Instituto Politécnico Nacional

Escuela Superior de Cómputo

Teoría computacional

Lenguajes regulares

Prof. Rafael Aguilar
raffaello.escom@gmail.com

Contenido

- Lenguajes regulares
- Definición formal de lenguaje regular
- Expresiones regulares
- Ejemplos

Lenguajes regulares

- Los lenguajes regulares se llaman así porque sus palabras contienen “regularidades” o repeticiones de los mismos componentes.
- v.g.
  - L_1 = {ab, abab, ababab, abababab, . . .}

    - En este ejemplo se aprecia que las palabras de 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.

Lenguajes regulares

- v.g.
  - L_2 ={abc, cc, abab, abccc, ababc, . . .}
    - La regularidad en 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 L_3 = {anita, lava, la, tina} es regular.

Lenguajes regulares

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

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

- También es regular una concatenación como L_3L_3 = {anitaanita,anitalava, anitala, anitatina, lavaanita, lavalava, lavala, lavatina, . . .}

Definición formal de Lenguaje Regular

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

i. 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.)*

ii. L es la unión o la concatenación de otros lenguajes regulares R_1 y R_2, $L = R_1 \cup R_2$; o $L = R_1R_2$ respectivamente

iii. L es la cerradura de Kleene de algún lenguaje regular, $L = R^*$

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**.

Expresiones regulares

- Las expresiones regulares se introducen para **describir** los **lenguajes regulares**, entonces las expresiones regulares serán <u>**metalenguajes**</u> i.e. *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.
  - v.g., 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.*

Expresiones regulares

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

A. El símbolo Φ (conjunto vacío) es una expresión regular y ***L(Φ)={ }***

B. El símbolo λ (palabra vacía) es una expresión regular y ***L(λ)={λ}***

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

Expresiones regulares

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

1. ***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.

Expresiones regulares

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

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

Expresiones regulares

2. ***Unión*** (se representa con el metacarácter **|**): Si *r* y *s* son expresiones regulares, entonces *r***|***s* también es una expresión regular y $L(r|s) = L(r) \cup L(s)$.

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

Expresiones regulares

3. ***Cierre o clausura estrella (cerradura de Kleene)*** (se representa con el metacarácter *****): Si *r* es una expresión regular, entonces *r** también es una expresiones regular y *L(r*******)=L(r) **.

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

Donde L^i es igual a la concatenación de L consigo mismo i veces y $L^0 = \lambda$.

Expresiones regulares

- Si a es una expresión regular, entonces a^* es una expresión regular que denota $\{a\}^*$ v.g.

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

- Ejemplo
  - El lenguaje generado por la expresión regular a^*ba^* es $L(a^*ba^*)=\{b,ab,ba,aba,aab,\ldots\}$

Expresiones regulares

4. ***Cierre positivo*** (se representa con el metacarácter $^+$): Si *r* es una expresión regular, entonces r^+ también es una expresiones regular y $L(r^+)=L(r)^+$.

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

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

Expresiones regulares

- Si a es una expresión regular, entonces a^+ es una expresión regular que denota $\{a\}^+$ v.e.

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

- Ejemplo:
  - El lenguaje generado por la expresión regular a^+ba^+ es $L(a^+ba^+)=\{aba,aaba,aabaa,aaaba,\ldots\}$

Expresiones regulares

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 neutro de la concatenación ***λ r = r λ = r***
- **ϕ** es el elemento neutro de la unión ***r* + ϕ = ϕ + *r* = *r***
- La relación entre ***λ*** y ***** es: ***r*=(r| λ)****
- ***r**** es idempotente ***r**=r****

Nota 1: λ* = λ

Nota 2: ▯* = λ

Expresiones regulares

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*

Ejemplos

Ejemplo 1

- $\sum = \{a, b\}$

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

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

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

iv. $L(a^*)=\{\lambda, a, aa, aaa, \ldots\}$, *todas las cadenas de cero o más a 's.*

Ejemplos

Ejemplo 2

- $(\boldsymbol{a})|((\boldsymbol{b})^*(\boldsymbol{c})) = \boldsymbol{a}|\boldsymbol{b}^*\boldsymbol{c}$
  - Conjunto de cadenas que son una sola a o cero o más b's seguidas por una c.

$$L(a|b^*c)=\{a,c,bc,bbc,bbbc,bbbbc,\ldots\}$$

Ejemplo 3

A={0,1}

- 0*1*Conjunto de palabras formadas por una sucesión de ceros seguida de una sucesión de unos. Ambas sucesiones pueden ser vacías.
- 00*11* Conjunto de palabras formadas por una sucesión de ceros seguida de una sucesión de unos. Ninguna de las sucesiones puede ser vacía.

A r*r se le denota como r^+. Por lo que en la ultima expresión regular quedaría como 0^+1^+

Ejemplo 5

∑ ={a, b}

- a|b denota el lenguaje L(a|b)={a, b}.
- (a|b)(a|b) denota a L((a|b)(a|b)), el lenguaje de todas las cadenas de longitud dos sobre el alfabeto ∑.
- (a|b)(a|b) =aa|ab|ba|bb
- L(a*)={λ, a, aa, aaa, …}, todas las cadenas de cero o más a 's.

Ejemplo 6

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

- Descripción:
  - Conjunto de cadenas que son una sola a o ninguna o más b's seguidas por una c.

- Algunas cadenas del lenguaje:
  - L(a|b*c)={a,c,bc,bbc,bbbc,bbbbc,...}

Ejercicios

*Describa los lenguajes generados por las siguientes **expresiones regulares** y enumere al menos 5 cadenas de cada lenguaje.* $\sum = \{a, b, c, d, e, \ldots, z\}$

1. $(ab)|(cz)|d^*$
2. $a^+b^+c \mid b^+d\, z$
3. $(abc)^*z$
4. $(a|b)^*|a$
5. $(ab)^+\,(bc)^+$