1.14. Expresiones regulares

Las expresiones regulares representan lenguajes regulares y su propósito es simplificar la escritura de los lenguajes regulares.

La siguiente es la definición recursiva de las **expresiones regulares** sobre un alfabeto Σ dado.

1. Expresiones regulares básicas:

   $\varnothing$ es una expresión regular que representa al lenguaje $\varnothing$.
   λ es una expresión regular que representa al lenguaje $\{\lambda\}$.
   a es una expresión regular que representa al lenguaje $\{a\}$, $a \in \Sigma$.

2. Si R y S son expresiones regulares sobre Σ, también lo son:

$$RS$$
$$R \cup S$$
$$R^*$$

   RS representa la concatenación de los lenguajes representados por R y S; $R \cup S$ representa su unión, y R^* representa la clausura de Kleene del lenguaje representado por R.

Ejemplo Dado el alfabeto $\Sigma = \{a, b, c\}$,

$$(a \cup b^*)a^*(bc)^*$$

es una expresión regular que representa al lenguaje

$$(\{a\} \cup \{b\}^*) \cdot \{a\}^* \cdot \{bc\}^*.$$

Ejemplo Dado el alfabeto $\Sigma = \{a, b\}$,

$$(\lambda \cup a)^*(a \cup b)^*(ba)^*$$

es una expresión regular que representa al lenguaje

$$(\{\lambda\} \cup \{a\})^* \cdot \{a, b\}^* \cdot \{ba\}^*.$$

Ejemplos Los tres primeros lenguajes de la sección 1.13 podemos representarlos con expresiones regulares:

1. El lenguaje A de todas las palabras que tienen exactamente una a:

$$A = b^*ab^*.$$

2. El lenguaje B de todas las palabras que comienzan con b:

$$B = b(a \cup b)^*.$$

3. El lenguaje C de todas las palabras que contienen la cadena ba:

$$C = (a \cup b)^*ba(a \cup b)^*.$$

Observación: La representación de lenguajes regulares por medio de expresiones regulares no es única. Es posible que haya varias expresiones regulares diferentes para el mismo lenguaje. Por ejemplo, $b(a \cup b)^*$ y $b(b \cup a)^*$ representan el mismo lenguaje. Otro ejemplo: las dos expresiones regulares

$$(a \cup b)^* \qquad (a^*b^*)^*$$

representan el mismo lenguaje por la igualdad (1.1) de la sección 1.11.

Ejemplos Encontrar expresiones regulares que representen los siguientes lenguajes, definidos sobre el alfabeto $\Sigma = \{a, b\}$.

1. Lenguaje de todas las palabras que comienzan con b y terminan con a.

   Solución: $b(a \cup b)^*a$.

2. Lenguaje de todas las palabras que tienen exactamente dos a's.

   Solución: $b^*ab^*ab^*$.

3. Lenguaje de todas las palabras que tienen un número par de símbolos (palabras de longitud par).

   Solución: $(aa \cup ab \cup ba \cup bb)^*$.

4. Lenguaje de todas las palabras que tienen un número impar de símbolos (palabras de longitud impar).

   Solución: $a(aa \cup ab \cup ba \cup bb)^* \cup b(aa \cup ab \cup ba \cup bb)^*$.

5. Lenguaje de todas las palabras que tienen un número par de $a's$.

   Soluciones:

$$\begin{array}{l} b^*(ab^*a)^*b^*. \\ (ab^*a \cup b)^*. \\ (b^*ab^*ab^*)^* \cup b^*. \\ b^*(b^*ab^*ab^*)^*b^*. \end{array}$$

Ejemplo Encontrar una expresión regular que represente el lenguaje de todas las palabras que no contienen la cadena bc, definido sobre el alfabeto $\Sigma = \{a, b, c\}$.

Solución: $c^*(b \cup ac^*)^*$.

Ejercicio Encontrar expresiones regulares para los siguientes lenguajes:

1. $\Sigma = \{a, b\}$. Lenguaje de todas las palabras que tienen la cadena ab un número par de veces.

2. $\Sigma = \{a, b\}$. Lenguaje de todas las palabras que tienen un número impar de a's.

3. $\Sigma = \{a, b\}$. Lenguaje de todas las palabras que tienen un número par de a's o un número impar de b's.

4. $\Sigma = \{a, b, c\}$. Lenguaje de todas las palabras que tienen un número par de símbolos.

5. $\Sigma = \{a, b, c\}$. Lenguaje de todas las palabras que tienen un número impar de símbolos.

6. $\Sigma = \{a, b, c\}$. Lenguaje de todas las palabras que comienzan con c y terminan con b.

7. $\Sigma = \{a, b, c\}$. Lenguaje de todas las palabras que no contienen la cadena cc.

8. (Opcional, ¡Difícil!) $\Sigma = \{a, b\}$. Lenguaje de todas las palabras que tienen un número par de a's y un número impar de b's.

Observación:

No todos los lenguajes sobre un alfabeto dado Σ son regulares. Más adelante se mostrará que el lenguaje

$$L = \{\lambda, ab, aabb, aaabbb, \dots\} = \{a^n b^n : n \geq 0\}$$

sobre $\Sigma = \{a, b\}$ no se puede representar por medio de una expresión regular, y por lo tanto, no es un lenguaje regular.