Теория автоматов и формальных языков

Синтаксический анализ

Автор: Екатерина Вербицкая

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»

4 декабря 2020

Синтаксический анализ

Задачи синтаксического анализа: по последовательности символов

- Определить, корректна ли она с точки зрения грамматики языка
  - Если нет, сообщить об ошибке
- Построить для нее некоторое структурное представление
  - Дерево вывода
  - Абстрактное синтаксическое дерево
  - Вычислить семантику (реже)

Безлексерный синтаксический анализ

Терминалами грамматики являются символы строки

$$
\begin{aligned}
E &\to E + T \mid T \\
T &\to T * F \mid F \\
F &\to (E) \mid N \mid Id \\
N &\to 0 \mid 1 \mid \cdots \mid 9 \mid 1N \mid 2N \mid \cdots \mid 9N \\
Id &\to a \mid b \mid \cdots \mid z \mid a\,Id \mid \cdots \mid z\,Id
\end{aligned}
$$

Недостатки безлексерного синтаксического анализа: грамматики очень быстро перестают быть читаемыми

$$
\begin{aligned}
E &\to Sp\ E\ Sp + Sp\ T\ Sp \mid Sp\ T\ Sp \\
T &\to Sp\ T\ Sp * Sp\ F\ Sp \mid Sp\ F\ Sp \\
F &\to Sp\ (Sp\ E\ Sp)\ Sp \mid Sp\ N\ Sp \mid Sp\ Id\ Sp \\
N &\to 0 \mid 1 \mid \cdots \mid 9 \mid 1N \mid 2N \mid \cdots \mid 9N \\
Id &\to a \mid b \mid \cdots \mid z \mid a\ Id \mid \cdots \mid z\ Id \\
Sp &\to \sqcup \mid \sqcup\ Sp
\end{aligned}
$$

Недостатки безлексерного синтаксического анализа: работает медленно

Большую часть времени синтаксический анализатор тратит на разбор регулярных фрагментов языка, что чревато накладными расходами

$$N \to 0 \mid 1 \mid \cdots \mid 9 \mid 1N \mid 2N \mid \cdots \mid 9N$$
$$Id \to a \mid b \mid \cdots \mid z \mid a\,Id \mid \cdots \mid z\,Id$$

Лексический анализ

Преобразование последовательности символов в последовательность *лексем* (или *токенов*)

Лексема — минимальная смысловая единица языка

Типы лексем:

- Ключевое слово
- Число
  - Целое
  - Целое в двоичной системе
  - С плавающей точкой
  - ...
- Идентификатор
- Оператор

Лексический анализ и регулярные языки

Часто языки разных типов лексем являются регулярными языками

Лексеры часто реализуются при помощи регулярных выражений

Лексеры также часто пропускают пробельные символы и выделяют комментарии

Синтаксический анализ после лексического

Терминалы грамматики — лексемы

$$E \to E\ PLUS\ T \mid T$$
$$T \to T\ MULT\ F \mid F$$
$$F \to LBR\ E\ RBR \mid NUMBER \mid IDENT$$