

# Теория формальных языков и компиляторов

## Часть 1. Порождающие грамматики и языки

### Лекция 6. Синтаксические деревья.

#### Эквивалентность и однозначность грамматик

##### 6.1 Синтаксические деревья

В предыдущих лекциях при решении прямой и обратной задачи использовалось понятие выводимости. Одной из задач языкового процессора является определение принадлежности заданной строки некоторому формальному языку. Например, компилятор с языка Си определяет принадлежность следующего текста языку Си:

```
#include <stdio.h>
int main(void)
{
    printf("Hello, world!\n");
    return 0;
}
```

Чтобы показать принадлежность строки заданному языку, необходимо построить цепочку вывода этой строки из начального символа грамматики.

**Пример 6.1.** Имеется грамматика  $G[Z]$ , заданная правилом:

$$P: Z \rightarrow ab \mid aZbb$$

Необходимо показать, что строка «aabb» принадлежит языку  $L(G[Z])$ .

Используя выводимость, легко построить строку по правилам грамматики:

$$Z \Rightarrow aZbb \Rightarrow aabb$$

Таким образом показано, что строка «aabb» принадлежит языку  $L(G[Z])$ .

Выводимость – не единственный способ анализа строк. Существуют другие методы, например, построение *синтаксических деревьев*.

Рассмотрим понятие геометрической интерпретации. *Геометрическая интерпретация* – это объяснение какого-либо математического процесса с помощью геометрических методов. Например, геометрическая интерпретация функции – это ее график.

Геометрической интерпретацией синтаксического разбора является изображение, показывающее последовательность вывода строки. Пример такого изображения – синтаксическое дерево.

**Определение 6.1.** Синтаксическим деревом называется граф, узлами которого являются нетерминалы из  $V_N$ , а дуги (ветви) показывают выводимость по правилам грамматики. Листья синтаксического дерева отражают цепочки языка  $\beta_i \in V_T^*$ , порождаемые заданной грамматикой с правилами вывода  $P$ .

На рис.6.1 представлено синтаксическое дерево для грамматики арифметических выражений  $G[\langle AB \rangle]$  (см. пример 4.6) с правилами:

- $P$ :
- 1)  $\langle AB \rangle \rightarrow T$  (Т — терм)
  - 2)  $\langle AB \rangle \rightarrow T + \langle AB \rangle$
  - 3)  $\langle AB \rangle \rightarrow T - \langle AB \rangle$
  - 4)  $T \rightarrow O$  (О — операнд)
  - 5)  $T \rightarrow O * T$
  - 6)  $T \rightarrow O / T$
  - 7)  $O \rightarrow (\langle AB \rangle) | a | b | c$

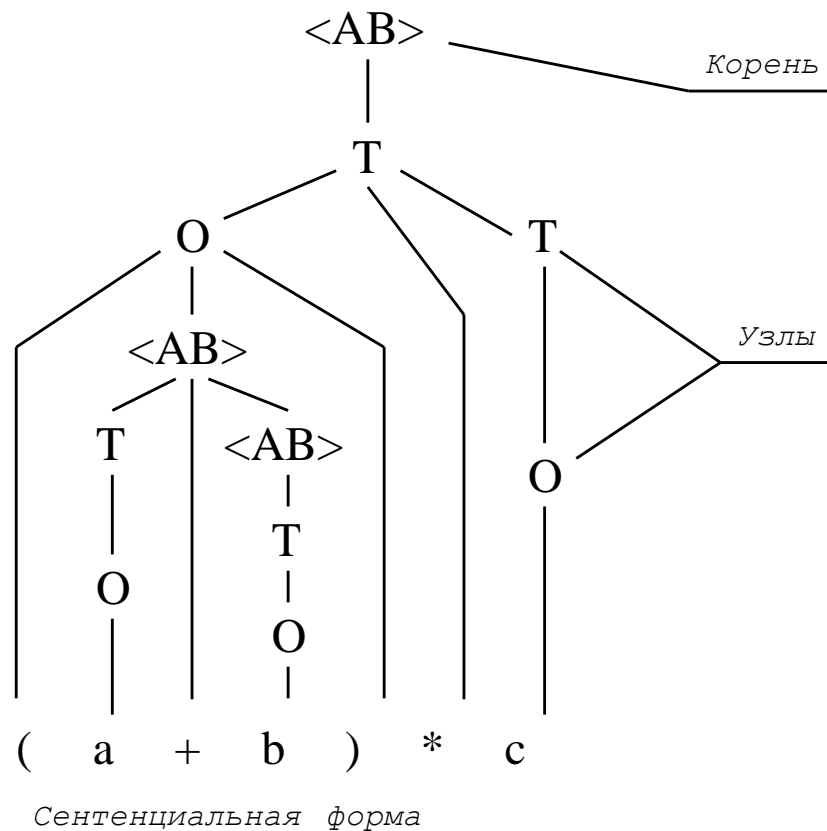


Рис 6.1. Синтаксическое дерево с грамматикой  $G[\langle AB \rangle]$

Дерево как бы перевернуто «вверх ногами» и корень находится вверху, хотя листья – цепочки терминальных символов  $\beta_i \in V_T^*$  – «свисают» с веток.

**Определение 6.2.** Терминальная цепочка  $\beta_i \in V_T^*$  называется сентенциальной формой, если цепочка является языковой конструкцией или выводима из начального нетерминала,  $\beta_i \in L(G[Z])$ .

Цепочка  $\beta_i \in V_T^*$  также называется «основа». При этом не требуется принадлежности  $\beta_i \in L(G[Z])$ , то есть в основе может быть допущена синтаксическая ошибка.

В примере 6.2 «(a+b)\*c» – сентенциальная форма или основа,  $\langle AB \rangle$  – корень дерева, O и T – узлы дерева.

На рис. 6.1. представлено дерево синтаксического разбора цепочки «(a+b)\*c», для которой в соответствии с грамматикой  $G[<AB>]$  и определением выводимости имеет место итерационная выводимость

$$\langle AB \rangle \Rightarrow^* (a+b)^*c$$

### Упражнения

Заданы грамматики, порождающие некоторые языки. Вывести по 3 строки, принадлежащих этим языкам и построить синтаксические деревья для этих строк.

1. P:    1)  $I \rightarrow AA$             2)  $A \rightarrow a$                     3)  $A \rightarrow aa$

2. P:    1)  $I \rightarrow aABc$     2)  $I \rightarrow \$$                     3)  $A \rightarrow Ab$   
          4)  $A \rightarrow cIB$     5)  $B \rightarrow bB$                     6)  $B \rightarrow a$

3. P:    1)  $I \rightarrow aM$             2)  $M \rightarrow A$                     3)  $A \rightarrow aA$   
          4)  $A \rightarrow B$             5)  $B \rightarrow bB$                     6)  $B \rightarrow b$

4. P:    1)  $I \rightarrow aA$             2)  $I \rightarrow Ic$                     3)  $I \rightarrow Ab$     4)  $A \rightarrow d$

### Список использованных источников

1. Шорников Ю.В. Теория и практика языковых процессоров.