

Linguagens Formais

UNICAP

Eduardo Araújo Oliveira
<http://sites.google.com/site/eaoufpe>



INTRODUÇÃO

slide 2

Estrutura

1. Motivação
2. Conceitos
 - Alfabeto e Linguagens
3. Operadores

slide 3

Linguagens Formais

- Parte da área de Teoria da Computação
- Estuda a definição formal da sintaxe das linguagens
- Sintaxe vs. Semântica
 - Sintaxe: é o formato
 - Semântica: é o significado

4

Linguagens Formais

- Aplicações práticas
 - Representar de maneira precisa a sintaxe de uma linguagem
 - Desafio: Tentar descrever C de maneira informal
 - Criação de compiladores mais eficientes

5

Conceitos

- Alfabeto
 - Conjunto finito
 - Elementos são chamados de **símbolos**
 - Notação: T ou Σ
- Símbolos
 - Analogia com valores do tipo "char" em C/C++/Java
 - Notação: usaremos letras minúsculas ou números – **a, b, c, ..., 0, 1**

6

Conceitos

- Palavras ou Cadeias
 - Sequência de símbolos justapostos
 - Analogia com strings em programação
 - "abaaa", "bbaa"

7

Conceitos

- Cadeias vazia
 - Analogia com a string "" de C ou Java
 - Não tem nenhum caractere
 - Notação: ε (ou λ)

8

Conceitos

- Conjunto de todas as cadeias sobre um alfabeto T dado
 - $T^* = \{ \varepsilon, a, b, aa, ab, ba, bb, aaa, \dots \}$
 - Inclui a cadeia vazia
 - Inclui cadeias de qualquer tamanho

9

Conceitos

- Linguagem L sobre alfabeto T
 - Qualquer conjunto de cadeias,
 - ou
 - Qualquer subconjunto de T^*

10

Conceitos

- C ou Java são linguagens finitas ou infinitas? É possível listar todos os programas válidos dessas linguagens?
- Como expressar de maneira finita linguagens (potencialmente) infinitas?

11

Linguagens Formais

- No decorrer da disciplina veremos como representar linguagens de maneira
 - Precisa, sem ambigüidade
 - Finita, mesmo que a linguagem seja infinita

12

Conceitos

- Exemplo
 - Alfabeto T_a : todos os caracteres ASCII
 - $T_a = \{ 'a', 'b', \dots, 'A', 'B', \dots, '*', '\n', ' ', \dots \}$
 - Cadeias: textos ASCII qualquer
 - $w = \text{"Aula de Linguagens Formais"}$
 - $y = \text{"39e72h0AG d"}$
 - T_a^* : todos os textos ASCII possíveis

13

Conceitos

- Linguagens sobre T_a
 - L_c = Linguagem C
 - As cadeias que descrevem programas C
 - Exemplo de cadeias/palavra da linguagem

```
#include <stdio.h>
int main() { printf("Hi!"); }
```
 - L_h = Linguagem HTML
 - Análogo...

14

Conceitos

- Inicialmente, serão vistas linguagens bem mais simples do que C ou HTML
 - A linguagem que só aceita o símbolo 'a'
 - A linguagem que reconhece identificadores válidos de C

15

Operações

- Veremos agora, algumas operações e relações sobre
 - **Cadeias/Palavras**
 - Tamanho, Concatenação, Reverso, Concatenação sucessiva, Subcadeia
 - Linguagens
 - União, Intersecção, Complemento e Concatenação

16

Tamanho

- Tamanho de uma cadeia: $|w|$
 - Número de ocorrências de símbolos
 - $|w| = n$, para $w = a_1a_2\dots a_n$
 - $|w| = 0$, para $w = \epsilon$

17

Reverso

- Reverso de uma cadeia: w^R
 - Símbolos aparecem na ordem inversa
 - Seja $w = a_1a_2\dots a_n$
 - Então teremos $w^R = a_na_{n-1}\dots a_1$

18

Concatenação

- O operador de concatenação justapõe os símbolos de duas cadeias formando uma nova cadeia

– Sejam $\mathbf{x} = "a_1...a_n"$ e $\mathbf{y} = "b_1...b_n"$

– $\mathbf{x.y} = \mathbf{x y} = "a_1...a_nb_1...b_n"$

x

y

x y

19

Concatenação

- Concatenação sucessiva w^n
 - A palavra w concatenada consigo n vezes

– $w^n = w w w w \dots w$

– Atenção: $w^0 \stackrel{\text{def}}{=} \epsilon$ n vezes

20

Concatenação

- Propriedades
 - Associatividade
$$v (w t) = (v w) t$$
 - Elemento neutro
$$w \varepsilon = w$$
$$\varepsilon w = w$$
 - $|x y| = |x| + |y|$

21

Subcadeia

- Subcadeia/subpalavra
 - Qualquer cadeia formada por uma parte contígua da seqüência de símbolos da palavra original
 - Dado $w = \text{"abacate"}$, são subpalavras de w :
 - "aba", "cat", "acate"
 - Não são subpalavras:
 - "abate", "bate", "ae"

22

Prefixo e Sufixo

- Prefixo
 - Subcadeia que acontece no começo
 - Exemplo: *"aba"* é prefixo de *"abacate"*
- Sufixo
 - Subcadeia que acontece no fim
 - Exemplo: *"te"* é sufixo de *"abacate"*

23

União/Intersecção

- Linguagens são conjuntos, então as todas as operações de conjuntos são válidas para elas
- Dadas $L_1 = \{ aa, b \}$ e $L_2 = \{ aa, c, dd \}$
 - $L_1 \cup L_2 = \{ aa, b, c, dd \}$
 - $L_1 \cap L_2 = \{ aa \}$

24

Complemento

- O conjunto T^* é tomado como conjunto universo para toda linguagem sobre um alfabeto T
- Assim, o complemento de L são todas as palavras de T^* que não estão em L

25

Complemento

- Sejam
 - $T = \{ a, b \}$
 - $T^* = \{ \varepsilon, a, b, aa, ab, ba, bb, aaa, aab, \dots \}$
 - $L = \{ w \mid w \text{ possui algum símbolo } a \}$
 $= \{ a, aa, ab, ba, aaa, aab, \dots \}$
- Então o complemento de L será
 - $\neg L = T^* - L$
 $= \{ w \mid w \text{ não tem nenhum símbolo } a \}$
 $= \{ \varepsilon, b, bb, \dots \}$

26

Referências

- Dois primeiros capítulos do livro em formato eletrônico
 - <http://teia.inf.ufrgs.br/library.html>

27

Linguagens Formais

UNICAP

Eduardo Araújo Oliveira
<http://sites.google.com/site/eaoufpe>