

Fordítóprogramok

7. előadás

Aszalós László

2015. október 25.

LR(k) elemzés

Definíció

Legyen $G = (N, T, P, S)$ nyelvtanhoz tartozó G' kiegészített nyelvtan a következő: $G' = (N \cup \{S'\}, T, P \cup \{S' \rightarrow S\}, S')$.

Megjegyzés

Sorszámozzuk meg a szabályokat, az $S' \rightarrow S$ legyen a nulladik szabály. Ha az elemzés során a nulladik szabály szerint kell redukálni, akkor ez az elemzés végét, és a szintaktikus helyességet jelzi.

Tétel

Egy G' kiegészített nyelvtan $LR(k)$ nyelvtan $k \geq 0$, ha $S' \xRightarrow{*} \alpha A w \Rightarrow \alpha \beta w$ és $S' \xRightarrow{*} \gamma B x \Rightarrow \gamma \delta x = \alpha \beta y$ és $First_k(w) = First_k(y)$ esetén $\alpha = \gamma$, $A = B$ és $x = y$.

Mutassa meg, hogy

- A $G'_1 = (\{S', S\}, \{a\}, \{S' \rightarrow S, S \rightarrow Sa|a\}, S')$ nem LR(0) nyelvtan.
- A $G'_3 = (\{S', S\}, \{a\}, \{S' \rightarrow S, S \rightarrow aSa|a\}, S')$ nem LR(k) nyelvtan egyetlen k -ra sem.

LL(k) és LR(k) kapcsolata

Tétel

Minden LL(k) nyelvtan LR(k) nyelvtan, de létezik olyan LR(k) nyelvtan, amelyik nem LL(k') nyelvtan egyetlen k' -re sem.

Tétel

Minden LR(k) ($k > 1$) nyelvtanhoz létezik vele ekvivalens LR(1) nyelvtan.

LR(0) elemzés

Definíció

Egy mondatforma legbaloldalibb egyszerű részmondatát a mondatforma *nyelének* nevezzük.

Definíció

Legyen az $\alpha\beta x$ mondatforma nyele β . Ekkor az $\alpha\beta$ jelsorozat prefixeit az $\alpha\beta x$ *járható prefixeinek* nevezzük.

Példa

Tekintsük $S' \rightarrow E$, $E \rightarrow T \mid E + T$, $T \rightarrow i \mid (E)$ nyelvtan $E + (i + i)$ mondatformáját. Adja meg a mondatforma nyelét és járható prefixeit!

Megoldás

Az első i a mondatforma nyele, a járható prefixei: E , $E+$, $E + ($ és $E + (i$.

LR(0)-elem

Definíció

Legyen a \bullet karakter egy metaszimbólum. Ha egy G' nyelvtan helyettesítési szabálya $A \rightarrow \alpha\beta$, akkor a nyelvtan egy *LR(0)-eleme* legyen $[A \rightarrow \alpha \bullet \beta]$.

Definíció

A G' nyelvtan $[A \rightarrow \alpha \bullet \beta]$ LR(0)-eleme *érvényes* a $\gamma\alpha$ járható prefixre nézve, ha $S' \xRightarrow{*} \gamma Ax \Rightarrow \gamma\alpha\beta x$.

closure halmazok

Definíció

Legyen az \mathcal{I} halmaz a nyelvtan egy LR(0)-elemhalmaza. Ekkor a $\text{closure}(\mathcal{I})$ halmaz a következő LR(0)-elemeket tartalmazza

- az \mathcal{I} minden eleme legyen a $\text{closure}(\mathcal{I})$ halmaznak is,
- ha $[A \rightarrow \alpha \bullet B\beta] \in \text{closure}(\mathcal{I})$ és $B \rightarrow \gamma$ a nyelvtan egy helyettesítési szabálya, akkor legyen $[B \rightarrow \bullet\gamma] \in \text{closure}(\mathcal{I})$.

Megjegyzés

Ezzel az eljárással nyert halmazok végesek.

read halmazok

Definíció

Legyen a \mathcal{I} halmaz egy nyelvtan egy LR(0)-elemhalmaza. Ekkor a $read(\mathcal{I}, X)$ halmaz a következő LR(0)-elemeket tartalmazza:

- ha $[A \rightarrow \alpha \bullet X \beta] \in \mathcal{I}$, akkor a $closure([A \rightarrow \alpha X \bullet \beta])$ minden eleme legyen a $read(\mathcal{I}, X)$ eleme,

Megjegyzés

Ezzel az eljárással nyert halmazok végesek.

Kanonikus halmazok

Definíció

- Jelöljük a nulladik kanonikus halmazt \mathcal{I}_0 -val, és legyen $\mathcal{I}_0 = \text{closure}([S' \rightarrow \bullet S])$.
- Ezután képezzük egy X szimbólumra a $\text{read}(\mathcal{I}_0, X)$ halmazt.
 - ▶ Ha az így kapott halmaz nem üres, és nem egyezik meg az korábbi kanonikus halmazok egyikével sem, akkor ez legyen a következő kanonikus halmaz, és indexe legyen eggyel nagyobb, mint a korábbi maximális index.
- Miután ezt végrehajtottuk az összes terminális és nemterminális X szimbólumra, ismételjük meg az összes kanonikus halmazra is.

Definíció

Az így létrehozott $\mathcal{I}_0, \mathcal{I}_1, \dots, \mathcal{I}_n$ halmazokat nevezzük a G nyelvtan $LR(0)$ -kanonikus halmazainak.

Kanonikus halmazok

Megjegyzés

A kanonikus halmazok darabszáma véges.

Tétel

Egy γ járható prefixre érvényes LR(0)-elemek halmaza az az \mathcal{I}_k kanonikus elemhalmaz, amelyik az elemző determinisztikus véges automatájának ahhoz a k állapotához tartozik, amelyikbe az automata a kezdőállapotból a γ hatására kerül.

Példa

Készítse el az alábbi nyelvtan kanonikus halmazait!

① $S' \rightarrow S$

② $S \rightarrow aAd$

③ $A \rightarrow bA$

④ $A \rightarrow c$

Megoldás.

- $[\mathcal{I}_0)] \{[S' \rightarrow \bullet S], [S \rightarrow \bullet aAd]\}$
- $[\mathcal{I}_1)] \{[S' \rightarrow S\bullet]\} = read(\mathcal{I}_0, S)$
- $[\mathcal{I}_2)] \{[S \rightarrow a \bullet Ad], [A \rightarrow \bullet bA], [A \rightarrow \bullet c]\} = read(\mathcal{I}_0, a)$
- $[\mathcal{I}_3)] \{[S \rightarrow aA \bullet d]\} = read(\mathcal{I}_2, A)$
- $[\mathcal{I}_4)] \{[A \rightarrow b \bullet A], [A \rightarrow \bullet bA], [A \rightarrow \bullet c]\} = read(\mathcal{I}_2, b) = read(\mathcal{I}_4, b)$
- $[\mathcal{I}_5)] \{[A \rightarrow c\bullet]\} = read(\mathcal{I}_2, c) = read(\mathcal{I}_4, c)$
- $[\mathcal{I}_6)] \{[S \rightarrow aAd\bullet]\} = read(\mathcal{I}_3, d)$
- $[\mathcal{I}_7)] \{[A \rightarrow bA\bullet]\} = read(\mathcal{I}_4, A)$

{LR(0) elemző táblázat

		S	A	a	b	c	d
0	s	1		2			
1	acc						
2	s		3		4	5	
3	s						6
4	s		7		4	5	
5	r_3						
6	r_1						
7	r_2						

Elemzés menete

- Az elemző verme egy dupla verem (push/pop két elemet mozgat). Az elemző állapotát a verem és az input még nem elemzett részének párosa adja. Így a kezdőállapot $(\#0, x\#)$, ahol x az input szó. Sikeres elemzésnél a végállapot $(\#0, \#)$. Jelölje a verem felső elemét i .
- Ha a táblázat i -dik sorának második eleme s , azaz shift, akkor a verembe a soron következő karakter, valamint a karakternek megfelelő mezőben szereplő szám kerüljön a verembe.
- Ha sor második eleme r_j , azaz redukálás a j -dik szabály szerint, akkor a j -dik szabály jobb oldalának megfelelő számú párost törölni kell a veremből, és a szabály bal oldalán szereplő nemterminálist, és az oszlopában szereplő számot kell a verembe írni.
- Ha a második elem acc , azaz elfogadás, az elemzés sikeres.

Példa

Elemezze az *abbcd#* inputot az előbbi táblázattal.

Megoldás

Az állapotátmenetek a következők: $(\#0, abbcd\#) \xrightarrow{s,2} (\#0a2, bbcd\#) \xrightarrow{s,4} (\#0a2b4, bcd\#) \xrightarrow{s,4} (\#0a2b4b4, cd\#) \xrightarrow{s,5} (\#0a2b4b4c5, d\#) \xrightarrow{r3} (\#0a2b4b4A7, d\#) \xrightarrow{r2} (\#0a2b4A7, d\#) \xrightarrow{r2} (\#0a2A3, d\#) \xrightarrow{s,6} (\#0a2A3d6, \#) \xrightarrow{r1} (\#0S1, \#) \xrightarrow{acc} (\#0, \#)$