Teoria da Computação
Conceitos Básicos
Gramáticas

Prof. Jefferson Magalhães de Morais

10 de março de 2021

Gramáticas

- São **dispositivos generativos e de síntese** (geração de cadeias)
- Baseiam-se em **regras de substituição**
- As regras **sintetizam** o conjunto das cadeias que compõem uma linguagem
- Construídas com **notações matemáticas rigorosas**
- As notações recebem o nome de **metalinguagens**

Definição

- Uma gramática G pode ser definida como sendo uma quádrupla

$$G = (V, \Sigma, P, S)$$

  - V: é o conjunto (finito e não-vazio) de símbolos, i.e., o vocabulário
  - Σ: é o conjunto (finito e não-vazio) dos símbolos terminais (alfabeto)
  - P: é o conjunto (finito e não-vazio) de produções ou regras de substituição
  - S: é a raiz (ou símbolo inicial), $S \in V$
  - $N = V - \Sigma$ é o conjunto dos símbolos não-terminais

Gramáticas

- Σ tem símbolos que por justaposição formam sentenças da linguagem
- N tem símbolos intermediários utilizados na geração de sentenças
- P é o conjunto das produções gramaticais, que obedecem à **forma geral**

$$\alpha \rightarrow \beta, \text{ com } \alpha \in V^*NV^* \text{ e } \beta \in V^*$$

> α é uma cadeia qualquer constituída por elementos de V, contendo pelo menos um símbolo não terminal, e β é uma cadeia qualquer, eventualmente vazia, de elementos de V

- "$\rightarrow$" é uma relação sobre os conjuntos V^*NV^* e V^*, uma vez que

$$P = \{(\alpha, \beta) \mid (\alpha, \beta) \in V^*NV^* \times V^*\}$$

Exemplo

- Seja $G_1 = (V_1, \Sigma_1, P_1, S)$, com
  - $V_1 = \{0, 1, 2, 3, S, A\}$
  - $\Sigma_1 = \{0, 1, 2, 3\}$
  - $N_1 = \{S, A\}$
  - $P_1 = \{S \to 0S33, S \to A, A \to 12, A \to \varepsilon\}$
- É fácil verificar que G_1 está formulada de acordo com as regras gerais acima enunciadas para a especificação de gramáticas

Forma sentencial

- É uma cadeia obtida pela aplicação das seguintes **regras de substituição**
  1. *S* é por definição uma forma sentencial
  2. Seja $\alpha\rho\beta$ uma forma sentencial, com α e β cadeias quaisquer de terminais e/ou não terminais, e seja $\rho \rightarrow \gamma$ uma produção da gramática. A aplicação dessa produção à forma sentencial produz uma nova forma sentencial $\alpha\gamma\beta$

Derivação direta

- A substituição anterior é chamada de **derivação direta**, na forma

$$\alpha\rho\beta \Rightarrow_G \alpha\gamma\beta$$

G informa que a produção aplicada $(\rho \to \gamma) \in P$

- Pode-se suprimir G quando a gramática é facilmente identificada

- Note a diferenciação dos símbolos
  - "$\to$" denota a produção da gramática (relação P)
  - "$\Rightarrow$" denota a derivação

Derivação, derivação não-trivial e sentença

- **Derivação** é uma sequência de zero ou mais derivações diretas $\alpha \Rightarrow \beta \Rightarrow \ldots \Rightarrow \mu$, e pode ser abreviada como $\alpha \Rightarrow^* \mu$

- **Derivação não-trivial** é aquela em que ocorre a aplicação de pelo menos uma produção, denotada por $\alpha \Rightarrow^+ \mu$

- Se da aplicação de uma derivação não-trivial à raiz S, for possível obter uma cadeia w formada exclusivamente de símbolos terminais, então
  - w é uma forma sentencial
  - w é uma **sentença**
- E a notação utilizada é da forma

$$S \Rightarrow^+ w$$

Gramáticas

- O processo de substituição começa na raiz S e finaliza assim que uma forma sentencial isenta de símbolos não-terminais seja obtida

- **Exemplo**
  $G_1 = \{\{0, 1, 2, 3, S, A\}, \{0, 1, 2, 3\}, \{S \rightarrow 0S33, S \rightarrow A, A \rightarrow 12, A \rightarrow \varepsilon\}, S\}$
  - S é por definição uma **forma sentencial**
  - $0S33$ é uma **forma sentencial**, pois $S \Rightarrow 0S33$
  - $00S3333$ e $00A3333$ são **formas sentenciais**, pois $0S33 \Rightarrow 00S3333 \Rightarrow 00A3333$ através da produção $S \rightarrow 0S33$ e $S \rightarrow A$
  - $S \Rightarrow 0S33$ é uma **derivação direta**
  - $00S3333 \Rightarrow^* 00S3333$ e $0S33 \Rightarrow^* 00A3333$ são exemplos de **derivações**
  - $S \Rightarrow^+ 00A3333$ e $S \Rightarrow^+ 0S33$ são **derivações não-triviais**
  - 12 e 00123333 são **sentenças**

Linguagem definida por uma gramática

- É o conjunto de todas as sentenças w geradas por uma gramática G
- Formalmente

$$L(G) = \{w \in \Sigma^* \mid S \Rightarrow^+ w\}$$

- **Exemplo**
  - Pela inspeção das produções da gramática G_1, pode concluir que

$$L_1(G_1) = \{0^m 1^n 2^n 3^{2m} \mid m >= 0 \text{ e } (n = 0 \text{ ou } n = 1)\}$$

  - São exemplos de sentenças pertencentes a $L_1 : \varepsilon, 12, 033, 01233, 003333, 00123333, \ldots$

Linguagem definida por uma gramática

- Definir $L(G)$ é uma tarefa que exige **abstração** e **prática**
- **Exemplo**: considere $G_2 = \{V_2, \Sigma_2, P_2, S\}$
  - $V_2 = \{a, b, c, S, B, C\}$
  - $\Sigma_2 = \{a, b, c\}$
  - $P_2 = \{S \rightarrow aSBC, S \rightarrow abC, CB \rightarrow BC, bB \rightarrow bb, bC \rightarrow bc, cC \rightarrow cc\}$

- A linguagem gerada por G_2 é $\{a^n b^n c^n \mid n >= 1\}$
  - Repetidas aplicações de $S \rightarrow aSBC$, uma aplicação de $S \rightarrow abC$, repetidas aplicações de $CB \rightarrow BC$ e $bB \rightarrow bb$ resulta em

    $$S \Rightarrow^i a^i S(BC)^i \Rightarrow a^i abC(BC)^i \Rightarrow^i a^{i+1} bB^i C^{i+1} \Rightarrow^i a^{i+1} b^{i+1} C^{i+1}$$

  - Uma aplicação de $bC \rightarrow bc$, repetidas aplicações de $cC \rightarrow cc$, resulta em

    $$\Rightarrow^i a^{i+1} b^{i+1} cC^i \Rightarrow^i a^{i+1} b^{i+1} c^{i+1}$$

- A forma sequencial

$$a^{i+1}b^{i+1}c^{i+1}$$

gera as sentenças *aabbcc*, *aaabbbccc*, etc. A sentença *aabbcc*, por exemplo, é derivada da seguinte forma nessa gramática

$$S \Rightarrow aSBC \Rightarrow aabCBC \Rightarrow aabBCC \Rightarrow aabbCC \Rightarrow aabbbcC \Rightarrow aabb$$

pela aplicação, respectivamente, das produções

$S \rightarrow aSBC$, $S \rightarrow abC$, $CB \rightarrow BC$, $bB \rightarrow bb$, $bC \rightarrow bc$ e $cC \rightarrow cc$

Gramáticas equivalentes

- Uma mesma linguagem pode ser definida por meio de duas ou mais gramáticas distintas
- As gramáticas que definem a mesma linguagem são **sintaticamente equivalentes** (ou **equivalentes**) uma à outra
- **Exemplo**
  - As gramáticas G_3 e G_4 a seguir são equivalentes

    $$G_3 = (\{a, b, S\}, \{a, b\}, \{S \to aS, S \to a, S \to bS, S \to b, S \to aSb\}, S$$
    $$G_4 = (\{a, b, S, X\}, \{a, b\}, \{S \to XS, S \to X, X \to a, X \to b\}, S)$$

  - Conclui-se que $L_3(G_3) = L_4(G_4) = \{a, b\}^+$

Notações

- Existem diversas notações (metalinguagens) para expressar **gramáticas**
- Utilizamos até o momento a **notação algébrica**
- Alguns exemplos
  - Expressões regulares
  - BFN (*Backus-Naur Form* ou Notação de Backus-Naur)
  - Notação de Wirth (expressões regulares estendidas)
  - Diagramas de sintaxe (diagramas ferroviários)

Exercício

- Obter gramáticas que geram as linguagens seguintes. Considerar $\Sigma = \{a, b\}$
    - Começam com *aa*
    - Não começam com *aa*
    - Contém a subcadeia *aabb*
    - Possuem comprimento maior ou igual a 3
    - Possuem comprimento par
    - Possuem comprimento ímpar
    - Possuem comprimento múltiplo de 4