rightarrow BabB \mid AA \mid Bab \mid abB$

$S \rightarrow aAb \mid c \mid baCab \mid baab$

$A \rightarrow aAb \mid c \mid ab \mid baCab \mid baab$

$B \rightarrow BabB \mid AA \mid ab \mid Bab \mid abB$

$B \rightarrow aAb \mid c \mid baCab \mid baab$

$C \rightarrow baCab \mid baab$

Skracovanie pravých strán

$\acute{S} \to \varepsilon \mid ABC \mid aSb \mid AB \mid AC \mid BC \mid ab$

$\acute{S} \to BabB \mid AA \mid Bab \mid abB$

$\acute{S} \to aAb \mid c \mid baCab \mid baab$

$S \to ABC \mid aSb \mid AB \mid AC \mid BC \mid ab$

$S \to BabB \mid AA \mid Bab \mid abB$

$S \to aAb \mid c \mid baCab \mid baab$

$A \to aAb \mid c \mid ab \mid baCab \mid baab$

$B \to BabB \mid AA \mid ab \mid Bab \mid abB$

$B \to aAb \mid c \mid baCab \mid baab$

$C \to baCab \mid baab$

Pravidlo $\acute{S} \to ABC$ rozdelíme do 2: $\acute{S} \to AA_1$ a $A_1 \to BC$, A_1 je nový neterminál.
Využijeme to aj pri skracovaní pravidla $S \to ABC$

Skracovanie $\acute{S} \rightarrow ABC, S \rightarrow ABC$

$\acute{S} \rightarrow \varepsilon \mid AA_1 \mid aSb \mid AB \mid AC \mid BC \mid ab$

$\acute{S} \rightarrow BabB \mid AA \mid Bab \mid abB$

$\acute{S} \rightarrow aAb \mid c \mid baCab \mid baab$

$S \rightarrow AA_1 \mid aSb \mid AB \mid AC \mid BC \mid ab$

$S \rightarrow BabB \mid AA \mid Bab \mid abB$

$S \rightarrow aAb \mid c \mid baCab \mid baab$

$A \rightarrow aAb \mid c \mid ab \mid baCab \mid baab$

$B \rightarrow BabB \mid AA \mid ab \mid Bab \mid abB$

$B \rightarrow aAb \mid c \mid baCab \mid baab$

$C \rightarrow baCab \mid baab$

$A_1 \rightarrow BC$

Skracovanie $\acute{S} \to aSb$, $S \to aSb$

$\acute{S} \to \varepsilon \mid AA_1 \mid aA_2 \mid AB \mid AC \mid BC \mid ab$

$\acute{S} \to BabB \mid AA \mid Bab \mid abB$

$\acute{S} \to aAb \mid c \mid baCab \mid baab$

$S \to AA_1 \mid aA_2 \mid AB \mid AC \mid BC \mid ab$

$S \to BabB \mid AA \mid Bab \mid abB$

$S \to aAb \mid c \mid baCab \mid baab$

$A \to aAb \mid c \mid ab \mid baCab \mid baab$

$B \to BabB \mid AA \mid ab \mid Bab \mid abB$

$B \to aAb \mid c \mid baCab \mid baab$

$C \to baCab \mid baab$

$A_1 \to BC$

$A_2 \to Sb$

$\acute{S} \to BabB, S \to BabB, B \to BabB$

$$\acute{S} \to \varepsilon \mid AA_1 \mid aA_2 \mid AB \mid AC \mid BC \mid ab$$
$$\acute{S} \to BA_3 \mid AA \mid Bab \mid abB$$
$$\acute{S} \to aAb \mid c \mid baCab \mid baab$$
$$S \to AA_1 \mid aA_2 \mid AB \mid AC \mid BC \mid ab$$
$$S \to BA_3 \mid AA \mid Bab \mid abB$$
$$S \to aAb \mid c \mid baCab \mid baab$$
$$A \to aAb \mid c \mid ab \mid baCab \mid baab$$
$$B \to BA_3 \mid AA \mid ab \mid Bab \mid abB$$
$$B \to aAb \mid c \mid baCab \mid baab$$
$$C \to baCab \mid baab$$
$$A_1 \to BC \qquad A_3 \to aA_4$$
$$A_2 \to Sb \qquad A_4 \to bB$$

$\acute{S} \to Bab, S \to Bab, B \to Bab$

$\acute{S} \to \varepsilon \mid AA_1 \mid aA_2 \mid AB \mid AC \mid BC \mid ab$

$\acute{S} \to BA_3 \mid AA \mid BA_5 \mid abB$

$\acute{S} \to aAb \mid c \mid baCab \mid baab$

$S \to AA_1 \mid aA_2 \mid AB \mid AC \mid BC \mid ab$

$S \to BA_3 \mid AA \mid BA_5 \mid abB$

$S \to aAb \mid c \mid baCab \mid baab$

$A \to aAb \mid c \mid ab \mid baCab \mid baab$

$B \to BA_3 \mid AA \mid ab \mid BA_5 \mid abB$

$B \to aAb \mid c \mid baCab \mid baab$

$C \to baCab \mid baab$

$A_1 \to BC \qquad A_3 \to aA_4 \qquad A_5 \to ab$

$A_2 \to Sb \qquad A_4 \to bB$

$\acute{S} \to abB, S \to abB, B \to abB$

$\acute{S} \to \varepsilon \mid AA_1 \mid aA_2 \mid AB \mid AC \mid BC \mid ab$

$\acute{S} \to BA_3 \mid AA \mid BA_5 \mid aA_4$

$\acute{S} \to aAb \mid c \mid baCab \mid baab$

$S \to AA_1 \mid aA_2 \mid AB \mid AC \mid BC \mid ab$

$S \to BA_3 \mid AA \mid BA_5 \mid aA_4$

$S \to aAb \mid c \mid baCab \mid baab$

$A \to aAb \mid c \mid ab \mid baCab \mid baab$

$B \to BA_3 \mid AA \mid ab \mid BA_5 \mid aA_4$

$B \to aAb \mid c \mid baCab \mid baab$

$C \to baCab \mid baab$

$A_1 \to BC \qquad A_3 \to aA_4 \qquad A_5 \to ab$

$A_2 \to Sb \qquad A_4 \to bB$

$\acute{S} \to aAb, S \to aAb, A \to aAb, B \to aAb$

$\acute{S} \to \varepsilon \mid AA_1 \mid aA_2 \mid AB \mid AC \mid BC \mid ab$

$\acute{S} \to BA_3 \mid AA \mid BA_5 \mid aA_4$

$\acute{S} \to aA_6 \mid c \mid baCab \mid baab$

$S \to AA_1 \mid aA_2 \mid AB \mid AC \mid BC \mid ab$

$S \to BA_3 \mid AA \mid BA_5 \mid aA_4$

$S \to aA_6 \mid c \mid baCab \mid baab$

$A \to aA_6 \mid c \mid ab \mid baCab \mid baab$

$B \to BA_3 \mid AA \mid ab \mid BA_5 \mid aA_4$

$B \to aA_6 \mid c \mid baCab \mid baab$

$C \to baCab \mid baab$

$A_1 \to BC \qquad A_3 \to aA_4 \qquad A_5 \to ab$

$A_2 \to Sb \qquad A_4 \to bB \qquad A_6 \to Ab$

$\acute{S} \to baCab, S \to baCab, A \to baCab, B \to baCab, C \to baCab$

$$\acute{S} \to \varepsilon \mid AA_1 \mid aA_2 \mid AB \mid AC \mid BC \mid ab$$

$$\acute{S} \to BA_3 \mid AA \mid BA_5 \mid aA_4 \mid aA_6 \mid c \mid bA_7 \mid baab$$

$$S \to AA_1 \mid aA_2 \mid AB \mid AC \mid BC \mid ab$$

$$S \to BA_3 \mid AA \mid BA_5 \mid aA_4 \mid aA_6 \mid c \mid bA_7 \mid baab$$

$$A \to aA_6 \mid c \mid ab \mid bA_7 \mid baab$$

$$B \to BA_3 \mid AA \mid ab \mid BA_5 \mid aA_4$$

$$B \to aA_6 \mid c \mid bA_7 \mid baab$$

$$C \to bA_7 \mid baab$$

$$A_1 \to BC \qquad A_3 \to aA_4 \qquad A_5 \to ab$$

$$A_2 \to Sb \qquad A_4 \to bB \qquad A_6 \to Ab$$

$$A_7 \to aA_8 \qquad A_8 \to CA_5$$

$\acute{S} \to baab, S \to baab, A \to baab, B \to baab, C \to baab$

$$\acute{S} \to \varepsilon \mid AA_1 \mid aA_2 \mid AB \mid AC \mid BC \mid ab$$
$$\acute{S} \to BA_3 \mid AA \mid BA_5 \mid aA_4 \mid aA_6 \mid c \mid bA_7 \mid bA_9$$
$$S \to AA_1 \mid aA_2 \mid AB \mid AC \mid BC \mid ab$$
$$S \to BA_3 \mid AA \mid BA_5 \mid aA_4 \mid aA_6 \mid c \mid bA_7 \mid bA_9$$
$$A \to aA_6 \mid c \mid ab \mid bA_7 \mid bA_9$$
$$B \to BA_3 \mid AA \mid ab \mid BA_5 \mid aA_4$$
$$B \to aA_6 \mid c \mid bA_7 \mid bA_9$$
$$C \to bA_7 \mid bA_9$$
$$A_1 \to BC \qquad A_3 \to aA_4 \qquad A_5 \to ab$$
$$A_2 \to Sb \qquad A_4 \to bB \qquad A_6 \to Ab$$
$$A_7 \to aA_8 \qquad A_8 \to CA_5 \qquad A_9 \to aA_5$$

Gramatika po skracovaní pravých strán

$$\acute{S} \to \varepsilon \mid AA_1 \mid aA_2 \mid AB \mid AC \mid BC \mid ab$$
$$\acute{S} \to BA_3 \mid AA \mid BA_5 \mid aA_4 \mid aA_6 \mid c \mid bA_7 \mid bA_9$$
$$S \to AA_1 \mid aA_2 \mid AB \mid AC \mid BC \mid ab$$
$$S \to BA_3 \mid AA \mid BA_5 \mid aA_4 \mid aA_6 \mid c \mid bA_7 \mid bA_9$$
$$A \to aA_6 \mid c \mid ab \mid bA_7 \mid bA_9$$
$$B \to BA_3 \mid AA \mid ab \mid BA_5 \mid aA_4$$
$$B \to aA_6 \mid c \mid bA_7 \mid bA_9$$
$$C \to bA_7 \mid bA_9$$

$$A_1 \to BC \qquad A_3 \to aA_4 \qquad A_5 \to ab$$
$$A_2 \to Sb \qquad A_4 \to bB \qquad A_6 \to Ab$$
$$A_7 \to aA_8 \qquad A_8 \to CA_5 \qquad A_9 \to aA_5$$

Úprava do tvaru $N \rightarrow NN$, $N \rightarrow T$

$\acute{S} \rightarrow \varepsilon \mid AA_1 \mid V_1 A_2 \mid AB \mid AC \mid BC \mid V_1 V_2$

$\acute{S} \rightarrow BA_3 \mid AA \mid BA_5 \mid V_1 A_4 \mid V_1 A_6 \mid c \mid V_2 A_7 \mid V_2 A_9$

$S \rightarrow AA_1 \mid V_1 A_2 \mid AB \mid AC \mid BC \mid V_1 V_2$

$S \rightarrow BA_3 \mid AA \mid BA_5 \mid V_1 A_4 \mid V_1 A_6 \mid c \mid V_2 A_7 \mid V_2 A_9$

$A \rightarrow V_1 A_6 \mid c \mid V_1 V_2 \mid V_2 A_7 \mid V_2 A_9$

$B \rightarrow BA_3 \mid AA \mid V_1 V_2 \mid BA_5 \mid V_1 A_4 \mid V_1 A_6 \mid c \mid V_2 A_7 \mid V_2 A_9$

$C \rightarrow V_2 A_7 \mid V_2 A_9$

$A_1 \rightarrow BC \qquad A_3 \rightarrow V_1 A_4 \qquad A_5 \rightarrow V_1 V_2$

$A_2 \rightarrow SV_2 \qquad A_4 \rightarrow V_2 B \qquad A_6 \rightarrow AV_2$

$A_7 \rightarrow V_1 A_8 \qquad A_8 \rightarrow CA_5 \qquad A_9 \rightarrow V_1 A_5$

$V_1 \rightarrow a \qquad V_2 \rightarrow b$

Výsledný CHNT

$$\acute{S} \to \varepsilon \mid AA_1 \mid V_1A_2 \mid AB \mid AC \mid BC \mid V_1V_2$$

$$\acute{S} \to BA_3 \mid AA \mid BA_5 \mid V_1A_4 \mid V_1A_6 \mid c \mid V_2A_7 \mid V_2A_9$$

$$S \to AA_1 \mid V_1A_2 \mid AB \mid AC \mid BC \mid V_1V_2$$

$$S \to BA_3 \mid AA \mid BA_5 \mid V_1A_4 \mid V_1A_6 \mid c \mid V_2A_7 \mid V_2A_9$$

$$A \to V_1A_6 \mid c \mid V_1V_2 \mid V_2A_7 \mid V_2A_9$$

$$B \to BA_3 \mid AA \mid V_1V_2 \mid BA_5 \mid V_1A_4 \mid V_1A_6 \mid c \mid V_2A_7 \mid V_2A_9$$

$$C \to V_2A_7 \mid V_2A_9$$

$$A_1 \to BC \qquad A_3 \to V_1A_4 \qquad A_5 \to V_1V_2$$

$$A_2 \to SV_2 \qquad A_4 \to V_2B \qquad A_6 \to AV_2$$

$$A_7 \to V_1A_8 \qquad A_8 \to CA_5 \qquad A_9 \to V_1A_5$$

$$V_1 \to a \qquad V_2 \to b$$

CYK algoritmus č. 1

Je daná gramatika:

$$S \to CB$$
$$B \to CD \mid EC$$
$$C \to DD \mid EE \mid ED$$
$$D \to a$$
$$E \to b$$

Zistite pomocou CYK algoritmu, či slovo $aabba \in L(G)$.

1. $N_{1,1} = \{D\}$
2. $N_{2,2} = \{D\}$
3. $N_{3,3} = \{E\}$
4. $N_{4,4} = \{E\}$
5. $N_{5,5} = \{D\}$

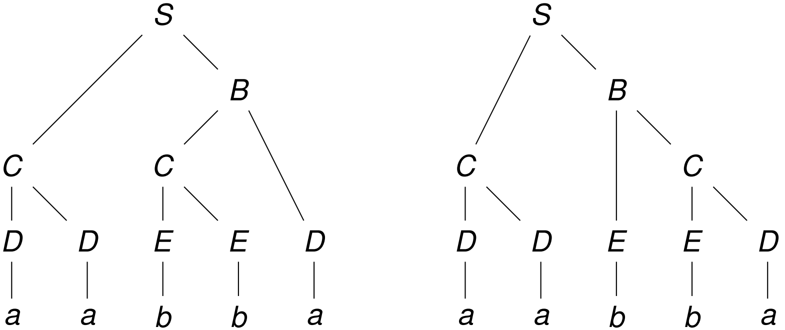
1. $N_{1,2} = \{C\}$, lebo v pravidlách $C \rightarrow DD$
2. $N_{2,3} = \emptyset$, lebo neexistuje pravidlo s pravou stranou DE
3. $N_{3,4} = \{C\}$, lebo v pravidlách $C \rightarrow EE$
4. $N_{4,5} = \{C\}$, lebo v pravidlách $C \rightarrow ED$

1. $N_{1,3} = \emptyset$
2. $N_{2,4} = \emptyset$,
3. $N_{3,5} = \{B\}$, lebo v pravidlách $B \to CD$ a taktiež $B \to EC$
   V prvom prípade $C \in N_{3,4}$ a $D \in N_{5,5}$
   V druhom prípade $E \in N_{3,3}$ a $C \in N_{4,5}$

1. $N_{1,4} = \emptyset$
2. $N_{2,5} = \emptyset$

1. $N_{1,5} = \{S\}$
   Keďže $S \in N_{1,5}$, t.j. pre celé slovo, tak *aabba* $\in L(G)$.

Na základe príslušných množín vieme teoreticky aj nájsť derivačný strom, resp. odvodenie:



T.j. táto gramatika **nie je jednoznačná.**

CYK algoritmus č. 2

Je daná gramatika $G = (\{S, A, B\}, \{a, b, c\}, P, S)$ ktorej pravidlá P:

- $S \to aSb \mid A$
- $A \to cB \mid cc$
- $B \to aSBc \mid \varepsilon$

Zistite pomocou CYK algoritmu, či majú v gramatike odvodenie reťazce *acaccb* a *acacb*.

Na to, aby sme mohli použiť CYK algoritmus, musíme previesť gramatiku do Chomského normálneho tvaru. Výsledok:

- $\acute{S} \rightarrow V_1 A_1 \mid V_3 B \mid V_3 V_3 \mid c$
- $S \rightarrow V_1 A_1 \mid V_3 B \mid V_3 V_3 \mid c$
- $B \rightarrow V_1 A_2 \mid V_1 A_4$
- $A_1 \rightarrow S V_2$
- $A_2 \rightarrow S A_3$
- $A_3 \rightarrow B V_3$
- $A_4 \rightarrow S V_3$
- $V_1 \rightarrow a$
- $V_2 \rightarrow b$
- $V_3 \rightarrow c$

acaccb

Priamo z pravidiel pre jednotlivé symboly:

- $N_{1,1} = \{V_1\}$
- $N_{2,2} = \{\acute{S}, S, V_3\}$
- $N_{3,3} = \{V_1\}$.
- $N_{4,4} = \{\acute{S}, S, V_3\}$
- $N_{5,5} = \{\acute{S}, S, V_3\}$
- $N_{6,6} = \{V_2\}$.

Podreťazce dĺžky 2:

- $N_{1,2} = \emptyset$
- $N_{2,3} = \emptyset$
- $N_{3,4} = \emptyset$
- $N_{4,5} = \{A_4, \acute{S}, S\}$
  - A_4 tam patrí pretože v pravidlách $A_4 \to SV_3$ a $S \in N_{4,4}$, $V_3 \in N_{5,5}$
  - $\acute{S}, S$ tam patria pretože v pravidlách $\acute{S}(S) \to V_3 V_3$ a $V_3 \in N_{4,4}$, $V_3 \in N_{5,5}$
- $N_{5,6} = \{A_1\}$, pretože v pravidlách $A_1 \to SV_2$ a $S \in N_{5,5}$, $V_2 \in N_{6,6}$

Podreťazce dĺžky 3:

- $N_{1,3} = \emptyset$
- $N_{2,4} = \emptyset$
- $N_{3,5} = \{B\}$, pretože v pravidlách $B \to V_1 A_4$ a $V_1 \in N_{3,3}, A_4 \in N_{4,5}$.
- $N_{4,6} = \{A_1\}$, pretože v pravidlách $A_1 \to SV_2$ a $S \in N_{4,5}, V_2 \in N_{6,6}$.

Podreťazce dĺžky 4:

- $N_{1,4} = \emptyset$
- $N_{2,5} = \{\acute{S}, S\}$, pretože v pravidlách $\acute{S} \to V_3 B$ a $V_3 \in N_{2,2}, B \in N_{3,5}$ (to isté $S \to V_3 B$)
- $N_{3,6} = \{\acute{S}, S\}$, pretože v pravidlách $\acute{S} \to V_1 A_1$ a $V_1 \in N_{3,3}, A_1 \in N_{4,6}$ (to isté $S \to V_1 A_1$)

acaccb

Podreťazce dĺžky 5:

- $N_{1,5} = \emptyset$
- $N_{2,6} = \{A_1\}$, pretože v pravidlách $A_1 \to SV_2$ a $S \in N_{2,5}$, $V_2 \in N_{6,6}$.

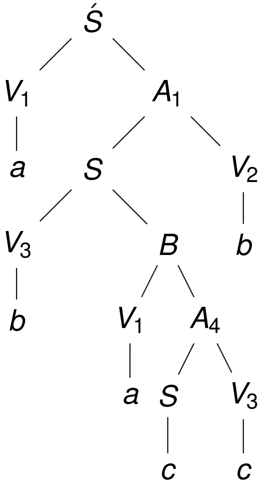
Podreťazce dĺžky 6:

- $N_{1,6} = \{\acute{S}, S\}$, pretože v pravidlách $\acute{S} \to V_1A_1$ a $V_1 \in N_{1,1}, A_1 \in N_{2,6}$ (to isté $S \to V_1A_1$)

Po zostrojení $N_{1,6}$ sa stačí pozrieť, či sa v tejto množine nachádza počiatočný neterminál $\acute{S}$:

- Ak áno, $\acute{S} \in N_{1,6}$, tak potom *acaccb* má v gramatike G deriváciu - **a to je tento prípad!**
- V prípade, že $\acute{S} \notin N_{1,6}$, tak *acaccb* by v gramatike G nemalo deriváciu. Ale to nie je tento prípad.

Pomocou množín $N_{i,j}$ a toho, ako vznikli, vieme odvodiť aj derivačný strom reťazca *acaccb* v Chomského normálnom tvare gramatiky G:



acacb

Priamo z pravidiel pre jednotlivé symboly:

- $N_{1,1} = \{V_1\}$
- $N_{2,2} = \{\acute{S}, S, V_3\}$
- $N_{3,3} = \{V_1\}$.
- $N_{4,4} = \{\acute{S}, S, V_3\}$
- $N_{5,5} = \{V_2\}$.

Podreťazce dĺžky 2:

- $N_{1,2} = \emptyset$
- $N_{2,3} = \emptyset$
- $N_{3,4} = \emptyset$
- $N_{4,5} = \{A_1\}$, pretože v pravidlách $A_1 \rightarrow SV_2$ a $S \in N_{4,4}$, $V_2 \in N_{5,5}$

Podreťazce dĺžky 3:

- $N_{1,3} = \emptyset$
- $N_{2,4} = \emptyset$
- $N_{3,5} = \{\acute{S}, S\}$, pretože v pravidlách $\acute{S} \rightarrow V_1 A_1$ a $V_1 \in N_{3,3}, A_1 \in N_{4,5}$ (to isté $S \rightarrow V_1 A_1$)

Podreťazce dĺžky 4:

- $N_{1,4} = \emptyset$
- $N_{2,5} = \emptyset$

acacb

Podreťazce dĺžky 5:

- $N_{1,5} = \emptyset$

Po zostrojení $N_{1,5}$ sa stačí pozrieť, či sa v tejto množine nachádza počiatočný neterminál $\acute{S}$:

- Vidíme, že množina $N_{1,5}$ je prázdna, teda počiatočný neterminál $\acute{S}$ sa v nej nenachádza. Z toho vyplýva, že reťazec *acacb* **nemá** v gramatike G odvodenie.

CYK algoritmus č. 3

Je daná gramatika $G = (\{S, A, B\}, \{a, b\}, P, S)$ ktorej pravidlá P:

- $S \rightarrow abA \mid BbA \mid BA$
- $A \rightarrow aAa \mid bAb \mid \varepsilon$
- $B \rightarrow a \mid S$

Zistite pomocou CYK algoritmu, či majú v gramatike odvodenie reťazce *ababba* a *abba*.

CHNT - ε-pravidlá

Prvým krokom pre použitie gramatiky v CYK je jej úprava do Chomského normálneho tvaru. Najprv odstránime ε-pravidlá ($A \to \varepsilon$):

Len pre info - v tejto gramatike $N_\varepsilon = \{A\}$. Po odstránení ε-pravidiel má gramatika tvar:

- $S \to abA \mid BbA \mid BA \mid ab \mid Bb \mid B$
- $A \to aAa \mid bAb \mid aa \mid bb$
- $B \to a \mid S$

CHNT - počiatočný neterminál

Počiatočný neterminál S je na pravej strane niektorého z pravidiel! Preto doplníme nový počiatočný neterminál $\acute{S}$ a pravidlo $\acute{S} \rightarrow S$.

- $\acute{S} \rightarrow S$
- $S \rightarrow abA \mid BbA \mid BA \mid ab \mid Bb \mid B$
- $A \rightarrow aAa \mid bAb \mid aa \mid bb$
- $B \rightarrow a \mid S$

CHNT - jednoduché pravidlá

- $\acute{S} \rightarrow S$
- $S \rightarrow abA \mid BbA \mid BA \mid ab \mid Bb \mid B$
- $A \rightarrow aAa \mid bAb \mid aa \mid bb$
- $B \rightarrow a \mid S$

Množiny $N_{\acute{S}} = \{\acute{S}, S, B\}$, $N_S = \{S, B\}$, $N_A = \{A\}$, $N_B = \{B, S\}$. Po odstránení jednoduchých pravidiel:

- $\acute{S} \rightarrow abA \mid BbA \mid BA \mid ab \mid Bb \mid a$
- $S \rightarrow abA \mid BbA \mid BA \mid ab \mid Bb \mid a$
- $A \rightarrow aAa \mid bAb \mid aa \mid bb$
- $B \rightarrow a \mid abA \mid BbA \mid BA \mid ab \mid Bb$

CHNT - redukovaná gramatika

Vo vzniknutej gramatike je nedostupný neterminál S. Preto jeho odstránením dostaneme redukovanú gramatiku:

- $\acute{S} \rightarrow abA \mid BbA \mid BA \mid ab \mid Bb \mid a$
- $A \rightarrow aAa \mid bAb \mid aa \mid bb$
- $B \rightarrow a \mid abA \mid BbA \mid BA \mid ab \mid Bb$

CHNT - skracovanie pravidiel

Upravíme gramatiku, aby na pravej strane boli reťazce dĺžky najviac 2:

- $\acute{S} \rightarrow aA_1 \mid BA_1 \mid BA \mid ab \mid Bb \mid a$
- $A \rightarrow aA_2 \mid bA_3 \mid aa \mid bb$
- $B \rightarrow a \mid aA_1 \mid BA_1 \mid BA \mid ab \mid Bb$
- $A_1 \rightarrow bA$
- $A_2 \rightarrow Aa$
- $A_3 \rightarrow Ab$

CHNT - úprava pravých strán na tvar *NN* alebo *T*

Upravíme gramatiku, aby na pravej strane boli alebo 2 neterminály, alebo 1 terminál:

- $\acute{S} \rightarrow V_1A_1 \mid BA_1 \mid BA \mid V_1V_2 \mid BV_2 \mid a$
- $A \rightarrow V_1A_2 \mid V_2A_3 \mid V_1V_1 \mid V_2V_2$
- $B \rightarrow a \mid V_1A_1 \mid BA_1 \mid BA \mid V_1V_2 \mid BV_2$
- $A_1 \rightarrow V_2A$
- $A_2 \rightarrow AV_1$
- $A_3 \rightarrow AV_2$
- $V_1 \rightarrow a$
- $V_2 \rightarrow b$

Výsledná verzia pôvodnej gramatiky v Chomského normálnom tvare je:

- $\acute{S} \to V_1A_1 \mid BA_1 \mid BA \mid V_1V_2 \mid BV_2 \mid a$
- $A \to V_1A_2 \mid V_2A_3 \mid V_1V_1 \mid V_2V_2$
- $B \to a \mid V_1A_1 \mid BA_1 \mid BA \mid V_1V_2 \mid BV_2$
- $A_1 \to V_2A$
- $A_2 \to AV_1$
- $A_3 \to AV_2$
- $V_1 \to a$
- $V_2 \to b$

ababba

Priamo z pravidiel pre jednotlivé symboly:

- $N_{1,1} = \{\acute{S}, B, V_1\}$
- $N_{2,2} = \{V_2\}$
- $N_{3,3} = \{\acute{S}, B, V_1\}$
- $N_{4,4} = \{V_2\}$
- $N_{5,5} = \{V_2\}$
- $N_{6,6} = \{\acute{S}, B, V_1\}$

Podreťazce dĺžky 2:

- $N_{1,2} = \{\acute{S}, B\}$ kvôli pravidlám $\acute{S} \to BV_2$ alebo $\acute{S} \to V_1 V_2$, resp. $B \to BV_2$ alebo $B \to V_1 V_2$.
- $N_{2,3} = \emptyset$
- $N_{3,4} = \{\acute{S}, B\}$ kvôli pravidlám $\acute{S} \to BV_2$ alebo $\acute{S} \to V_1 V_2$, resp. $B \to BV_2$ alebo $B \to V_1 V_2$.
- $N_{4,5} = \{A\}$ kvôli $A \to V_2 V_2$
- $N_{5,6} = \emptyset$

Podreťazce dĺžky 3:

- $N_{1,3} = \emptyset$
- $N_{2,4} = \emptyset$
- $N_{3,5} = \{\acute{S}, B\}$
  - $\acute{S}, B$ tam patria kvôli pravidlám $\acute{S}(B) \to BA$, $B \in N_{3,3}, A \in N_{4,5}$.
  - $\acute{S}, B$ tam patria aj kvôli pravidlám $\acute{S}(B) \to BV_2$, $B \in N_{3,4}, V_2 \in N_{5,5}$.
- $N_{4,6} = \{A_2\}$ kvôli pravidlo $A_2 \to AV_1$ a $A \in N_{4,5}, V_1 \in N_{6,6}$

Podreťazce dĺžky 4:

- $N_{1,4} = \emptyset$.
- $N_{2,5} = \emptyset$
- $N_{3,6} = \{A\}$ kvôli $A \to V_1 A_2$, $V_1 \in N_{3,3}, A_2 \in N_{4,6}$.

Podreťazce dĺžky 5:

- $N_{1,5} = \emptyset$
- $N_{2,6} = \{A_1\}$ kvôli pravidlo $A_1 \to V_2 A$, $V_2 \in N_{2,2}, A \in N_{3,6}$

Podreťazce dĺžky 6:

- $N_{1,6} = \{\acute{S}, B\}$, dokonca z viacerých dôvodov:
  - Pravidlo $\acute{S}(B) \to V_1 A_1, V_1 \in N_{1,1}, A_1 \in N_{2,6}$
  - Pravidlo $\acute{S}(B) \to BA_1, B \in N_{1,1}, A_1 \in N_{2,6}$
  - Pravidlo $\acute{S}(B) \to BA, B \in N_{1,2}, A \in N_{3,6}$

V každom prípade $\acute{S} \in N_{1,6}$ a teda pôvodná gramatika G **generuje** reťazec *ababba*, teda *ababba* $\in L(G)$.

Reťazec *abba*. Priamo z pravidiel pre jednotlivé symboly:

- $N_{1,1} = \{\acute{S}, B, V_1\}$
- $N_{2,2} = \{V_2\}$
- $N_{3,3} = \{V_2\}$
- $N_{4,4} = \{\acute{S}, B, V_1\}$

Podreťazce dĺžky 2:

- $N_{1,2} = \{\acute{S}, B\}$ kvôli pravidlám $\acute{S} \to BV_2$ alebo $\acute{S} \to V_1 V_2$, resp. $B \to BV_2$ alebo $B \to V_1 V_2$.
- $N_{2,3} = \{A\}$ kvôli $A \to V_2 V_2$, $V_2 \in N_{1,1}$, $V_2 \in N_{2,2}$
- $N_{3,4} = \emptyset$.

abba

Podreťazce dĺžky 3:

- $N_{1,3} = \{\acute{S}, B\}$ kvôli $\acute{S}(B) \to BA$, $B \in N_{1,1}$, $A \in N_{2,3}$. Rovnako aj kvôli $\acute{S}(B) \to BV_2$, $B \in B_{1,2}$, $V_2 \in N_{3,3}$.
- $N_{2,4} = \{A_2\}$ kvôli $A_2 \to AV_1$, $A \in N_{2,3}$, $V_1 \in N_{4,4}$.

Celý reťazec *abba*:

- $N_{1,4} = \{A\}$ kvôli $A \to V_1A_2$, $V_1 \in N_{1,1}$, $A_2 \in N_{2,4}$.

Hoci množina $N_{1,4}$ nie je prázdna, **neobsahuje počiatočný neterminál gramatiky v CHNT** a teda reťazec *abba* **nemá** v gramatike deriváciu, $abba \notin L(G)$.

FIRST príklad č. 1

Nech je daná redukovaná gramatika $G = (\{S, A, B, C\}, \{a, b, c\}, P, S)$ a pravidlá P:

$$S \rightarrow ABC \mid aSb$$
$$A \rightarrow aAb \mid c \mid C$$
$$B \rightarrow BabB \mid AA$$
$$C \rightarrow \varepsilon \mid baCab$$

Aká je množina $FIRST(X)$ pre symboly gramatiky?

FIRST príklad č. 1 - Pravidlá typu $A \to a\alpha$

Z pravidiel, ktorých pravá strana začína terminálom, vieme do množín *FIRST* pridať tieto symboly:

1. Z pravidla $S \to aSb$, ktorého pravá strana začína terminálom a vidíme, že do množiny $FIRST(S)$ patrí a.
2. Z pravidla $A \to aAb$, ktorého pravá strana začína terminálom a vidíme, že do množiny $FIRST(A)$ patrí a.
3. Z pravidla $A \to c$, ktorého pravá strana začína terminálom c vidíme, že do množiny $FIRST(A)$ patrí c.
4. Z pravidla $C \to baCab$, ktorého pravá strana začína terminálom b vidíme, že do množiny $FIRST(C)$ patrí b.

Teda po tomto kroku máme istotu o týchto prvkoch množín *FIRST*:

- $a \in FIRST(S)$
- $a, c \in FIRST(A)$
- $b \in FIRST(C)$

FIRST príklad č. 1 - množina N_ε

Najprv vyšetríme množinu N_ε. Na začiatku $N_\varepsilon = \emptyset$.

1. Ako prvý do N_ε patrí neterminál C, pretože priamo na základe pravidla $C \to \varepsilon$ vidíme, že $C \Rightarrow \varepsilon$, teda určite $N_\varepsilon = \{C\}$
2. Ďalej, keď vieme, že $C \in N_\varepsilon$, tak na základe pravidla $A \to C$ vidíme, že $A \Rightarrow C \Rightarrow \varepsilon$, teda $A \Rightarrow^* \varepsilon$, a teda $N_\varepsilon = \{A, C\}$.
3. Ďalej, keď vieme, že $A, C \in N_\varepsilon$, tak na základe pravidla $B \to AA$ vidíme, že $B \Rightarrow AA \Rightarrow A \Rightarrow \varepsilon$, teda $B \Rightarrow^* \varepsilon$, a teda $N_\varepsilon = \{B, A, C\}$.
4. Ďalej, keď vieme, že $B, A, C \in N_\varepsilon$, tak na základe pravidla $S \to ABC$ vidíme, že $S \Rightarrow ABC \Rightarrow BC \Rightarrow C \Rightarrow \varepsilon$, teda $S \Rightarrow^* \varepsilon$, a teda $N_\varepsilon = \{S, B, A, C\}$.

Výsledná množina $N_\varepsilon = \{S, A, B, C\}$, teda po tomto kroku vieme, že :

- $\varepsilon \in FIRST(S)$
- $\varepsilon \in FIRST(A)$
- $\varepsilon \in FIRST(B)$
- $\varepsilon \in FIRST(C)$

Po týchto prvých 2 krokoch teda vieme, že :

- $FIRST(S) = \{\varepsilon, a\}$.
- $FIRST(A) = \{\varepsilon, a, c\}$.
- $FIRST(B) = \{\varepsilon\}$.
- $FIRST(C) = \{\varepsilon, b\}$.

FIRST príklad č. 1 - Pravidlá typu $A \rightarrow \alpha$

Z pravidiel, ktorých pravá strana nezačína terminálom a je tvaru α vieme povedať, že $\mathit{FIRST}(\alpha) \subseteq \mathit{FIRST}(A)$. Množinu $\mathit{FIRST}(\alpha)$ určujeme na základe aktuálne známych množín $\mathit{FIRST}()$ pre neterminály.

Z pravidla $S \rightarrow ABC$ vieme, že $FIRST(ABC) \subseteq FIRST(S)$, kde $FIRST(ABC)$ určíme na základe aktuálnych znalostí o množinách $FIRST()$:

1. Keďže $a \in FIRST(A)$, tak vieme, že $A \Rightarrow^* a\ldots$ a teda $ABC \Rightarrow^* a\ldots$, čiže určite $a \in FIRST(ABC)$ a teda $a \in FIRST(S)$.
2. Podobne pre $c \in FIRST(A)$ vieme, že $A \Rightarrow^* c\ldots$ a teda $ABC \Rightarrow^* c\ldots$, čiže určite $c \in FIRST(ABC)$ a teda $c \in FIRST(S)$.
3. Keďže $\varepsilon \in FIRST(A)$, tak vieme, že $A \Rightarrow^* \varepsilon$ a teda pre $ABC \Rightarrow^* BC$. Teda keďže $\varepsilon \in FIRST(A)$, bude nás ďalej zaujímať $FIRST(B)$.
4. Keďže $\varepsilon \in FIRST(B)$, tak vieme, že $B \Rightarrow^* \varepsilon$ a teda pre $ABC \Rightarrow^* BC \Rightarrow^* C$ a musíme použiť aj $FIRST(C)$.
5. Keďže $b \in FIRST(C)$, tak vieme, že $C \Rightarrow^* c\ldots$ a teda pre $ABC \Rightarrow^* BC \Rightarrow^* C \Rightarrow^* c\ldots$ a teda $c \in FIRST(S)$.
6. Keďže $\varepsilon \in FIRST(C)$, tak vieme, že $C \Rightarrow^* \varepsilon$ a teda pre $ABC \Rightarrow^* BC \Rightarrow^* C \Rightarrow^* \varepsilon$ a teda $\varepsilon \in FIRST(S)$.

- Zistili sme teda, na základe analýzy pravidla $S \to ABC$ a na základe **aktuálnych** znalostí množín $FIRST(A)$, $FIRST(B)$, $FIRST(C)$, že $FIRST(ABC) = \{\varepsilon, a, b, c\}$. Keďže $FIRST(ABC) \subseteq FIRST(S)$, tak vieme, že $\{\varepsilon, a, b, c\} \subseteq FIRST(S)$.
- Podobne budeme analyzovať ďalšie pravidlá.

Z pravidla $A \to C$ dostávame, že $FIRST(C) \subseteq FIRST(A)$.

Keďže $FIRST(C) = \{\varepsilon, b\}$ a $FIRST(A) = \{\varepsilon, a, c\}$, vidíme, že aj $b \in FIRST(A)$.

Z pravidla $B \to AA$ dostávame, že $FIRST(AA) \subseteq FIRST(B)$.

Keďže aktuálne vieme, že $FIRST(A) = \{\varepsilon, a, b, c\}$ tak dokážeme vypočítať, že $FIRST(AA) = \{a, b, c, \varepsilon\}$ a teda $\{\varepsilon, a, b, c\} \subseteq FIRST(B)$. Keďže doteraz sme o $FIRST(B)$ vedeli len to, že tam patrí ε, zistili sme, že tam patria aj a, b, c.

- Tým pádom sme prešli jedenkrát cez pravidlá tvaru $A \to \alpha$, kde pravá strana α začína neterminálom. Každé pravidlo sme vyšetrili na základe vtedy známych množín *FIRST*.
- Je teda možné, že ak sa medzitým zmenili množiny *FIRST*, na základe ktorých sme niečo počítali, môžeme dostať po zmene iný výsledok, v zmysle taký, čo obsahuje viac symbolov.
- V prípade tejto gramatiky však takýto jav nenastane.

Množina *FIRST* - zhrnutie

$FIRST(X)$:

$FIRST$	S	A	B	C	a	b	c
Terminály:					a	b	c
Pravidlá $A \to a\alpha$	a	a, c		b			
Neterminály $A \in N_\varepsilon$	ε	ε	ε	ε			
1. iterácia:	c, b	b	a, b, c				
2. iterácia:							
Celkom:	a, b, c, ε	a, b, c, ε	a, b, c, ε	b, ε			

FOLLOW príklad č. 1

Nech je daná redukovaná gramatika $G = (\{S, A, B, C\}, \{a, b, c\}, P, S)$ a pravidlá P:

$$S \to ABC \mid aSb$$
$$A \to aAb \mid c \mid C$$
$$B \to BabB \mid AA$$
$$C \to \varepsilon \mid baCab$$

Už vieme, že:

$$FIRST(S) = \{\varepsilon, a, b, c\}$$
$$FIRST(A) = \{\varepsilon, a, b, c\}$$
$$FIRST(B) = \{\varepsilon, a, b, c\}$$
$$FIRST(C) = \{\varepsilon, b\}$$

Ako vyzerajú množiny *FOLLOW* pre neterminály gramatiky?

Množina *FOLLOW*

(podčiarknutý je neterminál, na základe ktorého je pravidlo aktuálne vyšetrované a zelenou farbou časť za príslušným neterminálom, t.j. na prednáške označená ako β)
(červenou sú označené terminály, ktoré sa už v danej množine nachádzajú)

FOLLOW	S	A	B	C	Dôvod
Počiatočný neterminál	ε				
1. iterácia:					
$S \to \underline{A}BC$ $S \to \underline{A}BC$		a, b, c ε			$a, b, c \in FIRST(BC)$ $\varepsilon \in FIRST(BC)$ a teda $FOLLOW(S) \subseteq FOLLOW(A)$
$S \to A\underline{B}C$ $S \to A\underline{B}C$			b ε		$b \in FIRST(C)$ $\varepsilon \in FIRST(C)$ a teda $FOLLOW(S) \subseteq FOLLOW(B)$
$S \to AB\underline{C}\varepsilon$				ε	$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(S) \subseteq FOLLOW(C)$
$S \to a\underline{S}b$	b				$FIRST(b) = \{b\}$
$A \to a\underline{A}b$		b			$FIRST(b) = \{b\}$
$A \to \underline{C}\varepsilon$				a, b, c, ε	$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(A) \subseteq FOLLOW(C)$
$B \to \underline{B}abB$			a		$FIRST(abB) = \{a\}$
$B \to Bab\underline{B}\varepsilon$			a, b, ε		$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(B) \subseteq FOLLOW(B)$
$B \to \underline{A}A$ $B \to \underline{A}A$		a, b, c a, b, ε			$a, b, c \in FIRST(A)$ $\varepsilon \in FIRST(A)$ a teda $FOLLOW(B) \subseteq FOLLOW(A)$
$B \to A\underline{A}\varepsilon$		a, b, ε			$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(B) \subseteq FOLLOW(A)$
$C \to ba\underline{C}ab$				a	$FIRST(ab) = \{a\}$

Množina *FOLLOW*

(podčiarknutý je neterminál, na základe ktorého je pravidlo aktuálne vyšetrované a zelenou farbou časť za príslušným neterminálom)
(červenou sú označené terminály, ktoré sa už v danej množine nachádzajú)

FOLLOW	S	A	B	C	Dôvod
Po 1. iterácii	ε, b	a, b, c, ε	a, b, ε	a, b, c, ε	
2. iterácia:					
$S \rightarrow \underline{A}BC$ $S \rightarrow \underline{A}BC$		a, b, c ε, b			$a, b, c \in FIRST(BC)$ $\varepsilon \in FIRST(BC)$ a teda $FOLLOW(S) \subseteq FOLLOW(A)$
$S \rightarrow A\underline{B}C$ $S \rightarrow A\underline{B}C$			b ε, b		$b \in FIRST(C)$ $\varepsilon \in FIRST(C)$ a teda $FOLLOW(S) \subseteq FOLLOW(B)$
$S \rightarrow AB\underline{C}\varepsilon$				ε, b	$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(S) \subseteq FOLLOW(C)$
$S \rightarrow a\underline{S}b$	b				$FIRST(b) = \{b\}$
$A \rightarrow a\underline{A}b$		b			$FIRST(b) = \{b\}$
$A \rightarrow \underline{C}\varepsilon$				a, b, c, ε	$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(A) \subseteq FOLLOW(C)$
$B \rightarrow \underline{B}abB$			a		$FIRST(abB) = \{a\}$
$B \rightarrow Bab\underline{B}\varepsilon$			a, b, ε		$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(B) \subseteq FOLLOW(B)$
$B \rightarrow \underline{A}A$ $B \rightarrow \underline{A}A$		a, b, c a, b, ε			$a, b, c \in FIRST(A)$ $\varepsilon \in FIRST(A)$ a teda $FOLLOW(B) \subseteq FOLLOW(A)$
$B \rightarrow A\underline{A}\varepsilon$		a, b, ε			$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(B) \subseteq FOLLOW(A)$
$C \rightarrow ba\underline{C}ab$				a	$FIRST(ab) = \{a\}$

Vidíme, že v druhej iterácii už nedošlo k nájdeniu nových symbolov, teda algoritmus končí a výsledok:

$$FOLLOW(S) = \{\varepsilon, b\}$$

$$FOLLOW(A) = \{\varepsilon, a, b, c\}$$

$$FOLLOW(B) = \{\varepsilon, a, b\}$$

$$FOLLOW(C) = \{\varepsilon, a, b, c\}$$

FIRST, *FOLLOW* príklad č. 2

Príklad:

Gramatika $G = (\{$ <program>, <príkaz>, <príkazy>$\}$, $\{$**begin, end, p, ;** $\}$, P,<program>), pravidlá P:

<program> $\rightarrow$ **begin**<príkazy>**end**

<príkazy> $\rightarrow$ <príkaz><príkazy>

<príkazy> $\rightarrow \varepsilon$

<príkaz> $\rightarrow$ **p;**

<príkaz> $\rightarrow$ **begin**<príkazy>**end**

Určte množiny *FIRST* a *FOLLOW* pre symboly gramatiky.

FIRST príklad č. 2 - množina N_ε

Najprv vyšetríme množinu N_ε. Na začiatku $N_\varepsilon = \emptyset$.

1. Ako prvý do N_ε patrí neterminál <príkazy>, pretože priamo na základe pravidla <príkazy> $\rightarrow \varepsilon$ vidíme, že <príkazy> $\Rightarrow \varepsilon$, teda určite $N_\varepsilon = \{$<príkazy>$\}$
2. Ďalej, keď vieme, že <príkazy> $\in N_\varepsilon$, tak by sme hľadali v pravidlách pravidlo, ktoré má na pravej strane len reťazec zložený z neterminálu <príkazy>. Keďže také pravidlo v gramatike nemáme, tak množina N_ε už nebude obsahovať ďalšie neterminály.

Výsledná množina $N_\varepsilon = \{$<príkazy>$\}$.

FIRST	<program>	<príkazy>	<príkaz>	**begin**	**end**	**p**	**;**	
Terminály:				**begin**	**end**	**p**	**;**	
Pravidlá $A \to a\alpha$:	**begin**		**p, begin**					
$A \in N_\varepsilon$		ε						
1. iterácia		**p, begin**						*FIRST*(<príkaz><príkazy>) $\subseteq$ $\subseteq$ *FIRST*(<príkazy>)
2. iterácia								
Celkom:	**begin**	**p, begin**, ε	**p, begin**	**begin**	**end**	**p**	**;**	

FOLLOW	<program>	<príkazy>	<príkaz>
Začiatočný symbol	ε		
1. iterácia <program> $\rightarrow$ **begin**<príkazy>**end**		**end**	
<príkazy> $\rightarrow$ <príkaz><príkazy>			**p, begin, end**
<príkazy> $\rightarrow$ <príkaz><príkazy>			
<príkaz> $\rightarrow$ **begin**<príkazy>**end**			
2. iterácia <program> $\rightarrow$ **begin**<príkazy>**end**			
<príkazy> $\rightarrow$ <príkaz><príkazy>			
<príkazy> $\rightarrow$ <príkaz><príkazy>			
<príkaz> $\rightarrow$ **begin**<príkazy>**end**			
Celkom:	ε	**end**	**p, begin, end**

FIRST, *FOLLOW* č. 3

Je daná redukovaná gramatika $G = (N, T, P, S)$. Zistite, ako vyzerá množina *FIRST* symbolov gramatiky, resp. *FOLLOW* neterminálov gramatiky. V gramatike $N = \{S, A, B, C\}$, $T = \{a, b, c\}$, S je počiatočný neterminál, pravidlá P:

- $S \to AbBB$
- $A \to CC \mid cSA$
- $B \to aCa \mid bb$
- $C \to \varepsilon \mid BC$

FIRST príklad č. 3 - množina N_ε

Najprv vyšetríme množinu N_ε. Na začiatku $N_\varepsilon = \emptyset$.

1. Ako prvý do N_ε patrí neterminál C, pretože priamo na základe pravidla $C \rightarrow \varepsilon$ vidíme, že $C \Rightarrow \varepsilon$, teda určite $N_\varepsilon = \{C\}$
2. Ďalej, keď vieme, že $C \in N_\varepsilon$, tak na základe pravidla $A \rightarrow CC$ vidíme, že $A \Rightarrow CC \Rightarrow C \Rightarrow \varepsilon$, teda $A \Rightarrow^* \varepsilon$, a teda $N_\varepsilon = \{A, C\}$.
3. Ďalej by sme hľadali v pravidlách pravidlo, ktoré ma na pravej strane reťazec zložený len z neterminálov A, C. Keďže také pravidlo tam nemáme, hľadanie množiny N_ε končí.

Výsledná množina $N_\varepsilon = \{A, C\}$.

(červenou sú označené terminály, ktoré by mali byť pridané do príslušnej množiny *FIRST* na základe aktuálne vyšetrovaného pravidla, ale už sa v danej množine nachádzajú)

$FIRST$	S	A	B	C	a	b	c	
Terminály:					a	b	c	
Pravidlá $A \rightarrow a\alpha$:		c	a, b					
$A \in N_\varepsilon$		ε		ε				
1. iterácia								
$S \rightarrow AbBB$	c, b							Aktuálne $FIRST(AbBB) = \{b, c\}$
$A \rightarrow CC$		ε						Aktuálne $FIRST(CC) = \{\varepsilon\}$
$C \rightarrow BC$				a, b				Aktuálne $FIRST(BC) = \{a, b\}$

(červenou sú označené terminály, ktoré by mali byť pridané do príslušnej množiny *FIRST* na základe aktuálne vyšetrovaného pravidla, ale už sa v danej množine nachádzajú)

$FIRST$	S	A	B	C	a	b	c	
Po 1. iterácii	b, c	c, ε	a, b	a, b, ε	a	b	c	
2. iterácia								
$S \to AbBB$	c, b							Aktuálne $FIRST(AbBB) = \{b, c\}$
$A \to CC$		a, b, ε						Aktuálne $FIRST(CC) = \{a, b, \varepsilon\}$
$C \to BC$				a, b				Aktuálne $FIRST(BC) = \{a, b\}$

(červenou sú označené terminály, ktoré by mali byť pridané do príslušnej množiny *FIRST* na základe aktuálne vyšetrovaného pravidla, ale už sa v danej množine nachádzajú)

$FIRST$	S	A	B	C	a	b	c	
Po 2. iterácii	b, c	a, b, c, ε	a, b	a, b, ε	a	b	c	
3. iterácia								
$S \to AbBB$	a, c, b							Aktuálne $FIRST(AbBB) = \{a, b, c\}$
$A \to CC$		a, b, ε						$FIRST(CC) = \{a, b, \varepsilon\}$
$C \to BC$				a, b				Aktuálne $FIRST(BC) = \{a, b\}$

(červenou sú označené terminály, ktoré by mali byť pridané do príslušnej množiny *FIRST* na základe aktuálne vyšetrovaného pravidla, ale už sa v danej množine nachádzajú)

FIRST	S	A	B	C	a	b	c
Po 3. iterácii	a, b, c	a, b, c, ε	a, b	a, b, ε	a	b	c
4. iterácia							
$S \to AbBB$	a, c, b						
$A \to CC$		a, b, ε					
$C \to BC$				a, b			

Vidíme, že v štvrtej iterácii sme nedostali žiaden nový symbol do žiadnej množiny *FIRST*, algoritmus teda končí.

Výsledné množiny *FIRST*:

1. $FIRST(S) = \{a, b, c\}$
2. $FIRST(A) = \{a, b, c, \varepsilon\}$
3. $FIRST(B) = \{a, b\}$
4. $FIRST(C) = \{a, b, \varepsilon\}$

Množina *FOLLOW*

(podčiarknutý je neterminál, na základe ktorého je pravidlo aktuálne vyšetrované a zelenou farbou časť za príslušným neterminálom)
(červenou sú označené terminály, ktoré by mali byť pridané do príslušnej množiny *FOLLOW* na základe aktuálne vyšetrovaného pravidla, ale už sa v danej množine nachádzajú)

FOLLOW	S	A	B	C	Dôvod
Počiatočný neterminál	ε				
1. iterácia:					
$S \to \underline{A}bBB$		b			$FIRST(bBB) = \{b\}$
$S \to Ab\underline{B}B$			a, b		$FIRST(B) = \{a, b\}$
$S \to AbB\underline{B}\varepsilon$			ε		$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(S) \subseteq FOLLOW(B)$
$A \to \underline{C}C$				a, b b	$a, b \in FIRST(C)$ $\varepsilon \in FIRST(C)$ a teda $FOLLOW(A) \subseteq FOLLOW(C)$
$A \to C\underline{C}\varepsilon$				b	$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(A) \subseteq FOLLOW(C)$
$A \to c\underline{S}A$	a, b, c b				$a, b, c \in FIRST(A)$ $\varepsilon \in FIRST(A)$ a teda $FOLLOW(A) \subseteq FOLLOW(S)$
$A \to aS\underline{A}\varepsilon$		b			$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(A) \subseteq FOLLOW(A)$
$B \to a\underline{C}a$			a		$FIRST(a) = \{a\}$
$C \to \underline{B}C$			a, b a, b		$a, b \in FIRST(C)$ $\varepsilon \in FIRST(C)$ a teda $FOLLOW(C) \subseteq FOLLOW(B)$
$C \to B\underline{C}\varepsilon$				a, b	$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(C) \subseteq FOLLOW(C)$

Množina *FOLLOW*

(podčiarknutý je neterminál, na základe ktorého je pravidlo aktuálne vyšetrované a zelenou farbou časť za príslušným neterminálom)
(červenou sú označené terminály, ktoré by mali byť pridané do príslušnej množiny *FOLLOW* na základe aktuálne vyšetrovaného pravidla, ale už sa v danej množine nachádzajú)

FOLLOW	S	A	B	C	Dôvod
Po 1. iterácii	ε, a, b, c	b	a, b, ε	a, b	
2. iterácia:					
$S \rightarrow \underline{A}bBB$		b			$FIRST(bBB) = \{b\}$
$S \rightarrow Ab\underline{B}B$			a, b		$FIRST(B) = \{a, b\}$
$S \rightarrow AbB\underline{B}\varepsilon$			c, ε, a, b		$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(S) \subseteq FOLLOW(B)$
$A \rightarrow \underline{C}C$				a, b b	$a, b \in FIRST(C)$ $\varepsilon \in FIRST(C)$ a teda $FOLLOW(A) \subseteq FOLLOW(C)$
$A \rightarrow C\underline{C}\varepsilon$				b	$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(A) \subseteq FOLLOW(C)$
$A \rightarrow c\underline{S}A$	a, b, c b				$a, b, c \in FIRST(A)$ $\varepsilon \in FIRST(A)$ a teda $FOLLOW(A) \subseteq FOLLOW(S)$
$A \rightarrow aS\underline{A}\varepsilon$		b			$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(A) \subseteq FOLLOW(A)$
$B \rightarrow a\underline{C}a$			a		$FIRST(a) = \{a\}$
$C \rightarrow \underline{B}C$			a, b a, b		$a, b \in FIRST(C)$ $\varepsilon \in FIRST(C)$ a teda $FOLLOW(C) \subseteq FOLLOW(B)$
$C \rightarrow B\underline{C}\varepsilon$				a, b	$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(C) \subseteq FOLLOW(C)$

Množina *FOLLOW*

(podčiarknutý je neterminál, na základe ktorého je pravidlo aktuálne vyšetrované a zelenou farbou časť za príslušným neterminálom)
(červenou sú označené terminály, ktoré by mali byť pridané do príslušnej množiny *FOLLOW* na základe aktuálne vyšetrovaného pravidla, ale už sa v danej množine nachádzajú)

FOLLOW	S	A	B	C	Dôvod
Po 2. iterácii	ε, a, b, c	b	a, b, c, ε	a, b	
3. iterácia:					
$S \rightarrow \underline{A}bBB$		b			$FIRST(bBB) = \{b\}$
$S \rightarrow Ab\underline{B}B$			a, b		$FIRST(B) = \{a, b\}$
$S \rightarrow AbB\underline{B}\varepsilon$			ε, a, b, c		$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(S) \subseteq FOLLOW(B)$
$A \rightarrow \underline{C}C$				a, b b	$a, b \in FIRST(C)$ $\varepsilon \in FIRST(C)$ a teda $FOLLOW(A) \subseteq FOLLOW(C)$
$A \rightarrow C\underline{C}\varepsilon$				b	$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(A) \subseteq FOLLOW(C)$
$A \rightarrow c\underline{S}A$	a, b, c b				$a, b, c \in FIRST(A)$ $\varepsilon \in FIRST(A)$ a teda $FOLLOW(A) \subseteq FOLLOW(S)$
$A \rightarrow aS\underline{A}\varepsilon$		b			$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(A) \subseteq FOLLOW(A)$
$B \rightarrow a\underline{C}a$			a		$FIRST(a) = \{a\}$
$C \rightarrow \underline{B}C$			a, b a, b		$a, b \in FIRST(C)$ $\varepsilon \in FIRST(C)$ a teda $FOLLOW(C) \subseteq FOLLOW(B)$
$C \rightarrow B\underline{C}\varepsilon$				a, b	$FIRST(\varepsilon) = \{\varepsilon\}$ a teda $FOLLOW(C) \subseteq FOLLOW(C)$

Vidíme, že v tretej iterácii sme nedostali žiaden nový symbol do žiadnej množiny *FOLLOW*, algoritmus teda končí.

Výsledné množiny *FOLLOW*:

1. $FOLLOW(S) = \{\varepsilon, a, b, c\}$
2. $FOLLOW(A) = \{b\}$
3. $FOLLOW(B) = \{\varepsilon, a, b, c\}$
4. $FOLLOW(C) = \{a, b\}$