Введение в синтаксический анализ

LL-грамматики

Автор: Григорьев Семён

Санкт-Петербургский государственный университет
Математико-Механический факультет
Кафедра системного программирования

22 февраля 2012г.

LL(k)-грамматики

Definition

Пусть $G = (V_N, V_T, P, S)$ – контекстно-свободная грамматика.
Определим функцию $FIRST_k^G(\alpha) = \{w \in V_T |$ либо $|w| < k$ и $\alpha \underset{G}{\overset{*}{\Rightarrow}} w$, либо $|w| = k$ и $\alpha \underset{G}{\overset{*}{\Rightarrow}} wx$ для некоторой цепочки $x \in V_T\}$.
Здесь $k \geq 0$ – целое, $\alpha \in (V_N \cup V_T)^*$.

LL(k)-грамматики

Definition

Пусть $G = (V_N, V_T, P, S)$ – контекстно-свободная грамматика. Говорят, что G есть $LL(k)$-грамматика для некоторого фиксированного k, если для любых двух левосторонних выводов вида

1. $S \underset{lm}{\overset{*}{\Rightarrow}} wA\alpha \underset{lm}{\Rightarrow} w\beta\alpha \underset{lm}{\overset{*}{\Rightarrow}} wx$
2. $S \underset{lm}{\overset{*}{\Rightarrow}} wA\alpha \underset{lm}{\Rightarrow} w\gamma\alpha \underset{lm}{\overset{*}{\Rightarrow}} wy$

в которых $FIRST_k^G(x) = FIRST_k^G(y)$, имеет место равенство $\beta = \gamma$.

Definition

Говорят, что контекстно-свободная грамматика G есть LL-грамматика, если она $LL(k)$ для некоторого $k \geq 0$.

Простые LL(1)-грамматики

Definition

Говорят, что контекстно-свободная грамматика G является простой $LL(1)$-грамматикой, если в ней нет ε -правил, и все альтернативы для каждого нетерминала начинаются с терминалов и притом различных.

Свойства $LL(k)$-грамматик

Theorem

Чтобы контекстно-свободная грамматика $G = (V_N, V_T, P, S)$ была $LL(k)$-грамматикой, необходимо и достаточно, чтобы

$$FIRST_k^G(\beta\alpha) \cap FIRST_k^G(\gamma\alpha) = \varnothing$$

для всех α, β, γ , таких, что существуют правила $A \to \beta, A \to \gamma \in P, \beta \neq \gamma$ и существует вывод $S \underset{lm}{\overset{*}{\Rightarrow}} wA\alpha$.

$FOLLOW_k^G$

Definition

Пусть $G = (V_N, V_T, P, S)$ – контекстно-свободная грамматика, и $\beta \in V^*$. Определим функцию

$$FOLLOW_k^G(\beta) = \{w \in V_T | S \overset{*}{\underset{G}{\Rightarrow}} \gamma\beta\alpha, w \in FIRST_k^G(\alpha)\}$$

Здесь $k \geq 0$ – целое.

Свойство $LL(1)$-грамматик

Theorem

Чтобы контекстно-свободная грамматика $G = (V_N, V_T, P, S)$ была $LL(1)$-грамматикой, необходимо и достаточно, чтобы

$$FIRST_1^G(\beta FOLLOW_1^G(A)) \cap FIRST_1^G(\gamma FOLLOW_1^G(A)) = \varnothing$$

для всех $A \in V_N, \beta, \gamma \in (V_N \cup V_T)^*$, таких, что существуют правила $A \to \beta, A \to \gamma \in P, \beta \neq \gamma$ и существует вывод $S \underset{lm}{\overset{*}{\Rightarrow}} wA\alpha$.

Пояснение: $FIRST_1^G(W) = \bigcup_{s \in W} FIRST_1^G(s)$.

Свойство $LL(1)$-грамматик

Consequence

Эту теорему можно переформулировать следующим образом: КС-грамматика G является $LL(1)$-грамматикой тогда и только тогда, когда для каждого множества A-правил: $A \rightarrow \alpha_1|\alpha_2|...|\alpha_n$ – выполняются следующие условия:

1. $FIRST_1^G(\alpha_i) \cap FIRST_1^G(\alpha_j) = \varnothing$, при $i \neq j, 0 \leq i, j \leq n$;
2. если $\alpha_i = \varepsilon$, то $FIRST_1^G(\alpha_j) \cap FIRST_1^G(A) = 0$, для всех $j : 0 \leq j \leq n, j \neq i$.

Леворекурсивные и $LL(k)$-грамматики

Theorem

Если $G = (V_N, V_T, P, S)$ – контекстно-свободная грамматика, и G – леворекурсивна, то G – не $LL(k)$-грамматика ни при каком k.

Леворекурсивные и $LL(k)$-грамматики

Theorem

Если $G = (V_N, V_T, P, S)$ – контекстно-свободная грамматика, и G – леворекурсивна, то G – не $LL(k)$-грамматика ни при каком k.

Consequence

Мы имеем два достаточных признака для того, чтобы считать КС-грамматику не LL-грамматикой. Это – неоднозначность и леворекурсивность.