

Análise sintática

O papel do analisador sintático

Prof. Edson Alves

Faculdade UnB Gama

Sumário

1. Gramáticas livre de contexto

Gramática livre de contexto

Definição

Uma gramática livre de contexto é composta por terminais, não-terminais, um símbolo de partida e produções, onde

1. os terminais (tokens) são símbolos básicos para a formação de cadeias;
2. os não-terminais são variáveis sintáticas que identificam cadeias de caracteres e que impõem uma estrutura hierárquica na linguagem;
3. um dentre os não-terminais é designado como símbolo de partida e o conjunto de cadeias geradas por ele é a linguagem definida pela gramática; e
4. as produções estabelecem as relações entre terminais e não-terminais e como novas cadeias podem ser formadas. Cada produção é composta por um não-terminal seguido de uma seta, a qual é sucedida por uma cadeia de terminais e não-terminais.

Convenções de notação

1. São terminais:

- (i) Letras minúsculas do alfabeto (por exemplo, a, b, c, \dots)
- (ii) Símbolos de operadores (por exemplo, $+$, $-$, \times , etc)
- (iii) Símbolos de pontuação, parêntesis, vírgulas, etc
- (iv) Os dígitos decimais 0, 1, 2, ..., 9
- (v) Cadeias em negrito (por exemplo, **if**, **else**, **for**, etc)

Convenções de notação

1. São terminais:

- (i) Letras minúsculas do alfabeto (por exemplo, a, b, c, \dots)
- (ii) Símbolos de operadores (por exemplo, $+$, $-$, \times , etc)
- (iii) Símbolos de pontuação, parêntesis, vírgulas, etc
- (iv) Os dígitos decimais 0, 1, 2, ..., 9
- (v) Cadeias em negrito (por exemplo, **if**, **else**, **for**, etc)

2. São não-terminais:

- (i) Letras maiúsculas do início do alfabeto (por exemplo, A, B, C, \dots)
- (ii) A letra S , em geral indicado o símbolo de partida
- (iii) Nomes em itálico formados por letras minúsculas (por exemplo, *cmd* e *expr*)

Convenções de notação

1. São terminais:

- (i) Letras minúsculas do alfabeto (por exemplo, a, b, c, \dots)
- (ii) Símbolos de operadores (por exemplo, $+$, $-$, \times , etc)
- (iii) Símbolos de pontuação, parêntesis, vírgulas, etc
- (iv) Os dígitos decimais 0, 1, 2, ..., 9
- (v) Cadeias em negrito (por exemplo, **if**, **else**, **for**, etc)

2. São não-terminais:

- (i) Letras maiúsculas do início do alfabeto (por exemplo, A, B, C, \dots)
- (ii) A letra S , em geral indicado o símbolo de partida
- (iii) Nomes em itálico formados por letras minúsculas (por exemplo, *cmd* e *expr*)

3. Letras maiúsculas do final do alfabeto (X, Y, Z) representam símbolos gramaticais, isto é, terminais ou não-terminais

Convenções de notação

4. Letras minúsculas do fim do alfabeto (x, y, z) representam cadeias de terminais

Convenções de notação

4. Letras minúsculas do fim do alfabeto (x, y, z) representam cadeias de terminais
5. Letras gregas minúsculas (por exemplo, $\alpha, \beta, \gamma, \dots$) representam cadeias de símbolos gramaticais (por exemplo, $A \rightarrow \alpha$ seria uma produção)

Convenções de notação

4. Letras minúsculas do fim do alfabeto (x, y, z) representam cadeias de terminais
5. Letras gregas minúsculas (por exemplo, $\alpha, \beta, \gamma, \dots$) representam cadeias de símbolos gramaticais (por exemplo, $A \rightarrow \alpha$ seria uma produção)
6. Se $A \rightarrow \alpha_1, A \rightarrow \alpha_2, \dots, A \rightarrow \alpha_N$ são produções com A à esquerda (denominadas produções- A) então estas produções podem ser grafadas em uma só linha como

$$A \rightarrow \alpha_1 \mid \alpha_2 \mid \dots \mid \alpha_N,$$

sendo cada α_i uma alternativa para A

Convenções de notação

4. Letras minúsculas do fim do alfabeto (x, y, z) representam cadeias de terminais
5. Letras gregas minúsculas (por exemplo, $\alpha, \beta, \gamma, \dots$) representam cadeias de símbolos gramaticais (por exemplo, $A \rightarrow \alpha$ seria uma produção)
6. Se $A \rightarrow \alpha_1, A \rightarrow \alpha_2, \dots, A \rightarrow \alpha_N$ são produções com A à esquerda (denominadas produções- A) então estas produções podem ser grafadas em uma só linha como

$$A \rightarrow \alpha_1 \mid \alpha_2 \mid \dots \mid \alpha_N,$$

sendo cada α_i uma alternativa para A

7. O lado esquerdo da primeira produção é o símbolo de partida, salvo indicação contrária

Exemplo de gramática sem e com as convenções de notação

$$expr \rightarrow expr\ op\ expr$$
$$expr \rightarrow (expr)$$
$$expr \rightarrow -\ expr$$
$$expr \rightarrow \mathbf{id}$$
$$op \rightarrow +$$
$$op \rightarrow -$$
$$op \rightarrow \times$$
$$op \rightarrow \div$$
$$op \rightarrow \uparrow$$
$$E \rightarrow E\ A\ E \mid (E) \mid -\ E \mid \mathbf{id}$$
$$A \rightarrow + \mid - \mid \times \mid \div \mid \uparrow$$

Derivações

Definição de derivação

Seja E um não-terminal, $E \rightarrow \alpha_1 \mid \alpha_2 \mid \dots \mid \alpha_N$ produções- E e $\sigma = \beta E \gamma$. É dito que σ deriva $\beta \alpha_i \gamma$, e notamos $\sigma \Rightarrow \beta \alpha_i \gamma$, se uma das instâncias de E em σ é substituída por uma das alternativas α_i das produções- E .

Uma sequência de substituições em σ que resulte em X é chamada derivação de X a partir de σ .

Derivações em zero ou mais passos

Definição em zero ou mais passos

Se $\alpha_1 \Rightarrow \alpha_2 \Rightarrow \dots \Rightarrow \alpha_n$, então α_1 deriva α_n . O símbolo \Rightarrow significa “deriva em um passo”. O símbolo $\overset{*}{\Rightarrow}$ significa “deriva em zero ou mais passos” e o símbolo $\overset{+}{\Rightarrow}$ significa “deriva em um ou mais passos”.

A derivação em zero ou mais passos tem duas importantes propriedades:

1. $\alpha \overset{*}{\Rightarrow} \alpha$ para qualquer cadeia α , e
2. se $\alpha \overset{*}{\Rightarrow} \beta$ e $\beta \overset{*}{\Rightarrow} \gamma$, então $\alpha \overset{*}{\Rightarrow} \gamma$

Exemplo de derivação da expressão $-(\text{id} + \text{id}) \times \text{id}$

$$\begin{aligned} E &\Rightarrow E \times E \\ &\Rightarrow - E \times E \\ &\Rightarrow - (E) \times E \\ &\Rightarrow - (E + E) \times E \\ &\Rightarrow - (E + E) \times \text{id} \\ &\Rightarrow - (\text{id} + E) \times \text{id} \\ &\Rightarrow - (\text{id} + \text{id}) \times \text{id} \end{aligned}$$

Linguagem gerada por G

Definição

Seja G uma gramática e S um símbolo de partida. O conjunto $L(G)$, denominado linguagem gerada por G , contém uma cadeia w se, e somente se, w contém apenas não-terminais e $S \xRightarrow{+} w$. A cadeia w é denominada uma sentença de G . Uma linguagem que pode ser gerada por uma gramática é chamada linguagem livre de contexto.

Se duas gramáticas G_1 e G_2 geram a mesma linguagem, então G_1 e G_2 são ditas equivalentes.

Se $S \xRightarrow{*} \alpha$, onde α pode conter não-terminais, α é denominada uma forma sentencial de G .

Árvores gramaticais e derivações

- ▶ A ordem de substituição em uma derivação é arbitrária

Árvores gramaticais e derivações

- ▶ A ordem de substituição em uma derivação é arbitrária
- ▶ Convém, portanto, assumir uma ordem de substituição, sendo as mais comuns substituir sempre o não-terminal mais à esquerda (gerando a forma sentencial mais à esquerda) ou mais à direita (derivações canônicas)

Árvores gramaticais e derivações

- ▶ A ordem de substituição em uma derivação é arbitrária
- ▶ Convém, portanto, assumir uma ordem de substituição, sendo as mais comuns substituir sempre o não-terminal mais à esquerda (gerando a forma sentencial mais à esquerda) ou mais à direita (derivações canônicas)
- ▶ Uma árvore gramatical pode ser vista como uma representação gráfica de uma derivação que determine uma ordem de substituição

Árvores gramaticais e derivações

- ▶ A ordem de substituição em uma derivação é arbitrária
- ▶ Convém, portanto, assumir uma ordem de substituição, sendo as mais comuns substituir sempre o não-terminal mais à esquerda (gerando a forma sentencial mais à esquerda) ou mais à direita (derivações canônicas)
- ▶ Uma árvore gramatical pode ser vista como uma representação gráfica de uma derivação que determine uma ordem de substituição
- ▶ A ordem escolhida definirá o formato da árvore

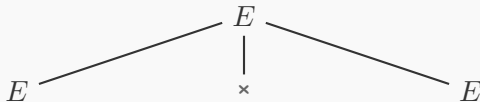
Árvores gramaticais e derivações

- ▶ A ordem de substituição em uma derivação é arbitrária
- ▶ Convém, portanto, assumir uma ordem de substituição, sendo as mais comuns substituir sempre o não-terminal mais à esquerda (gerando a forma sentencial mais à esquerda) ou mais à direita (derivações canônicas)
- ▶ Uma árvore gramatical pode ser vista como uma representação gráfica de uma derivação que determine uma ordem de substituição
- ▶ A ordem escolhida definirá o formato da árvore
- ▶ Uma gramática que produz mais de uma árvore gramatical para alguma sentença é dita ambígua

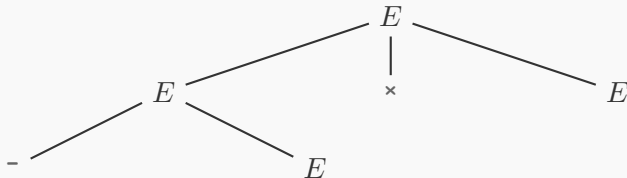
Árvore gramatical da derivação da expressão $-(id + id) \times id$

E

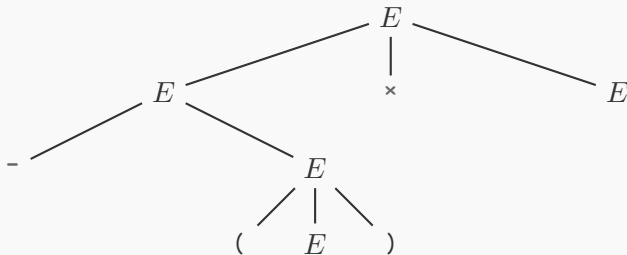
Árvore gramatical da derivação da expressão $-(\text{id} + \text{id}) \times \text{id}$



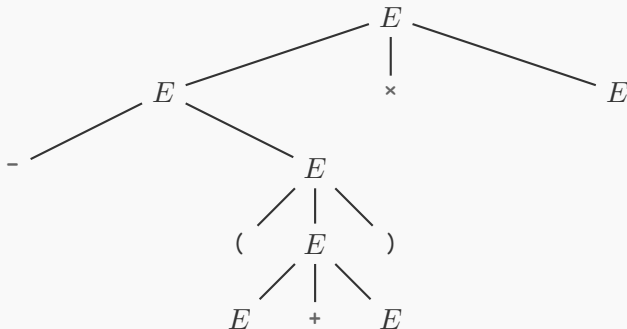
Árvore gramatical da derivação da expressão $-(\text{id} + \text{id}) \times \text{id}$



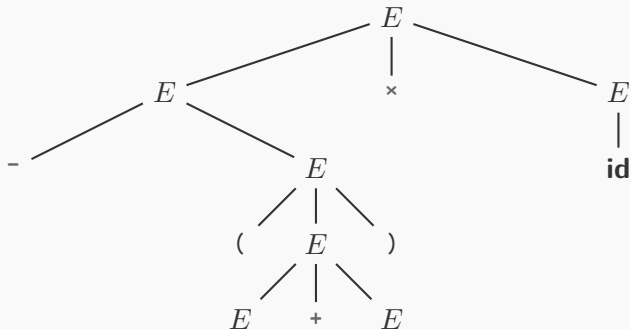
Árvore gramatical da derivação da expressão $-(id + id) \times id$



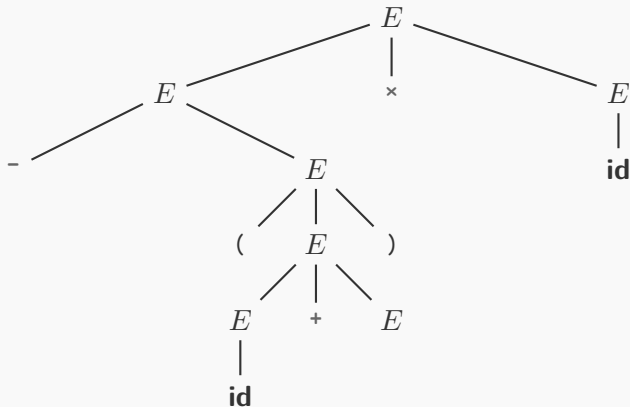
Árvore gramatical da derivação da expressão $-(id + id) \times id$



Árvore gramatical da derivação da expressão $-(\text{id} + \text{id}) \times \text{id}$



Árvore gramatical da derivação da expressão $-(\text{id} + \text{id}) \times \text{id}$



Árvore gramatical da derivação da expressão $-(\text{id} + \text{id}) \times \text{id}$

