■■ série livros didáticos informática ufrgs ■■

bookman

grupo a
Conhecimento que transforma.

.Inf
INSTITUTO
DE INFORMÁTICA
UFRGS

Linguagens Formais e Autômatos

Paulo Blauth Menezes

Linguagens Formais e Autômatos

P. Blauth Menezes

2 – Linguagens e Gramáticas

2 Linguagens e Gramáticas

Linguagem: Dicionário Aurélio

o uso da palavra articulada ou escrita como
meio de expressão e comunicação entre pessoas

Não é suficientemente precisa para

- desenvolvimento matemático de uma teoria baseada em linguagens

Linguagem

- conceito fundamental em computação e informática

Para definir linguagem

- alfabeto
- palavra ou cadeia de caracteres

2 – Linguagens e Gramáticas

2.1 Alfabeto

Símbolo ou caractere

- entidade abstrata básica, não definida formalmente
- base para definições
- exemplos: letras e dígitos

Def: Alfabeto

Conjunto finito de símbolos ou caracteres

Portanto

- conjunto infinito *não* é alfabeto
- $\varnothing$ *é* um alfabeto

Exp: Alfabeto

São alfabetos

- { a, b, c }
- ∅ (conjunto vazio)

Não são alfabetos (por quê?)

- $\mathbb{N}$ (conjunto dos números naturais)
- { a, b, aa, ab, ba, bb, aaa,… }

Exp: Alfabeto: linguagem de programação

Alfabeto de uma linguagem de programação como Pascal

- o conjunto de todos os símbolos usados nos programas
  - ∗ letras
  - ∗ dígitos
  - ∗ caracteres especiais como “**>**”, “**/**”, etc
  - ∗ espaço ou “branco”

Alfabeto binário { a, b }

- domínio de valores de um bit
- analogia com a representação interna dos computadores reais
- poucos símbolos: simplifica as diversas abordagens desenvolvidas

2 – Linguagens e Gramáticas

2.1 Alfabeto
2.2 Palavra
2.3 Linguagem Formal
2.4 Gramática

2.2 Palavra

Def: Palavra, cadeia de caracteres, sentença

Sobre um alfabeto

- sequência finita de símbolos justapostos

Cadeia sem símbolos

ε - cadeia vazia ou palavra vazia

Def: Prefixo, sufixo, subpalavra

Prefixo (Sufixo)

- qualquer sequência inicial (final) de símbolos da palavra

Subpalavra

- qualquer sequência de símbolos contíguos da palavra

Exp: Palavra, prefixo, sufixo, subpalavra

abcb palavra sobre o alfabeto { a, b, c }

- ε, a, ab, abc, abcb são todos os prefixos
- ε, b, cb, bcb, abcb são todos os sufixos
- qualquer prefixo ou sufixo é uma subpalavra

Exp: Palavra: linguagem de programação

Em uma linguagem de programação como Pascal

uma palavra é um programa

Def: Concatenação de palavras

Concatenação de palavras ou simplesmente concatenação

- operação binária sobre um conjunto de palavras
- associa a cada par de palavras
  - palavra formada pela justaposição da primeira com a segunda

Notação

- justaposição dos símbolos que representam as palavras componentes

Propriedades

- Elemento neutro: $\varepsilon\, w = w = w\, \varepsilon$
- Associativa: $v(w\,t) = (v\,w)t$

Associatividade – parênteses podem ser omitidos: $v\,w\,t$

Exp: Concatenação de palavras

$\Sigma = \{ a, b \}$ um alfabeto. Para $v = baaaa$ e $w = bb$

- $v w = baaaabb$
- $v \varepsilon = v = baaaa$

Def: Concatenação sucessiva de uma palavra

Concatenação sucessiva de uma palavra (com ela mesma) ou simplesmente concatenação sucessiva

w^n onde n é o número de concatenações sucessivas

indutivamente a partir da operação de concatenação

- $w^0 = \varepsilon$
- $w^n = w\, w^{n-1}$, para $n > 0$

Exp: Concatenação sucessiva

w palavra e a símbolo

- $w^3 = w\,w\,w$
- $w^1 = w$
- $a^5 = aaaaa$
- $a^n = aaa\ldots a$ (o símbolo a repetido n vezes)

Se Σ é um alfabeto

- Σ^* conjunto de todas as palavras possíveis sobre Σ
- $\Sigma^+ = \Sigma^* - \{\varepsilon\}$

Def: Conjunto de todas as palavras

Σ alfabeto. Σ^* é indutivamente definido

Base de indução

- $\varepsilon \in \Sigma^*$
- para qualquer $x \in \Sigma$, vale $x \in \Sigma^*$

Passo de indução

- Se u e v são palavras de Σ^*,

- então a concatenação $u\,v$ é uma palavra de Σ^*

Definição alternativa para palavra sobre um alfabeto Σ

- qualquer elemento w de Σ^*

$$w \in \Sigma^*$$

Exp: Conjunto de todas as palavras

Se $\Sigma = \{ a, b \}$, então:

- $\Sigma^+ = \{ a, b, aa, ab, ba, bb, aaa, \ldots \}$
- $\Sigma^* = \{ \varepsilon, a, b, aa, ab, ba, bb, aaa, \ldots \}$

Def: Comprimento, tamanho de uma palavra

De uma palavra w, representado por $|w|$

- número de símbolos que compõem a palavra
- função com domínio em Σ^* e codomínio em $\mathbb{N}$

Exp: Palavra, prefixo, sufixo, comprimento

$|abcb| = 4$

$|\varepsilon| = 0$

2 – Linguagens e Gramáticas

2.1 Alfabeto
2.2 Palavra
2.3 Linguagem formal
2.4 Gramática

2.3 Linguagem formal

Def: Linguagem formal

Linguagem formal ou simplesmente linguagem L sobre um alfabeto Σ

$$L \subseteq \Sigma^*$$

Exp: Linguagem formal

∅ e {ε} são linguagens sobre qualquer alfabeto

$$\varnothing \neq \{\varepsilon\}$$

Σ^* e Σ^+ são linguagens sobre um Σ qualquer

$$\Sigma^* \neq \Sigma^+$$

Conjunto de palíndromos sobre $\Sigma = \{a, b\}$

ε, a, b, aa, bb, aaa, aba, bab, bbb, aaaa,…

Exp: Conjunto de todas as linguagens sobre um alfabeto

Conjunto das partes de Σ^*

$$2^{\Sigma^*}$$

Exp: Linguagem formal: linguagem de programação

Linguagem de programação como Pascal

conjunto de todos os programas (palavras) da linguagem

2 – Linguagens e Gramáticas

2.4 Gramática

Linguagem de programação

- definida pelo conjunto de todos os programas (palavras)

Linguagem de propósitos gerais como Pascal

- conjunto de todos os programas é *infinito*
- não é definição adequada para implementação em computador

Formalismo gramática

- uma maneira de especificar de forma *finita* linguagens (eventualmente) *infinitas*

Gramática é, basicamente

- conjunto *finito* de regras
- quando aplicadas sucessivamente, geram palavras
- conjunto de todas as palavras geradas por uma gramática
  - define a linguagem

Gramáticas para linguagens naturais como Português

- as mesmas que as usadas para linguagens artificiais como Pascal

Gramáticas também são usadas para definir semântica

- entretanto, em geral, são usados outros formalismos

Def: Gramática

Gramática de Chomsky, Gramática irrestrita ou gramática

$$G = (V, T, P, S)$$

- V, conjunto *finito* de símbolos variáveis ou não terminais
- T, conjunto *finito* de símbolos terminais disjunto de V
- P: $(V \cup T)^+ \rightarrow (V \cup T)^*$, relação *finita*: produções
    - par da relação: regra de produção ou produção
- S, elemento distinguido de V: símbolo inicial ou variável inicial

Representação de uma regra de produção (α, β)

$$\alpha \rightarrow \beta$$

Representação abreviada para $\alpha \rightarrow \beta_1$, $\alpha \rightarrow \beta_2$, ..., $\alpha \rightarrow \beta_n$

$$\alpha \to \beta_1 \mid \beta_2 \mid \ldots \mid \beta_n$$

Derivação

- aplicação de uma regra de produção é denominada derivação

- aplicação sucessiva de regras de produção
  - fecho transitivo da relação de derivação
  - permite derivar palavras da linguagem

Def: Relação de Derivação

$G = (V, T, P, S)$ gramática

Derivação é um par da relação de derivação denotada por $\Rightarrow$

- domínio em $(V \cup T)^+$ e codomínio em $(V \cup T)^*$
- $\langle \alpha, \beta \rangle$ é representado de forma infixada

$$\alpha \Rightarrow \beta$$

$\Rightarrow$ é indutivamente definida como segue:

- para toda produção da forma $S \rightarrow \beta$ (S é o símbolo inicial de G)

$$S \Rightarrow \beta$$

- para todo par $\eta \Rightarrow \rho \alpha \sigma$ da relação de derivação
  - se $\alpha \rightarrow \beta$ é regra de P, então

$$\eta \Rightarrow \rho \beta \sigma$$

Portanto, derivação

- substituição de uma subpalavra
- de acordo com uma regra de produção

Sucessivos passos de derivação

- $\Rightarrow^*$ fecho transitivo e reflexivo da relação $\Rightarrow$
  - zero ou mais passos de derivações sucessivos
- $\Rightarrow^+$ fecho transitivo da relação $\Rightarrow$
  - um ou mais passos de derivações sucessivos
- $\Rightarrow^i$
  - exatos i passos de derivações sucessivos (i natural)

Gramática é um formalismo

- aximático
- de geração
  - permite derivar ("gerar") todas as palavras da linguagem

Def: Linguagem gerada

$G = (V, T, P, S)$ gramática

Linguagem gerada por G: $L(G)$ ou $GERA(G)$

- palavras de símbolos terminais deriváveis a partir de S

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

Exp: Gramática, derivação, linguagem gerada: números naturais

$G = (V, T, P, N)$

- $V = \{ N, D \}$
- $T = \{ 0, 1, 2, \ldots, 9 \}$
- $P = \{ N \rightarrow D, N \rightarrow DN, D \rightarrow 0 \mid 1 \mid \ldots \mid 9 \}$

Gera, sintaticamente, o conjunto dos números naturais

- se distinguem os zeros à esquerda
- exemplo: 123 de 0123

Exp: Gramática, derivação, linguagem gerada: números naturais

$G = (V, T, P, N)$

- $V = \{ N, D \}$
- $T = \{ 0, 1, 2, \ldots, 9 \}$
- $P = \{ N \rightarrow D, N \rightarrow DN, D \rightarrow 0 \mid 1 \mid \ldots \mid 9 \}$

Uma derivação do número 243

- $N \Rightarrow$ $\qquad N \rightarrow DN$
- $DN \Rightarrow$ $\qquad D \rightarrow 2$
- $2N \Rightarrow$ $\qquad N \rightarrow DN$
- $2DN \Rightarrow$ $\qquad D \rightarrow 4$
- $24N \Rightarrow$ $\qquad N \rightarrow D$
- $24D \Rightarrow$ $\qquad D \rightarrow 3$
- 243

Portanto

- $S \Rightarrow^* 243$
- $S \Rightarrow^+ 243$
- $S \Rightarrow^6 243$

Interpretação indutiva da gramática

- *Base de indução*: todo dígito é natural
- *Passo de indução*: se n é natural, então a concatenação com qualquer dígito também é natural

Exp: Gramática, derivação, linguagem gerada: palavra duplicada

$$G = (\{ S, X, Y, A, B, F \}, \{ a, b \}, P, S)$$

na qual:

- $P = \{ S \to XY,$
- $X \to XaA \mid XbB \mid F$
- $Aa \to aA,\ Ab \to bA,\ AY \to Ya,$
- $Ba \to aB,\ Bb \to bB,\ BY \to Yb,$
- $Fa \to aF,\ Fb \to bF,\ FY \to \varepsilon \}$

gera a linguagem

$$\{ ww \mid w \text{ é palavra de } \{ a, b \}^* \}$$

Derivação de baba

- $S \Rightarrow$
- $XY \Rightarrow$
- $XaAY \Rightarrow$
- $XaYa \Rightarrow$
- $XbBaYa \Rightarrow$
- $XbaBYa \Rightarrow$
- $XbaYba \Rightarrow$
- $FbaYba \Rightarrow$
- $bFaYba \Rightarrow$
- $baFYba \Rightarrow$
- $baba$

$S \rightarrow XY$
$X \rightarrow XaA$
$AY \rightarrow Ya$
$X \rightarrow XbB$
$Ba \rightarrow aB$
$BY \rightarrow Yb$
$X \rightarrow F$
$Fb \rightarrow bF$
$Fa \rightarrow aF$
$FY \rightarrow \varepsilon$

Existe mais alguma derivação de baba?

Def: Gramáticas equivalentes

G_1 e G_2 são gramáticas equivalentes se e somente se

$$GERA(G_1) = GERA(G_2)$$

Convenções

- A, B, C,…, S, T para símbolos variáveis
- a, b, c,…, s, t para símbolos terminais
- u, v, w, x, y, z para palavras de símbolos terminais
- α, β,… para palavras de símbolos variáveis ou terminais

Linguagens Formais e Autômatos

P. Blauth Menezes