Linguagens Formais e Autômatos

Conceitos Básicos

Cristiano Lehrer, M.Sc.

Introdução (1/3)

- A **Teoria das Linguagens Formais** foi originariamente desenvolvida na década de 1950 com o objetivo de desenvolver teorias relacionadas com as linguagens naturais.
- Entretanto, logo foi verificado que esta teoria era importante para o estudo das linguagens artificiais e, em especial, para as linguagens originárias na Ciência da Computação.

Introdução (2/3)

- A Teoria dos Autômatos lida com as definições e propriedades de modelos matemáticos de computação.
- Esses modelos desempenham um papel em diversas áreas aplicadas da Ciência da Computação.
- Um modelo, chamado **Autômato Finito**, é usado em processamento de texto, compiladores e projeto de hardware.
- Outro modelo, denominado **Gramática Livre do Contexto**, é utilizado em linguagens de programação e inteligência artificial.

Introdução (3/3)

- O estudo das Linguagens Formais desenvolveu-se significativamente e com diversos enfoques, com destaque para:
  - Aplicações em análise léxica e sintática de linguagens de programação.
  - Modelos de sistemas biológicos.
  - Desenho de hardware.
  - Relacionamentos com linguagens naturais.
- Recentemente, inclui-se animações, hipertextos e hipermídias.

Alfabeto (1/2)

- Cadeias de caracteres são blocos básicos fundamentais em Ciência da Computação.
- O alfabeto sobre o qual as cadeias são definidas pode variar com a aplicação.
- Definimos um **alfabeto** como sendo qualquer conjunto finito não vazio.
- Os membros do alfabeto são os **símbolos** do alfabeto.
- Geralmente usamos a letra grega Σ para designar alfabetos.

Alfabeto (2/2)

- Alfabeto binário:
  - $\Sigma = \{0, 1\}$
- Alfabeto romano:
  - $\Sigma = \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z\}$
- Alfabeto decimal:
  - $\Sigma = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

Palavras (1/2)

- Uma **palavra**, **cadeia de caracteres** ou **sentença** sobre um alfabeto é uma sequência finita de símbolos daquele alfabeto, geralmente escritos um seguido do outro e não separados por vírgulas.
- A **palavra vazia**, representada pelo símbolo ε, é uma palavra sem símbolos.
- Exemplos:
  - Se $\Sigma = \{0, 1\}$ então 01001 é uma palavra sobre Σ.
  - Se $\Sigma = \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z\}$, então abracadabra é uma palavra sobre Σ.

Palavras (2/2)

- O conjunto de todas as palavras possíveis sobre Σ é representado por Σ^*, enquanto que Σ^+ representa o conjunto de todas as palavras possíveis sobre Σ com exceção da palavra vazia, ou seja:
  - $\Sigma^+ = \Sigma^* - \{\varepsilon\}$

Tamanho

- O **tamanho** ou **comprimento** de uma palavra é o seu comprimento como uma sequência, logo, o comprimento da palavra casa é 4.
- Denotamos o comprimento de uma palavra *w* por |*w*|, portanto:
  - |futebol| = 7
  - $|\varepsilon| = 0$

Prefixo e Sufixo

- O **prefixo** de uma palavra é qualquer sequência de símbolos iniciais da palavra:
  - Por exemplo, os possíveis prefixos da palavra livro são:
    - ε, l, li, liv, livr, livro
- O **sufixo** de uma palavra é qualquer sequência de símbolos finais da palavra:
  - Por exemplo, os possíveis sufixos da palavra livro são:
    - ε, o, ro, vro, ivro, livro

Subpalavra

- A **subpalavra** de uma palavra é qualquer sequência de símbolos contígua da palavra.
  - Por exemplo, as possíveis subpalavras da palavra livro são:
    - ε, l, i, v, r, o, li, iv, vr, ro, liv, ivr, vro, livr, ivro, livro
  - Observe que qualquer prefixo ou sufixo de uma palavra é uma subpalavra dessa palavra.

Reverso

- O **reverso** de uma palavra $w \in \Sigma^*$, escrito w^R, é a palavra obtida escrevendo-se w na ordem inversa, isto é:
  - $w_n w_{n-1} \ldots w_1$
- Por exemplo, se w = teoria então w^R = airoet
- Quando $w = w^R$, dizemos que w é uma palavra palíndroma, como por exemplo, $w = w^R$ = arara.

Concatenação (1/3)

- Se temos a palavra *x* de comprimento *m* e a palavra *y* de comprimento *n*, a **concatenação** de *x* e *y*, escrito *xy*, é a palavra obtida concatenando-se *y* ao final de *x*, como $x_1...x_m y_1...y_n$.
- Para concatenar uma cadeia com si própria, utilizamos a notação de expoente, como w^n, onde *w* é uma palavra e *n* indica o número de concatenações sucessivas.

Concatenação (2/3)

- A operação de concatenação satisfaz às seguintes propriedades (suponha v, w e t palavras):
  - Associatividade → $v(wt) = (vw)t$
  - Elemento neutro à esquerda e à direita → $\varepsilon w = w = w\varepsilon$

Concatenação (3/3)

- A operação de concatenação sucessiva é definida indutivamente a partir da concatenação binária, como segue:
  - Caso 1 $\rightarrow$ $w \neq \varepsilon$
    - $w^0 = \varepsilon$
    - $w^n = w^{n-1}w$, para $n > 0$
  - Caso 2 $\rightarrow$ $w = \varepsilon$
    - w^0 é indefinida
    - $w^n = \varepsilon$, para $n > 0$

Ordenação Lexicográfica

- A **ordenação lexicográfica** de palavras é a mesma que a ordenação familiar do dicionário, exceto que as palavras mais curtas precedem as palavras mais longas.
- Por conseguinte, a ordenação lexicográfica de todas as palavras sobre o alfabeto $\Sigma = \{0, 1\}$ é
  - $\{\varepsilon, 0, 1, 00, 01, 10, 11, 000, ...\}$

Linguagem Formal

- Uma **Linguagem Formal** ou simplesmente **Linguagem** é um conjunto de palavras sobre um alfabeto.
- Suponha o alfabeto $\Sigma = \{a, b\}$, então:
  - O conjunto vazio e o conjunto formado pela palavra vazia são linguagens sobre Σ:
    - Obviamente, $\varnothing \neq \{\varepsilon\}$
  - O conjunto de palíndromos (palavras que têm a mesma leitura da esquerda para a direita e vice-versa) sobre Σ é um exemplo de linguagem infinita.
    - Assim: ε, a, b, aa, bb, aaa, aba, bab, bbb, aaaa, abba, ...