Script generated by TTT

Title: Simon: Compilerbau (03.06.2013)

Date: Mon Jun 03 14:19:02 CEST 2013

Duration: 32:39 min

Pages: 34

$S' \to \bullet E, \varepsilon$

The canonical LR(1)-automaton

The canonical $LR(1)$-automaton $LR(G,1)$ is created from $c(G,1)$, by performing arbitrarily many ϵ-transitions and then making the resulting automaton deterministic ...
But again, it can be constructed directly from the grammar
Analoguously to $LR(0)$, we need a helper function:

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

Then, we define:

States: Sets of $LR(1)$-items;

Start state: $\delta^*_\epsilon\, \{[S' \to \bullet\, S, \epsilon]\}$

Final states: $\{q \mid \exists A \to \alpha \in P : [A \to \alpha \bullet, x] \in q\}$

Transitions: $\delta(q, X) = \delta^*_\epsilon\, \{[A \to \alpha\, X \bullet \beta, x] \mid [A \to \alpha \bullet X\, \beta, x] \in q\}$

The canonical LR(1)-automaton

The canonical $LR(1)$-automaton $LR(G,1)$ is created from $c(G,1)$, by performing arbitrarily many ϵ-transitions and then making the resulting automaton deterministic ...
But again, it can be constructed directly from the grammar
Analoguously to $LR(0)$, we need a helper function:

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

Then, we define:

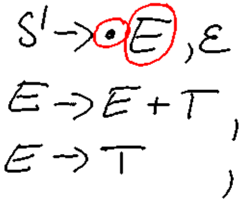
States: Sets of $LR(1)$-items;

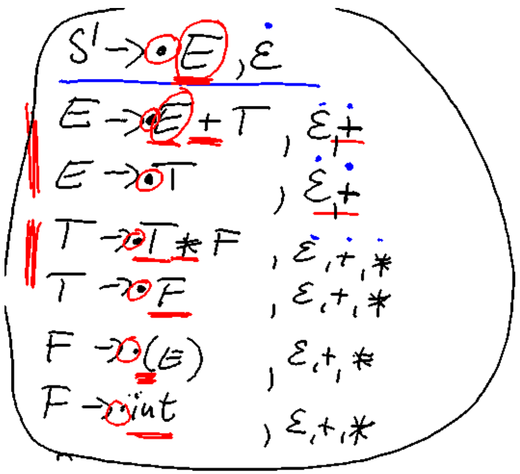
Start state: $\delta^*_\epsilon\, \{[S' \to \bullet\, S, \epsilon]\}$

Final states: $\{q \mid \exists A \to \alpha \in P : [A \to \alpha \bullet, x] \in q\}$

Transitions: $\delta(q, X) = \delta^*_\epsilon\, \{[A \to \alpha\, X \bullet \beta, x] \mid [A \to \alpha \bullet X\, \beta, x] \in q\}$

134 / 150





The canonical LR(1)-automaton

The canonical $LR(1)$-automaton $LR(G,1)$ is created from $c(G,1)$, by performing arbitrarily many ϵ-transitions and then making the resulting automaton deterministic ...
But again, it can be constructed directly from the grammar
Analoguously to $LR(0)$, we need a helper function:

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

Then, we define:

States: Sets of $LR(1)$-items;

Start state: $\delta_\epsilon^* \{[S' \to \bullet\, S,\, \epsilon]\}$

Final states: $\{q \mid \exists A \to \alpha \in P : \; [A \to \alpha \bullet,\, x] \in q\}$

Transitions: $\delta(q, X) = \delta_\epsilon^* \{[A \to \alpha\, X \bullet\, \beta,\, x] \mid [A \to \alpha \bullet X\, \beta,\, x] \in q\}$

The canonical LR(1)-automaton

The canonical $LR(1)$-automaton $LR(G,1)$ is created from $c(G,1)$, by performing arbitrarily many ϵ-transitions and then making the resulting automaton deterministic ...
But again, it can be constructed directly from the grammar
Analoguously to $LR(0)$, we need a helper function:

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

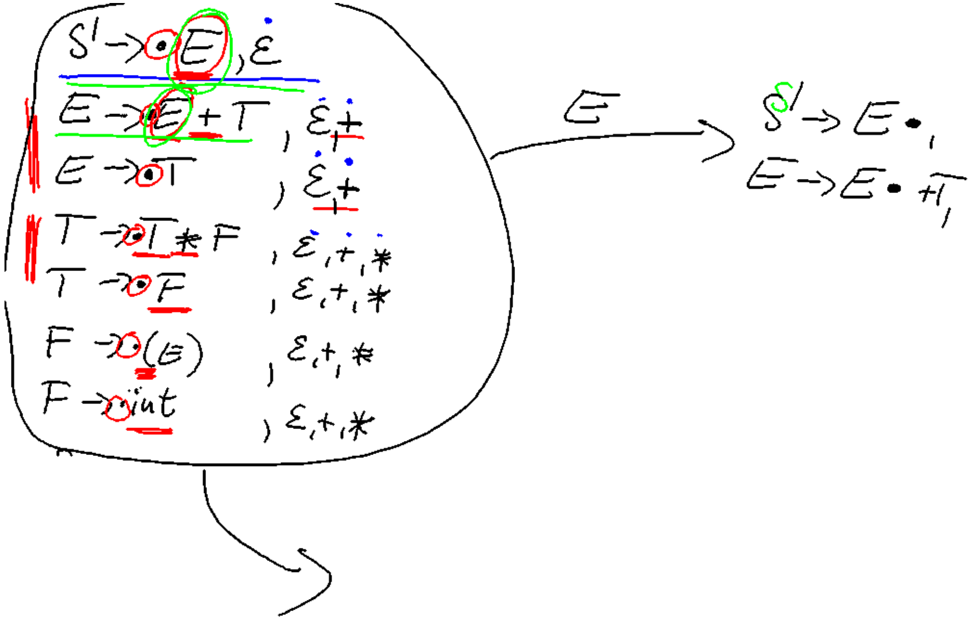
Then, we define:

States: Sets of $LR(1)$-items;

Start state: $\delta^*_\epsilon\,\{[S'\to\bullet\,S,\epsilon]\}$

Final states: $\{q \mid \exists A\to\alpha\in P: \ [A\to\alpha\,\bullet, x]\in q\}$

Transitions: $\delta(q,X) = \delta^*_\epsilon\,\{[A\to\alpha\,X\bullet\,\beta, x] \mid [A\to\alpha\bullet X\,\beta, x]\in q\}$



The canonical LR(1)-automaton

The canonical $LR(1)$-automaton $LR(G,1)$ is created from $c(G,1)$, by performing arbitrarily many ϵ-transitions and then making the resulting automaton deterministic ...
But again, it can be constructed directly from the grammar
Analoguously to $LR(0)$, we need a helper function:

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

Then, we define:

States: Sets of $LR(1)$-items;

Start state: $\delta^*_\epsilon\,\{[S'\to\bullet\,S,\epsilon]\}$

Final states: $\{q \mid \exists A\to\alpha\in P: \ [A\to\alpha\,\bullet, x]\in q\}$

Transitions: $\delta(q,X) = \delta^*_\epsilon\,\{[A\to\alpha\,X\bullet\,\beta, x] \mid [A\to\alpha\bullet X\,\beta, x]\in q\}$

134 / 150

The Canonical LR(1)-Automaton

for example:

$$\begin{array}{lllll}
q_0 & = & & & \{[S' \to \bullet E,\ \{\epsilon\}], \\
& & & & [E \to \bullet E+T,\ \{\epsilon,+\}], \\
& & & & [E \to \bullet T,\ \{\epsilon,+\}], \\
& & & & [T \to \bullet T*F,\ \{\epsilon,+,*\}], \\
& & & & [T \to \bullet F,\ \{\epsilon,+,*\}], \\
& & & & [F \to \bullet (E),\ \{\epsilon,+,*\}], \\
& & & & [F \to \bullet \text{int},\ \{\epsilon,+,*\}]\} \\
q_1 & = & \delta(q_0,E) & = & \{[S' \to E\bullet,\ \{\epsilon\}], \\
& & & & [E \to E\bullet +T,\ \{\epsilon,+\}]\} \\
q_2 & = & \delta(q_0,T) & = & \{[E \to T\bullet,\ \{\epsilon,+\}], \\
& & & & [T \to T\bullet *F,\ \{\epsilon,+,*\}]\}
\end{array}$$

$$\begin{array}{lllll}
q_3 & = & \delta(q_0,F) & = & \{[T \to F\bullet,\ \{\epsilon,+,*\}]\} \\
q_4 & = & \delta(q_0,\text{int}) & & \{[F \to \text{int}\bullet,\ \{\epsilon,+,*\}]\} \\
q_5 & = & \delta(q_0,\ (\ ) & = & \{[F \to (\bullet E),\ \{\epsilon,+,*\}], \\
& & & & [E \to \bullet E+T,\ \{),+\}], \\
& & & & [E \to \bullet T,\ \{),+\}], \\
& & & & [T \to \bullet T*F,\ \{),+,*\}], \\
& & & & [T \to \bullet F,\ \{),+,*\}], \\
& & & & [F \to \bullet (E),\ \{),+,*\}], \\
& & & & [F \to \bullet \text{int},\ \{),+,*\}]\}
\end{array}$$

The Canonical LR(1)-Automaton

for example:

$$\begin{array}{lllll}
q_0 & = & & & \{[S' \to \bullet E,\ \{\epsilon\}], \\
& & & & [E \to \bullet E+T,\ \{\epsilon,+\}], \\
& & & & [E \to \bullet T,\ \{\epsilon,+\}], \\
& & & & [T \to \bullet T*F,\ \{\epsilon,+,*\}], \\
& & & & [T \to \bullet F,\ \{\epsilon,+,*\}], \\
& & & & [F \to \bullet (E),\ \{\epsilon,+,*\}], \\
& & & & [F \to \bullet \text{int},\ \{\epsilon,+,*\}]\} \\
q_1 & = & \delta(q_0,E) & = & \{[S' \to E\bullet,\ \{\epsilon\}], \\
& & & & [E \to E\bullet +T,\ \{\epsilon,+\}]\} \\
q_2 & = & \delta(q_0,T) & = & \{[E \to T\bullet,\ \{\epsilon,+\}], \\
& & & & [T \to T\bullet *F,\ \{\epsilon,+,*\}]\}
\end{array}$$

$$\begin{array}{lllll}
q_3 & = & \delta(q_0,F) & = & \{[T \to F\bullet,\ \{\epsilon,+,*\}]\} \\
q_4 & = & \delta(q_0,\text{int}) & & \{[F \to \text{int}\bullet,\ \{\epsilon,+,*\}]\} \\
q_5 & = & \delta(q_0,\ (\ ) & = & \{[F \to (\bullet E),\ \{\epsilon,+,*\}], \\
& & & & [E \to \bullet E+T,\ \{),+\}], \\
& & & & [E \to \bullet T,\ \{),+\}], \\
& & & & [T \to \bullet T*F,\ \{),+,*\}], \\
& & & & [T \to \bullet F,\ \{),+,*\}], \\
& & & & [F \to \bullet (E),\ \{),+,*\}], \\
& & & & [F \to \bullet \text{int},\ \{),+,*\}]\}
\end{array}$$

135 / 150

The canonical LR(1)-Automaton

for example:

$$\begin{array}{lllll}
q_5' & = & \delta(q_5,\ (\ ) & = & \{[F \to (\bullet E)\ ], \\
& & & & [E \to \bullet E+T\ ], \\
& & & & [E \to \bullet T\ ], \\
& & & & [T \to \bullet T*F\ ], \\
& & & & [T \to \bullet F\ ], \\
& & & & [F \to \bullet (E)\ ], \\
& & & & [F \to \bullet \text{int}\ ]\} \\
q_6 & = & \delta(q_1,+) & = & \{[E \to E+\bullet T\ ], \\
& & & & [T \to \bullet T*F\ ], \\
& & & & [T \to \bullet F\ ], \\
& & & & [F \to \bullet (E)\ ], \\
& & & & [F \to \bullet \text{int}\ ]\}
\end{array}$$

$$\begin{array}{lllll}
q_7 & = & \delta(q_2,*) & = & \{[T \to T*\bullet F\ ], \\
& & & & [F \to \bullet (E)\ ], \\
& & & & [F \to \bullet \text{int}\ ]\} \\
q_8 & = & \delta(q_5,E) & = & \{[F \to (E\bullet)\ ]\} \\
& & & & [E \to E\bullet +T\ ]\} \\
q_9 & = & \delta(q_6,T) & = & \{[E \to E+T\bullet\ ], \\
& & & & [T \to T\bullet *F\ ]\} \\
q_{10} & = & \delta(q_7,F) & = & \{[T \to T*F\bullet\ ]\} \\
q_{11} & = & \delta(q_8,\ )\ ) & = & \{[F \to (E)\bullet\ ]\}
\end{array}$$

The canonical LR(1)-Automaton

for example:

$$\begin{array}{lllll}
q_5' & = & \delta(q_5,\ (\ ) & = & \{[F \to (\bullet E)\ ], \\
& & & & [E \to \bullet E+T\ ], \\
& & & & [E \to \bullet T\ ], \\
& & & & [T \to \bullet T*F\ ], \\
& & & & [T \to \bullet F\ ], \\
& & & & [F \to \bullet (E)\ ], \\
& & & & [F \to \bullet \text{int}\ ]\} \\
q_6 & = & \delta(q_1,+) & = & \{[E \to E+\bullet T\ ], \\
& & & & [T \to \bullet T*F\ ], \\
& & & & [T \to \bullet F\ ], \\
& & & & [F \to \bullet (E)\ ], \\
& & & & [F \to \bullet \text{int}\ ]\}
\end{array}$$

$$\begin{array}{lllll}
q_7 & = & \delta(q_2,*) & = & \{[T \to T*\bullet F\ ], \\
& & & & [F \to \bullet (E)\ ], \\
& & & & [F \to \bullet \text{int}\ ]\} \\
q_8 & = & \delta(q_5,E) & = & \{[F \to (E\bullet)\ ]\} \\
& & & & [E \to E\bullet +T\ ]\} \\
q_9 & = & \delta(q_6,T) & = & \{[E \to E+T\bullet\ ], \\
& & & & [T \to T\bullet *F\ ]\} \\
q_{10} & = & \delta(q_7,F) & = & \{[T \to T*F\bullet\ ]\} \\
q_{11} & = & \delta(q_8,\ )\ ) & = & \{[F \to (E)\bullet\ ]\}
\end{array}$$

136 / 150

The canonical LR(1)-Automaton

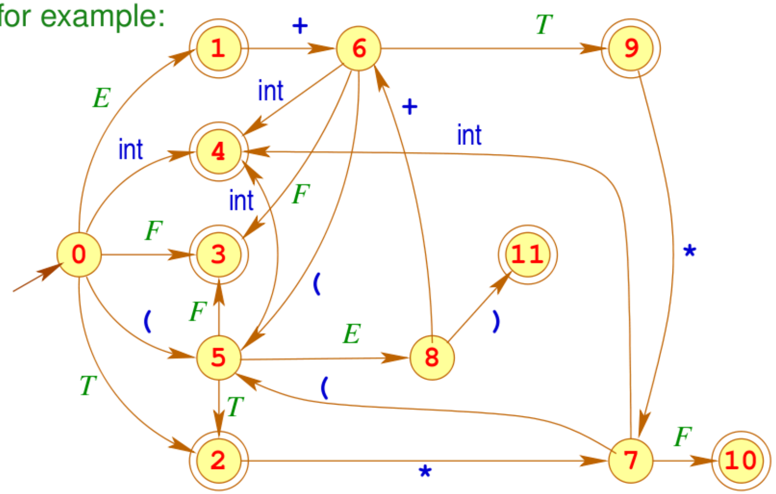
for example:

$$
\begin{array}{lllll}
q_5' & = & \delta(q_5, (\,) & = & \{[F \to (\bullet E\,),\ \{\,)\,,+,*\}],\\
 & & & & [E \to \bullet E+T,\ \{\,)\,,+\}],\\
 & & & & [E \to \bullet T,\ \{\,)\,,+\}],\\
 & & & & [T \to \bullet T*F,\ \{\,)\,,+,*\}],\\
 & & & & [T \to \bullet F,\ \{\,)\,,+,*\}],\\
 & & & & [F \to \bullet (E\,),\ \{\,)\,,+,*\}],\\
 & & & & [F \to \bullet \text{int},\ \{\,)\,,+,*\}]\}\\
q_6 & = & \delta(q_1, +) & = & \{[E \to E+\bullet T,\ \{\epsilon,+\}],\\
 & & & & [T \to \bullet T*F,\ \{\epsilon,+,*\}],\\
 & & & & [T \to \bullet F,\ \{\epsilon,+,*\}],\\
 & & & & [F \to \bullet (E\,),\ \{\epsilon,+,*\}],\\
 & & & & [F \to \bullet \text{int},\ \{\epsilon,+,*\}]\}
\end{array}
$$

$$
\begin{array}{lllll}
q_7 & = & \delta(q_2, *) & = & \{[T \to T*\bullet F \quad],\\
 & & & & [F \to \bullet (E\,) \quad],\\
 & & & & [F \to \bullet \text{int} \quad]\}\\
q_8 & = & \delta(q_5, E) & = & \{[F \to (E\bullet) \quad]\}\\
 & & & & [E \to E\bullet + T \quad]\}\\
q_9 & = & \delta(q_6, T) & = & \{[E \to E+T\bullet \quad],\\
 & & & & [T \to T\bullet * F \quad]\}\\
q_{10} & = & \delta(q_7, F) & = & \{[T \to T*F\bullet \quad]\}\\
q_{11} & = & \delta(q_8,)\,) & = & \{[F \to (E\,)\bullet \quad]\}
\end{array}
$$

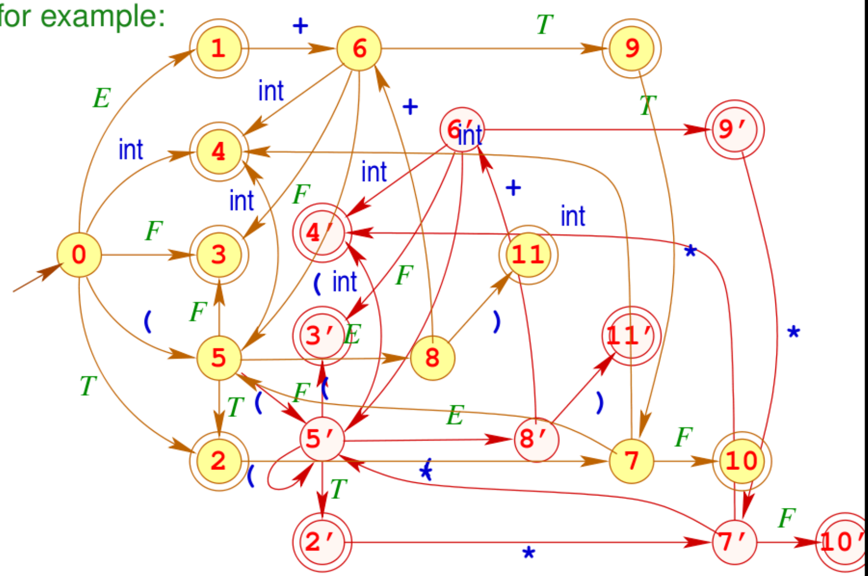
Thee canonical LR(1)-Automaton

for example:



Thee canonical LR(1)-Automaton

for example:



The canonical LR(1)-Automaton

Discussion:

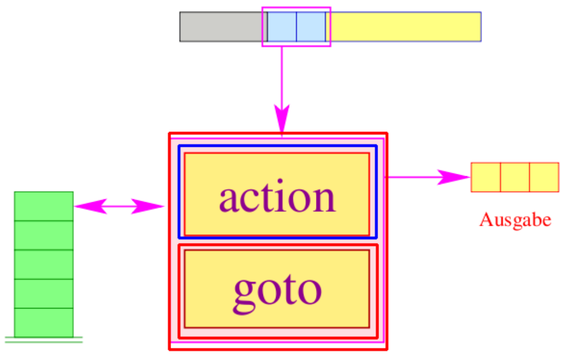
- In the example, the number of states was almost doubled
  ... and it can become even worse
- The conflicts in states q_1, q_2, q_9 are now resolved !
  e.g. we have for:

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

with:

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

The LR(1)-Parser:



- The goto-table encodes the transitions:

$$\text{goto}[q, X] = \delta(q, X) \quad \in \quad Q$$

- The action-table describes for every state q and possible lookahead w the necessary action.

The LR(1)-Parser:

Possible actions are:

shift	//	Shift-operation
reduce $(A \to \gamma)$	//	Reduction with callback/output
error	//	Error

... for example:

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

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}'				$F, 0$	$F, 0$	$F, 0$

The Canonical LR(1)-Automat

In general: We identify two conflicts:

Reduce-Reduce-Conflict:

$[A \to \gamma \bullet, x]\,,\ [A' \to \gamma' \bullet, x] \ \in \ q$ with $A \neq A' \vee \gamma \neq \gamma'$

Shift-Reduce-Conflict:

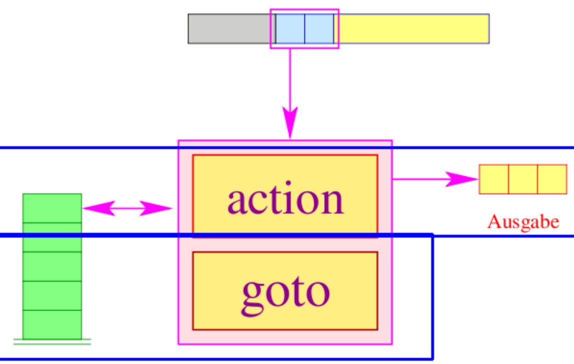
$[A \to \gamma \bullet, x]\,,\ [A' \to \alpha \bullet a\beta, y] \ \in \ q$

with $a \in T$ und $x \in \{a\}$.

for a state $q \in Q$.

Such states are now called $LR(1)$-unsuited

The LR(1)-Parser:



- The goto-table encodes the transitions:

$$\text{goto}[q, X] = \delta(q, X) \quad \in \quad Q$$

- The action-table describes for every state q and possible lookahead w the necessary action.

The Canonical LR(1)-Automat

In general: We identify two conflicts:

Reduce-Reduce-Conflict:

$[A \to \gamma \bullet, x]\,,\ [A' \to \gamma' \bullet, x] \ \in\ q$ with $A \neq A' \vee \gamma \neq \gamma'$

Shift-Reduce-Conflict:

$[A \to \gamma \bullet, x]\,,\ [A' \to \alpha \bullet a\,\beta, y] \ \in\ q$

with $a \in T$ und $x \in \{a\} \odot_k \mathsf{First}_k(\beta) \odot_k \{y\}$.

for a state $q \in Q$.

Such states are now called $LR(k)$-unsuited

Special LR(k)-Subclasses

Theorem:

A reduced contextfree grammar G is called $LR(k)$ iff the canonical $LR(k)$-automaton $LR(G, k)$ has no $LR(k)$-unsuited states.

Special LR(k)-Subclasses

Theorem:

A reduced contextfree grammar G is called $LR(k)$ iff the canonical $LR(k)$-automaton $LR(G, k)$ has no $LR(k)$-unsuited states.

Discussion:

- Our example apparently is $LR(1)$
- In general, the canonical $LR(k)$-automaton has much more states then $LR(G) = LR(G, 0)$
- Therefore in practice, subclasses of $LR(k)$-grammars are often considered, which only use $LR(G)$ ...

Special LR(k)-Subclasses

Theorem:

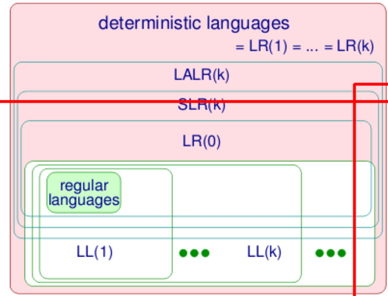
A reduced contextfree grammar G is called $LR(k)$ iff the canonical $LR(k)$-automaton $LR(G, k)$ has no $LR(k)$-unsuited states.

Discussion:

- Our example apparently is $LR(1)$
- In general, the canonical $LR(k)$-automaton has much more states then $LR(G) = LR(G, 0)$
- Therefore in practice, subclasses of $LR(k)$-grammars are often considered, which only use $LR(G)$ ...
- For resolving conflicts, the items are assigned special lookahead-sets:
  1. independently on the state itself $\Longrightarrow$ Simple $LR(k)$
  2. dependent on the state itself $\Longrightarrow$ $LALR(k)$

Chapter 5:
Summary

Parsing Methods

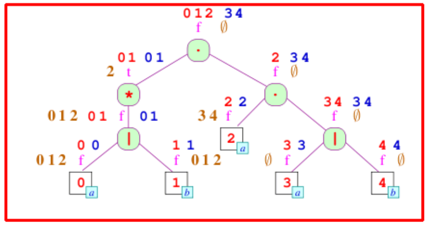


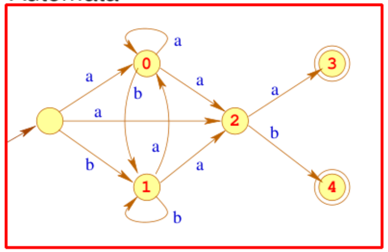
Discussion:

- All contextfree languages, that can be parsed with a deterministic pushdown automaton, can be characterized with an LR(1)-grammar.
- LR(0)-grammars describe all prefixfree deterministic contextfree languages
- The language-classes of LL(k)-grammars form a hierarchy within the deterministic contextfree languages.

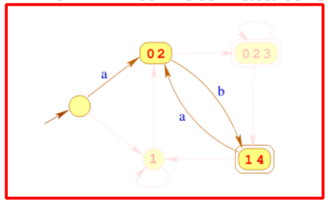
Lexical and Syntactical Analysis:

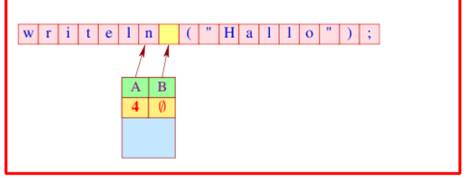
From Regular Expressions to Finite Automata





From Finite Automata to Scanners

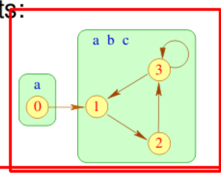


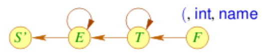


Lexical and Syntactical Analysis:

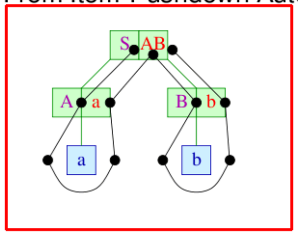
Computation of lookahead sets:

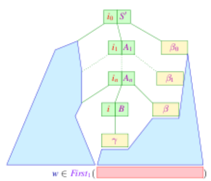
$$F_\epsilon(S') \supseteq F_\epsilon(E) \quad F_\epsilon(E) \supseteq F_\epsilon(E)$$
$$F_\epsilon(E) \supseteq F_\epsilon(T) \quad F_\epsilon(T) \supseteq F_\epsilon(T)$$
$$F_\epsilon(T) \supseteq F_\epsilon(F) \quad F_\epsilon(F) \supseteq \{(,\text{name},\text{int}\}$$

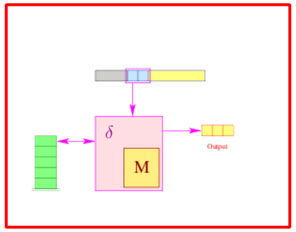




From Item-Pushdown Automata to LL(1)-Parsers:

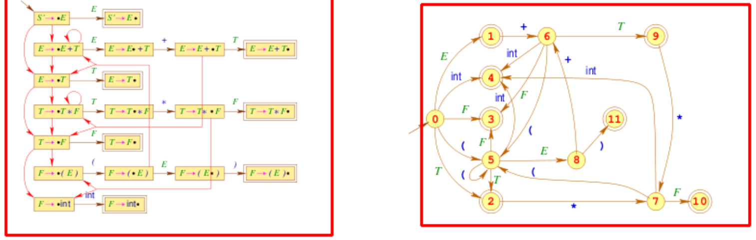




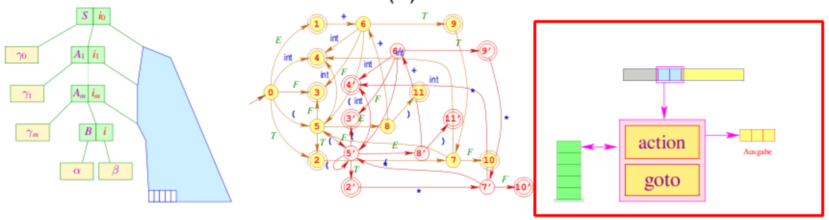


Lexical and syntactical Analysis:

From characteristic to canonical Automata:



From Shift-Reduce-Parsers to LR(1)-Parsers:



Ende der Präsentation. Klicken Sie zum Schließen.