Lenguajes Regulares y Expresiones Regulares

Conjuntos regulares

- Un conjunto es regular si:

1. Es el conjunto vacío, $\emptyset$, ó el conjunto cuyo elemento es la palabra vacía, $\{\varepsilon\}$, ó es un subconjunto simple (sólo un elemento) del alfabeto.

2. Puede ser generado a partir de $\emptyset$ ó de $\{\varepsilon\}$ ó de un subconjunto simple utilizando las operaciones de **unión**, **concatenación** y **cerradura** de Kleene.

Definición recursiva de conjunto regular.

Sea Σ un alfabeto. Los conjuntos regulares sobre Σ se definen recursivamente como:

- **Base**: $\emptyset$, $\{\varepsilon\}$ y $\{a\}$, para toda $a \in \Sigma$, son conjuntos regulares sobre Σ.
- **Paso recursivo**: Si X e Y son conjuntos regulares sobre Σ, entonces los conjuntos $X \cup Y$, $X.Y$ y X^* también lo son.
- **Cerradura**: X es un conjunto regular sobre Σ sólo si puede ser obtenido a partir de los elementos base mediante un número finito de aplicaciones del paso recursivo.

Ejemplos de conjuntos regulares

- $\{a, b\}^*$
- $\{\mathrm{aa}, \mathrm{bb}\}$
- $\{a, b\}^*\{\mathrm{bb}\}\ \{a, b\}^*$

Son regulares sobre $\Sigma=\{a, b\}$

- El conjunto de cadenas que empiezan y terminan con una a y contienen al menos una b es regular sobre $\{a, b\}$.

$$\{a\}\ \{a, b\}^*\{b\}\ \{a, b\}^*\{a\}$$

Expresiones regulares (ER)

Las expresiones regulares (ER) se utilizan para abreviar la descripción de conjuntos regulares:

- El conjunto regular $\{a\}$ es representado por a.
- Las operaciones:
  - **unión** representado por +
  - **concatenación** representado por yuxtaposición
  - **cerradura** de Kleene representado por *

- Definición: Sea Σ un alfabeto. Las expresiones regulares sobre Σ se definen recursivamente como:

  - **Base**: $\emptyset$, ε y a, para toda $a \in \Sigma$, son expresiones regulares sobre Σ.

  - **Paso recursivo**: Si u y v son expresiones regulares sobre Σ, entonces las expresiones $(u+v)$, (uv) y (u^*) también lo son y representan a los conjuntos $\{u\} \cup \{v\}$, $\{u\}\{v\}$ y $\{u\}^*$, respectivamente.

  - **Cerradura**: u es una expresión regular sobre Σ sólo si puede ser obtenido a partir de los elementos base mediante un número finito de aplicaciones del paso recursivo.

Ejemplos:

Lenguaje	Expresión regular
$\{\varepsilon\}$	ε
$\{0\}$	0
$\{001\} = \{0\}\{0\}\{1\}$	001
$\{0, 1\} = \{0\} \cup \{1\}$	$0 + 1$
$\{0, 10\} = \{0\} \cup \{10\}$	$0 + 10$
$\{1, \varepsilon\}\{001\}$	$(1 + \varepsilon)001$
$\{110\}^*\{0, 1\}$	$(110)^*(0 + 1)$
$\{1\}^*\{10\}$	1^*10

Expresiones regulares

Entonces, las expresiones regulares se pueden simplificar más reduciendo el número de paréntesis:

- $\{a, b\}^* \{bb\} \{a, b\}^* = (a + b)^* bb(a + b)^*$
- $\{a\} \{a, b\}^* \{b\} \{a, b\}^* \{a\} = a(a + b)^* b(a + b)^* a$

- Notación
  - $u^+ = uu^*$
  - $u^2 = uu, \quad u^3 = u^2 u, \ \ldots$

Ejemplo

- El conjunto $\{bawab \mid w \in \{a, b\}^*\}$ es regular sobre $\{a, b\}$

Demostración:

Conjunto	Expresión	Justificación
1. $\{a\}$	a	Base
2. $\{b\}$	b	Base
3. $\{a\}\{b\}=\{ab\}$	ab	Concatenación de 1 y 2
4. $\{a\} \cup \{b\}=\{a,b\}$	$a+b$	Unión de 1 y 2
5. $\{b\}\{a\}=\{ba\}$	ba	Concatenación de 2 y 1
6. $\{a,b\}^*$	$(a+b)^*$	Cerradura Kleene de 4
7. $\{ba\}\{a,b\}^*$	$ba(a+b)^*$	Concatenación de 5 y 6
8. $\{ba\}\{a,b\}^*\{ab\}$	$ba(a+b)^*ab$	Concatenación de 7 y 3

Lenguajes regulares

- Definición:

  Un lenguaje es regular si se puede representar por una expresión regular o conjunto regular.

Equivalencias

- Una expresión regular define un patrón; una palabra **pertenece** al lenguaje definido por esa expresión regular si y sólo si sigue el patrón.

- Una expresión regular que represente un lenguaje debe cumplir dos condiciones:

  - **Correcta**: todas las palabras representadas por la expresión regular deben ser parte del lenguaje.

  - **Completa**: toda palabra del lenguaje debe ser representada por la expresión regular.

Equivalencias

- **Concatenación** indica orden de los símbolos
- **Cerradura de Kleene** permite repeticiones
- + indica **selección**.
- Dos expresiones que representan al mismo conjunto son llamadas **equivalentes**.

$$\mathbf{ER1} = (a+b)^* aa (a+b)^* + (a+b)^* bb (a+b)^*$$

$$= \{\mathbf{aa, bb, aaa, baa, abb, bbb, \ldots..}\}$$

Representa al conjunto de cadenas sobre $\{a, b\}$ que contienen a la subcadena *aa* o a la subcadena *bb*.

$$ER2 = a^*ba^*ba^*$$

$$= \{\mathbf{bb, abb, bab, bba, ...}\}$$

Representa al conjunto de cadenas sobre $\{a, b\}$ que contienen exactamente dos b's

$$\boldsymbol{ER3 = a^*(a^*ba^*ba^*)^* \quad \text{y} \quad \text{ER4} = a^*(ba^*ba^*)^*}$$

$$= \{\varepsilon, \text{bb}, \text{abb}, \text{bba}, \text{bab}, \ldots\ldots\}$$

Representan cadenas con un número par de *b*'s.

ER5= $(a+ab)^*$

Expresión regular para el lenguaje sobre $\{a, b\}$ en cuyas palabras inmediatamente antes de toda b aparece una a

ER6$= (ba+bc+a+c)^{*}bb(a+c+ab+cb)^{*}$

Expresión regular que representa a las palabras que contienen exactamente una vez dos *b*'s contiguas

Ejercicio

- Escriba una expresión regular para el lenguaje sobre {0 ,1} que consiste de las palabras en las que no hay dos símbolos iguales contiguos, es decir, los 0's y los 1's se alternan.

  - $(01)^* + (10)^* + 0(10)^* + 1(01)^*$
  - $(\varepsilon + 1)(01)^*(\varepsilon + 0)$
  - $(\varepsilon + 0)(10)^*(\varepsilon + 1)$

Identidades

- $\varnothing u = u\varnothing = \varnothing$
- $\varepsilon\, u = u\, \varepsilon = u$
- $\varnothing^* = \varepsilon$
- $\varepsilon^* = \varepsilon$
- $u+v = v+u$
- $u+\varnothing = u$
- $u+u = u$
- $u^* = u^* u^* = (u^*)^*$
- $u(v+w) = uv + uw$
- $(u+v)w = uw+vw$
- $(uv)^*u = u(vu)^*$
- $(u+v)^* = (u^*+v)^* = u^*(u+v)^* = (u+vu^*)^* = (u^*v^*)^* = u^*(vu^*)^* = (u^*v)^*u^*$
- $u^*(u + \varepsilon) = u^*$
- $u^*u^* = u^*$
- $u^* + v^* = v^* + u^*$
- $(u^*v^*)^* = (u + v)^* = (u + v)^*uv(u + v)^* + v^*u^*$

Ejemplos

- Expresión que representa las cadenas sobre $\{a, b\}$ que no contienen la subcadena aa:

$$
\begin{aligned}
& b^*(ab^+)^*+b^*(ab^+)^*a = \\
& \quad b^*(ab^+)^*(\varepsilon +a) = \\
& \quad b^*(abb^*)^*(\varepsilon +a) = \\
& \quad (b+ab)^*(\varepsilon +a)
\end{aligned}
$$

- Expresión regular que representa las cadenas sobre $\{a, b, c\}$ que contienen la subcadena bc:

  $(a+b+c)^*bc(a+b+c)^*$

- $c^*(b+ac^*)^*$

  Representa las cadenas que no contienen la subcadena bc.

Ejemplos

- Cadenas sobre $\{0, 1\}$ de longitud igual a 5:
  $(0 + 1)(0 + 1)(0 + 1)(0 + 1)(0 + 1)$
  $= (0 + 1)^5$

- Cadenas sobre $\{0, 1\}$ de longitud mayor o igual a 6:
  $(0 + 1)(0 + 1)(0 + 1)(0 + 1)(0 + 1)(0 + 1)(0 + 1)^*$
  $= (0 + 1)^6(0 + 1)^*$

- Cadenas sobre $\{0, 1\}$ de longitud menor o igual a 6:
  $(0 + 1 + \varepsilon)(0 + 1 + \varepsilon)(0 + 1 + \varepsilon)(0 + 1 + \varepsilon)(0 + 1 + \varepsilon)(0 + 1 + \varepsilon)$
  $= (0 + 1 + \varepsilon)^6$

Ejercicios

- Obtenga una expresión regular para el conjunto de palabras sobre $\{a, b, c\}$ que tienen al menos una a y al menos una b.

$$(a + \mathrm{b} + c)^* a(a + b + c)^* \, b(a + b + c)^*$$

- Obtenga una expresión regular para el lenguaje sobre $\{0 , 1\}$ que consiste de las palabras cuyo décimo símbolo contado de la derecha a la izquierda es un 1.
  - $(0+1)^*1(0+1)\ (0+1)\ (0+1)\ (0+1)\ (0+1)\ (0+1)\ (0+1)\ (0+1)\ (0+1)$ $= (0 + 1)^*1(0 + 1)^9$