Linguagens Formais e Autômatos

Conceitos Básicos de Linguagens

Eduardo Furlan Miranda

Baseado em: GARCIA, A. de V.; HAEUSLER, E. H. Linguagens Formais e Autômatos. Londrina: EDA, 2017.

Alfabetos, palavras, cadeias, linguagens

- Linguagem - expressão de ideias. Aspectos:

foco da disciplina

- Léxicos ← análise léxica
  - Símbolos
- Sintáticos ← análise sintática
  - Regras
- Semânticos ← análise semântica
  - Significado

Alfabeto

- Alfabeto (ou vocabulário) ← representado pela letra Σ (sigma)
  - Conjunto finito não vazio de símbolos

- Linguagem sobre esse alfabeto
  - Conjunto de sequências finitas de símbolos do alfabeto

- Exemplos de alfabeto
  - $\Sigma_1 = \{ \alpha , \beta , \gamma , \delta , \dots , \omega \}$
    $\Sigma_2 = \{ 0 , 1 \}$

símbolos

Cadeia

$\Sigma_1 = \{ \alpha , \beta , \gamma , \delta , \ldots , \omega \}$
$\Sigma_2 = \{ 0, 1 \}$

- Cadeia (ou palavra, ou cadeia de símbolos)
  - Sequência qualquer de caracteres
    - de Σ_1 : "ψω"
    - de Σ_2 : "10001"
  - Cadeia com nenhum símbolo (ou vazia) = ε = “”
    - ε é uma cadeia sobre qualquer alfabeto

(épsilon)

  - Comprimento:
    - quantidade de símbolos
      - Ex.: (n = 2)

comprimento 2

- “ε”, “ϵ”, ou “”
- Imaginar ε como se fosse uma variável criada (ela existe), porém sem nenhum elemento dentro
  - Ex.: a = “”

Exemplo

- "10001" = comprimento 5
- Cadeia vazia = “” = ε
- Justaposição = concatenação de palavras
  - "ab" + "c" = "abc"
- Gramática = conjunto de regras
  - $S \rightarrow a\, S\, b$; $S \rightarrow \varepsilon$

cadeia vazia

Concatenação

- Dadas duas cadeias, definimos sua concatenação como a justaposição de seus valores
  - Ex.: se ω_1 ="101" e ω_2 ="000", sua concatenação é "101000"

- Representamos a concatenação como:
  - $\omega_1 \circ \omega_2$
  - $\omega_1 \bullet \omega_2$
  - $\omega_1 \, \omega_2$

neste caso a representação não é
“vezes”, é “concatenação”

Concatenação

- A concatenação é um operador associativo
  - $\omega_1 \circ (\omega_2 \circ \omega_3) = (\omega_1 \circ \omega_2) \circ \omega_3$
- Concatenação da cadeia vazia
  - $\varepsilon \circ \omega$
    - ω = "abc" , $\varepsilon \circ \omega$ = $\varepsilon \circ$ "abc" = "abc"
- Prefixos de uma cadeia (ex. a cadeia “abcdef”)
  - “ε”, “ab”, “abc”, “abcd”, etc.
    - começa do início e inclui a cadeia vazia

Prefixo e sufixo

- Prefixo
  - Dadas 2 cadeias ω_1 e ω_2 dizemos que
    - ω_1 é prefixo de ω_2 , se existe uma cadeia ω_3 tal que $\omega_1 \circ \omega_3 = \omega_2$
  - Ex.: a cadeia "101" possui os prefixos ε , "1", "10" e "101"

- Sufixo
  - Do final para o início
  - Ex.: a cadeia "100" possui os sufixos: “ε”, "0", "00" e "100"

- Dado um alfabeto Σ ,
  - definimos uma linguagem L sobre este alfabeto como
    - um conjunto de cadeias sobre este alfabeto. Ex.:
      - Alfabeto: $\Sigma = \{ a , b \}$
      - Linguagem: $L = \{ \varepsilon , a , b , a a , a b , b a , b b \}$

- Linguagem vazia = ∅
  - Não contém nenhuma cadeia
  - É simplesmente um conjunto vazio sem qualquer elemento

∅ = linguagem vazia (não possui nem ε)
ε = cadeia vazia

- A linguagem que contém apenas a cadeia vazia ε
  - contém exatamente uma palavra, que é a cadeia vazia ε

- Isso significa que o conjunto tem um elemento
  - esse elemento é uma cadeia que não contém nenhum símbolo

- O alfabeto $\Sigma_2 = \{ 0, 1 \}$ tem infinitas linguagens possíveis
  - $L_1 = \emptyset$
    - não possui nenhuma cadeia, nem a vazia ε
  - $L_2 = \{ \varepsilon \}$
    - possui apenas a cadeia vazia ε
  - $L_3 = \{ \varepsilon , 0, 1, 00, 01, 10, 11, 000, \dots \}$
    - possui todas as cadeias possíveis com símbolos do alfabeto Σ_2 (até o infinito)

Concatenação

- A concatenação de duas linguagens é uma linguagem cujas cadeias são todas as possíveis concatenações entre
  - cadeias da primeira linguagem

    com as

    cadeias da segunda linguagem

- Dadas as linguagens L_1 e L_2, definimos sua concatenação como a linguagem
  - $L_1 \circ L_2 = \{ \omega_1 \circ \omega_2 \mid \omega_1 \in L_1 \text{ e } \omega_2 \in L_2 \}$

“L_1 concatenado com L_2 é igual ao conjunto de todas as cadeias ω_1 concatenado com ω_2, tal que ω_1 pertence a L_1 e ω_2 pertence a L_2 “

Exemplo

- Se $L_1 = \{a, ab\}$ $(n = 2)$, e
  $L_2 = \{bb, c\}$ $(m = 2)$, então $L_1 \circ L_2$ será:

  $\{abb, ac, abbb, abc\}$

- O número de elementos em $L_1 \circ L_2$ é $2 * 2 = 4$

- Que linguagem é $L \circ \varnothing$?
  - $\varnothing \rightarrow (n = 0)$ não possui nenhuma cadeia, nem a vazia ε
  - $0 \cdot m = 0$
  - Portanto é uma linguagem vazia

não confundir

- Qual é a melhor definição para L^0 ?
  - $L^0 = \{ \varepsilon \}$
    - Linguagem que contém apenas a cadeia vazia ε
  - $L^1 = L$
    - L elevado a 1 é a própria linguagem L

Notação de potência

- $L^3 = L \circ L \circ L$ (L elevado a 3 é a concatenação de L com L com L)
- Definição
  - $L^0 = \{\varepsilon\}$ (caso base)
  - $L^n = L^{n-1} \circ L$, para $n > 0$ (passo recursivo)
    - Para calcular L^n , usamos L^{n-1} e concatenamos com L
    - Esse processo se repete até atingir a base da recursão, que é L^0

Exemplo

- Seja $L_1 = \{a, b\}$
  - $L_1^0 = \{\varepsilon\}$
  - $L_1^1 = \{a, b\}$
  - $L_1^2 = \{aa, ab, ba, bb\}$
  - $L_1^3 = \{aaa, aab, aba, abb, baa, bab, bba, bbb\}$
    - $L_1^3 = \{aa, ab, ba, bb\} \circ \{a, b\} =$
      $= \{aaa, aab, aba, abb, baa, bab, bba, bbb\}$
  - ...

ab ≠ ba

Fecho de Kleene (“*”)

- O fecho de Kleene é representado por *
  - $\{a, b\}^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, aab, aba, abb, baa, \ldots\}$
    - Conjunto infinito
    - Todas as combinações possíveis de símbolos
    - Incluindo o conjunto vazio ε
- L^* = fecho de Kleene da linguagem L
- $L^* = L^0 \cup L^1 \cup L^3 \cup \ldots$ $(L^0 = \varepsilon)$

$$\{a, b\} \cup \{b, c\} = \{a, b, c\}$$
$$\{a, b\} \circ \{c, d\} = \{ac, ad, bc, bd\}$$

Exemplo

- Seja L = {a, b}

- $L^* = L^0 \cup L^1 \cup L^3 \cup \ldots$
  - $L^0 = \{\varepsilon\}$
  - $L^1 = \{a, b\}$
  - $L^2 = \{aa, ab, ba, bb\}$
  - $L^3 = \{aaa, aab, aba, abb, baa, bab, bba, bbb\}$
  - ...

- Resultado parcial apenas para $L^0 \cup L^1 \cup L^2 \cup L^3 =$
  - $\{\varepsilon, a, b, aa, ab, ba, bb, aaa, aab, aba, abb, baa, bab, bba, bbb\}$

- Usando a concatenação de conjuntos, a união, e o fecho de Kleene, podemos especificar algumas linguagens simples
  - Considerando o alfabeto $\Sigma = \{ 0 , 1 \}$ podemos especificar algumas linguagens formadas por cadeias que representam números na base binária de numeração
    - L = { nº na base 2 que são múltiplos de 4 } = $\{ 0 , 1 \}^* \circ \{ 00 \}$:
      - 0100 4
        1000 8
        1100 12

Concatenação	= $\circ$	
União	= $\cup$	(ou $+$)
Fecho de Kleene	= *	

Fecho positivo (ou transitivo) (“$^+$”)

- $L^+ = L \circ L^*$

  Ex.: Seja $L = \{ a , b \}$

  - $L^* = \{\varepsilon, a, b, aa, ab, \ldots \}$
    - todas as combinações de a e b incluindo ε
  - $L \circ L^* = \{a\varepsilon, aa, ab, \ldots , b\varepsilon, ba, bb, \ldots \}$
    - o ε acaba “sumindo”
- $^+$ não possui a cadeia ε , a não ser que ε já pertença a L
- $L^+ = L^1 \cup L^2 \cup L^3 \cup \ldots$ (não tem o L^0)
- $L^* = L^0 \cup L^+$ (a união acrescenta o ε)

Monoide

- É uma estrutura algébrica possuindo as propriedades:
  - Fechamento
    - a°b concatenado com b°a resulta em a°b°b°a
      - o resultado é outra cadeia dentro do mesmo conjunto
  - Associatividade
    - (a°a)°b = a°(a°b)
      - a ordem das operações não altera o resultado
  - Elemento neutro (identidade)
    - ε°a = a°ε = a
      - elemento combinado com qualquer outro, resulta no elemento

Exemplo - Monoide

Σ^* = conjunto de todas as cadeias

- Seja o alfabeto $\Sigma = \{a, b\}$
  - Fechamento
    - a concatenação de duas cadeias em Σ^* resulta em outra cadeia também em Σ^*
      - "ab" ($\in \Sigma^*$) ° "ba" ($\in \Sigma^*$) = "abba" ($\in \Sigma^*$)
      - "a" ($\in \Sigma^*$) ° ε ($\in \Sigma^*$) = "a" ($\in \Sigma^*$)
  - Associatividade
    - ordem de agrupamento das operações não altera o resultado final
      - ("ab" ° "ba") ° "a" = "abba" ° "a" = "abbaa"
  - Elemento identidade
    - quando concatenado com qualquer cadeia, não a modifica
      - "ab" ° ε = "ab"
      - ε ° ε = ε

Definição de linguagem simples

- Seja o alfabeto $\Sigma = \{ n , + , \times \}$
  - n um número qualquer, + a soma, e × a multiplicação
- Definir uma linguagem L sobre Σ como a linguagem de todas as 'expressões' bem formadas usando-se as duas operações:
  - $L = \{ n, n+n, n\times n, n+n+n, n+n\times n, n\times n+n, n\times n\times n, \ldots \}$
  - Temos uma definição de linguagem que diz que a linguagem é simplesmente um conjunto de sentenças (cadeias)
    - Porém não define o aspecto estrutural da linguagem

- A escolha da forma de especificar uma linguagem depende do contexto e dos objetivos da aplicação

- Ao escolher um método, deve ser considerado a complexidade da linguagem, o objetivo da especificação e as ferramentas disponíveis

Exemplo

- Os numerais decimais são a linguagem definida pelo conjunto $\{ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 \}^+$

  $^+$ não tem ε

- Para restringir os numerais decimais com a impossibilidade de termos zeros à esquerda
  - $\{ 1, 2, 3, 4, 5, 6, 7, 8, 9 \} \circ \{ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 \}^*$

    * tem ε
    - $1 \circ \varepsilon = 1$, $2 \circ 0 = 2$, $3 \circ 45 = 345$, $1 \circ 000 = 1000$