Expressões regulares

Definição: Expressões regulares sobre um alfabeto
Seja I um conjunto finito. O conjunto das expressões regulares sobre I representa-se por R_I e define-se indutivamente como se segue

- $a \in R_I$ para cada $a \in I$
- $\epsilon \in R_I$
- $\emptyset \in R_I$
- $(\alpha_1 + \alpha_2) \in R_I$ se $\alpha_1, \alpha_2 \in R_I$ (soma)
- $(\alpha_1\alpha_2) \in R_I$ se $\alpha_1, \alpha_2 \in R_I$ (concatenação/justaposição)
- $(\alpha^*) \in R_I$ se $\alpha \in R_I$ (fecho de Kleene)

Usualmente, para que se possam eliminar alguns parênteses de modo a não haver grande sobrecarga, assume-se que o fecho de Kleene tem precedência sobre a concatenção e esta sobre a soma. Omitem-se também os parênteses mais exteriores. Uma vez que a soma é associativa (ver abaixo), escreve-se $\alpha_1 + \alpha_2 + \alpha_3$ em vez de $(\alpha_1 + \alpha_2) + \alpha_3$ ou $\alpha_1 + (\alpha_2 + \alpha_3)$. O mesmo acontece se existirem mais parcelas e idênticas observações se podem fazer relativamente à concatenação.

Definição: Linguagem denotada por expressão regular
Sendo $\alpha \in R_I$, a linguagem representada (ou denotada) por α representa-se por $L(\alpha)$ e define-se indutivamente como se segue

- $L(a) = \{a\}$ para cada $a \in I$
- $L(\epsilon) = \{\epsilon\}$
- $L(\emptyset) = \emptyset$
- $L(\alpha_1 + \alpha_2) = L(\alpha_1) \cup L(\alpha_2)$
- $L(\alpha_1\alpha_2) = L(\alpha_1).L(\alpha_2) = \{w_1w_2 : w_1 \in L(\alpha_1), w_2 \in L(\alpha_2)\}$
- $L(\alpha^*) = (L(\alpha))^*$
  $= \{\epsilon\} \cup \{w_1 \dots w_n : n \geq 1 \text{ e } w_i \in L(\alpha) \text{ para cada } 1 \leq i \leq n\}$

Definição: Equivalência de expressões regulares
Sendo $\alpha, \beta \in R_I$, $\alpha = \beta$ sse $L(\alpha) = L(\beta)$.

É usual usar α^+ como abreviatura para expressão regular $\alpha\alpha^*$.

Proposição: Sendo $\alpha, \beta, \gamma \in R_I$,

$$\begin{array}{ll}
\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma & \alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma \\
\alpha + \beta = \beta + \alpha & (\alpha + \beta)\gamma) = \alpha\gamma + \beta\gamma \\
\alpha + \alpha = \alpha & \emptyset^* = \epsilon \\
\alpha + \emptyset = \alpha & (\alpha^*)^* = \alpha^* \\
\alpha(\beta\gamma) = (\alpha\beta)\gamma & \alpha\alpha^* = \alpha^*\alpha \\
\alpha\epsilon = \alpha = \epsilon\alpha & \alpha^*\alpha^* = \alpha^* \\
\alpha\emptyset = \emptyset = \emptyset\alpha & \alpha^* + \epsilon = \alpha^* \\
 & \alpha^* + \alpha\alpha^* = \alpha^* \\
 & \epsilon + \alpha\alpha^* = \alpha^* \\
 & \alpha^+ + \epsilon = \alpha^* \\
 & (\alpha\beta)^* = \epsilon + \alpha(\beta\alpha)^*\beta
\end{array}$$

Exemplo: A linguagem denotada pela expressão regular $01^*(0 + \epsilon)$, isto é $L(01^*(0 + \epsilon))$, é obtida do seguinte modo

$L(01^*(0 + \epsilon)) =$
$L(0).L(1^*).L(0 + \epsilon) =$
$\{0\}.(L(1))^*.(L(0) \cup L(\epsilon)) =$
$\{0\}.\{1\}^*.(\{\epsilon\} \cup \{0\}) =$
$\{0\}.\{1\}^*.\{\epsilon, 0\} =$
$\{0x0 : x \in \{1\}^*\} \cup \{0x : x \in \{1\}^*\}$

Exemplo: A linguagem denotada pela expressão regular $(0+1)^*1(0+1)^*$, isto é $L((0 + 1)^*1(0 + 1)^*)$, é obtida do seguinte modo

$L((0 + 1)^*1(0 + 1)^*) =$
$L((0 + 1)^*).L(1).L((0 + 1)^*) =$
$(L(0 + 1))^*.\{1\}.(L(0 + 1))^* =$
$(L(0) \cup L(1))^*.\{1\}.(L(0) \cup L(1))^* =$
$(\{0\} \cup \{1\})^*.\{1\}.(\{0\} \cup \{1\})^* =$
$\{0, 1\}^*.\{1\}.\{0, 1\}^* =$
$\{x1y : x, y \in \{0, 1\}^*\}$

Exemplo: Seja L a linguagem constituída pelas sequências formadas por elementos de $\{a, ..., z, 0, ..., 9\}$ que começam por uma letra. Uma expressão regular α tal que $L(\alpha) = L$ é $(a + b + \ldots + z)(a + b + \ldots + z + 0 + \ldots + 9)^*$

Exemplo: Uma expressão regular que denota o conjunto das sequências de 0's e 1's que têm dois 0's consecutivos: $(0 + 1)^*00(0 + 1)^*$

Exemplo: Uma expressão regular que denota o conjunto das sequências de 0's e 1's que comecem e terminem em 1 é $1 + 1(0 + 1)^*1$

Exemplo: Uma expressão regular que denota o conjunto das sequências de 0's e 1's que comecem em 0 ou terminem em 1 é $0(0+1)^* + (0+1)^*1$

Exemplo: Uma expressão regular que denota o conjunto das sequências de a's, b's e c's que têm pelo menos dois b's é $(a+c)^*b(a+c)^*b(a+b+c)^*$

Exemplo: Uma expressão regular que denota o conjunto das sequências de 0's e 1's que têm um número par de 0's é $(1+01^*0)^*$