Математическая статистика. Теория

Александр Сергеев

1 Введение

Процесс компиляции

вход $\xrightarrow{\text{лексический анализ}}$ токены $\xrightarrow{\text{парсинг}}$ дерево разбора $\xrightarrow{\text{вычисление/компиляция}}$ результат

Определение

Токен – неделимая единица парсинга

Синтаксически управляемая трансляция – технология написания парсеров, когда одновременно задаются и зависят друг от друга парсинг и вычисление (правила вычисления применяются прямо во время разбора)

Напоминание

Контекстно-свободная грамматика:

Алфавит Σ

Нетерминалы N

Стартовый нетерминал $S \in N$

Правила $P \subset N \times (N \cup \Sigma)^* : \langle A, \alpha \rangle \in P$ или $A \to \alpha$

$\alpha \Rightarrow \beta$ – из α выводится за 1 шаг β, если $\alpha = \alpha_1 A \alpha_2, \beta = \alpha_1 \xi \alpha_2$ и есть правило $A \to \xi \in P$

Язык грамматики $L(\Gamma) = \{x | S \Rightarrow^* x, x \in \Sigma^*\}$

Определение

Грамматика $\Gamma \in LL(1)$, если из

$S \Rightarrow^* xA\alpha \Rightarrow x\xi\alpha \Rightarrow^* xcy$

$S \Rightarrow^* xA\beta \Rightarrow x\eta\beta \Rightarrow^* xcz$

$c \in \Sigma$ или $c = \varepsilon, y = \varepsilon, z = \varepsilon$

следует $\xi = \eta$

Замечание

Буквы из конца латинского алфавита – строки из терминалов

Буквы из греческого алфавита – любые строки (возможно, содержащие нетерминалы)

Замечание

Другими словами, если мы хотим, чтобы из нетерминала A получилась строка, начинающаяся на c, то у нас есть только одно правило для достижения этого

Определение

$LL(k)$ – вместо символа c у нас k символов

Из

$S \Rightarrow^* xA\alpha \Rightarrow x\xi\alpha \Rightarrow^* xcy$

$S \Rightarrow^* xA\beta \Rightarrow x\eta\beta \Rightarrow^* xcz$

$c \in \Sigma^k$ или $c = \Sigma^{\leq k}, y = \varepsilon, z = \varepsilon$

следует $\xi = \eta$

Замечание

$LL(0)$-грамматики задают линейные программы (обобщение архиваторов)

Утверждение

$LL(1)$-грамматики – это грамматики, для которых можно написать рекурсивный спуск

Определение

$FIRST : (N \cup \Sigma)^* \to 2^{\Sigma \cup \{\varepsilon\}}$

$FOLLOW : N \to 2^{\Sigma \cup \{\$\}}$

Пока будем считать, что *бесполезных* символов нет – из любого нетерминала можно вывести терминал

$FIRST(\alpha) = \{c | \alpha \Rightarrow^* c\beta\} \cup \{\varepsilon | \alpha \Rightarrow^* \varepsilon\}$

$FOLLOW(A) = \{c | S \Rightarrow^* \alpha Ac\beta\} \cup \{\$ | S \Rightarrow^* \alpha A\}$

Теорема

Грамматика $\Gamma \in LL(1) \Leftrightarrow \forall\, A \to \alpha, A \to \beta$ выполнено

1. $FIRST(\alpha) \cap FIRST(\beta) = \varnothing$

2. $\varepsilon \in FIRST(\alpha) \Rightarrow FIRST(\beta) \cap FOLLOW(A) = \varnothing$

Лемма

$\alpha = c\beta \Rightarrow FIRST(\alpha) = \{c\}$

$\alpha = \varepsilon \Rightarrow FIRST(\alpha) = \{\varepsilon\}$

$\alpha = A\beta \Rightarrow FIRST(\alpha) = FIRST(A) \setminus \varepsilon \cup (FIRST(\beta)\ if\ \varepsilon \in FIRST(A))$