Teoria de Linguagem

Alfabetos, Palavras e Linguagens Formais

Vinicius H. S. Durelli
✉ durelli@ufsj.edu.br

UFSJ
UNIVERSIDADE FEDERAL DE SÃO JOÃO DEL-REI

Organização

O que é linguagem?

O que é linguagem?

Dicionário Aurélio

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

O que é linguagem?

Dicionário Aurélio

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

➟ Claramente, **a definição acima não é suficiente** para permitir o desenvolvimento matemático de uma teoria baseada em linguagens.

Alfabetos

Definição → *Alfabeto*

Um conjunto finito de símbolos ou caracteres (Sipser 2012; Menezes 2011).[a]

[a] Os membros de um alfabeto são os **símbolos** do alfabeto.

Iremos utilizar letras gregas maiúsculas Σ e Γ para denotar alfabetos. Exemplos de alfabetos:

$$\Sigma_1 = \{0, 1\}$$

$$\Sigma_2 = \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z\}$$

$$\Gamma = \{0, 1, x, y\}$$

Quiz

Pergunta ①**:** O conjunto $\mathbb{N}$ é um exemplo de alfabeto? Por quê?

Pergunta ②**:** O conjunto vazio (i.e., $\emptyset$) é um alfabeto?

Quiz

Pergunta ①: O conjunto $\mathbb{N}$ é um exemplo de alfabeto? Por quê?

Não, pois o conjunto $\mathbb{N}$ (conjunto dos números naturais) é infinito.

Pergunta ②: O conjunto vazio (i.e., $\emptyset$) é um alfabeto?

Quiz

Pergunta ①: O conjunto $\mathbb{N}$ é um exemplo de alfabeto? Por quê?

Não, pois o conjunto $\mathbb{N}$ (conjunto dos números naturais) é infinito.

Pergunta ②: O conjunto vazio (i.e., $\emptyset$) é um alfabeto?

Sim.

Palavras

Definição $\rightarrow$ *Palavra*

Uma *string*, cadeia, sentença ou palavra é uma sequência finita de símbolos do alfabeto justapostos (Menezes 2002).

Exemplos:

- Se $\Sigma_1 = \{0, 1\}$, então 01001 é uma cadeia de símbolos em Σ_1.
- Se $\Sigma_2 = \{a, b, c, \ldots, x, y, z\}$, então *abracadabra* é uma palavra sobre Σ_2.

Tamanho de uma palavra...

Se w é uma palavra sobre Σ, o tamanho (ou comprimento) de w, denotado $|w|$, é o número de símbolos que formam a palavra.
Exemplos:

$$\begin{aligned} |1| &= 1 \\ |10| &= 2 \\ |0000| &= 4 \\ |001101| &= 6 \end{aligned}$$

Palavra vazia: uma palavra de tamanho zero é denominada **palavra vazia**. O símbolo usado para representar tal palavra é ε.
Portanto:

$$|\varepsilon| = 0$$

Mais sobre palavras...

- Se uma palavra w tem tamanho n, podemos dizer que $w = w_1, w_2, w_3, \ldots, w_n$, onde cada $w_i \in \Sigma$.
- O **reverso** de uma palavra w, denotado por w^R, é a palavra obtida quando inverte-se a ordem dos símbolos de w: $w_n, \ldots, w_3, w_2, w_1$.
  Exemplos:
  $banana^R = ananab$ $\qquad$ $teoria^R = airoet$
- Dado um alfabeto Σ, é possível denotar o conjunto de todas as palavras de tamanho k usando a notação Σ^k.
  Por exemplo, dado $\Sigma = \{0, 1\}$, então:
  $\Sigma^1 = \{0, 1\}$ $\qquad$ $\Sigma^2 = \{00, 01, 10, 11\}$

Mais sobre palavras e alfabetos: fechamento recursivo e transitivo

O **fechamento recursivo e transitivo** definido como o conjunto de todas as palavras sobre um alfabeto Σ é normalmente denotado por Σ^*.

Definição $\rightarrow$ *Fechamento Recursivo e Transitivo*

O fechamento recursivo[a] e transitivo é formalmente definido como a seguir:

$$\Sigma^* = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \cup \cdots = \bigcup_{i=0}^{\infty} \Sigma^i$$

[a] Às vezes chamado fechamento flexivo e transitivo.

Mais sobre palavras e alfabetos: fechamento transitivo

Em alguns casos, é conveniente remover a palavra vazia do conjunto de palavras. O **fechamento transitivo** representa o conjunto de palavras **não vazias** sobre um alfabeto Σ e é denotado por Σ^+.

Definição $\rightarrow$ *Fechamento Transitivo*

O fechamento transitivo é formalmente definido como a seguir:

$$\Sigma^+ = \Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \cup \cdots = \bigcup_{i=1}^{\infty} \Sigma^i$$

Portanto, como se pode perceber:

$$\Sigma^* = \Sigma^+ \cup \{\varepsilon\}$$

Concatenação de palavras (1)...

Dado duas palavras α e β de comprimentos n e m respectivamente. Então, pode-se dizer que $\alpha\beta$ representa a concatenação de α e β, ou seja, representa a palavra formada por uma cópia de α seguida por uma cópia da palavra β

- Mais precisamente, se α é uma palavra composta de i símbolos $\alpha = a_1 a_2 a_3 a_4 \cdots a_i$ e se β é uma palavra composta de j símbolos $\alpha = b_1 b_2 b_3 b_4 \cdots b_j$, então $\alpha\beta$ tem comprimento $i + j$: $\alpha\beta = a_1 a_2 a_3 a_4 \cdots a_i b_1 b_2 b_3 b_4 \cdots b_j$.

Exemplo: Dado $\alpha = 01101$ e $\beta = 110$, então $\alpha\beta = 01101110$ e $\beta\alpha = 11001101$.

Concatenação de palavras (2)...

A notação de expoente é utilizada para concatenar uma palavra com si própria. Dado uma palavra w, w^n representa a concatenação de w com si própria n vezes:

$$\overbrace{ww w \cdots w}^{n}$$

Exemplo: Dado $w = 0011$, então $w^3 = 001100110011$

Propriedades → *Concatenação*

A operação de concatenação satisfaz as seguintes propriedades:

- Associatividade[a]
- Elemento neutro à esquerda e à direita

[a] Embora associativa, a concatenação não é comutativa.

Concatenação de palavras (2)...

A notação de expoente é utilizada para concatenar uma palavra com si própria. Dado uma palavra w, w^n representa a concatenação de w com si própria n vezes:

$$\overbrace{www\cdots w}^{n}$$

Exemplo: Dado $w = 0011$, então $w^3 = 001100110011$

Propriedades → *Concatenação*

A operação de concatenação satisfaz as seguintes propriedades:

- Associatividade[a] → $v(wt) = (vw)t$
- Elemento neutro à esquerda e à direita

[a] Embora associativa, a concatenação não é comutativa.

Concatenação de palavras (2)...

A notação de expoente é utilizada para concatenar uma palavra com si própria. Dado uma palavra w, w^n representa a concatenação de w com si própria n vezes:

$$\overbrace{ww w \cdots w}^{n}$$

Exemplo: Dado $w = 0011$, então $w^3 = 001100110011$

Propriedades → *Concatenação*

A operação de concatenação satisfaz as seguintes propriedades:

- Associatividade[a] → $v(wt) = (vw)t$
- Elemento neutro à esquerda e à direita → $\varepsilon w = w = w\varepsilon$

[a] Embora associativa, a concatenação não é comutativa.

Prefixos, sufixos e subpalavras

- Diz-se que uma palavra α é **prefixo** de uma outra palavra β se é possível escrever β como $\alpha\gamma$.
- Uma palavra α é **sufixo** de outra palavra β se é possível escrever β como $\gamma\alpha$.
- Dado quatro palavras, α, β, γ e δ, uma palavra α é denominada **subpalavra** de uma palavra β quando $\beta = \gamma\alpha\delta$. (Se γ e δ ou ambos forem vazios, a definição ainda se aplica.)

Linguagens (formais)

Definição $\rightarrow$ *Linguagem*

Uma **linguagem formal** é um conjunto (finito ou infinito) de cadeias de comprimento finito formadas pela concatenação de um alfabeto finito e não vazio.
Mais formalmente, uma linguagem pode ser vista como o conjunto de todas as palavras que podem ser selecionadas de um dado alfabeto Σ, i.e., Σ^* (Hopcroft et al. 2006).

Exemplo:

A linguagem de todas as palavras que consistem de n 0s seguidos de n 1s para todo $n \geqslant 0$: $\{\varepsilon, 01, 0011, 000111, \ldots\}$

Quiz

O conjunto vazio (i.e., $\emptyset$) e o conjunto formado pela palavra vazia (i.e., $\{\varepsilon\}$) são linguagens sobre qualquer alfabeto.

Pergunta ③: $\emptyset = \{\varepsilon\}$? Por quê?

Pergunta ④: Qual é o conjunto resultante de Σ^0?

Quiz

O conjunto vazio (i.e., $\emptyset$) e o conjunto formado pela palavra vazia (i.e., $\{\varepsilon\}$) são linguagens sobre qualquer alfabeto.

Pergunta ③: $\emptyset = \{\varepsilon\}$? Por quê?

Obviamente, $\emptyset \neq \{\varepsilon\}$.

Pergunta ④: Qual é o conjunto resultante de Σ^0?

Quiz

O conjunto vazio (i.e., $\emptyset$) e o conjunto formado pela palavra vazia (i.e., $\{\varepsilon\}$) são linguagens sobre qualquer alfabeto.

Pergunta ③: $\emptyset = \{\varepsilon\}$? Por quê?

Obviamente, $\emptyset \neq \{\varepsilon\}$.

Pergunta ④: Qual é o conjunto resultante de Σ^0?

Independentemente de qual seja o alfabeto, $\Sigma^0 = \varepsilon$, pois ε é a única palavra de tamanho 0.

Como descrever linguagens formais?

No decorrer da disciplina, usaremos a notação para construção de conjuntos (*set-builder notation*) para descrever linguagens complexas ou que contém muitas palavras. Considere o seguinte exemplo:

$$L = \{w | \text{ w contém o mesmo número de 0s e 1s}\}$$

ou

$$L = \{w : \text{ w contém o mesmo número de 0s e 1s}\}$$

1. Contextualização. . .
2. Alfabetos
3. Palavras
4. Linguagens
5. Considerações finais

Considerações finais. . .

Na aula de hoje nós vimos:

- Alfabetos;
- Palavras; e
- Linguagens.

Na **próxima aula**:

- Autômatos finitos determinísticos.

Referências

Menezes, Paulo Blauth (2002). *Linguagens Formais e Autômatos*. 3rd ed. Livros Didáticos do Instituto de Informática da UFRGS. Sagra Luzzatto, p. 165.

Hopcroft, John E., Rajeev Motwani, & Jeffrey D. Ullman (2006). *Introduction to Automata Theory, Languages, and Computation*. 3rd ed. Pearson, p. 750.

Menezes, Paulo Blauth (2011). *Linguagens Formais e Autômatos*. 6th ed. Livros Didáticos Informática da UFRGS. Bookman, p. 256.

Sipser, Michael (2012). *Introduction to the Theory of Computation*. 3rd ed. Cengage Learning, p. 480.

"Quiz" icon by iconsphere, from the Noun Project (`https://thenounproject.com/`).