LENGUAJES REGULARES

CONTENIDO

- Lenguajes regulares
- Definición formal de lenguaje regular
- Expresiones regulares
  - Propiedades algebraicas de las expresiones regulares
  - Precedencia de las operaciones con las expresiones regulares
- Ejemplos

LENGUAJES REGULARES

- Un **<u>lenguaje es un conjunto de palabras</u>** *(cadenas) de un determinado alfabeto **Σ***.

- Los lenguajes más sencillos que formalmente se consideran son los **lenguajes regulares**.

- Un lenguaje regular se puede generar a partir de lenguajes básicos, con la aplicación de las operaciones de **unión**, **concatenación** y *** de Kleene** un número finito de veces.

- Los lenguajes regulares se llaman así porque sus palabras contienen "**regularidades**" o repeticiones de los mismos componentes.

Ejemplo:

$$L_1 = \{ab, abab, ababab, abababab, \ldots\}$$

- En este ejemplo se aprecia que las palabras de $\mathbf{L_1}$ son simplemente repeticiones de *"ab"* cualquier número de veces.
- Aquí la *"regularidad"* consiste en que las palabras contienen *"ab"* algún número de veces.

- Un **lenguaje regular** es un tipo de **lenguaje formal** que satisface las siguientes propiedades:

Puede ser reconocido por:

- Un autómata finito determinista
- Un autómata finito no determinista
- Un autómata de pila
- Un autómata finito alterno
- Una máquina de turing de solo lectura

Es generado por:

- Una gramática regular
- Una gramática de prefijos

Es descrito por:

- Una expresión regular

Ejemplo:

$$\mathbf{L_2} = \{abc, cc, abab, abccc, ababc, \ldots\}$$

La regularidad en $\mathbf{L_2}$ consiste en que sus palabras comienzan con 0 o más repeticiones de “ab”, seguidas de repeticiones de 0 o más “c”.

- Similarmente es posible definir muchos otros lenguajes basados en la idea de repetir esquemas simples.
- Adicionalmente a las repeticiones de esquemas simples, vamos a considerar que los lenguajes finitos son también regulares por definición.
- Por ejemplo, el lenguaje $\mathbf{L_3}$ = {anita, lava, la, tina} es regular.

- Finalmente, al combinar lenguajes regulares uniéndolos o concatenándolos, también se obtiene un lenguaje regular.

Ejemplo:

$\mathbf{L_1}$ = {ab, abab, ababab, abababab, . . .}

$\mathbf{L_3}$ = {anita, lava, la, tina}

$\mathbf{L_1} \cup \mathbf{L_3}$ = {anita, lava, la, tina, ab, abab, ababab, abababab, . . .}

- También es regular una concatenación como $\mathbf{L_3L_3}$

$\mathbf{L_3L_3}$ = {anitaanita, anitalava, anitala, anitatina, lavaanita, lavalava, lavala, lavatina, laanita, lalava, lala, latina, tinaanita, tinalava, tinala, tinatina}

DEFINICIÓN FORMAL DE LENGUAJE REGULAR

- Un lenguaje regular sobre un alfabeto $\boldsymbol{\Sigma}$ dado se define recursivamente como:
  - El lenguaje vacío $\emptyset$ es un lenguaje regular.
  - El lenguaje cadena vacía $\{\boldsymbol{\lambda}\}$ es un lenguaje regular.
  - Para todo símbolo $\mathbf{a} \in \boldsymbol{\Sigma}$ entonces $\{\mathbf{a}\}$ es un lenguaje regular.
  - Si $\mathbf{A}$ y $\mathbf{B}$ son lenguajes regulares entonces $\mathbf{A} \cup \mathbf{B}$ (unión), $\boldsymbol{AB}$ (concatenación) y $\boldsymbol{A}^*$ (cerradura de kleene) son lenguajes regulares.
  - Si $\mathbf{A}$ es un lenguaje regular entonces $\{\mathbf{A}\}$ es el mismo lenguaje regular.
  - No existen más lenguajes regulares sobre Σ.

DEFINICIÓN FORMAL DE LENGUAJE REGULAR

- Un lenguaje **L** es regular **si y solo si se cumple al menos** una de las condiciones siguientes:

1. El lenguaje **L** es finito *(estos son lenguajes obviamente regulares y uno podría crear expresiones regulares que serían la unión de todas las palabras del lenguaje que definirían dicho lenguaje.)*

2. El lenguaje **L** es la unión o la concatenación de otros lenguajes regulares $\mathbf{R_1}$ y $\mathbf{R_2}$, $\mathbf{L=R_1 \cup R_2}$ o $\mathbf{L=R_1R_2}$ respectivamente

3. El lenguaje **L** es la cerradura de Kleene de algún lenguaje regular, **L=R***

- Todo **lenguaje formal finito** constituye un **lenguaje regular**.

- Un lenguaje formal infinito puede ser regular o no regular.

  - El lenguaje **L = {a^n, n > 0}** es regular porque puede ser representado, por ejemplo, mediante la **expresión regular a^+**.

  - El lenguaje **L = {a^n b^n, n > 0}** es un lenguaje no regular dado que no es reconocido por ninguna de las formas de representación anteriormente enumeradas.

EXPRESIONES REGULARES

- Una **expresión regular** es una forma abreviada de representar cadenas de caracteres que se ajustan a un determinado patrón.

- Al **conjunto de cadenas representado por la expresión** ***r*** se le llama *lenguaje generado por la expresión regular* ***r*** y se escribe ***L(r)***.

- Una expresión regular se define sobre un alfabeto **Σ** y **es una cadena formada por caracteres** de dicho alfabeto y por una serie de **operadores** también llamados **metacaracteres.**

- Las expresiones regulares se introducen para **describir** los **lenguajes regulares**, entonces las expresiones regulares serán **metalenguajes.**
- Es decir, *las expresiones regulares son un* ***metalenguaje para describir los lenguajes regulares****.*

- Una **expresión regular**, a menudo es llamada también **patrón**, y es una expresión que describe un conjunto de cadenas sin enumerar sus elementos.

Ejemplo:

- El grupo formado por las cadenas *handel*, *händel* y *haendel* se describe mediante el patrón : *h(a|ä|ae)ndel*

Metalenguaje*: lenguaje para hablar de otro lenguaje.*

- Las expresiones regulares básicas se definen de la siguiente forma:

1. El símbolo $\emptyset$ (conjunto vacío) es una expresión regular y $\mathbf{L}(\emptyset) = \{ \quad \}$

2. El símbolo $\boldsymbol{\lambda}$ (palabra vacía) es una expresión regular y **L(λ)={λ}**

3. Cualquier símbolo $\boldsymbol{a} \in \boldsymbol{\Sigma}$ es una expresión regular y ***L(a)={a}***

OPERACIONES DE LAS EXPRESIONES REGULARES

A partir de las expresiones regulares básicas pueden construirse expresiones regulares más complejas aplicando las siguientes operaciones:

- *Concatenación*
- *Unión*
- *Cierre o cerradura estrella (Cerradura de Kleene)*

CONCATENACIÓN DE EXPRESIONES REGULARES

- *La **concatenación*** se representa con el metacarácter ·

- Si **r** y **s** son expresiones regulares, entonces **r · s** también es una expresión regular y **L(r · s) = L(r) · L(s)**

- El operador "·" puede omitirse de modo que ***rs*** también representa la concatenación.

- La concatenación de dos lenguajes $\mathbf{L_1}$ y $\mathbf{L_2}$, se obtiene concatenando cada cadena de $\mathbf{L_1}$ con todas las cadenas de $\mathbf{L_2}$.

Ejemplo:

$\Sigma=\{0,1\}$
$L_1=\{00,1\}$
$L_2=\{11,0,10\}$
$L_1L_2=\{0011,000,0010,111,10,110\}$

L(ab)={ab}

UNIÓN DE EXPRESIONES REGULARES

- *La **unión*** se representa con el metacarácter **|**

- Si **r** y **s** son expresiones regulares, entonces **r | s** también es una expresión regular y **L(r | s) = L(r) ∪ L(s)**

Ejemplo:

El lenguaje generado por la expresión regular **ab|c** es **L(ab|c) = {ab, c}**

CIERRE O CLAUSURA ESTRELLA (CERRADURA DE KLEENE) EN EXPRESIONES REGULARES

- *La* ***cerradura*** se representa con el metacarácter *

- Si **r** es una expresión regular, entonces **r*** también es una expresión regular y **L(r*) = L(r)***

$$L^* = \bigcup_{i=0}^{\infty} L^i$$

- Donde $\boldsymbol{L^i}$ es igual a la concatenación de $\boldsymbol{L}$ consigo mismo $\boldsymbol{i}$ veces y $\boldsymbol{L^0 = \lambda}$

- Si **a** es una expresión regular, entonces **a*** es una expresión regular que denota **{a}***

- Es decir,

$$\mathbf{L(a^{*}) = \{\lambda, a, aa, aaa, \ldots, aaaaaa\ldots a\}}$$

Ejemplo:

El lenguaje generado por la expresión regular

a*ba*

es

L(a*ba*) = {b, ab, ba, aba, aab,…}

CIERRE POSITIVO EN EXPRESIONES REGULARES

- *La* ***cerradura*** se representa con el metacarácter $^{\mathbf{+}}$

- Si **r** es una expresión regular, entonces $\mathbf{r}^{\mathbf{+}}$ también es una expresión regular y $\mathbf{L(r^{+}) = L(r)^{+}}$

$$L^{+} = \bigcup_{i=1}^{\infty} L^{i}$$

- Donde $\boldsymbol{L^{i}}$ es igual a la concatenación de $\boldsymbol{L}$ consigo mismo $\boldsymbol{i}$ veces y no se incluye $\boldsymbol{L^{0} = \lambda}$

- Si **a** es una expresión regular, entonces $\mathbf{a}^{+}$ es una expresión regular que denota $\{\mathbf{a}\}^{+}$

- Es decir,

$$\mathbf{L(a^{+}) = \{a, aa, aaa, \ldots, aaaaaa\ldots a\}}$$

Ejemplo:

El lenguaje generado por la expresión regular

$\mathbf{a^{+}ba^{+}}$

es

$$\mathbf{L(a^{+}\,ba^{+}) = \{aba, aaba, aabaa, aaaba, \ldots\}}$$

PROPIEDADES ALGEBRAICAS DE LAS EXPRESIONES REGULARES

- La **concatenación** es asociativa **(rs)t = r(st)**
- La **concatenación** se distribuye sobre: **r(s|t) = rs|rt** y **(s|t)r = sr|tr**
- La **unión** es conmutativa: **r | s = s | r**
- La **unión** es asociativa: **(r | s) | t = r | (s | t)**
- **λ** es el elemento identidad para la concatenación **λ r = r λ = r**
- La relación entre **λ** y * es: **r* = (r | λ)***
- **r*** es idempotente **r** = r***

PRECEDENCIA DE LAS OPERACIONES CON LAS EXPRESIONES REGULARES

- Se permite el uso de paréntesis para indicar la precedencia de las operaciones, pero cuando no se utilizan paréntesis para evaluar una expresión regular, hay que tener en cuenta el siguiente orden de precedencia:

1. *Uso de paréntesis*
2. *Operación cierre y cierre positivo*
3. *Operación concatenación*
4. *Operación unión o alternativa*

EJEMPLO 1

Σ={a, b}

i. **a|b** denota el lenguaje **L(a|b) = {a, b}.**

ii. **(a|b)(a|b)** denota a **L((a|b)(a|b)) = {aa, ab, ba, bb}**, es el lenguaje de todas las cadenas de longitud dos sobre el alfabeto **Σ**.

iii. **(a|b)(a|b) = aa | ab | ba | bb**

iv. **L(a*) = {λ, a, aa, aaa, …}**, todas las cadenas de cero o más **a's**.

EJEMPLO 2

Σ={a, b}

(a) | ((b)*(c)) = a|b*c

- *Descripción:*
  - Conjunto de cadenas que son una sola $\boldsymbol{a}$ o cero o más $\boldsymbol{b's}$ seguidas por una $\boldsymbol{c}$.
- *Algunas cadenas del lenguaje:*

L(a|b*c) = {a, c, bc, bbc, bbbc, bbbbc, …**}**