Script generated by TTT

Title: Petter: Compilerbau (08.06.2015)

Date: Mon Jun 08 14:20:58 CEST 2015

Duration: 94:57 min

Pages: 51

LR(0)-Parser

Attention:
Unfortunately, the $LR(0)$-parser is in general non-deterministic.

We identify two reasons:

Reduce-Reduce-Conflict:

$$[A \to \gamma \bullet], \; [A' \to \gamma' \bullet] \quad \in \quad q \quad \text{with} \quad A \neq A' \vee \gamma \neq \gamma'$$

Shift-Reduce-Conflict:

$$\boxed{[A \to \gamma \bullet]}, \; \boxed{[A' \to \alpha \bullet a\, \beta]} \quad \in \quad q \quad \text{with} \quad \boxed{a \in T}$$

for a state $q \in Q$.

Those states are called $\boxed{LR(0)\text{-unsuited.}}$

Revisiting the Conflicts of the LR(0)-Automaton

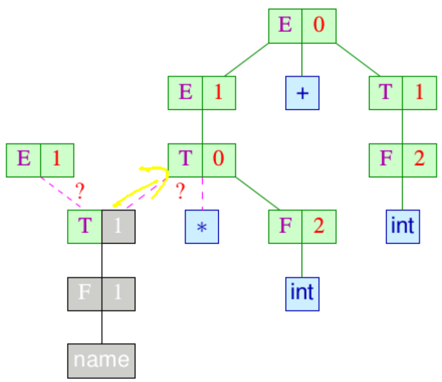
What differenciates the particular Reductions and Shifts?

Input:

$* \, 2 + 40$

Pushdown:

$(\, q_0 \; T \,)$



$$\begin{array}{lcll} E & \to & E + T & \mid \; T \\ T & \to & T * F & \mid \; F \\ F & \to & (\, E \,) & \mid \; \text{int} \end{array}$$

Revisiting the Conflicts of the LR(0)-Automaton

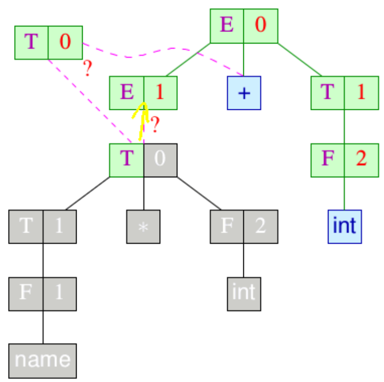
What differenciates the particular Reductions and Shifts?

Input:

$+ \, 40$

Pushdown:

$(\, q_0 \; T \,)$



$$\begin{array}{lcll} E & \to & E + T & \mid \; T \\ T & \to & T * F & \mid \; F \\ F & \to & (\, E \,) & \mid \; \text{int} \end{array}$$

Revisiting the Conflicts of the LR(0)-Automaton

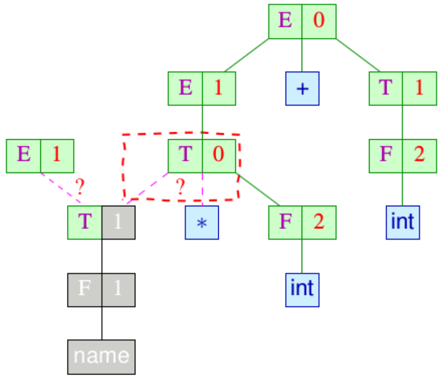
Idea: Matching lookahead with *right context* matters!

Input:

$\boxed{*}\, 2 + 40$

Pushdown:

$(\, q_0\, T\,)$



$$
\begin{array}{lcl}
E & \rightarrow & E + T \quad | \quad T \\
T & \rightarrow & \boxed{T *} F \quad | \quad F \\
F & \rightarrow & (\, E\,) \quad | \quad \text{int}
\end{array}
$$

Revisiting the Conflicts of the LR(0)-Automaton

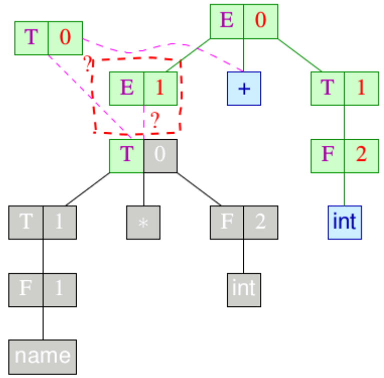
Idea: Matching lookahead with *right context* matters!

Input:

$\boxed{+}\, 40$

Pushdown:

$(\, q_0\, T\,)$



$$
\begin{array}{lcl}
\boxed{E} & \rightarrow & E \boxed{+} T \quad | \quad T \\
T & \rightarrow & T * F \quad | \quad F \\
F & \rightarrow & (\, E\,) \quad | \quad \text{int}
\end{array}
$$

LR(k)-Grammars

Idea: Consider k-lookahead in conflict situations.

Definition:

The reduced contextfree grammar G is called $LR(k)$-grammar, if for $\boxed{\text{First}_k(w)} = \text{First}_k(x)$ with:

$$
\left.\begin{array}{lcccl}
S & \rightarrow^*_{\boxed{R}} & \alpha\, A\, w & \rightarrow & \alpha\, \beta\, \boxed{w} \\
S & \rightarrow^*_R & \alpha'\, A'\, w' & \rightarrow & \alpha\, \beta\, \boxed{x}
\end{array}\right\} \text{follows: } \boxed{\alpha = \alpha'} \wedge \boxed{A = A'} \wedge w' = x
$$

LR(k)-Grammars

Idea: Consider k-lookahead in conflict situations.

Definition:

The reduced contextfree grammar G is called $LR(k)$-grammar, if for $\text{First}_k(w) = \text{First}_k(x)$ with:

$$
\left.\begin{array}{lcccl}
S & \rightarrow^*_R & \alpha\, A\, w & \rightarrow & \alpha\, \beta\, w \\
S & \rightarrow^*_R & \alpha'\, A'\, w' & \rightarrow & \alpha\, \beta\, x
\end{array}\right\} \text{follows: } \alpha = \alpha' \wedge A = A' \wedge w' = x
$$

Strategy for testing Grammars for $LR(k)$-property

1. Focus iteratively on all rightmost derivations $\boxed{S \rightarrow^*_R \alpha\, X\, w \rightarrow \alpha\, \beta\, w}$
2. Identify handle $\alpha\, \beta$ in sentence forms $\alpha\, \beta\, w$
3. Determine minimal k, such that $\text{First}_k(w)$ associates β with a unique $X \rightarrow \beta$ for non-prefixfree $\alpha\, \beta$s

LR(k)-Grammars

for example:

(1) $S \to A \mid B \qquad \boxed{A \to a\,A\,b} \mid 0 \qquad B \to a\,B\,b\,b \mid 1$

LR(k)-Grammars

for example:

(1) $S \to A \mid B \qquad A \to a\,A\,b \mid 0 \qquad B \to a\,B\,b\,b \mid 1$

... is not $LL(k)$ for any k :

Let $S \to^*_R \alpha\,X\,w \to \boxed{\alpha\,\beta\,w}$. Then $\alpha\,\beta$ is of one of these forms:

A , B , $a\,A\,b$, 0 $a^n\,a\,B\,b^2\,b\,b^{2n}$

$a\,a\,A\,b\,b$

$a\,a\,a\,A\,b\,b\,b$

$a^n\,a\,A\,b\,b^n$

$a^n\,0\,b^n$

$a^n\,1\,b^{2n}$

LR(k)-Grammars

for example:

(1) $S \to A \mid B \qquad A \to a\,A\,b \mid 0 \qquad B \to a\,B\,b\,b \mid 1$

... is not $LL(k)$ for any k :

Let $S \to^*_R \alpha\,X\,w \to \alpha\,\beta\,w$. Then $\alpha\,\beta$ is of one of these forms:

$$\underline{A}\,,\ \underline{B}\,,\ a^n\,\underline{a\,A\,b}\,,\ a^n\,\underline{a\,B\,b\,b}\,,\ a^n\,\underline{0}\,,\ a^n\,\underline{1} \qquad (n \geq 0)$$

LR(k)-Grammars

for example:

(1) $S \to A \mid B \qquad A \to a\,A\,b \mid 0 \qquad B \to a\,B\,b\,b \mid 1$

... is not $LL(k)$ for any k — but $LR(0)$:

Let $S \to^*_R \alpha\,X\,w \to \alpha\,\boxed{\beta}\,w$. Then $\alpha\,\beta$ is of one of these forms:

$$\underline{A}\,,\ \underline{B}\,,\ a^n\,\underline{a\,A\,b}\,,\ a^n\,\underline{a\,B\,b\,b}\,,\ a^n\,\underline{0}\,,\ a^n\,\underline{1} \qquad (n \geq 0)$$

LR(k)-Grammars

for example:

(1) $S \to A \mid B \qquad A \to a\,A\,b \mid 0 \qquad B \to a\,B\,b\,b \mid 1$

... is not $LL(k)$ for any k — but $LR(0)$:

Let $S \to_R^* \alpha\,X\,w \to \alpha\,\beta\,w$. Then $\alpha\,\underline{\beta}$ is of one of these forms:

$$\underline{A}\,,\ \underline{B}\,,\ a^n\,\underline{a\,A\,b}\,,\ a^n\,\underline{a\,B\,b\,b}\,,\ a^n\,\underline{0}\,,\ a^n\,\underline{1} \qquad (n \geq 0)$$

(2) $S \to a\,A\,c \qquad \boxed{A \to A\,b\,b} \mid b$

LR(k)-Grammars

for example:

(1) $S \to A \mid B \qquad A \to a\,A\,b \mid 0 \qquad B \to a\,B\,b\,b \mid 1$

... is not $LL(k)$ for any k — but $LR(0)$:

Let $S \to_R^* \alpha\,X\,w \to \alpha\,\beta\,w$. Then $\alpha\,\underline{\beta}$ is of one of these forms:

$$\underline{A}\,,\ \underline{B}\,,\ a^n\,\underline{a\,A\,b}\,,\ a^n\,\underline{a\,B\,b\,b}\,,\ a^n\,\underline{0}\,,\ a^n\,\underline{1} \qquad (n \geq 0)$$

(2) $S \to a\,A\,c \qquad A \to A\,b\,b \mid b$

... is also not $LL(k)$ for any k :

Let $S \to_R^* \alpha\,X\,w \to \alpha\,\beta\,w$. Then $\alpha\,\underline{\beta}$ is of one of these forms:

$$a\,\underline{b}\,,\ a\,\underline{A\,b\,b}\,,\ \underline{a\,A\,c}$$

LR(k)-Grammars

for example:

(3) $S \to a\,A\,c \qquad A \to b\,b\,A \mid b$

LR(k)-Grammars

for example:

(3) $S \to a\,A\,c \qquad A \to b\,b\,A \mid b$

Let $S \to_R^* \alpha\,X\,w \to \alpha\,\beta\,w$ with $\{y\} = \text{First}_k(w)$ then $\alpha\,\underline{\beta}\,y$ is of one of these forms:

$$a\,b^{2n}\,\underline{b}\,c\,,\ a\,b^{2n}\,\underline{b\,b\,A}\,c\,,\ \underline{a\,A\,c}$$

LR(k)-Grammars

for example:

(3) $S \to a\,A\,c \qquad A \to b\,b\,A \mid b$ ... is not $LR(0)$, but $LR(1)$:

Let $S \to_R^* \alpha\,X\,w \to \alpha\,\beta\,w$ with $\{y\} = \text{First}_k(w)$ then $\alpha\,\underline{\beta}\,y$ is of one of these forms:

$$a\,b^{2n}\,\underline{b}\,c\,,\quad a\,b^{2n}\,\underline{b\,b\,A}\,c\,,\quad \underline{a\,A\,c}$$

LR(k)-Grammars

for example:

(3) $S \to a\,A\,c \qquad A \to b\,b\,A \mid b$

Let $S \to_R^* \alpha\,X\,w \to \alpha\,\beta\,w$ with $\{y\} = \text{First}_k(w)$ then $\alpha\,\underline{\beta}\,y$ is of one of these forms:

$$a\,b^{2n}\,\underline{b}\,c\,,\quad a\,b^{2n}\,\underline{b\,b\,A}\,c\,,\quad \underline{a\,A\,c}$$

LR(k)-Grammars

for example:

(3) $S \to a\,A\,c \qquad A \to b\,b\,A \mid b$ ... is not $LR(0)$, but $LR(1)$:

Let $S \to_R^* \alpha\,X\,w \to \alpha\,\beta\,w$ with $\{y\} = \text{First}_k(w)$ then $\alpha\,\underline{\beta}\,y$ is of one of these forms:

$$a\,b^{2n}\,\underline{b}\,c\,,\quad a\,b^{2n}\,\underline{b\,b\,A}\,c\,,\quad \underline{a\,A\,c}$$

(4) $S \to a\,A\,c \qquad A \to b\,A\,b \mid b$

LR(k)-Grammars

for example:

(3) $S \to a\,A\,c \qquad A \to b\,b\,A \mid b$ ... is not $LR(0)$, but $LR(1)$:

Let $S \to_R^* \alpha\,X\,w \to \alpha\,\beta\,w$ with $\{y\} = \text{First}_k(w)$ then $\alpha\,\underline{\beta}\,y$ is of one of these forms:

$$a\,b^{2n}\,\underline{b}\,c\,,\quad a\,b^{2n}\,\underline{b\,b\,A}\,c\,,\quad \underline{a\,A\,c}$$

(4) $S \to a\,A\,c \qquad A \to b\,A\,b \mid b$

Consider the rightmost derivations:

$$S \to_R^* a\,b^n\,A\,b^n\,c \to \boxed{a\,b^n\,\underline{b}\,b^n\,c}$$

LR(k)-Grammars

for example:

(3) $S \to a\,A\,c \qquad A \to b\,b\,A \mid b$ ... is not $LR(0)$, but $LR(1)$:

Let $S \to_R^* \alpha\,X\,w \to \alpha\,\beta\,w$ with $\{y\} = \text{First}_k(w)$ then $\alpha\,\underline{\beta}\,y$ is of one of these forms:

$$a\,b^{2n}\,\underline{b\,c}\;,\quad a\,b^{2n}\,\underline{b\,b\,A}\,c\;,\quad \underline{a\,A\,c}$$

(4) $S \to a\,A\,c \qquad A \to b\,A\,b \mid b$ ... is not $LR(k)$ for any $k \geq 0$:

Consider the rightmost derivations:

$$S \to_R^* a\,b^n\,A\,b^n\,c \to a\,b^n\,\underline{b}\,b^n\,c$$

Revisiting the Conflicts of the LR(0)-Automaton

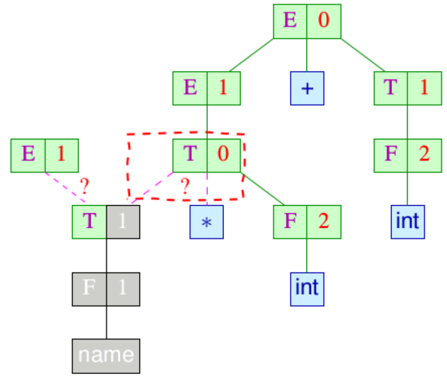
Idea: Matching lookahead with *right context* matters!

Input:

$*\,2 + 40$

Pushdown:

$(\,q_0\,T\,)$



$$\begin{array}{lcll} E & \to & E+T & \mid\; T \\ T & \to & T*F & \mid\; F \\ F & \to & (\,E\,) & \mid\; \text{int} \end{array}$$

LR(1)-Parsing

Idea: Let's equip items with 1-lookahead

Definition LR(1)-Item

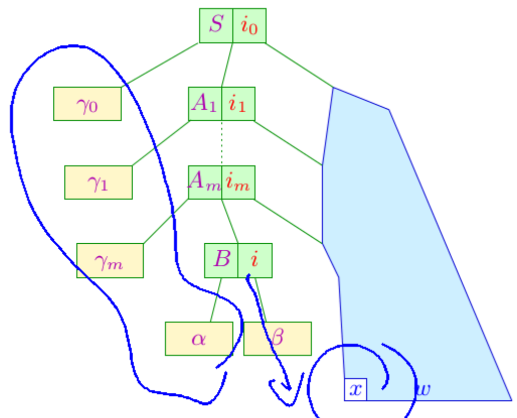
An $LR(1)$-item is a pair $[B \to \alpha \bullet \beta,\; x]$ with

$$x \in \text{Follow}_1(B) = \bigcup\{\text{First}_1(\nu) \mid S \to^* \mu\,B\,\nu\}$$

Admissible LR(1)-Items

The item $[B \to \alpha \bullet \beta,\; x]$ is *admissable* for $\gamma\,\alpha$ if:

$$S \to_R^* \gamma\,B\,w \qquad \text{with} \qquad \{x\} = \text{First}_1(w)$$



... with $\gamma_0 \ldots \gamma_m = \gamma$

The Characteristic LR(1)-Automaton

The set of admissible $LR(1)$-items for viable prefixes is again computed with the help of the finite automaton $c(G,1)$.

The automaton $c(G,1)$:

States: $LR(1)$-items

Start state: $[S' \to \bullet\, S,\ \epsilon]$

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

Transitions:

(1) $([A \to \alpha \bullet X\,\beta,\ x], X, [A \to \alpha\, X \bullet \beta,\ x]),\quad X \in (N \cup T)$

(2) $([A \to \alpha \bullet B\,\beta,\ x], \epsilon,\ [B \to \bullet\,\gamma,\ x']),$
$A \to \alpha\, B\, \beta\ ,\quad B \to \gamma \in P, x' \in \text{First}_1(\beta) \odot_1 \{x\};$

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 ...

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; analogously to $LR(0)$, we need the ϵ-closure δ^*_ϵ as a helper function:

$$\delta^*_\epsilon(q) = q \cup \{[C \to \bullet\gamma, x] \mid \exists\, [A \to \alpha \bullet B\,\beta',\ x'] \in q, \beta \in (N \cup T)^* :\ B \to^* C\,\beta\ \wedge\ x \in \text{First}_1(\beta\,\beta') \odot_1 \{x'\}\}$$

The Characteristic LR(1)-Automaton

The set of admissible $LR(1)$-items for viable prefixes is again computed with the help of the finite automaton $c(G,1)$.

The automaton $c(G,1)$:

States: $LR(1)$-items

Start state: $[S' \to \bullet\, S,\ \epsilon]$

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

Transitions:

(1) $([A \to \alpha \bullet X\,\beta,\ x], X, [A \to \alpha\, X \bullet \beta,\ x]),\quad X \in (N \cup T)$

(2) $([A \to \alpha \bullet B\,\beta,\ x], \epsilon,\ [B \to \bullet\,\gamma,\ x']),$
$A \to \alpha\, B\, \beta,\quad B \to \gamma \in P, x' \in \text{First}_1(\beta) \odot \{x\};$

This automaton works like $c(G)$ — but additionally manages a 1-prefix from Follow_1 of the left-hand sides.

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 the ϵ-closure δ^*_ϵ as a helper function:

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

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

The Canonical LR(1)-Automaton

For example:

$$\begin{array}{lcll} E & \to & E+T & \mid\ T \\ T & \to & T*F & \mid\ F \\ F & \to & (\,E\,) & \mid\ \mathsf{int} \end{array}$$

$\mathsf{First}_1(S') = \mathsf{First}_1(E) = \mathsf{First}_1(T) = \mathsf{First}_1(F) = \mathsf{name}, \mathsf{int}, ($

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

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

The Canonical LR(1)-Automaton

For example:

$$\begin{array}{lcll} E & \to & E+T & \mid\ T \\ T & \to & T*F & \mid\ F \\ F & \to & (\,E\,) & \mid\ \mathsf{int} \end{array}$$

$\mathsf{First}_1(S') = \mathsf{First}_1(E) = \mathsf{First}_1(T) = \mathsf{First}_1(F) = \mathsf{name}, \mathsf{int}, ($

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

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

cb.pdf

File Edit View Go Bookmarks Help

136 (308 of 511) 383,05%

$F \to (\,E\,) \mid \mathsf{int}$

$\mathsf{First}_1(S') = \mathsf{First}_1(E) = \mathsf{First}_1(T) = \mathsf{First}_1($

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

$q_3 = \delta(q_0, F$

$q_4 = \delta(q_0, \mathsf{ir}$

$q_5 = \delta(q_0, ($

cb.pdf

File Edit View Go Bookmarks Help

136 (309 of 511) 383,05%

$\mathsf{First}_1(S') = \mathsf{First}_1(E) = \mathsf{First}_1(T) = \mathsf{First}_1($

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

$q_3 = \delta(q_0, F$

$q_4 = \delta(q_0, \mathsf{in}$

$q_5 = \delta(q_0, ($

cb.pdf

File Edit View Go Bookmarks Help

136 (309 of 511) 383,05%

$\mathsf{st}_1(E) = \mathsf{First}_1(T) = \mathsf{First}_1(F) = \mathsf{name}, \mathsf{int}, ($

$\to \bullet\, E,\ \{\epsilon\}],$
$\to \bullet\, E + T,\ \{\epsilon, +\}],$
$\to \bullet\, T,\ \{\epsilon, +\}],$
$\to \bullet\, T * F,\ \{\epsilon, +, *\}],$
$\to \bullet\, F,\ \{\epsilon, +, *\}],$
$\to \bullet\, (\ E\ ),\ \{\epsilon, +, *\}],$
$\to \bullet\, \mathsf{int},\ \{\epsilon, +, *\}]\}$

$\to E\, \bullet,\ \{\epsilon\}],$
$\to E\, \bullet + T,\ \{\epsilon, +\}]\}$

$\to T\, \bullet,\ \{\epsilon, +\}],$
$\to T\, \bullet * F,\ \{\epsilon, +, *\}]\}$

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

cb.pdf

File Edit View Go Bookmarks Help

- ✓ Continuous
- Dual
- Side Pane F9
- Fullscreen F11
- Presentation F5
- Zoom In Ctrl++
- Zoom Out Ctrl+-
- Odd Pages Left
- Inverted Colors Ctrl+I
- Reload Ctrl+R

383,05%

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

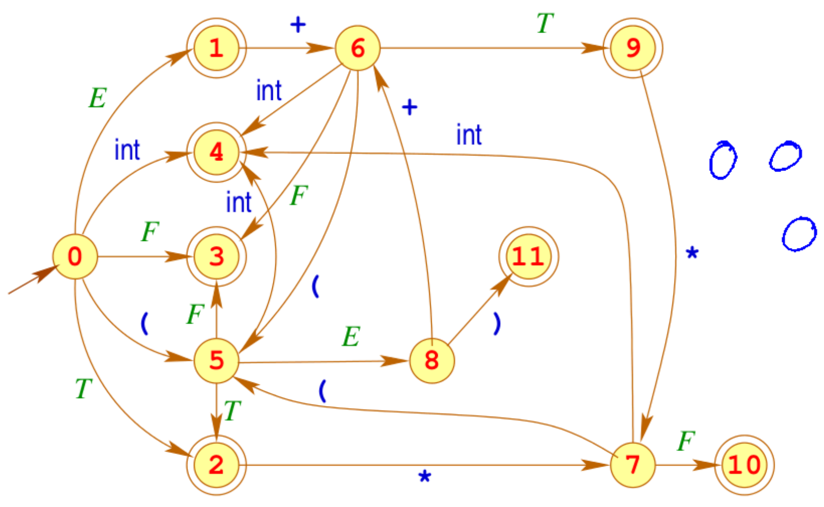
$\to E\, \bullet,\ \{\epsilon\}],$
$\to E\, \bullet + T,\ \{\epsilon, +\}]\}$

$\to T\, \bullet,\ \{\epsilon, +\}],$
$\to T\, \bullet * F,\ \{\epsilon, +, *\}]\}$

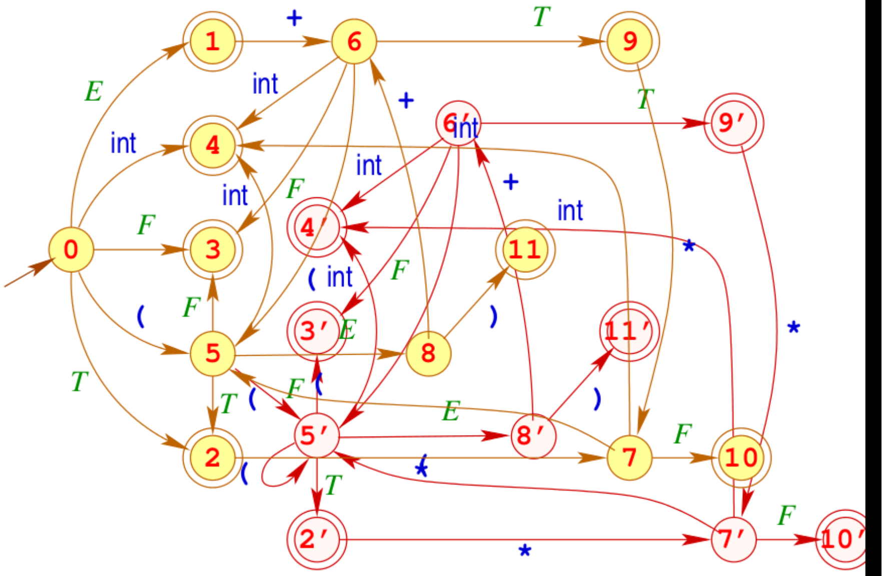
136 / 238

)-Automaton

The Canonical LR(1)-Automaton



The Canonical LR(1)-Automaton



The Canonical LR(1)-Automaton

Discussion:

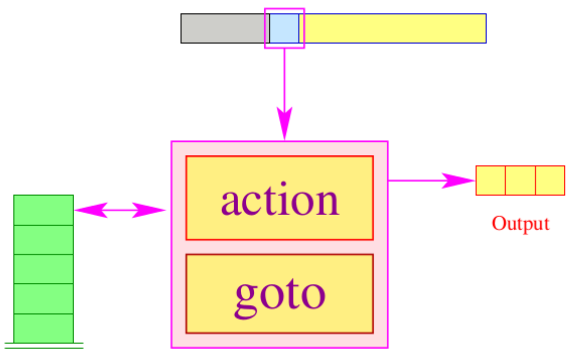
- In the example, the number of states was almost doubled
  ... and it can become even worse
- The conflicts in states q_1, q_2, q_9 are now resolved !
  e.g. we have for:

$$q_9 = \boxed{\begin{array}{l}\{[E \to E + T\bullet,\ \{\epsilon, +\}], \\ \ [T \to T\bullet * F,\ \{\epsilon, +, *\}]\}\end{array}}$$

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:

$$\begin{array}{lllll} E & \to & E + T^{\,0} & \mid & T^{\,1} \\ T & \to & T * F^{\,0} & \mid & F^{\,1} \\ F & \to & (\ E\ )^{\,0} & \mid & \text{int}^{\,1} \end{array}$$

action	ϵ	int	(	)	$\boxed{+}$	$\boxed{*}$
q_1	$S', 0$					s
q_2	$E, 1$					s
q_2'				$E, 1$		s
$\boxed{q_3}$	$T, 1$				$\boxed{T, 1}$	$T, 1$
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
$\boxed{q_9'}$				$E, 0$	$E, 0$	$\boxed{\text{s}}$
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 LR(1)-Parser

The construction of the $LR(1)$-parser:

States: $Q \cup \{f\}$ $\quad$ (f fresh)
Start state: q_0
Final state: f

Transitions:

Shift: $(p, a, p\, q)$ if $q = \text{goto}[q, a]$, $s = \text{action}[p, w]$

Reduce: $(p\, q_1 \ldots q_{|\beta|}, \epsilon, p\, q)$ if $[A \to \beta\, \bullet] \in q_{|\beta|}$, $q = \text{goto}(p, A)$, $[A \to \beta\, \bullet] = \text{action}[q_{|\beta|}, w]$

Finish: $(q_0\, p, \epsilon, f)$ if $[S' \to S \bullet] \in p$

with $LR(G, 1) = (Q, T, \delta, q_0, F)$.

The Canonical LR(1)-Automaton

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 Canonical LR(1)-Automaton

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 $x \in \{a\} \odot_k \text{First}_k(\beta) \odot_k \{y\}$.

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

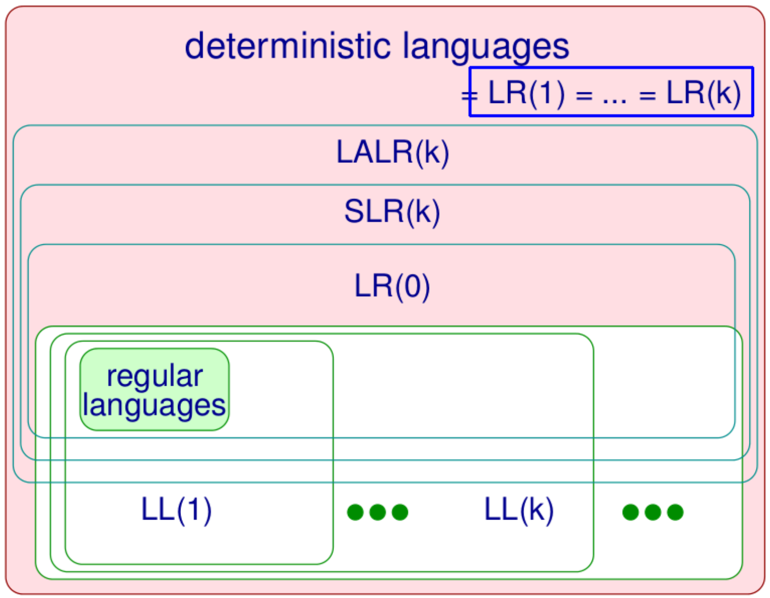
Theorem:

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:

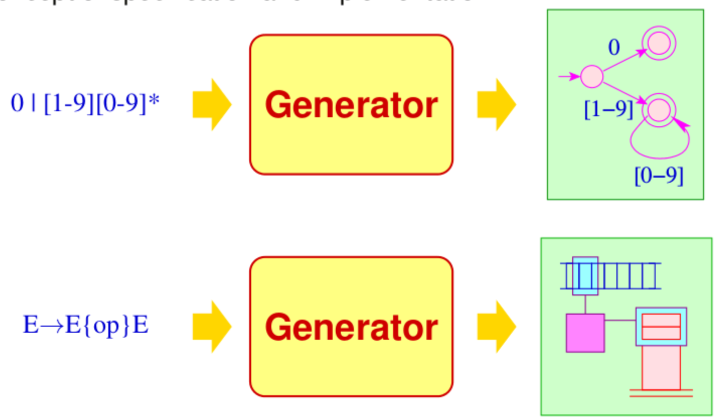
- Our example apparently is $LR(1)$
- In general, the canonical $LR(k)$-automaton has much more states then $LR(G) = LR(G, 0)$
- Therefore in practice, subclasses of $LR(k)$-grammars are often considered, which only use $LR(G)$ ...
- For resolving conflicts, the items are assigned special lookahead-sets: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

Parsing Methods



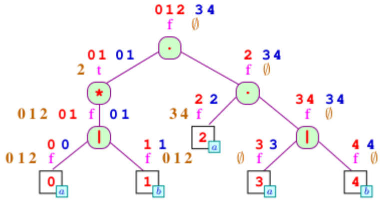
Lexical and Syntactical Analysis:

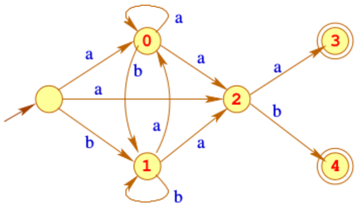
Concept of specification and implementation:



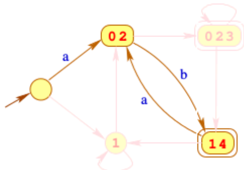
Lexical and Syntactical Analysis:

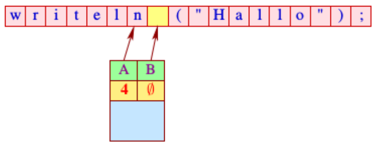
From Regular Expressions to Finite Automata





From Finite Automata to Scanners

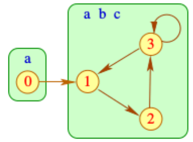


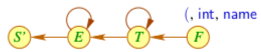


Lexical and Syntactical Analysis:

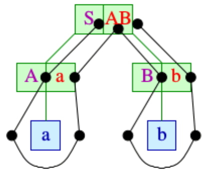
Computation of lookahead sets:

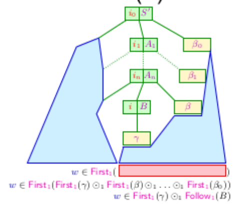
$$\begin{array}{lll} F_\epsilon(S') & \supseteq & F_\epsilon(E) \\ F_\epsilon(E) & \supseteq & F_\epsilon(T) \\ F_\epsilon(T) & \supseteq & F_\epsilon(F) \end{array} \qquad \begin{array}{lll} F_\epsilon(E) & \supseteq & F_\epsilon(E) \\ F_\epsilon(T) & \supseteq & F_\epsilon(T) \\ F_\epsilon(F) & \supseteq & \{(, \text{name}, \text{int}\} \end{array}$$

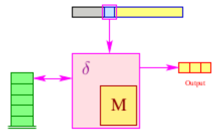




From Item-Pushdown Automata to LL(1)-Parsers:

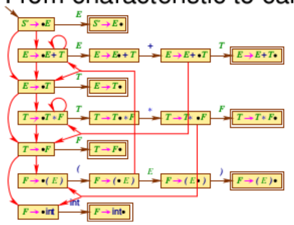


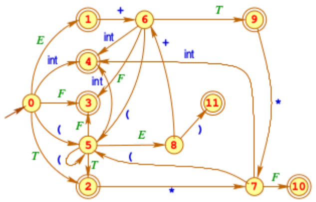




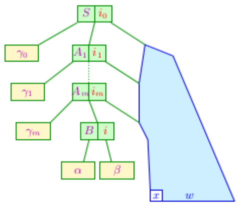
Lexical and Syntactical Analysis:

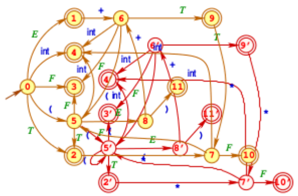
From characteristic to canonical Automata:

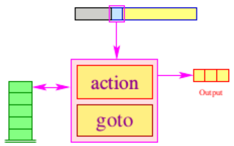




From Shift-Reduce-Parsers to LR(1)-Parsers:







Topic:

Semantic Analysis