Teoria da Computação
Linguagens Regulares (Parte 1)
Conjuntos e Expressões Regulares

Prof. Jefferson Magalhães de Morais

22 de março de 2021

Conjuntos regulares

- São **notações alternativas** que representam linguagens regulares
- **Conjuntos regulares** sobre um alfabeto Σ são linguagens definidas **recursivamente** da forma

  1. $\emptyset$ é um conjunto regular sobre Σ
  2. $\{\varepsilon\}$ é um conjunto regular sobre Σ
  3. $\{\sigma\}, \forall\sigma \in \Sigma$, é um conjunto regular sobre Σ

  Se X e Y são conjuntos regulares sobre Σ, então também são conjuntos regulares sobre Σ:

  4. (X)
  5. $X \cup Y$
  6. $X \cdot Y$, também denotado por XY
  7. X^*

Um subconjunto de Σ^* é um conjunto regular se ele puder ser formulado pela combinação das regras acima

Exemplo 1

- Seja $L = \{0^m 1^n \mid m \geq 0, n \geq 0\}$ sobre $\Sigma = \{0, 1\}$. Portanto

$$L = \{\varepsilon, 0, 1, 00, 01, 11, \ldots\}$$

- Considere as linguagens sobre Σ

$L_1 = \{0\}$	onde L_1 é um conjunto regular
$L_2 = \{1\}$	onde L_2 é um conjunto regular
$L_3 = \{0^i \mid i \geq 0\}$	onde $L_3 = L_1^*$ é um conjunto regular
$L_4 = \{1^i \mid i \geq 0\}$	onde $L_4 = L_2^*$ é um conjunto regular
$L_5 = \{0^p 1^q \mid p \geq 0, q \geq 0\}$	onde $L_5 = L_3 L_4 = L$ é um conjunto regular

Na notação de conjuntos regulares L é da forma $\{0\}^*\{1\}^*$

Exemplo 2

- A linguagem N formada pelos números naturais decimais é um **conjunto regular** sobre o alfabeto dos algarismos arábicos representada por

$$\{0,1,2,3,4,5,6,7,8,9\}\{0,1,2,3,4,5,6,7,8,9\}^*$$

  Se $D = \{0,1,2,3,4,5,6,7,8,9\}$, então $\mathbb{N} = DD^*$

- O conjunto R dos números reais decimais sem sinal é um **conjunto regular** sobre $D \cup \{.\}$ representado por

$$DD^*\{.\}D^* \cup D^*\{.\}DD^*$$

  Obs.: o conjunto acima exclui a cadeia "." e inclui números iniciando ou terminando com o caractere "." (e.g., $47.$ ou $.315$)

Exemplo 3

- O conjunto P dos números em ponto flutuante com exponente (denotado por "E") e sinal opcional ("$+$" ou "-")

$$\{+,-,\varepsilon\}(DD^*\{.\}D^* \cup D^*\{.\}DD^*)\{E\}\{+,-,\varepsilon\}DD^*$$

  Exemplos

  - $27 \in N$
  - $915.4 \in R$
  - $-211.56\mathrm{E}{+}10 \in P$

- Nota-se que $N \subset R \subset P, P \neq R$ e $R \neq N$

Expressões regulares

- A **expressão regular** é uma alternativa aos conjuntos regulares

- Foi desenvolvida por Stephen Cole Kleene na década de 50

- Também são definidas recursivamente da forma
  1. $\emptyset$ é uma expressão regular e denota o conjunto regular $\emptyset$
  2. ε é uma expressão regular e denota o conjunto regular $\{\varepsilon\}$
  3. $\sigma, \forall\sigma \in \Sigma$, é uma expressão regular e denota o conjunto $\{\sigma\}, \forall\sigma \in \Sigma$

     Se x e y são expressões regulares sobre Σ que denotam, respectivamente, os conjuntos regulares X e Y, então também são expressões regulares:
  4. (x)
  5. $x \mid y$ ou $x + y$
  6. $x \cdot y$ ou xy
  7. x^*

Expressões regulares

- As expressões regulares eliminam o uso dos símbolos "{" e "}", e substituiu o símbolo "$\cup$" por "|" ou "+"
- Deve-se obedecer as precedências

Precedência	Operador	Representação
Mais alta	Fechamento	x^*
Intermediária	Concatenação	$x \cdot y$ ou xy
Mais baixa	União	$x \mid y$ ou $x + y$

  Usa-se parênteses para modificar a precedência localmente
- É possível abreviar a expressão xx^* por x^+, denotando o conjunto regular correspondente ao fechamento transitivo de X

Exemplos

- A expressão regular $(ab \mid c^*) = ((ab) \mid c^*) = ((ab) \mid (c^*))$ representam o conjunto $\{\varepsilon, ab, c, cc, ccc \ldots\}$
- $a(b \mid c)^*$ representa o conjunto $\{a, ab, ac, abc, abb, acc, \ldots\}$
- $(ab \mid c)^*$ representa o conjunto $\{\varepsilon, ab, c, abc, cab, abab, cc, \ldots\}$
- $L = \{0^m 1^n \mid m \geq 0, n \geq 0\}$ pode ser reescrita como $((0)^*(1)^*)$ ou 0^*1^*. Para $m \geq 0$ e $n \geq 1$, a expressão correspondente seria $0^*11^* = 0^*1^+$

Exemplos: Conjuntos e Expressões Regulares

- $\Sigma = \{a, b, c, d\}$ e os dois subconjuntos $A = \{a\}, B = \{b, c\}$
  - Sentenças que possuem no mínimo um símbolo a

$$\Sigma^* A \Sigma^* \text{ ou } (a \mid b \mid c \mid d)^* a (a \mid b \mid c \mid d)^*$$

  - Sentenças que possuem exatamente dois símbolos a

$$(\Sigma - A)^* A (\Sigma - A)^* A (\Sigma - A)^* \text{ ou } (b \mid c \mid d)^* a (b \mid c \mid d)^* a (b \mid c \mid d)^*$$

  - Sentenças que possuem um número par de símbolos a

$$((\Sigma - A)^* A (\Sigma - A)^* A (\Sigma - A)^*)^* \text{ ou } ((b \mid c \mid d)^* a (b \mid c \mid d)^* a (b \mid c \mid d$$

  - Sentenças que iniciam com o símbolo a e terminam com o símbolo b ou c

$$A \Sigma^* B \text{ ou } a(a \mid b \mid c \mid d)^* (b \mid c)$$

  - Sentenças contendo apenas os símbolos a, b, c, com no mínimo um símbolo

$$(A \cup B)^+ \text{ ou } (a \mid b \mid c)^+$$

Leis algébricas das expressões regulares

- Associatividade:
  - União: $(\alpha \mid \beta) \mid \gamma = \alpha \mid (\beta \mid \gamma)$
  - Concatenação: $(\alpha\beta)\gamma = \alpha(\beta\gamma)$
- Comutatividade:
  - União: $\alpha \mid \beta = \beta \mid \alpha$
  - Concatenação: não se aplica
- Elemento neutro:
  - União: $\alpha \mid \emptyset = \emptyset \mid \alpha = \alpha$
  - Concatenação: $\alpha\varepsilon = \varepsilon\alpha = \alpha$
- Distributividade da concatenação sobre a união:
  - União: $\alpha(\beta \mid \gamma) = \alpha\beta \mid \alpha\gamma$
  - Concatenação: $(\beta \mid \gamma)\alpha = \beta\alpha \mid \gamma\alpha$

Relações de identidade para x, y, z

- $x \mid y = y \mid x$
- $\emptyset^* = \varepsilon$
- $x \mid (y \mid z) = (x \mid y) \mid z$
- $x(yz) = (xy)z$
- $x(y \mid z) = xy \mid xz$
- $(x \mid y)z = xz \mid yz$
- $x\varepsilon = \varepsilon x = x$
- $x\emptyset = \emptyset x = \emptyset$
- $\varepsilon\emptyset = \emptyset\varepsilon = \emptyset$
- $x^* = x \mid x^*$
- $(x^*)^* = x^*$
- $x \mid x = x$
- $x \mid \emptyset = x$
- $(xy)^*x = x(yx)^*$

Exercícios

- Obter expressões regulares que representam as linguagens sobre o alfabeto $\Sigma = \{a, b\}$ cujas sentenças estão descritas a seguir
  - Começam com aa
  - Não começam com aa
  - Contém a subcadeia $aabbb$
  - Possuem comprimento maior ou igual a 3
  - Possuem comprimento par
  - Possuem comprimento ímpar
  - Possuem comprimento múltiplo de 4