MATEMÁTICAS PARA LA COMPUTACIÓN

CAPÍTULO 9. INTRODUCCIÓN A LOS LENGUAJES FORMALES

MANIPULACIÓN DE CADENAS Y LENGUAJES.

AUTOR: JOSÉ ALFREDO JIMÉNEZ MURILLO

MANIPULACIÓN DE CADENAS Y LENGUAJES.

Las gramáticas regulares son parte esencial de los lenguajes regulares. Los autómatas finitos (AF) son una representación gráfica de los lenguajes regulares. Es conveniente antes de abordar el tema de AF tener en cuenta algunos conceptos importantes que permiten la manipulación correcta de los lenguajes regulares y los autómatas finitos, por esa razón se tratarán algunos aspectos que se consideran importantes antes de entrar de lleno con el tema de autómatas finitos.

Una palabra que pertenece a un lenguaje L(G) realmente es una cadena de símbolos o caracteres y por eso la relación existente entre símbolos, cadenas, lenguajes, alfabetos y gramáticas es importante.

Cadena. Es una secuencia de símbolos yuxtapuestos (se coloca uno seguido del otro).

Sea $\sum = \{0, 1, 2\}$
$w = 0$ $x = 02$ $y = 011$ $z = 12012$

En este caso w, x, y, z son cadenas formadas con símbolos del alfabeto $\sum$.

Longitud de una cadena $|w|$. Es el número de símbolos que integran dicha cadena.

$\sum = \{a, b, c, d, e, \ldots\ldots, z\}$
w = hola x = manzana y = limón

$|w| = 4$ $|x| = 7$ $|y| = 5$

Concatenación (wx). Es la cadena que se forma al escribir la primera cadena w, seguida de la segunda x sin espacios entre ellas.

Sea. w = árbol x = grande.

La concatenación de w con x es: wx = árbolgrande

Cadena vacía (ε). Es aquella cadena que no tiene símbolos. En donde $\varepsilon w = w\varepsilon$. También $\varepsilon\varepsilon = \varepsilon$.

Si w = pera entonces εw = pera = wε

Cadena elevada a una potencia (w^n). Si n es un elemento del conjunto de números naturales ($n \in N$) y w es una cadena. Entonces:

$w^0 = \varepsilon$ (cadena vacía)
$w^1 = ww^0$
$w^2 = ww^1 = www^0$
$w^3 = w\,w^2 = www^1 = wwww^0$
$w^n = ww^{n-1}$

Sea w = Hola

$w^4 = ww^3 = www^2 = wwww^1 = wwwww^0 =$ HolaHolaHolaHolaε = HolaHolaHolaHola

Sea w = Pera x = Verde

$(wx)^2 = wxwxw^0x^0 =$ PeraVerdePeraVerde
$w^2x^2 = www^0xxx^0 =$ PeraPeraVerdeVerde

Prefijo de una cadena (p). Se dice que p es el prefijo de una cadena w si para alguna cadena x se obtiene w = px. Ejemplo.

Sea w = Bimestre donde p = Bi x = mestre
w = Bianual p = Bi x = anual

O bien
w = Pera El conjunto de prefijos es {P, Pe, Per} en donde x es {era, ra, a}

Sufijo de una cadena (s). Se dice que s es el sufijo de una cadena w, si para alguna cadena x se obtiene w = xs

Sea w = Suspensión donde x = Suspen s = sión
w = División x = Divi s = sión

O bien
w = Mango El conjunto de sufijos es {o, go, ngo, ango}
en donde x es {Mang, Man, Ma, M}.

Subcadena. La cadena x es una subcadena de la cadena w si existen las cadenas p y s en las cuales w = pxs.

Sean w = Pera
El conjunto de subcadenas de w es {P, e, r, a, Pe, er, ra, Per, era, Pera, ε}

Inversa de una cadena (w^I). Es la cadena que se obtiene al escribir los caracteres en forma invertida. Para la inversa de w se cumple que:

$w^I = w$ si $w = \varepsilon$
$w^I = x^Ia$ si $w = ax$ Donde: $(a \in \sum)$ y $(x \in L)$

Sea w = Hola

$w^I = (Hola)^I = (ola)^I H = (la)^I oH = (a)^I loH = (\varepsilon)^I aloH = aloH$

Se sabe que $w^I = (xy)^I = y^Ix^I$.

Ejemplo.

Sea w = Morelia y sean x = More y = lia

$w^I = (xy)^I = (\text{lia})^I(\text{More})^I = (\text{ia})^I\text{l}(\text{ore})^I\,\text{M} = (\text{a})^I\,\text{il}\,(\text{re})^I\,\text{oM}$
$= (\varepsilon)^I\,\text{ail}\,(\text{e})^I\,\text{roM} = \text{ail}\,(\varepsilon)^I\,\text{eroM} = \text{aileroM}$

La inversa se puede anular a si misma aplicando nuevamente la inversa, de tal forma que: $w = (w^I)^I$.

Cerradura de Kleene o cerradura estrella (∑*). Es el lenguaje con todas las cadenas que se pueden formar con el alfabeto (∑).

Sea $\Sigma = \{0, 1\}$
$\Sigma^* = \{0, 1, 00, 01, 10, 11, 000, 100, 101, 110, 111, 0000, 0001, 0010, \ldots\ldots\}$

∑ es un conjunto finito, pero ∑* es infinito, ya que el número de cadenas diferentes que se pueden formar es infinita.

Operaciones con lenguajes.

La operación con lenguajes algunas veces provoca un poco de dificultad. Con la finalidad de simplificar y hacer más nítido el manejo de lenguajes se se indicarán los lenguajes por medio de una sola letra mayúscula, de tal manera que L, M, N pueden ser lenguajes en lugar de L(G), M(G) y N(G).

Los principios de concatenación, potencia, inversa y cerradura de Kleene usados en cadenas, son aplicables también a lenguajes.

Concatenación de los lenguajes L y M (LM).

La concatenación de lenguajes se forma combinando todas las cadenas del lenguaje L con las cadenas del lenguaje M. Ejemplo.

Sean: $L = \{a, b, c\}$ y $M = \{1, 2\}$
$LM = \{a1, a2, b1, b2, c1, c2\}$

Observar que si L es lenguaje del alfabeto Σ_1 y M es lenguaje del alfabeto Σ_2, entonces LM será lenguaje del alfabeto $(\Sigma_1 \cup \Sigma_2)$.

Si se concatena un lenguaje L con otro que tiene como único elemento la cadena vacía, el lenguaje L no sufre cambio alguno. Así mismo; si se concatena un lenguaje L con el lenguaje vacío (∅) el resultado es ∅.

$$L\{\varepsilon\} = \varepsilon L = L$$
$$L\varnothing = \varnothing L = \varnothing$$

Potenciación de lenguajes. De la misma manera que se comporta la potencia con cadenas de símbolos, así se comporta cuando se trata de lenguajes. De tal manera que:

$L^n = \{ \varepsilon \}$ Si $n = 0$
$L^n = LL^{n-1}$ Si $n > 0$ $n \in N$

Ejemplo.

Sean:
L={011} sobre ∑={0, 1} entonces:

$L^0 = \{ \varepsilon \}$ Si $n = 0$
$L^4 = LL^3 = LLL^2 = LLLL^1 = LLLLL^0 = 011011011011\varepsilon =$ 011011011011

Inversa de un lenguaje (L^I). Es el lenguaje que se obtiene al escribir los elementos de un lenguaje en forma invertida.

Sea: L = {Morelia, Michoacán}; ∑={a, b, c,……..z}

L^I = {aileroM, nácaohciM}

Cerradura estrella sobre un lenguaje (L*). Es la unión de todos los lenguajes potencia de L desde n=0 hasta infinito ($n \in N$), que se pueden formar con el alfabeto (∑).

Sea ∑= {a, b}, L = {abb};
$L^* = L^0 \cup L^1 \cup \ldots\ldots\ldots \cup L^\infty = \{ \varepsilon \} \cup \{abb\} \cup \{abbabb\} \cup \ldots\ldots\ldots \cup \{abb\}^\infty$
$L^* = \{\varepsilon, abb, abbabb, abbabbabb, \ldots\ldots\ldots \}$

Cerradura Positiva de un lenguaje (L^+). Es la unión de todos los lenguajes potencia de L, desde n=1 hasta infinito, que se pueden formar con el alfabeto (∑).

Sea ∑= {a, b}, L = {abb};
$L^+ = L^1 \cup L^2 \cup \ldots\ldots\ldots \cup L^\infty = \{abb\} \cup \{abbabb\} \cup \ldots\ldots\ldots \cup \{abb\}^\infty$
$L^+ = \{abb, abbabb, abbabbabb, \ldots\ldots\ldots \}$

Cerradura elevada a una potencia ($(L^+)^i$). Es la concatenación i veces del lenguaje L^+.

$(L^+)^i = L^+L^+\ldots\ldots L^+$ i veces L^+

Sublenguaje. Si L y M son lenguajes, se dice que L es sublenguaje de M si todos los elementos de L están también contenidos en M. Se expresa como.

$$L \subseteq M$$

Observar que la diferencia entre la cerradura estrella y la cerradura positiva únicamente es la cadena vacía (ε). De tal manera que se puede afirmar que $L^{+} \subseteq L^{*}$.

Ejemplo.

Sean: $\sum = \{0, 1\}$
$L = \{011, 011011, 011011011011\}$;
$M = \{a^{n} \,/\, a=011;\ n \in N\} = \{011, 011011, 011011011, 011011011011, \ldots\}$

De aquí se puede observar que:
$L \subseteq M$

Considerando que un lenguaje es un conjunto de símbolos o palabras sobre un alfabeto, se puede llevar a cabo algunas operaciones de conjuntos como la unión, intersección, diferencia y complementación.

Sean L y M lenguajes entonces:

$(L \cup M) = \{x \,/ x \in L \text{ ó } x \in M \text{ ó } x \in (L \cap M)\}$
$(L \cap M) = \{x \,/ x \in L \text{ y } x \in M\}$
$(L - M) = \{x \,/ x \in L \text{ y } x \notin M\}$
$L^{c} = \sum^{*} - L = \{x \,/ x \in \sum^{*} \text{ y } x \notin L\}$

Sean: $\sum = \{a, b, c\}$
$L = \{\varepsilon, abc, aab, aabb, bbc, cb, cab\}$; $M = \{aa, ab, aabb, bab, cb, cc\}$

Entonces:
$(L \cup M) = \{\varepsilon, aa, aab, aabb, ab, abc, bab, bbc, cab, cb, cc\}$
$(L \cap M) = \{aabb, cb\}$
$(L - M) = \{\varepsilon, abc, aab, bbc, cab\}$
$L^{c} = \sum^{*} - L = \{x \,/ x \in \sum^{*} \text{ y } x \notin \{\varepsilon, abc, aab, aabb, bbc, cb, cab\}\}$

A continuación se muestra un pequeño resumen de algunas de las propiedades de las operaciones con lenguajes:

$(L^{I})^{I} = L$
$(LM)^{I} = M^{I} L^{I}$
$(L \cup M)^{I} = L^{I} \cup M^{I}$
$(L \cap M)^{I} = L^{I} \cap M^{I}$
$L^{*} = L^{0} \cup L^{+} = \{\varepsilon\} \cup L^{+}$
$(\varepsilon)^{+} = \{\varepsilon\} = \{\varepsilon\}^{*}$
$L\varepsilon = \varepsilon L\ L$
$L\varnothing = \varnothing L = \varnothing$
$L^{*} = (L^{*})^{*}$
$L^{+} = LL^{*} = L^{*}L$
$L^{+} = (L^{+})^{+}$
$L^{+} = LL^{*} = L^{*}L$

$$(L^c)^I = (L^I)^c$$
$$(L^+)^I = (L^I)^+$$
$$(L^*)^I = (L^I)^*$$
$$\varnothing^* = \varepsilon$$
$$(\varepsilon \cup L)^* = L^*$$
$$(L \cup M)^* = (L^*M^*)^*$$
$$(L^+)^0 = \{\varepsilon\}$$
$$(L^*)^0 = \{\varepsilon\}$$

Donde: L y M son lenguajes, ε la cadena vacía y ∅ el lenguaje vacío.

Lenguajes regulares.

Con el alfabeto ∑ es posible formar una gran cantidad de lenguajes, pero no existe un método efectivo para saber cuántos de ellos son regulares. Todos los lenguajes sobre ∑ son sublenguajes del lenguaje universo ∑*, en lugar de ello se utilizarán algunas propiedades de los lenguajes para determinar cuales de ellos son regulares. El conjunto de lenguajes regulares sobre ∑ son:.

- El lenguaje vacío ∅.
- El lenguaje que contiene la cadena vacía como único elemento {ε}.
- El lenguaje unitario {a} es regular $\forall$ a $\in$ ∑.
- Si L y M son lenguajes regulares entonces también lo son: El lenguaje que resulta de la unión (L∪M), la concatenación (LM), la potenciación (L^n o M^n) y la cerradura de Kleene (L*).
- Ningún otro lenguaje sobre ∑. Es regular.

Ejemplo sean ∑.={0, 1} L={01, 11, 001}; M={1, 11, 00} entonces:

- ∅ es regular.
- { ε} es regular.
- {0} y {1} son regulares ya que son conjuntos unitarios sobre ∑.
- L y M son lenguajes regulares.
- La unión (L∪M) = {1, 01, 00, 11, 001}. También es un lenguaje regular.
- La concatenación (LM) = { 011, 0111, 0100, 111, 1111, 1100, 0011, 00111, 00100}, es un lenguaje regular.
- La potenciación $L^2 = LLL^0 =$ {01, 11, 001}{01, 11, 001}{ ε) es lenguaje regular.
- La cerradura de Kleene $L^* = L^0 \cup L^1 \cup \ldots\ldots\ldots \cup L^\infty$. también es lenguaje regular.
- $\{0^n / n \in N\}$={0, 00, 000, 0000, …..} También es lenguaje regular (potenciación).
- $\{0^n1^m / n\geq 0; m\geq 0\}$. Potenciación y concatenación también es un lenguaje regular.
- $\{(011)^n / n \in N\}$. potenciación es lenguaje regular.

Se puede concluir que a partir del alfabeto ∑, los lenguajes obtenidos con las operaciones unión, concatenación, potenciación y cerradura de Kleene son lenguajes regulares, además de los lenguajes de un solo elemento, incluyendo la cadena vacía y el lenguaje vació.

Expresiones Regulares. Es una nueva forma de expresar los lenguajes regulares y tiene como finalidad facilitar la manipulación y simplificación de los mismos. La equivalencia entre leguajes regulares y expresiones regulares es como se indica en la siguiente tabla:

Lenguaje regular	Expresión regular
$\{a\}$	a
$\{\varepsilon\}$	ε
$\emptyset^*$	ε
$\{a\}^+$	a^+
$\{a\}^*$	a^*
$\{ab\}$	ab
$\{a\}\cup\{b\}=\{a,b\}$	$a\cup b$

De tal manera que el lenguaje $\{a, bc\}^*\{\varepsilon, aab\}\emptyset\{c\}$ sobre $\sum=\{a, b, c\}$, se puede indicar por medio de una expresión regular de la siguiente manera:

$$(a\cup bc)^*(\varepsilon\cup aab)\,\varepsilon c$$

La jerarquía de operación es como se indica en la siguiente tabla:

Operación	Jerarquía
paréntesis	1ª
*, +, n	2ª
Concatenación (·)	3ª
$\cup$	4ª

Los paréntesis solamente se usarán en casos extremos, ya que la finalidad de las expresiones regulares es la simplificación en la representación de los lenguajes regulares.

Muchas veces se tienen expresiones regulares muy complejas, las cuales se pueden simplificar para obtener una expresión más compacta. A continuación se tiene una lista de equivalencias de las expresiones regulares **a, b** y **c** sobre el alfabeto $\sum$, que es conveniente tener en cuenta en el momento de la simplificación.

No.	Equivalencia
1	$a\cup b=b\cup a$
2	$a\cup a = a$
3	$a\cup\emptyset = \emptyset\cup a = a$
4	$\emptyset^*=\varepsilon$
5	$(a\cup b)\cup c = a\cup(b\cup c)$
6	$a\varepsilon = \varepsilon a = a$
7	$a\emptyset = \emptyset a = \emptyset$
8	$b^* = \varepsilon \cup b^+$
9	$a^+ = aa^*$
10	$a(b\cup c) = ab\cup ac$; $(a\cup b)c = ac\cup bc$
11	$(ab)c = a(bc)$

12	$a^* = a^*a^* = a^{**} = (\varepsilon\cup a)^* = (a\cup\varepsilon)a^* = a^*(a\cup\varepsilon) = \varepsilon\cup aa^*$
13	$(a\cup b)^* = (a^*\cup b^*)^* = (a^*b^*)^* = (a^*b)^*a^* = a^*(ab^*)^*$
14	$a(ba)^* = (ab)^*a$
15	$(a^*b)^* = \varepsilon\cup(a\cup b)^*b$
16	$(ab^*)^* = \varepsilon\cup a(a\cup b)^*$
17	$b(a\cup\varepsilon)^*(a\cup\varepsilon)\cup b = ba^*$
18	$aa^* = a^*a$

Sean r, s y t expresiones regulares sobre ∑. Probrar que:

$$(rt^*)^*(\varepsilon\cup r(r\cup t)^*)(r(t\cup\varepsilon)^*(t\cup\varepsilon)\cup r)^* \text{ es equivalente a } (rt^*)^*(r((t\cup\varepsilon)^+\cup s))^*$$

Procedimiento :

$(rt^*)^*(\varepsilon\cup r(r\cup t)^*)(r(t\cup\varepsilon)^*(t\cup\varepsilon)\cup rs)^*$
$(rt^*)^*(rt^*)^*(r(t\cup\varepsilon)^*(t\cup\varepsilon)\cup rs)^*$ (16)
$(rt^*)^*(rt^*)^*(r(t\cup\varepsilon)^+\cup rs)^*$ (9)
$(rt^*)^*(rt^*)^*(r((t\cup\varepsilon)^+\cup s))^*$ (10)
$(rt^*)^*(r((t\cup\varepsilon)^+\cup s))^*$ (12)

Se puede observar que una expresión regular compacta es más clara y sencilla.