

Jméno:

UČO:



líst

učo

body

Oblast strojově snímaných informací. Svě učo a číslo lístu vyplňte
zleva dle vzoru číslic. Jinak do této oblasti nezasahujte.

0 1 2 3 4 5 6 7 8 9

1. [0,5 bodu] Uvažte gramatiku $\mathcal{G} = (\{N, M, L, V\}, \{s, a, i, k\}, P, N)$ s následujícími pravidly:

$$\begin{aligned} P = \{ & N \rightarrow saLiNa \mid siska, \\ & M \rightarrow VLas \mid LiNka \mid sLiNa, \\ & L \rightarrow MasiNka \mid VLak \mid sisiNka, \\ & V \rightarrow sLasiNka \mid sLiMak\}. \end{aligned}$$

Převedte gramatiku \mathcal{G} na jazykově ekvivalentní gramatiku v Chomského normální formě (CNF).

Vstupní podmínkou algoritmu pro převedení gramatiky do CNF je, aby byla vstupní gramatika vlastní a bez jednoduchých pravidel. Gramatika \mathcal{G} neobsahuje žádná jednoduchá pravidla ani ϵ -pravidla, stačí nám zkontrolovat, že všechny symboly z \mathcal{G} jsou použitelné.

Normované neterminály můžeme spočítat pomocí algoritmu 3.1. Postupně budeme získávat množiny $N_0 = \emptyset$, $N_1 = \{N\}$, $N_2 = \{N, L\}$, $N_3 = \{N, L, M, V\} = N_4 = N_e$. Jelikož množina N_e obsahuje všechny neterminály gramatiky \mathcal{G} , nemusíme odstraňovat žádné symboly.

Dosažitelné symboly určíme pomocí algoritmu 3.2, kdy v jednotlivých krocích dostaneme množiny $V_0 = \{N\}$, $V_1 = \{N, s, a, L, i, k\}$, $V_2 = \{N, s, a, L, i, k, M, V\} = V_3$. Zjistili jsme tedy, že všechny symboly gramatiky \mathcal{G} jsou použitelné a \mathcal{G} je redukováná.

Pro převod \mathcal{G} do CNF bychom mohli použít přímo algoritmus 3.6, ale tento postup může vést ke gramatice s příliš velkým počtem neterminálů. Při zjednodušování pravých stran pravidel však není nutné vždy oddělit jeden nejlevější znak (např. $saLiNa \rightsquigarrow s'\langle aLiNa \rangle$), ale můžeme řetězec na pravé straně pravidla rozdělit i na jiném místě (např. $saLiNa \rightsquigarrow \langle sa \rangle \langle LiNa \rangle$). Pokud rozdělení provedeme vždy pouze na jednom místě, i tímto způsobem dostaneme gramatiku v CNF (např. nesmíme provést rozdělení $saLiNa \rightsquigarrow \langle sa \rangle \langle Li \rangle \langle Na \rangle$).

Abychom počet neterminálů co nejvíce snížili, můžeme si všimnout opakujících se podčástí pravých stran pravidel. Nejprve si pravděpodobně uvědomíme, že $LiNka$, $MasiNka$, $sisiNka$ a $sLasiNka$ mají stejnou koncovku $iNka$. Dále si můžeme všimnout, že $VLas$ a $sLasiNka$ obsahují stejný podřetězec Las , $siska$ i $sisiNka$ začínají na sis , a $saLiNa$ i $sLiNa$ končí na $LiNa$. Z toho získáme následující pravidla:

$$\begin{aligned} N &\rightarrow \langle sa \rangle \langle LiNa \rangle \mid \langle sis \rangle \langle ka \rangle \\ M &\rightarrow V \langle Las \rangle \mid L \langle iNka \rangle \mid s' \langle LiNa \rangle \\ L &\rightarrow \langle Mas \rangle \langle iNka \rangle \mid V \langle Lak \rangle \mid \langle sis \rangle \langle iNka \rangle \\ V &\rightarrow \langle sLas \rangle \langle iNka \rangle \mid s' \langle LiMak \rangle \end{aligned}$$

Dále si můžeme všimnout, že řetězce $LiMak$ a $LiNa$ sdílejí začátek Li , a že se několikrát opakuje koncovka ak a as . Nově získané neterminály tedy rozdělíme tímto způsobem:

$$\begin{aligned} \langle sa \rangle &\rightarrow s'a' \\ \langle LiNa \rangle &\rightarrow \langle Li \rangle \langle Na \rangle \\ \langle sis \rangle &\rightarrow s' \langle is \rangle \\ \langle ka \rangle &\rightarrow k'a' \\ \langle Las \rangle &\rightarrow L \langle as \rangle \\ \langle iNka \rangle &\rightarrow \langle iN \rangle \langle ka \rangle \end{aligned}$$

Jméno:

UČO:

0007

líst

2

učo

body

0

Oblast strojově snímaných informací. Svě učo a číslo lístu vyplňte zleva dle vzoru číslic. Jinak do této oblasti nezasahujte.

0123456789

$$\langle Mas \rangle \rightarrow M \langle as \rangle$$

$$\langle Lak \rangle \rightarrow L \langle ak \rangle$$

$$\langle sLas \rangle \rightarrow \langle sL \rangle \langle as \rangle$$

$$\langle LiMak \rangle \rightarrow \langle Li \rangle \langle Mak \rangle$$

V dalším kroku rozdělíme zbylé neterminály.

$$\langle Li \rangle \rightarrow Li'$$

$$\langle Na \rangle \rightarrow Na'$$

$$\langle is \rangle \rightarrow i' s'$$

$$\langle as \rangle \rightarrow a' s'$$

$$\langle iN \rangle \rightarrow i' N$$

$$\langle ak \rangle \rightarrow a' k'$$

$$\langle sL \rangle \rightarrow s' L$$

$$\langle Mak \rangle \rightarrow M \langle ak \rangle$$

Nakonec přidáme pravidla pro získání terminálů.

$$i' \rightarrow i$$

$$a' \rightarrow a$$

$$s' \rightarrow s$$

$$k' \rightarrow k$$

Jazykově ekvivalentní gramatika v CNF, jejíž neterminály jsou velká písmena anglické abecedy, je čtveřice $(N', \{s, a, i, k\}, P', N)$. Při přejmenovávání neterminálů zachováváme význam původních neterminálů N, M, L, V a neterminály s', a', i', k' přejmenujeme na odpovídající velká písmena S, A, I, K . Ostatní přejmenujeme v abecedním pořadí podle jejich prvního výskytu. Množiny N' a P' tedy definujeme následovně (pravý sloupec v definici množiny pravidel obsahuje odpovídající pravidla před přejmenováním):

$$N' = \{N, M, L, V, B, C, D, E, F, G, H, J, O, P, Q, R, T, U, W, X, Y, Z, I, A, S, K\}$$

Jméno:

UČO:

0007

list

3

učo

body

0

Oblast strojově snímaných informací. Svě učo a číslo listu vyplňte
zleva dle vzoru číslic. Jinak do této oblasti nezasahujte.

0123456789

$$P' = \{N \rightarrow BC \mid DE$$

$$M \rightarrow VF \mid LG \mid SC$$

$$L \rightarrow HG \mid VJ \mid DG$$

$$V \rightarrow OG \mid SP$$

$$N \rightarrow \langle sa \rangle \langle LiNa \rangle \mid \langle sis \rangle \langle ka \rangle$$

$$M \rightarrow V \langle Las \rangle \mid L \langle iNka \rangle \mid s' \langle LiNa \rangle$$

$$L \rightarrow \langle Mas \rangle \langle iNka \rangle \mid V \langle Lak \rangle \mid \langle sis \rangle \langle iNka \rangle$$

$$V \rightarrow \langle sLas \rangle \langle iNka \rangle \mid s' \langle LiMak \rangle$$

$$B \rightarrow SA$$

$$\langle sa \rangle \rightarrow s'a'$$

$$C \rightarrow QR$$

$$\langle LiNa \rangle \rightarrow \langle Li \rangle \langle Na \rangle$$

$$D \rightarrow ST$$

$$\langle sis \rangle \rightarrow s' \langle is \rangle$$

$$E \rightarrow KA$$

$$\langle ka \rangle \rightarrow k'a'$$

$$F \rightarrow LU$$

$$\langle Las \rangle \rightarrow L \langle as \rangle$$

$$G \rightarrow WE$$

$$\langle iNka \rangle \rightarrow \langle iN \rangle \langle ka \rangle$$

$$H \rightarrow MU$$

$$\langle Mas \rangle \rightarrow M \langle as \rangle$$

$$J \rightarrow LX$$

$$\langle Lak \rangle \rightarrow L \langle ak \rangle$$

$$O \rightarrow YU$$

$$\langle sLas \rangle \rightarrow \langle sL \rangle \langle as \rangle$$

$$P \rightarrow QZ$$

$$\langle LiMak \rangle \rightarrow \langle Li \rangle \langle Mak \rangle$$

$$Q \rightarrow LI$$

$$\langle Li \rangle \rightarrow Li'$$

$$R \rightarrow NA$$

$$\langle Na \rangle \rightarrow Na'$$

$$T \rightarrow IS$$

$$\langle is \rangle \rightarrow i's'$$

$$U \rightarrow AS$$

$$\langle as \rangle \rightarrow a's'$$

$$W \rightarrow IN$$

$$\langle iN \rangle \rightarrow i'N$$

$$X \rightarrow AK$$

$$\langle ak \rangle \rightarrow a'k'$$

$$Y \rightarrow SL$$

$$\langle sL \rangle \rightarrow s'L$$

$$Z \rightarrow MX$$

$$\langle Mak \rangle \rightarrow M \langle ak \rangle$$

$$I \rightarrow i$$

$$i' \rightarrow i$$

$$A \rightarrow a$$

$$a' \rightarrow a$$

$$S \rightarrow s$$

$$s' \rightarrow s$$

$$K \rightarrow k$$

$$k' \rightarrow k\}.$$