

Title: Simon: Compilerbau (23.05.2012)

Date: Wed May 23 14:13:06 CEST 2012

Duration: 78:44 min

Pages: 33

LR(k)-Grammatik

253

Idee: Benutze k -Vorausschau, um Konflikte zu lösen.

Definition:

Die reduzierte kontextfreie Grammatik G heißt $LR(k)$ -Grammatik, falls für $\text{First}_k(w) = \text{First}_k(x)$ aus:

$$\left. \begin{array}{l} S \xrightarrow{*}_R \alpha A w \rightarrow \alpha \beta w \\ S \xrightarrow{*}_R \alpha' A' w' \rightarrow \alpha \beta x \end{array} \right\} \text{folgt: } \alpha = \alpha' \wedge A = A' \wedge w' = x$$

LR(0)-Parser

Achtung:

Leider ist der $LR(0)$ -Parser im allgemeinen nicht-deterministisch

Wir identifizieren zwei Gründe:

Reduce-Reduce-Konflikt:

$$[A \rightarrow \gamma \bullet], [A' \rightarrow \gamma' \bullet] \in q \text{ mit } A \neq A' \vee \gamma \neq \gamma'$$

Shift-Reduce-Konflikt:

$$[A \rightarrow \gamma \bullet], [A' \rightarrow \alpha \bullet a \beta] \in q \text{ mit } a \in T \text{ für einen Zustand } q \in Q.$$

Solche Zustände nennen wir $LR(0)$ -ungeeignet.

LR(k)-Grammatik

254

im Beispiel:

$$(1) \quad S \rightarrow A \mid B \quad A \rightarrow aAb \mid 0 \quad B \rightarrow aBbb \mid 1$$

... ist nicht $LL(k)$ für jedes k — aber $LR(0)$:

Sei $S \xrightarrow{*}_R \alpha X w \rightarrow \alpha \beta w$. Dann ist $\alpha \beta$ von einer der Formen:

$$\underline{A}, \underline{B}, \underline{a^n aAb}, \underline{a^n aBbb}, \underline{a^n 0}, \underline{a^n 1} \quad (n \geq 0)$$

b b b b

LR(k)-Grammatik

im Beispiel:

- (1) $S \rightarrow A \mid B \quad A \rightarrow aAb \mid 0 \quad B \rightarrow aBbb \mid 1$
 ... ist nicht $LL(k)$ für jedes k — aber $LR(0)$:

Sei $S \xrightarrow{*}_R \alpha X w \rightarrow \alpha \beta w$. Dann ist $\alpha \underline{\beta}$ von einer der Formen:

$$\underline{A}, \underline{B}, a^n \underline{aAb}, a^n \underline{aBbb}, a^n \underline{0}, a^n \underline{1} \quad (n \geq 0)$$

- (2) $S \rightarrow aAc \quad A \rightarrow Abb \mid b$
 ... ist ebenfalls $LR(0)$:

Sei $S \xrightarrow{*}_R \alpha X w \rightarrow \alpha \beta w$. Dann ist $\alpha \underline{\beta}$ von einer der Formen:

$$\underline{ab}, \underline{aAbb}, \underline{aAc}$$

255

129/160

LR(k)-Grammatik

im Beispiel:

- (3) $S \rightarrow aAc \quad A \rightarrow bbA \mid b$... ist nicht $LR(0)$, aber $LR(1)$:

Für $S \xrightarrow{*}_R \alpha X w \rightarrow \alpha \beta w$ mit $\{y\} = \text{First}_k(w)$ ist $\alpha \underline{\beta} y$ von einer der Formen:

$$ab^{2n} \underline{bc}, ab^{2n} \underline{bbA}c, \underline{aAc}$$

256

130/160

LR(k)-Grammatik

im Beispiel:

- (3) $S \rightarrow aAc \quad A \rightarrow bbA \mid b$... ist nicht $LR(0)$, aber $LR(1)$:

Für $S \xrightarrow{*}_R \alpha X w \rightarrow \alpha \beta w$ mit $\{y\} = \text{First}_k(w)$ ist $\alpha \underline{\beta} y$ von einer der Formen:

$$ab^{2n} \underline{bc}, ab^{2n} \underline{bbA}c, \underline{aAc}$$

- (4) $S \rightarrow aAc \quad A \rightarrow bAb \mid b$... ist nicht $LR(k)$ für jedes $k \geq 0$:

Betrachte einfach die Rechtsableitungen:

$$S \xrightarrow{*}_R ab^n Ab^n c \rightarrow ab^n \underline{bb^n} c$$

257

130/160

LR(k)-Grammatik

im Beispiel:

- (1) $S \rightarrow A \mid B \quad A \rightarrow aAb \mid 0 \quad B \rightarrow aBbb \mid 1$
 ... ist nicht $LL(k)$ für jedes k — aber $LR(0)$:

Sei $S \xrightarrow{*}_R \alpha X w \rightarrow \alpha \beta w$. Dann ist $\alpha \underline{\beta}$ von einer der Formen:

$$\underline{A}, \underline{B}, a^n \underline{aAb}, a^n \underline{aBbb}, a^n \underline{0}, a^n \underline{1} \quad (n \geq 0)$$

- (2) $S \rightarrow aAc \quad A \rightarrow Abb \mid b$
 ... ist ebenfalls $LR(0)$:

Sei $S \xrightarrow{*}_R \alpha X w \rightarrow \alpha \beta w$. Dann ist $\alpha \underline{\beta}$ von einer der Formen:

$$\underline{ab}, \underline{aAbb}, \underline{aAc}$$

255

129/160

LR(1)-Items

Sei $k = 1$.

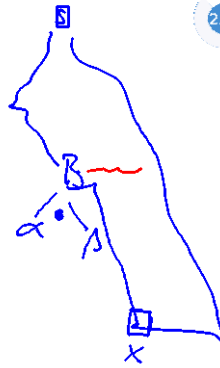
Idee: Wir staten Items mit 1-Vorausschau aus

Ein LR(1)-Item ist dann ein Paar:

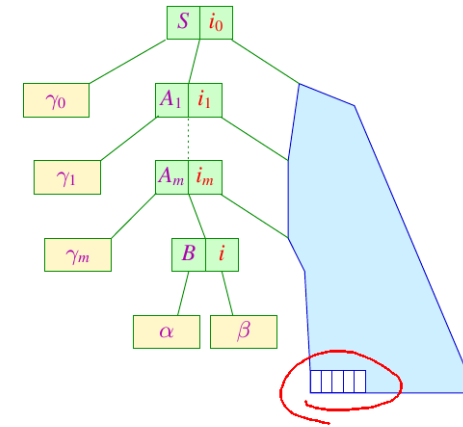
$$[B \rightarrow \alpha \bullet \beta, x], \quad x \in \text{Follow}_1(B) = \bigcup \{ \text{First}_1(\nu) \mid S \rightarrow^* \mu B \nu \}$$

Dieses Item ist gültig für $\gamma \alpha$ falls:

$$S \rightarrow_R^* \gamma B w \quad \text{mit} \quad \{x\} = \text{First}_1(w)$$



LR(1)-Items



... wobei $\gamma_0 \dots \gamma_m = \gamma$

Der charakteristische LR(1)-Automat

Der Automat $c(G, 1)$:

Zustände: LR(1)-Items

Anfangszustand: $[S' \rightarrow \bullet S, \epsilon]$

Endzustände: $\{ [B \rightarrow \gamma \bullet, x] \mid B \rightarrow \gamma \in P, x \in \text{Follow}_1(B) \}$

Übergänge:

- (1) $([A \rightarrow \alpha \bullet X \beta, x], X, [A \rightarrow \alpha X \bullet \beta, x]), \quad X \in (N \cup T)$
- (2) $([A \rightarrow \alpha \bullet B \beta, x], \epsilon, [B \rightarrow \bullet \gamma, x']),$
 $A \rightarrow \alpha B \beta, B \rightarrow \gamma \in P, x' \in \text{First}_1(\beta) \odot \{x\};$

Der kanonische LR(1)-Automat

Den kanonischen LR(1)-Automaten $LR(G, 1)$ erhält man aus $c(G, 1)$, indem man nach jedem Übergang beliebig viele ϵ liest und dann den Automaten deterministisch macht ...

Der kanonische LR(1)-Automat

Den kanonischen LR(1)-Automaten $LR(G, 1)$ erhält man aus $c(G, 1)$, indem man nach jedem Übergang beliebig viele ϵ liest und dann den Automaten **deterministisch** macht ...

Man kann ihn aber auch **direkt** aus der Grammatik konstruieren

Wie bei LR(0) benötigt man eine Hilfsfunktion:

$$\delta_\epsilon^*(q) = q \cup \{ [B \rightarrow \bullet \gamma, x] \mid \exists [A \rightarrow \alpha \bullet B' \beta', x'] \in q, \beta \in (N \cup T)^* : B' \rightarrow^* B \beta \wedge x \in \text{First}_1(\beta \beta') \odot \{x'\} \}$$

Der kanonische LR(1)-Automat

Den kanonischen LR(1)-Automaten $LR(G, 1)$ erhält man aus $c(G, 1)$, indem man nach jedem Übergang beliebig viele ϵ liest und dann den Automaten **deterministisch** macht ...

Man kann ihn aber auch **direkt** aus der Grammatik konstruieren

Wie bei LR(0) benötigt man eine Hilfsfunktion:

$$\delta_\epsilon^*(q) = q \cup \{ [B \rightarrow \bullet \gamma, x] \mid \exists [A \rightarrow \alpha \bullet B' \beta', x'] \in q, \beta \in (N \cup T)^* : B' \rightarrow^* B \beta \wedge x \in \text{First}_1(\beta \beta') \odot \{x'\} \}$$

Dann definiert man:

- Zustände:** Mengen von LR(1)-Items;
- Anfangszustand:** $\delta_\epsilon^* \{ [S' \rightarrow \bullet S, \epsilon] \}$
- Endzustände:** $\{ q \mid \exists A \rightarrow \alpha \in P : [A \rightarrow \alpha \bullet, x] \in q \}$
- Übergänge:** $\delta(q, X) = \delta_\epsilon^* \{ [A \rightarrow \alpha X \bullet \beta, x] \mid [A \rightarrow \alpha \bullet X \beta, x] \in q \}$

Der kanonische LR(1)-Automat

im Beispiel:

Handwritten notes: $\bullet B + T_i +$
 $E \rightarrow \bullet T_i +$

$q_0 = \{ [S' \rightarrow \bullet E, \epsilon], [E \rightarrow \bullet E, \epsilon], [E \rightarrow \bullet T, \epsilon], [T \rightarrow \bullet T, \epsilon], [T \rightarrow \bullet F, \epsilon], [F \rightarrow \bullet (E), \epsilon], [F \rightarrow \bullet \text{int}, \epsilon] \}$
 $q_1 = \delta(q_0, E) = \{ [S' \rightarrow E \bullet, \epsilon], [E \rightarrow E \bullet + T, \epsilon] \}$
 $q_2 = \delta(q_0, T) = \{ [E \rightarrow T \bullet, \epsilon], [T \rightarrow T \bullet * F, \epsilon] \}$
 $q_3 = \delta(q_0, F) = \{ [T \rightarrow F \bullet, \epsilon] \}$
 $q_4 = \delta(q_0, \text{int}) = \{ [F \rightarrow \text{int} \bullet, \epsilon] \}$
 $q_5 = \delta(q_0, ()) = \{ [F \rightarrow (\bullet E), \epsilon], [E \rightarrow \bullet E + T, \epsilon], [E \rightarrow \bullet T, \epsilon], [T \rightarrow \bullet T * F, \epsilon], [T \rightarrow \bullet F, \epsilon], [F \rightarrow \bullet (E), \epsilon], [F \rightarrow \bullet \text{int}, \epsilon] \}$

Der kanonische LR(1)-Automat

im Beispiel:

$q_0 = \{ [S' \rightarrow \bullet E, \{\epsilon\}], [E \rightarrow \bullet E + T, \{\epsilon, +\}], [E \rightarrow \bullet T, \{\epsilon, +\}], [T \rightarrow \bullet T * F, \{\epsilon, +, *\}], [T \rightarrow \bullet F, \{\epsilon, +, *\}], [F \rightarrow \bullet (E), \{\epsilon, +, *\}], [F \rightarrow \bullet \text{int}, \{\epsilon, +, *\}] \}$
 $q_1 = \delta(q_0, E) = \{ [S' \rightarrow E \bullet, \{\epsilon\}], [E \rightarrow E \bullet + T, \{\epsilon, +\}] \}$
 $q_2 = \delta(q_0, T) = \{ [E \rightarrow T \bullet, \{\epsilon, +\}], [T \rightarrow T \bullet * F, \{\epsilon, +, *\}] \}$
 $q_3 = \delta(q_0, F) = \{ [T \rightarrow F \bullet, \{\epsilon, +, *\}] \}$
 $q_4 = \delta(q_0, \text{int}) = \{ [F \rightarrow \text{int} \bullet, \{\epsilon, +, *\}] \}$
 $q_5 = \delta(q_0, ()) = \{ [F \rightarrow (\bullet E), \{\epsilon, +, *\}], [E \rightarrow \bullet E + T, \{\epsilon, +\}], [E \rightarrow \bullet T, \{\epsilon, +\}], [T \rightarrow \bullet T * F, \{\epsilon, +, *\}], [T \rightarrow \bullet F, \{\epsilon, +, *\}], [F \rightarrow \bullet (E), \{\epsilon, +, *\}], [F \rightarrow \bullet \text{int}, \{\epsilon, +, *\}] \}$

Der kanonische LR(1)-Automat

268

im Beispiel:

$$\begin{aligned}
 q_0 &= \{[S' \rightarrow \bullet E, \{\epsilon\}], [E \rightarrow \bullet E+T, \{\epsilon, +\}], [E \rightarrow \bullet T, \{\epsilon, +\}], [T \rightarrow \bullet T*F, \{\epsilon, +, *\}], [T \rightarrow \bullet F, \{\epsilon, +, *\}], [F \rightarrow \bullet (E), \{\epsilon, +, *\}], [F \rightarrow \bullet \text{int}, \{\epsilon, +, *\}]\} \\
 q_1 &= \delta(q_0, E) = \{[S' \rightarrow E \bullet, \{\epsilon\}], [E \rightarrow E \bullet + T, \{\epsilon, +\}]\} \\
 q_2 &= \delta(q_0, T) = \{[E \rightarrow T \bullet, \{\epsilon, +\}], [T \rightarrow T \bullet * F, \{\epsilon, +, *\}]\} \\
 q_3 &= \delta(q_0, F) = \{[T \rightarrow F \bullet, \{\epsilon, +, *\}]\} \\
 q_4 &= \delta(q_0, \text{int}) = \{[F \rightarrow \text{int} \bullet, \{\epsilon, +, *\}]\} \\
 q_5 &= \delta(q_0, () = \{[F \rightarrow (\bullet E), \{\epsilon, +, *\}], [E \rightarrow \bullet E+T, \{\epsilon, +\}], [E \rightarrow \bullet T, \{\epsilon, +\}], [T \rightarrow \bullet T*F, \{\epsilon, +, *\}], [T \rightarrow \bullet F, \{\epsilon, +, *\}], [F \rightarrow \bullet (E), \{\epsilon, +, *\}], [F \rightarrow \bullet \text{int}, \{\epsilon, +, *\}]\}
 \end{aligned}$$

135/160

Der kanonische LR(1)-Automat

269

im Beispiel:

$$\begin{aligned}
 q_0 &= \{[S' \rightarrow \bullet E, \{\epsilon\}], [E \rightarrow \bullet E+T, \{\epsilon, +\}], [E \rightarrow \bullet T, \{\epsilon, +\}], [T \rightarrow \bullet T*F, \{\epsilon, +, *\}], [T \rightarrow \bullet F, \{\epsilon, +, *\}], [F \rightarrow \bullet (E), \{\epsilon, +, *\}], [F \rightarrow \bullet \text{int}, \{\epsilon, +, *\}]\} \\
 q_1 &= \delta(q_0, E) = \{[S' \rightarrow E \bullet, \{\epsilon\}], [E \rightarrow E \bullet + T, \{\epsilon, +\}]\} \\
 q_2 &= \delta(q_0, T) = \{[E \rightarrow T \bullet, \{\epsilon, +\}], [T \rightarrow T \bullet * F, \{\epsilon, +, *\}]\} \\
 q_3 &= \delta(q_0, F) = \{[T \rightarrow F \bullet, \{\epsilon, +, *\}]\} \\
 q_4 &= \delta(q_0, \text{int}) = \{[F \rightarrow \text{int} \bullet, \{\epsilon, +, *\}]\} \\
 q_5 &= \delta(q_0, () = \{[F \rightarrow (\bullet E), \{\epsilon, +, *\}], [E \rightarrow \bullet E+T, \{\epsilon, +\}], [E \rightarrow \bullet T, \{\epsilon, +\}], [T \rightarrow \bullet T*F, \{\epsilon, +, *\}], [T \rightarrow \bullet F, \{\epsilon, +, *\}], [F \rightarrow \bullet (E), \{\epsilon, +, *\}], [F \rightarrow \bullet \text{int}, \{\epsilon, +, *\}]\}
 \end{aligned}$$

135/160

Der kanonische LR(1)-Automat

273

im Beispiel:

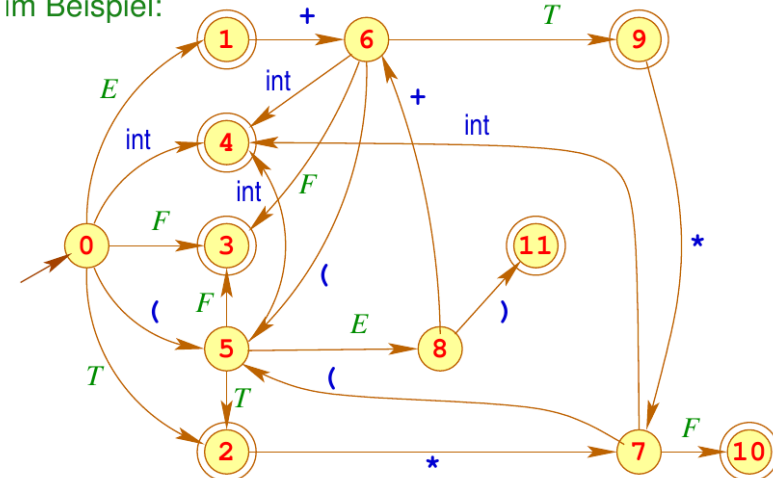
$$\begin{aligned}
 q'_5 &= \delta(q_5, () = \{[F \rightarrow (\bullet E), \{\epsilon, +, *\}], [E \rightarrow \bullet E+T, \{\epsilon, +\}], [E \rightarrow \bullet T, \{\epsilon, +\}], [T \rightarrow \bullet T*F, \{\epsilon, +, *\}], [T \rightarrow \bullet F, \{\epsilon, +, *\}], [F \rightarrow \bullet (E), \{\epsilon, +, *\}], [F \rightarrow \bullet \text{int}, \{\epsilon, +, *\}]\} \\
 q_6 &= \delta(q_1, +) = \{[E \rightarrow E+ \bullet T, \{\epsilon, +\}], [T \rightarrow \bullet T*F, \{\epsilon, +, *\}], [T \rightarrow \bullet F, \{\epsilon, +, *\}], [F \rightarrow \bullet (E), \{\epsilon, +, *\}], [F \rightarrow \bullet \text{int}, \{\epsilon, +, *\}]\} \\
 q_7 &= \delta(q_2, *) = \{[T \rightarrow T* \bullet F, \{\epsilon, +, *\}], [F \rightarrow \bullet (E), \{\epsilon, +, *\}], [F \rightarrow \bullet \text{int}, \{\epsilon, +, *\}]\} \\
 q_8 &= \delta(q_5, E) = \{[F \rightarrow (E \bullet), \{\epsilon, +, *\}], [E \rightarrow E+ \bullet T, \{\epsilon, +\}]\} \\
 q_9 &= \delta(q_6, T) = \{[E \rightarrow E+T \bullet, \{\epsilon, +\}], [T \rightarrow T* \bullet F, \{\epsilon, +, *\}]\} \\
 q_{10} &= \delta(q_7, F) = \{[T \rightarrow T*F \bullet, \{\epsilon, +, *\}]\} \\
 q_{11} &= \delta(q_8,) = \{[F \rightarrow (E) \bullet, \{\epsilon, +, *\}]\}
 \end{aligned}$$

136/160

Der kanonische LR(1)-Automat

275

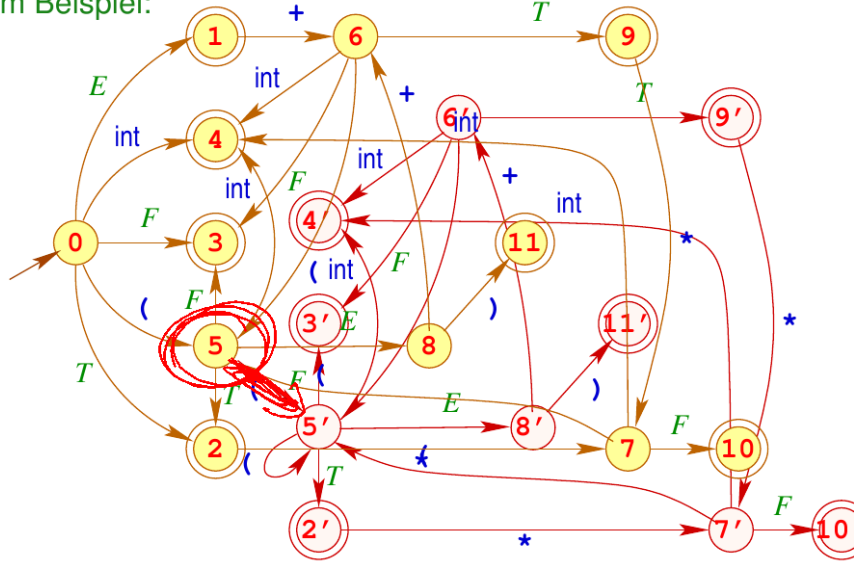
im Beispiel:



138/160

Der kanonische LR(1)-Automat

im Beispiel:



Der kanonische LR(1)-Automat

Diskussion:

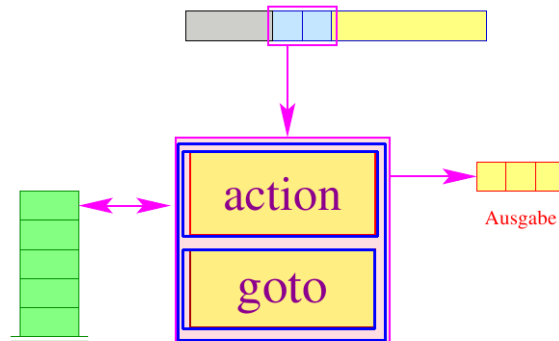
- Im Beispiel hat sich die Anzahl der Zustände fast verdoppelt ... und es kann noch schlimmer kommen
- Die Konflikte in den Zuständen q_1, q_2, q_9 sind nun aufgelöst ! z.B. haben wir für:

$$q_9 = \{ [E \rightarrow E + T \bullet, \{\epsilon, +\}], [T \rightarrow T \bullet * F, \{\epsilon, +, *\}] \}$$

mit:

$$\{\epsilon, +\} \cap (\text{First}_1(*F) \odot \{\epsilon, +, *\}) = \{\epsilon, +\} \cap \{*\} = \emptyset$$

Der LR(1)-Parser:



- Die **goto**-Tabelle kodiert die Zustandsübergänge:

$$\text{goto}[q, X] = \delta(q, X) \in Q$$
- Die **action**-Tabelle beschreibt für jeden Zustand q und möglichen Look-ahead w die erforderliche Aktion.

Der LR(1)-Parser:

Mögliche Aktionen sind:

- shift** // Shift-Operation
- reduce** ($A \rightarrow \gamma$) // Reduktion mit Ausgabe
- error** // Fehler

... im Beispiel:

- $E \rightarrow E + T^0 \mid T^1$
- $T \rightarrow T * F^0 \mid F^1$
- $F \rightarrow (E)^0 \mid \text{int}^1$

action	ϵ	int	()	+	*
q_1	S, 0					S
q_2	E, 1					S
q'_2			E, 1			S
q_3	T, 1				T, 1	T, 1
q'_3			T, 1		T, 1	T, 1
q_4	F, 1				F, 1	F, 1
q'_4			F, 1		F, 1	F, 1
q_9	E, 0				E, 0	S
q'_9			E, 0		E, 0	S
q_{10}	T, 0				T, 0	T, 0
q'_{10}			T, 0		T, 0	T, 0
q_{11}	F, 0				F, 0	F, 0
q'_{11}			F, 0		F, 0	F, 0

Der kanonische LR(1)-Automat

280

Allgemein: Wir identifizieren zwei Konflikte:

Reduce-Reduce-Konflikt:

$[A \rightarrow \gamma \bullet x], [A' \rightarrow \gamma' \bullet x] \in q$ mit $A \neq A' \vee \gamma \neq \gamma'$

Shift-Reduce-Konflikt:

$[A \rightarrow \gamma \bullet, x], [A' \rightarrow \alpha \bullet a\beta, y] \in q$
mit $a \in T$ und $x \in \{a\}$.

für einen Zustand $q \in Q$.

Solche Zustände nennen wir jetzt **LR(1)-ungeeignet**

142/160

Spezielle LR(k)-Teilklassen

282

Satz:

Eine reduzierte kontextfreie Grammatik G ist genau dann **LR(k)** wenn der kanonische **LR(k)**-Automat $LR(G, k)$ keine **LR(k)**-ungeeigneten Zustände besitzt.

143/160

Spezielle LR(k)-Teilklassen

284

Satz:

Eine reduzierte kontextfreie Grammatik G ist genau dann **LR(k)** wenn der kanonische **LR(k)**-Automat $LR(G, k)$ keine **LR(k)**-ungeeigneten Zustände besitzt.

Diskussion:

- Unser Beispiel ist offenbar **LR(1)**
- Im Allgemeinen hat der kanonische **LR(k)**-Automat sehr viel mehr Zustände als $LR(G) = LR(G, 0)$
- Man betrachtet darum i.a. **Teilklassen** von **LR(k)**-Grammatiken, bei denen man nur $LR(G)$ benutzt ...
- Zur Konflikt-Auflösung ordnet man den Items in den Zuständen Vorausschau-Mengen zu:
 - 1 Die Zuordnung ist unabhängig vom Zustand \implies **Simple LR(k)**
 - 2 Die Zuordnung hängt vom Zustand ab \implies **LALR(k)**

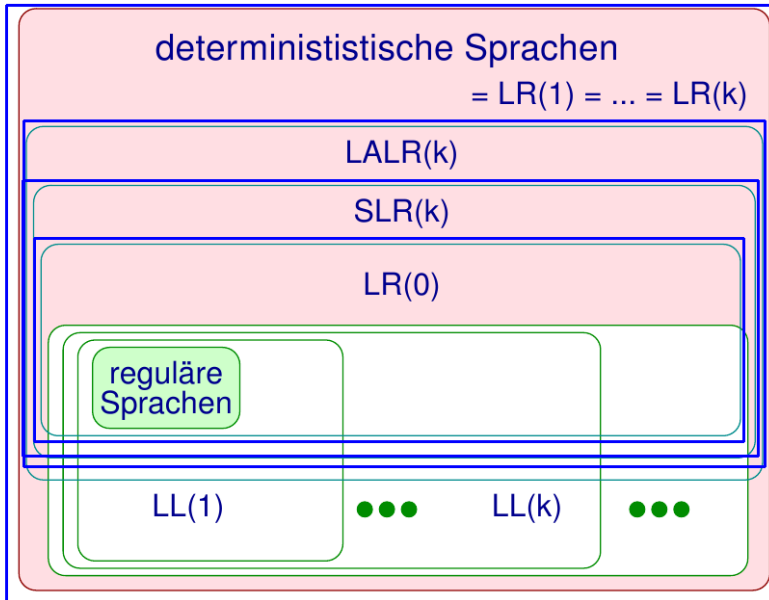
143/160

Syntaktische Analyse

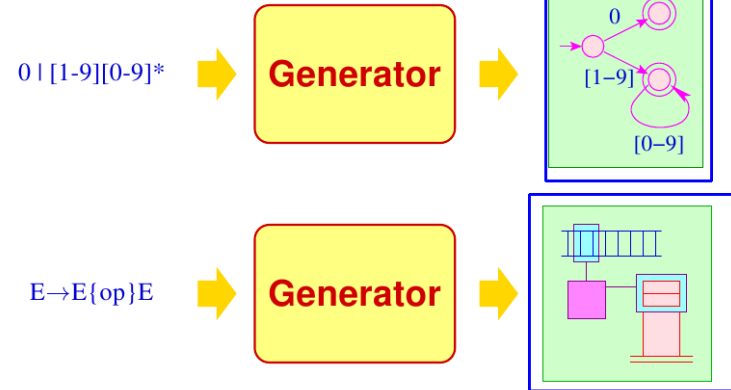
285

Kapitel 5: Zusammenfassung

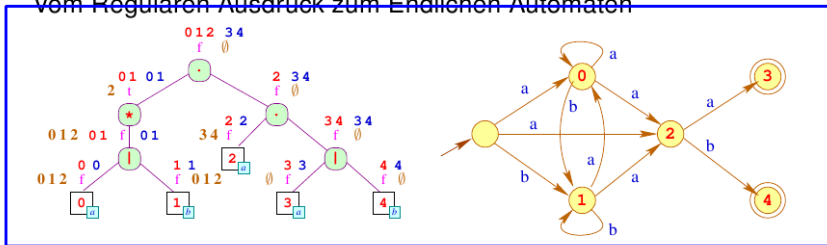
144/160



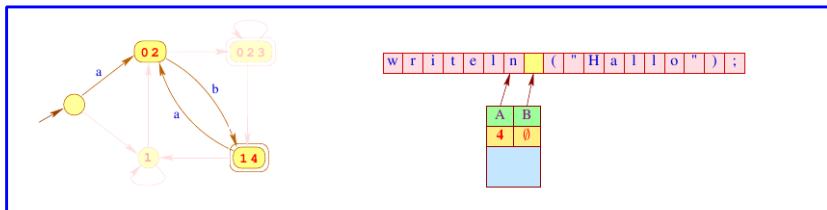
Prinzip von Spezifikation und Realisierung:



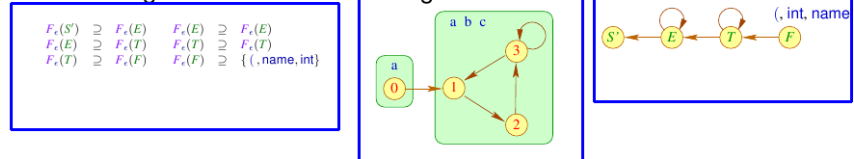
Vom Regulären Ausdruck zum Endlichen Automaten



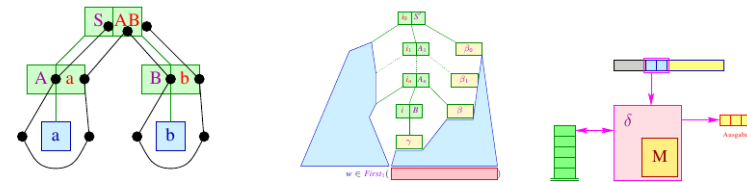
Vom Endlichen Automaten zum Scanner

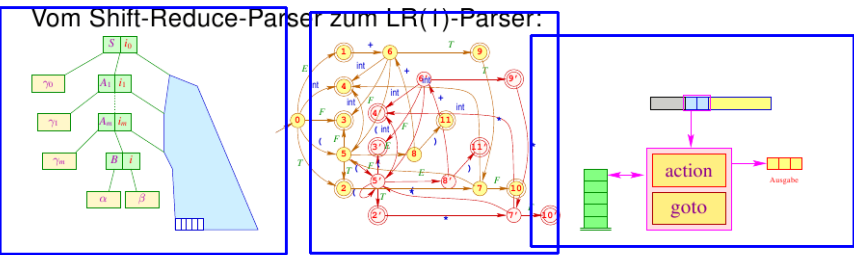
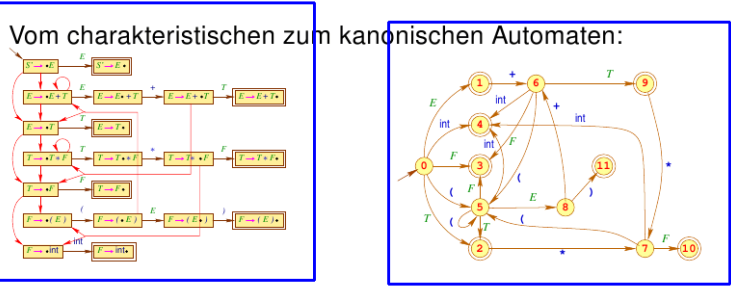


Berechnung von Vorausschau-Mengen:



Vom Item-Kellerautomat zum LL(1)-Parser:





Themengebiet: Die semantische Analyse