Script generated by TTT

Title: Petter: Compiler Construction (14.05.2020) - 14: First(1) Computation

Date: Mon May 04 16:27:13 CEST 2020

Duration: 44:49 min

Pages: 19

Lookahead Sets

Definition: First_1-Sets

For a set $L \subseteq T^*$ we define:

$$\text{First}_1(L) = \{\epsilon \mid \epsilon \in L\} \cup \{u \in T \mid \exists v \in T^* : uv \in L\}$$

Example: $S \to \epsilon \mid a\,S\,b$

$\text{First}_1(\llbracket S \rrbracket)$
ϵ
$a\,b$
$a\,a\,b\,b$
$a\,a\,a\,b\,b\,b$
$\dots$

$= \{ \varepsilon, a \}$

$\equiv$ the yield's prefix of length 1

Lookahead Sets

Arithmetics:

$\text{First}_1(\_)$ is distributive with union and concatenation:

$$\begin{array}{lll}
\text{First}_1(\emptyset) & = & \emptyset \\
\text{First}_1(L_1 \cup L_2) & = & \text{First}_1(L_1) \cup \text{First}_1(L_2) \\
\text{First}_1(L_1 \cdot L_2) & = & \text{First}_1(\text{First}_1(L_1) \cdot \text{First}_1(L_2)) \\
 & := & \text{First}_1(L_1) \odot_1 \text{First}_1(L_2)
\end{array}$$

$\odot_1$ being $1-$concatenation

Lookahead Sets

Arithmetics:

$\text{First}_1(\_)$ is distributive with union and concatenation:

$$\begin{array}{lll}
\text{First}_1(\emptyset) & = & \emptyset \\
\text{First}_1(L_1 \cup L_2) & = & \text{First}_1(L_1) \cup \text{First}_1(L_2) \\
\text{First}_1(L_1 \cdot L_2) & = & \text{First}_1(\text{First}_1(L_1) \cdot \text{First}_1(L_2)) \\
 & := & \text{First}_1(L_1) \odot_1 \text{First}_1(L_2)
\end{array}$$

$\odot_1$ being $1-$concatenation

Definition: 1-concatenation

Let $L_1, L_2 \subseteq T \cup \{\epsilon\}$ with $L_1 \neq \emptyset \neq L_2$. Then:

$$L_1 \odot_1 L_2 = \begin{cases} L_1 & \text{if } \epsilon \notin L_1 \\ (L_1 \setminus \{\epsilon\}) \cup L_2 & \text{otherwise} \end{cases}$$

If all rules of G are productive, then all sets $\text{First}_1(A)$ are non-empty.

Lookahead Sets

For $\alpha \in (N \cup T)^*$ we are interested in the set:

$$\text{First}_1(\alpha) \;=\; \text{First}_1(\{w \in T^* \mid \alpha \rightarrow^* w\})$$

Lookahead Sets

For $\alpha \in (N \cup T)^*$ we are interested in the set:

$$\text{First}_1(\alpha) \;=\; \text{First}_1(\{w \in T^* \mid \alpha \rightarrow^* w\})$$

Idea: Treat ϵ separately: $\text{First}_1(A) = F_\epsilon(A) \cup \{\epsilon \mid A \rightarrow^* \epsilon\}$

- Let $\text{empty}(X) = \text{true}$ iff $X \rightarrow^* \epsilon$.
- $F_\epsilon(X_1 \dots X_m) = F_\epsilon(X_1) \cup \dots \cup F_\epsilon(X_{j-1})$ if $\neg\text{empty}(X_j) \;\wedge\; \bigwedge_{i=1}^{j-1} \text{empty}(X_i)$

Lookahead Sets

For $\alpha \in (N \cup T)^*$ we are interested in the set:

$$\text{First}_1(\alpha) \;=\; \text{First}_1(\{w \in T^* \mid \alpha \rightarrow^* w\})$$

Lookahead Sets

For $\alpha \in (N \cup T)^*$ we are interested in the set:

$$\text{First}_1(\alpha) \;=\; \text{First}_1(\{w \in T^* \mid \alpha \rightarrow^* w\})$$

Lookahead Sets

For $\alpha \in (N \cup T)^*$ we are interested in the set:

$$\text{First}_1(\alpha) \;=\; \text{First}_1(\{w \in T^* \mid \alpha \to^* w\})$$

Idea: Treat ϵ separately: $\text{First}_1(A) = F_\epsilon(A) \cup \{\epsilon \mid A \to^* \epsilon\}$

- Let $\text{empty}(X) = \text{true}$ iff $X \to^* \epsilon$.
- $F_\epsilon(X_1 \dots X_m) = \bigcup_{i=1}^{j} F_\epsilon(X_i)$ if $\neg\text{empty}(X_j) \;\wedge\; \bigwedge_{i=1}^{j-1} \text{empty}(X_i)$

Lookahead Sets

For $\alpha \in (N \cup T)^*$ we are interested in the set:

$$\text{First}_1(\alpha) \;=\; \text{First}_1(\{w \in T^* \mid \alpha \to^* w\})$$

Idea: Treat ϵ separately: $\text{First}_1(A) = F_\epsilon(A) \cup \{\epsilon \mid A \to^* \epsilon\}$

- Let $\text{empty}(X) = \text{true}$ iff $X \to^* \epsilon$.
- $F_\epsilon(X_1 \dots X_m) = \bigcup_{i=1}^{j} F_\epsilon(X_i)$ if $\neg\text{empty}(X_j) \;\wedge\; \bigwedge_{i=1}^{j-1} \text{empty}(X_i)$

Lookahead Sets

For $\alpha \in (N \cup T)^*$ we are interested in the set:

$$\text{First}_1(\alpha) \;=\; \text{First}_1(\{w \in T^* \mid \alpha \to^* w\})$$

Idea: Treat ϵ separately: $\text{First}_1(A) = F_\epsilon(A) \cup \{\epsilon \mid A \to^* \epsilon\}$

- Let $\text{empty}(X) = \text{true}$ iff $X \to^* \epsilon$.
- $F_\epsilon(X_1 \dots X_m) = \bigcup_{i=1}^{j} F_\epsilon(X_i)$ if $\neg\text{empty}(X_j) \;\wedge\; \bigwedge_{i=1}^{j-1} \text{empty}(X_i)$

We characterize the ϵ-free First_1-sets with an inequality system:

$$\begin{array}{lcll} F_\epsilon(a) & = & \{a\} & \text{if } a \in T \\ F_\epsilon(A) & \supseteq & F_\epsilon(X_j) & \text{if } A \to X_1 \dots X_m \in P, \;\; \text{empty}(X_1) \wedge \dots \wedge \text{empty}(X_{j-1}) \end{array}$$

Lookahead Sets

for example...

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

with $\text{empty}(E) = \text{empty}(T) = \text{empty}(F) = \text{false}$

$F_\epsilon(E) \supseteq F_\epsilon(E)$ $\quad$ $F_\epsilon(E) \supseteq F_\epsilon(T)$

$F_\epsilon(T) \supseteq F_\epsilon(T)$ $\quad$ $F_\epsilon(T) \supseteq F_\epsilon(F)$

$F_\epsilon(F) \supseteq \{\text{int}\}$ $\quad$ $F_\epsilon(F) \supseteq \{\text{name}\}$ $\quad$ $F_\epsilon(F) \supseteq \{\,(\,\}$

Lookahead Sets

for example...

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

with $\text{empty}(E) = \text{empty}(T) = \text{empty}(F) = \text{false}$

... we obtain:

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

Fast Computation of Lookahead Sets

Observation:

- The form of each inequality of these systems is:

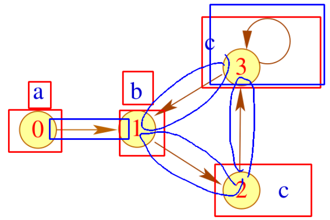
$$x \sqsupseteq y \quad \text{resp.} \quad x \sqsupseteq d$$

for variables x, y und $d \in \mathbb{D}$.

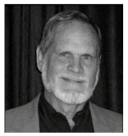
- Such systems are called pure unification problems
- Such problems can be solved in linear space/time.

for example: $\mathbb{D} = 2^{\{a,b,c\}}$

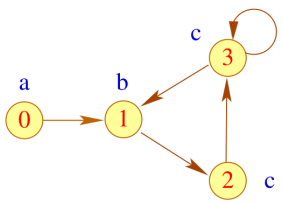
$$\begin{array}{lll} x_0 \supseteq \{a\} & x_1 \supseteq x_0 & x_1 \supseteq x_3 \\ x_1 \supseteq \{b\} & x_2 \supseteq x_1 & \\ x_2 \supseteq \{c\} & & \\ x_3 \supseteq \{c\} & x_3 \supseteq x_2 & x_3 \supseteq x_3 \end{array}$$



Fast Computation of Lookahead Sets



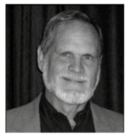
Frank DeRemer
& Tom Pennello



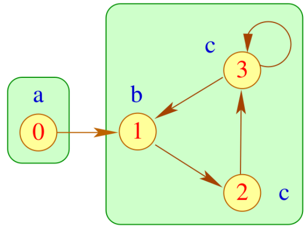
Proceeding:

- Create the Variable Dependency Graph for the inequality system.

Fast Computation of Lookahead Sets



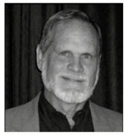
Frank DeRemer
& Tom Pennello



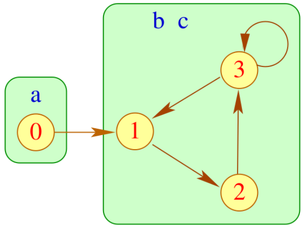
Proceeding:

- Create the Variable Dependency Graph for the inequality system.
- Whithin a Strongly Connected Component (→ Tarjan) all variables have the same value

Fast Computation of Lookahead Sets



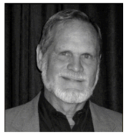
Frank DeRemer
& Tom Pennello



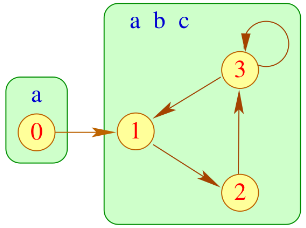
Proceeding:

- Create the Variable Dependency Graph for the inequality system.
- Whithin a Strongly Connected Component (→ Tarjan) all variables have the same value
- Is there no ingoing edge for an SCC, its value is computed via the smallest upper bound of all values within the SCC

Fast Computation of Lookahead Sets



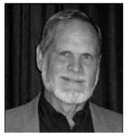
Frank DeRemer
& Tom Pennello



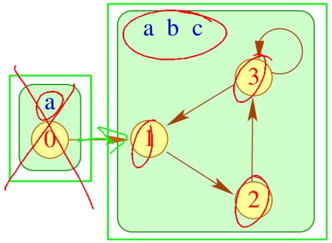
Proceeding:

- Create the Variable Dependency Graph for the inequality system.
- Whithin a Strongly Connected Component (→ Tarjan) all variables have the same value
- Is there no ingoing edge for an SCC, its value is computed via the smallest upper bound of all values within the SCC
- In case of ingoing edges, their values are also to be considered for the upper bound

Fast Computation of Lookahead Sets



Frank DeRemer
& Tom Pennello



Proceeding:

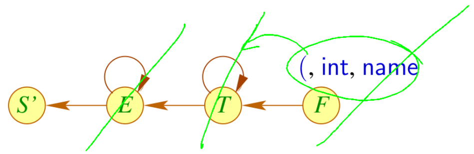
- Create the Variable Dependency Graph for the inequality system.
- Whithin a Strongly Connected Component (→ Tarjan) all variables have the same value
- Is there no ingoing edge for an SCC, its value is computed via the smallest upper bound of all values within the SCC
- In case of ingoing edges, their values are also to be considered for the upper bound

Fast Computation of Lookahead Sets

$First_1(X) = F_\epsilon(X) \cup \{\epsilon \mid empty(X)\}$

... for our example grammar:

$First_1$:



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