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:

$$(R)(S)$$
$$(R \cup S)$$
$$(R)^*$$

   $(R)(S)$ 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. Los paréntesis (y) son símbolos de agrupación y se pueden omitir si no hay peligro de ambigüedad.

Para una expresión regular R cualquiera se utiliza en ocasiones la siguiente notación:

$$L(R) := \text{lenguaje representado por } R.$$

Utilizando esta notación y la definición de expresión regular podemos escribir, para R y S expresiones regulares arbitrarias:

$$\begin{aligned}
&L(\varnothing) = \varnothing.\\
&L(\lambda) = \{\lambda\}.\\
&L(a) = \{a\}, \quad a \in \Sigma.\\
&L(RS) = L(R)L(S).\\
&L(R \cup S) = L(R) \cup L(S).\\
&L(R^*) = L(R)^*.
\end{aligned}$$

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

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

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

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

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

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

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

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

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

✎ 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 $(a \cup b)^*$ y $(a^*b^*)^*$ representan el mismo lenguaje por la igualdad establecida en el ejercicio final 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 cadenas que comienzan con b y terminan con a.
   Solución: $b(a \cup b)^*a$.

2. Lenguaje de todas las cadenas que tienen un número par de símbolos (cadenas de longitud par).
   Solución: $(aa \cup ab \cup ba \cup bb)^*$.

3. Lenguaje de todas las cadenas que tienen un número par de $a$es.
   Soluciones:

$$b^*(b^*ab^*ab^*)^*.$$
$$(ab^*a \cup b)^*.$$
$$(b^*ab^*ab^*)^* \cup b^*.$$

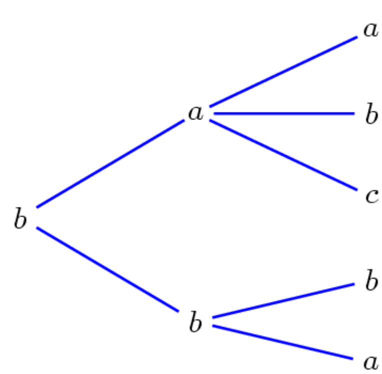
Ejemplos Encontrar expresiones regulares que representen los siguientes lenguajes, definidos sobre el alfabeto $\Sigma = \{0, 1\}$.

1. Lenguaje de todas las cadenas que tienen exactamente dos ceros.
   Solución: $1^*01^*01^*$.

2. Lenguaje de todas las cadenas cuyo penúltimo símbolo, de izquierda a derecha, es un 0.
   Solución: $(0 \cup 1)^*0(0 \cup 1)$.

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

Solución: Una b puede estar seguida solamente de otra b o de una a, mientras que las $a$es y las $c$es pueden estar seguidas de cualquier símbolo. Esto se puede visualizar por medio del siguiente diagrama:



Teniendo en cuenta todas las restricciones y posibilidades, arribamos a la siguiente expresión: $(a \cup c \cup b^+a)^*b^*$. Una expresión regular más sencilla para este lenguaje es $c^*(b \cup ac^*)^*$.

Ejercicios Encontrar expresiones regulares para los lenguajes descritos a continuación:

1. $\Sigma = \{0, 1, 2\}$. Lenguaje de todas las cadenas que comienzan con 2 y terminan con 1.

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

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

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

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

6. $\Sigma = \{0, 1\}$. Lenguaje de todas las cadenas que tienen por lo menos un 0 y por lo menos un 1.

7. $\Sigma = \{0, 1\}$. Lenguaje de todas las cadenas que tienen a lo sumo dos ceros consecutivos.

8. $\Sigma = \{0, 1\}$. Lenguaje de todas las cadenas cuyo quinto símbolo, de izquierda a derecha, es un 1.

9. $\Sigma = \{0, 1\}$. Lenguaje de todas las cadenas de longitud par ≥ 2 formadas por ceros y unos alternados.

10. $\Sigma = \{0, 1\}$. Lenguaje de todas las cadenas cuya longitud es ≥ 4.

11. $\Sigma = \{0, 1\}$. Lenguaje de todas las cadenas de longitud impar que tienen unos únicamente en las posiciones impares.

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

13. $\Sigma = \{a, b\}$. Lenguaje de todas las cadenas que tienen un número par de $a$es o un número impar de $b$es.

14. $\Sigma = \{0, 1\}$. Lenguaje de todas las cadenas cuya longitud es un múltiplo de tres.

15. $\Sigma = \{0, 1, 2\}$. Lenguaje de todas las cadenas que no contienen dos unos consecutivos.

✎ 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.