

IBM1088 Linguagens Formais e Teoria da Computação

Linguagens e Gramáticas

Evandro Eduardo Seron Ruiz
evandro@usp.br

UNIVERSIDADE DE SÃO PAULO

Sofremos muito com o pouco que nos falta e gozamos pouco o muito que temos.

William Shakespeare^a

^aDramaturgo e poeta inglês (1564 – 1616).

Objetivo deste tópico:

- Definir formalmente dois conceitos, **linguagem** e **gramática**
- Preparar outros conceitos e propriedades para o estudo das linguagens formais

- 1 Alfabeto
- 2 Palavra
- 3 Linguagem formal
- 4 Gramática

Linguagem

O que é:

Para CC^a sabemos que é um conceito (+ a ferramenta embutida) fundamental em Computação, no processamento da informação e na Informática.

^aMesmo antes de definir formalmente o que é linguagem.

Linguagem segundo Aurélio

Definição

A linguagem é o uso da palavra articulada, ou escrita, como meio de expressão, ou comunicação entre pessoas.

Linguagem segundo Aurélio

Definição

A linguagem é o uso da palavra articulada, ou escrita, como meio de expressão, ou comunicação entre pessoas.

- Definição ampla, pouco clara para Ciências Exatas
- Insuficiente para o estudo de linguagens em Computação
- Falta formalismo (Álgebra, Matemática, um modelo)
- Não permite o desenvolvimento lógico (algébrico, matemático)

Linguagem: início da definição

Vamos então definir linguagem através de dois outros conceitos:

Alfabeto

Palavra

Alfabeto

Definição

Um **alfabeto** é um conjunto finito de símbolos ou caracteres.

Alfabeto

Definição

Um **alfabeto** é um conjunto finito de símbolos ou caracteres.

Portanto:

- Um conjunto infinito não é um alfabeto
- Um conjunto vazio ($\emptyset$) é um alfabeto

Exemplos

São exemplos de alfabetos:

- $\{a, b, c\}$
- $\{\}$
- $A = \{\forall x \in \mathbb{N} \mid x \leq 10\}$

Exemplos

São exemplos de alfabetos:

- $\{a, b, c\}$
- $\{\}$
- $A = \{\forall x \in \mathbb{N} \mid x \leq 10\}$

Não são exemplos de alfabetos:

- $\mathbb{N}$
- $\{a, ab, aa, ab, bb, aaa, \dots\}$

Exemplos

São exemplos de alfabetos:

- $\{a, b, c\}$
- $\{\}$
- $A = \{\forall x \in \mathbb{N} \mid x \leq 10\}$

Não são exemplos de alfabetos:

- $\mathbb{N}$
- $\{a, ab, aa, ab, bb, aaa, \dots\}$

Representação

Σ é normalmente usado para representar um alfabeto, o conjunto dos símbolos usados na linguagem.

Exercício em sala: 1 minuto

Qual é o alfabeto da linguagem de programação **C**?

Alfabeto e linguagem de programação

O alfabeto de uma linguagem de programação, como a linguagem **C**, é o conjunto de todos os símbolos usados na construção de programas, tais como:

- Letras;
- Dígitos;
- Caracteres especiais; e
- Espaço em branco

Alfabeto e linguagem de programação

O alfabeto de uma linguagem de programação, como a linguagem **C**, é o conjunto de todos os símbolos usados na construção de programas, tais como:

- Letras;
- Dígitos;
- Caracteres especiais; e
- Espaço em branco

Alfabeto binário

$\Sigma = \{a, b\}$. Usaremos muito este alfabeto. Similar ao alfabeto real de hardware binário baseado em 0 e 1. Facilita o desenvolvimento de abordagens relacionadas a LFA.

Palavra

Definição

Uma **palavra** (ou cadeia de caracteres) sobre um alfabeto é uma seqüência finita de símbolos justapostos do alfabeto.

Notem que, pela definição acima, uma cadeia sem símbolos é uma palavra válida.

Notação: ϵ = palavra vazia ou cadeia vazia

Prefixo e sufixo

Definição

Um **prefixo** de uma palavra é qualquer seqüência inicial de símbolos da palavra.

Prefixo e sufixo

Definição

Um **prefixo** de uma palavra é qualquer seqüência inicial de símbolos da palavra.

Definição

Um **sufixo** de uma palavra é qualquer seqüência final de símbolos da palavra.

Exemplos

- **abcd** é uma palavra sobre $\Sigma = \{a, b, c, d\}$
- $\epsilon, a, ab, abc, abcd$ são prefixos na palavra *abcd*
- $\epsilon, d, cd, bcd, abcd$ são sufixos na palavra *abcd*

Exemplos

- **abcd** é uma palavra sobre $\Sigma = \{a, b, c, d\}$
- $\epsilon, a, ab, abc, abcd$ são prefixos na palavra *abcd*
- $\epsilon, d, cd, bcd, abcd$ são sufixos na palavra *abcd*

Definição

Uma **subpalavra** de uma palavra é qualquer seqüência de símbolos contíguos da palavra.

Portanto, qualquer prefixo ou sufixo é uma subpalavra.

Outro exemplo!!!

Palavra

Considerando uma linguagem de programação como **C**, um programa é uma palavra.

Concatenação de palavras

Definição

A **concatenação de palavras**, ou simplesmente *concatenação*, é uma operação binária, definida sobre um conjunto de palavras, a qual associa a cada par de palavras uma palavra formada pela justaposição da primeira com a segunda.

Exemplo de concatenação

Suponha o alfabeto $\Sigma = \{a, b\}$. Suponha as palavras $v = baaaa$ e $w = bb$.

Vale que:

- $vw = baaaa**bb**$

Exemplo de concatenação

Suponha o alfabeto $\Sigma = \{a, b\}$. Suponha as palavras $v = baaaa$ e $w = bb$.

Vale que:

- $vw = baaaa**bb**$
- $v\epsilon = v = baaaa$

Exemplo de concatenação

Suponha o alfabeto $\Sigma = \{a, b\}$. Suponha as palavras $v = baaaa$ e $w = bb$.

Vale que:

- $vw = baaaa\textcolor{red}{bb}$
- $v\epsilon = v = baaaa$

Usamos aqui duas propriedades da concatenação:

Elemento neutro $\epsilon w = w\epsilon = w$

Associatividade $v(wt) = (vw)t$

Concatenação sucessiva

Definição

A **concatenação sucessiva** de uma palavra, representada na forma de um expoente de uma palavra (e.g.: 'w'), ou seja:

w^n , tal que, n é o número de concatenações sucessivas,

é definida indutivamente como:

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

Percebam como ϵ é realmente uma palavra em Σ .

Exemplo

Sejam: a um símbolo de Σ , w uma palavra e $n \in \mathbb{N}$. Então:

- $w^3 = www$
- $w^1 = w$
- $a^n = \underbrace{aaaa \dots a}_n$

Importante!

Notações importantes

Se Σ representa um alfabeto, então:

- Σ^* denota o conjunto de todas as palavras possíveis em Σ ; e
- $\Sigma^+ = \Sigma^* - \{\epsilon\}$

Conjunto de todas as palavras

Definição

Seja Σ um alfabeto. Então o conjunto de todas as palavras, Σ^* , é indutivamente definido como:

Base da indução:

$$\epsilon \in \Sigma^*$$

$$\forall x \in \Sigma, x \in \Sigma^*$$

Passo da indução

Se u e v são palavras em Σ^* então a concatenação uv é uma palavra em Σ^* .

Definição alternativa

Definição de palavra

Assim, uma definição alternativa para **palavra** ou **sentença** sobre um alfabeto Σ é qualquer elemento **w** de Σ^* , ou seja,

$$w \in \Sigma^*$$

Exemplo

Considerando $\Sigma = \{a, b\}$, então:

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

Comprimento de uma palavra

Definição

O **comprimento**, ou **tamanho**, de uma palavra **w**, representado por $|w|$, é o número de símbolos que compõem a palavra.

Comprimento de uma palavra

Definição

O **comprimento**, ou **tamanho**, de uma palavra **w**, representado por $|w|$, é o número de símbolos que compõem a palavra.

- Representação do comprimento da palavra w : $|w|$
- Domínio: Σ^*
- Imagem: $\mathbb{N}$

Exemplos

- $|abcd| = 4$
- $|\epsilon| = 0$

Linguagem formal

Definição

Uma **linguagem formal**, ou simplesmente uma **linguagem** L definida sobre um alfabeto Σ , é um conjunto de palavras sobre Σ , ou seja,

$$L \subseteq \Sigma^*$$

Lembrando... Σ^* são as palavras possíveis, incluindo ϵ .

Lembrando...

Palíndromo

Um palíndromo é uma palavra, ou qualquer outra sequência símbolos, cujo significado pode ser interpretado igualmente se lida tanto da direita para a esquerda como da esquerda para a direita.

Exemplos: anilina, Renner, 'A cara rajada da jararaca'.

Exemplos de linguagens

- Σ^* e Σ^+ são linguagens sobre Σ . Lembrando que

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

Exemplos de linguagens

- Σ^* e Σ^+ são linguagens sobre Σ . Lembrando que

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

- Seja $\Sigma = \{a, b\}$, então o conjunto de palíndromos sobre Σ é um exemplo de linguagem infinita.

São palavras desta linguagem $\epsilon, a, b, aa, bb, aba, bab, aaa, \dots$

Exemplos de linguagens

- Σ^* e Σ^+ são linguagens sobre Σ . Lembrando que

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

- Seja $\Sigma = \{a, b\}$, então o conjunto de palíndromos sobre Σ é um exemplo de linguagem infinita.
São palavras desta linguagem $\epsilon, a, b, aa, bb, aba, bab, aaa, \dots$
- Uma linguagem de programação

Exemplos de linguagens

- Σ^* e Σ^+ são linguagens sobre Σ . Lembrando que

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

- Seja $\Sigma = \{a, b\}$, então o conjunto de palíndromos sobre Σ é um exemplo de linguagem infinita.
São palavras desta linguagem $\epsilon, a, b, aa, bb, aba, bab, aaa, \dots$
- Uma linguagem de programação
- $\emptyset$ e $\{\epsilon\}$ são linguagens sobre qualquer alfabeto. Saibam que:

$$\emptyset \neq \{\epsilon\}$$

Diferença conceitual entre $\emptyset$ e $\{\epsilon\}$

- Imaginem uma palavra: *abab*

Diferença conceitual entre $\emptyset$ e $\{\epsilon\}$

- Imaginem uma palavra: $abab$
- $abab = abab\epsilon$

Diferença conceitual entre $\emptyset$ e $\{\epsilon\}$

- Imaginem uma palavra: $abab$
- $abab = abab\epsilon$
- Imaginem a leitura, caracter a caracter, desta palavra
- Existirá o momento da leitura de ϵ , o final da cadeia
- $\emptyset$ é a ausência de palavra
- $w = \epsilon$ é a palavra vazia, mas é uma palavra

Diferença conceitual entre $\emptyset$ e $\{\epsilon\}$, continua

- Imaginem a linguagem $L = \{a, ab, abab\}$
- $\epsilon \notin L$, ou seja, a cadeia vazia não faz parte de L

Diferença conceitual entre $\emptyset$ e $\{\epsilon\}$, continua

- Imaginem a linguagem $L = \{a, ab, abab\}$
- $\epsilon \notin L$, ou seja, a cadeia vazia não faz parte de L
- Mas, como em todo conjunto, $\emptyset \in L$
- $2^L =$
 $\{\emptyset, \{a\}, \{ab\}, \{abab\}, \{aab\}, \{aabab\}, \{ababab\}, \{a, ab, abab\}\}$

Gramática: definição informal

É basicamente um conjunto **finito** de regras para gerar palavras. O conjunto de todas as palavras geradas pela gramática define uma linguagem.

Assim. . .

- Um programa = palavra
- Conjunto de todas as palavras de uma gramática de uma LP = Linguagem de programação

Assim. . .

- Um programa = palavra
- Conjunto de todas as palavras de uma gramática de uma LP = Linguagem de programação
- Conjunto de todos os programas (e.g.: Na linguagem **C**)
- **Porém**, esta não é uma definição formalmente adequada, ou algebricamente ideal para o estudo de LFA

Gramática

Definição

Uma gramática de Chomsky, ou uma gramática **irrestrita**, aqui denotada por G , é uma quádrupla ordenada

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

em que

- V é um conjunto de símbolos **não terminais**
- T é um conjunto de símbolos **terminais**, para T disjunto de V
- $P : (V \cup T)^+ \rightarrow (V \cup T)^*$ é uma relação finita chamada de **relação de produções**. Cada par é denominado **regra de produção** ou *produção*
- S é um elemento de V denominado **símbolo inicial**.

Em termos práticos... Nem tanto!

Uma regra de produção pode ser representada como:

$$\alpha \rightarrow \beta$$

Um grupo de regras de produção da forma

$$\alpha \rightarrow \beta_1; \alpha \rightarrow \beta_2; \dots; \alpha \rightarrow \beta_n$$

e pode ser abreviado como:

$$\alpha \rightarrow \beta_1 \mid \beta_2 \mid \dots \mid \beta_n$$

Em termos práticos... Nem tanto!

Uma regra de produção pode ser representada como:

$$\alpha \rightarrow \beta$$

Um grupo de regras de produção da forma

$$\alpha \rightarrow \beta_1; \alpha \rightarrow \beta_2; \dots; \alpha \rightarrow \beta_n$$

e pode ser abreviado como:

$$\alpha \rightarrow \beta_1 \mid \beta_2 \mid \dots \mid \beta_n$$

Importante!

As regras de produção definem as condições de geração das palavras da linguagem.

Derivação: informalmente

Derivação

A aplicação de uma regra de produção é denominada **derivação** de uma **palavra** e é formalmente definida como um par de uma relação.

Derivação: informalmente

Derivação

A aplicação de uma regra de produção é denominada **derivação** de uma **palavra** e é formalmente definida como um par de uma relação.

Nota:

A aplicação sucessiva de regras de produção (*Fecho Transitivo da Relação de Derivação*) permite derivar as palavras da linguagem representada pela gramática.

Relação de derivação

Definição

Seja $G = (V, T, P, S)$ uma gramática. Uma **derivação** é um par da *relação de derivação* denotada por $\Rightarrow$ com domínio em $(V \cup T)^+$ e imagem (codomínio) $(V \cup T)^*$.^a

Um par $\langle \alpha, \beta \rangle$ da relação de derivação é representado na forma infixa como:

$$\alpha \Rightarrow \beta.$$

Esta relação é definida como:

$$^a(V \cup T)^+ \Rightarrow (V \cup T)^*$$

próximo slide. . .

Relação de derivação: continua...

Definição

Esta relação é definida como:

- Para toda produção na forma $S \rightarrow \beta$, (S símbolo inicial em G) o seguinte par pertence a relação de derivação

$$S \Rightarrow \beta$$

- Para todo par $\eta \Rightarrow \rho \alpha \sigma$ da relação de derivação, se $\alpha \rightarrow \beta$ é regra de P , então o seguinte par também pertence à relação de derivação

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

Vemos que uma derivação é a substituição de uma subpalavra de acordo com uma regra de produção.

Notem que:

- Existe uma **derivação** que é um par de $(V \cup T)^+ \Rightarrow (V \cup T)^*$
- Um par da relação de derivação $\langle \alpha, \beta \rangle$ é representado por $\alpha \Rightarrow \beta$;
- Esta relação é definida através de **regras de produção** na forma inicial $S \rightarrow \beta$

Uma coisa é uma coisa, outra coisa é...

Derivação A derivação de uma cadeia é consequência da aplicação de regras de produção que transformam um símbolo inicial nesta cadeia.

A derivação prova que a cadeia pertence a linguagem descrita pela gramática.

Regras de produção É a definição da regra de derivação.

Passos de derivação

Os sucessivos passos de derivação são definidos como:

- $\Rightarrow^*$ Fecho transitivo e reflexivo da relação $\Rightarrow$, ou seja, **zero** ou mais passos sucessivos de derivação;
- $\Rightarrow^+$ Fecho transitivo da relação $\Rightarrow$, ou seja, **um** ou mais passos sucessivos de derivação;
- $\Rightarrow^i$ Exatos i passos de derivações sucessivas, $i \in \mathbb{N}$.

Linguagem gerada

Definição

Seja $G = (V, T, P, S)$ uma gramática. A **linguagem gerada** por G , denotada por $L(G)$, é composta por todas as palavras de símbolos terminais deriváveis a partir do símbolo inicial S , ou seja,

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

Exemplo

Suponha $G = (V, T, P, S)$ uma gramática capaz de gerar qualquer número natural válido. Assim G , na qual:

- $V = \{N, D\}$, símbolos não terminais
- $T = \{0, 1, 2, \dots, 9\}$, símbolos terminais
- $P = \{N \rightarrow D, N \rightarrow DN, D \rightarrow 0 \mid 1 \mid \dots \mid 9\}$

gera, sintaticamente, $\mathbb{N}$.

Perceba que G gera distintamente 0123 e 123

Regra $\times$ Passo da derivação

Passo da derivação	Regra usada
$N \Rightarrow$	$N \rightarrow DN$
$DN \Rightarrow$	$D \rightarrow 2$
$2N \Rightarrow$	$N \rightarrow DN$
$2DN \Rightarrow$	$D \rightarrow 4$
$24N \Rightarrow$	$N \rightarrow D$
$24D \Rightarrow$	$D \rightarrow 3$
243	

Existe alguma outra derivação para 243?

Regra $\times$ Passo da derivação

Passo da derivação	Regra usada
$N \Rightarrow$	$N \rightarrow DN$
$DN \Rightarrow$	$D \rightarrow 2$
$2N \Rightarrow$	$N \rightarrow DN$
$2DN \Rightarrow$	$D \rightarrow 4$
$24N \Rightarrow$	$N \rightarrow D$
$24D \Rightarrow$	$D \rightarrow 3$
243	

Existe alguma outra derivação para 243? Portanto, indica-se que:

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

Outro exemplo

Seja $G = (\{S, X, Y, A, B, F\}, \{a, b\}, P, S)$ uma gramática na qual:

$$\begin{aligned}
 P = \{ & S \rightarrow XY, \\
 & X \rightarrow XaA \mid XbB \mid F \\
 & Aa \rightarrow aA, Ab \rightarrow bA, AY \rightarrow Ya, \\
 & Ba \rightarrow aB, Bb \rightarrow bB, BY \rightarrow Yb, \\
 & Fa \rightarrow aF, Fb \rightarrow bF, FY \rightarrow \epsilon \}.
 \end{aligned}$$

G gera a linguagem cujas palavras são tais que a primeira metade é igual a segunda metade, ou seja,

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

Veja a seguir:

Regra $\times$ Passo da derivação para a palavra **baba**

Passo da derivação	Regra usada
$S \Rightarrow$	$S \rightarrow XY$
$XY \Rightarrow$	$X \rightarrow XaA$
$XaAY \Rightarrow$	$AY \rightarrow Ya$
$XaYa \Rightarrow$	$X \rightarrow XbB$
$XbBaYa \Rightarrow$	$Ba \rightarrow aB$
$XbaBYa \Rightarrow$	$BY \rightarrow Yb$
$XbaYba \Rightarrow$	$X \rightarrow F$
$FbaYba \Rightarrow$	$Fb \rightarrow bF$
$bFaYba \Rightarrow$	$Fa \rightarrow aF$
$baFYba \Rightarrow$	$FY \rightarrow \epsilon$
<i>baba</i>	

Exercício em sala: 5 minutos

Existe alguma outra derivação para **baba**?

Caros,

Teremos mais na próxima aula, inclusive exercícios.

Estudem, pois novos conceitos dependerão dos conceitos estudados hoje.