

Lenguajes Formales y Teoría de Autómatas (0901-028) TEORÍA PARA LA HOJA DE TRABAJO 6

INSTRUCCIONES: Agregue a su cuaderno las siguientes definiciones y teoremas. Imprima esta hoja y téngala a mano a fin de resolver la hoja de trabajo adjunta.

Lenguajes, Autómatas y Gramáticas Capítulo 3

21. AFABETO, PALABRAS Y SEMIGRUPO LIBRE

Escolio: En este capítulo se estudian tres temas que tienen una estrecha relación entre sí: *lenguajes*, *autómatas* y *gramáticas*. En los lenguajes que se usan aquí se utilizan las letras a, b, \ldots para codificar los datos, a diferencia de los dígitos 0 y 1 que se usan en otros textos.

- —**Definición 42** (**Palabras de un Alfabeto**): Sea $A \neq \emptyset$, el cual llamaremos *alfabeto*.
 - (i) $palabra_A(x, n) \stackrel{\Delta}{\Leftrightarrow} x: \{1, ..., n\} \longrightarrow A, \text{ para algún } n \ge 1.$
 - (ii) $palabra_A(\lambda, 0) \stackrel{\Delta}{\Leftrightarrow} \lambda : \emptyset \longrightarrow A \iff \lambda = \emptyset \text{ (por teorema 55(v))}$
 - (iii) $x_n \triangleq x(n)$, si $palabra_A(x, n)$, para algún $n \ge 1$.
 - (iv) $x_{A,n} \triangleq x_1 x_2 \dots x_n \triangleq x$, si $palabra_A(x,n)$, para algún $n \geq 1$.
 - (v) Se usarán *u*, *v*, *w*, *x*, *y*, *z* como metavariables que toman valores a partir de las palabras de un alfabeto dado.

Escolio: Es decir, una palabra o cadena (de longitud n), $w_{A,n}$, sobre la clase A es una sucesión finita de elementos de A. A dicha clase A se le llama alfabeto.

Ejemplo: Sea el alfabeto $A = \{a, b, c\}$. Entonces las siguientes sucesiones son palabras sobre dicho alfabeto A.

$$u_{A,5} = ababb$$

 $v_{A,7} = accbaaa$

—**Definición 43 (Letra):** $letra_A(a) \stackrel{\Delta}{\Leftrightarrow} a \in A$, si $A \neq \emptyset$.

Escolio: Sea el alfabeto $A \neq \emptyset$. A los elementos de A se les denomina *letras*.

—**Definición 44 (Notación):** Sea el alfabeto $A \neq \emptyset$. Denotaremos por λ (o bien por ε) a la sucesión vacía de letras y se le denominará *palabra vacía*. Formalmente,

- (i) $a^0 \triangleq \lambda \triangleq \varepsilon \triangleq \emptyset$, si a es una letra de A.
- (ii) $a^1 \triangleq a$, si a es una letra de A.
- (iii) $a^{n+1} \triangleq a^n a, \forall n \geq 0$, si a es una letra de A.

—**Teorema 60:** $a^n = aa \dots a, n - veces$, es una palabra de A, si a es una letra del alfabeto $A \neq \emptyset$.

Demostración (por inducción finita sobre n):

Sea a es una letra del alfabeto $A \neq \emptyset$.

Si n=0 entonces $a^n=a^0=\lambda$ (por hipótesis y la definición 44(i)) es una palabra vacía de A (por definición 44).

Si $a^n = aa ... a, n - veces$, es una palabra de A, entonces

$$a^{n+1} = a^n a$$
 por definición 44(iii)
= $\underbrace{aa \dots a}_{n-veces} a, n+1-veces$ por hipótesis de inducción

La cual, a su vez, es una palabra sobre A.

Escolio: En base al teorema anterior se abrevia la notación y se escribe a^2 en lugar de aa, a^3 por aaa, etc. de modo que las palabras del ejemplo anterior se pueden reescribir como:

$$u_{A,5} = ababb = abab^2$$

 $v_{A,7} = accbaaa = ac^2ba^3$

—Definición 45 (La Clase Estrella):

 $A^* \triangleq \{x | (\forall n) palabra_A(x, n)\} = \{x_{A,n} | (\forall n) palabra_A(x_{A,n}, n)\}, \text{ si } A \neq \emptyset.$

Escolio: A la clase de todas las palabras (de distinta longitud) sobre A se denota por A^* .

—Definición 46 (La Longitud de una Palabra):

- (i) $l(x_{A,n}) \triangleq n$, si $palabra_A(x,n)$, para algún $n \geq 1$.
- (ii) $l(\lambda) \triangleq 0$

Ejemplos: $l(abab^2) = 5$, $l(ac^2ba^3) = 7$.

—Definición 47 (Concatenación de Palabras):

- (i) $x_{A,m}y_{A,n} \triangleq c(x_{A,m} y_{A,n})$, si $c: (A^*)^2 \longrightarrow A^*$
- (ii) $x_{A,m}^0 \triangleq \lambda$
- (iii) $x_{A,m}^1 \triangleq x_{A,m}$
- (iv) $x_{A,m}^{n+1} \triangleq x_{A,m}^n x_{A,m}$

—**Teorema 61:** Sea el alfabeto $A \neq \emptyset$.

- (i) $oper_{A^*}(c)$
- (ii) $(x_{A,m}y_{A,n})z_{A,k} = x_{A,m}(y_{A,n}z_{A,k}), \forall x_{A,m}, y_{A,n}, z_{A,k} \in A^*$
- (iii) $\lambda \in A^*$
- (iv) $x_{A,m}^n = x_{A,m} x_{A,m} \dots x_{A,m}, n veces$, es una palabra de longitud $m \times n$ sobre el alfabeto $A \neq \emptyset$.
- (i) Demostración:

Ya que $c: (A^*)^2 \to A^*$ (definición 47(i)), se sigue que $\operatorname{oper}_{A^*}(c)$, a partir de la definición 32(i).

(ii) Demostración:

Sean $x_{A,m}$, $y_{A,n}$, $z_{A,k} \in A^*$ tales que, a partir de la definición 42(iii), son posibles de reescribir como:

$$x_{A,m} = x_1 x_2 \dots x_m$$
 (1)

$$y_{A,n} = y_1 y_2 \dots y_n$$
 (2)

$$z_{A,n} = z_1 z_2 \dots z_k$$
 (3)

Entonces:

$$(x_{A,m}y_{A,n})z_{A,k} = (x_1x_2 ... x_m y_1y_2 ... y_n)z_1z_2 ... z_k \quad a \ partir \ de \ (1), (2) \ y \ (3)$$

$$= x_1x_2 ... x_m y_1y_2 ... y_nz_1z_2 ... z_k \quad por \ definición \ 47(i)$$

$$= x_{A,m}(y_{A,n}z_{A,k}) \qquad por \ la \ misma \ razón$$

(iii) Demostración:

∴ palabra_A(
$$\lambda$$
, 0) por definición 42(ii)
∴ $\lambda \in A^*$ por definiciones 45, 5(i)

(iv) Demostración (por inducción finita sobre n):

Sea
$$x_{A,m} \in A^*$$
.

Si n=0 entonces $x_{A,m}^n=x_{A,m}^0=\lambda$ (por hipótesis y la definición 47(ii)) es una palabra vacía de A de longitud $0=m\times 0=m\times n$ (por definición 44). Si $x_{A,m}^n=x_{A,m}x_{A,m}\dots x_{A,m}, n-veces$, es una palabra de longitud $m\times n$ sobre A, entonces

$$x_{A,m}^{n+1} = x_{A,m}^n x_{A,m}$$
 por definición 47(iv)
= $\underbrace{x_{A,m} x_{A,m} \dots x_{A,m}}_{n-veces} x_{A,m}, n+1-veces$ por hipótesis de inducción

La cual, a su vez, es una palabra sobre A de longitud $m \times (n+1)$.

Ejemplo: la *concatenación* de $u_{A,5}$ y $v_{A,7}$, por ejemplo, se escribe

$$u_{A,5}v_{A,7} = abab^2ac^2ba^3$$

Además.

$$v_{A7}u_{A5} = ac^2ba^3abab^2 = ac^2ba^4bab^2$$

Puede notarse que la operación concatenación NO es conmutativa.

Escolio: En este caso, como ocurre con las letras, para cualquier palabra $u_{A,n}$, a partir del teorema 61(ii) se tiene que $u_{A,n}^2 = u_{A,n}u_{A,n}$, $u_{A,n}^3 = u_{A,n}u_{A,n}u_{A,n}$, etc.

Resulta evidente, además, que para palabras arbitrarias $u_{A,m}$, $v_{A,n}$, $w_{A,k}$ las palabras $(u_{A,m}v_{A,n})w_{A,k}$ y $u_{A,m}(v_{A,m}w_{A,n})$ son idénticas, ya que sólo consta de las letras de $u_{A,m}$, $v_{A,n}$, $w_{A,k}$ escritas una después de la otra.

También, al adjuntar una palabra vacía antes o después de una palabra $u_{A,m}$ no se modifica la palabra. Es decir,

—**Teorema 62:** $Semigr(A^*, c, \lambda)$, si $A \neq \emptyset$.

Demostración:

Claramente, a partir de la definición 33,

$$mon(A^*,c) \qquad (a)$$
ya que $\sigma per_{A^*}(c)$ (por teorema 61(i)) y
$$(x_{A,m}y_{A,n})z_{A,k} = x_{A,m}(y_{A,n}z_{A,k}), \forall x_{A,m}, y_{A,n}, z_{A,k} \in A^* \text{ (por el teorema 61(ii))}.$$
Por otro lado, siendo $\lambda \in A^*$ (por teorema 61(iii)
$$\lambda x_{A,m} = x_{A,m}\lambda = x_{A,m}, \forall x_{A,m} \in A^* \qquad (b) \qquad \text{(a partir de la definición 44 (i))}$$

De (a) y (b) se sigue la tesis (a partir de la definición 34).

Escolio: la clase de todas las palabras de A inclusive la palabra vacía λ , A^* , es llamado también *semigrupo libre (sobre A)*.

—Corolario 1.62: $mon(A^* - \{\lambda\}, c)$, si $A \neq \emptyset$.

Escolio: la clase de todas las palabras no vacías de A, $A^* - \{\lambda\}$, es llamado también *monoide libre (sobre A o generado por A)*. Resulta fácil demostrar que $A^* - \{\lambda\}$ satisface las leyes de cancelación por la izquierda y por la derecha, a pesar de que (como ya se vio anteriormente) no es conmutativo cuando $A^* - \{\lambda\}$ posee más de un elemento.

—Corolario 2.62 (Leyes de Cancelación): Si $A \neq \emptyset$,

- (i) $x_{A,m}y_{A,n} = z_{A,m}y_{A,n} \Longrightarrow x_{A,m} = z_{A,m}, \forall x_{A,m}, y_{A,n}, z_{A,k} \in A^*$
- (ii) $y_{A,n}x_{A,m} = y_{A,n}z_{A,m} \Longrightarrow x_{A,m} = z_{A,m}, \forall x_{A,m}, y_{A,n}, z_{A,k} \in A^*$

—Definición 48 (Subpalabras):

- (i) $subpalabra_{x_{A}n}(y_{A,k-j+1}) \stackrel{\Delta}{\Leftrightarrow} x = y \land 1 \le j \le k \le n, \text{ si } A \ne \emptyset$
- (ii) $subpalabra_{x_{A,n}}(\lambda), \forall x_{A,n}, \text{ si } A \neq \emptyset$

Escolio: es decir, cualquier sucesión de la forma $w_{A,k-j+1} = a_j a_{j+1} \dots a_k$ es una subpalabra de $v_{A,n} = a_1 a_2 \dots a_n$ ssi $1 \le j \le k \le n$. En otras palabras, w es una subpalabra de u ssi es de la forma $u = v_1 w v_2$. Observe que ambas λ y u son subpalabras de u puesto que $u = \lambda u$.

Ejemplo: considere la palabra $u_{A,4} = abca$. Las subpalabras de $u_{A,4}$ son las siguientes: λ , a, b, c, ab, bc, ca, abc, bca, $u_{A,4}$ (10 subpalabras).

—Definición 49 (Segmentos iniciales):

- (i) $segmentoinicial_{x_{A,n}}(y_{A,k}) \stackrel{\Delta}{\Leftrightarrow} x = y \land x_1 = y_1 \land 1 \le k \le n, \text{ si } A \ne \emptyset$
- (ii) segmentoinicial_{$x_{A,n}$}(λ), $\forall x_{A,n}$, si $A \neq \emptyset$

Escolio: es decir, cualquier sucesión de la forma $w_{A,k} = a_1 a_2 \dots a_k$ es un segmento inicial de $v_{A,n} = a_1 a_2 \dots a_n$ ssi $1 \le k \le n$. En otras palabras, w es un segmento inicial de u ssi u = wv. Note que todo segmento inicial es una subpalabra.

Ejemplo: considere la palabra $u_{A,4} = abca$. Los segmentos iniciales de $u_{A,4}$ son los siguientes: λ , a, ab, abc, $u_{A,4}$ (5 subpalabras).

—**Teorema 63:** $segmentoinicial_{x_{A,n}}(y_{A,k}) \Rightarrow subpalabra_{x_{A,n}}(y_{A,k-j+1})$, si $A \neq \emptyset$ **Escolio:** Es decir, todo segmento inicial es una subpalabra.

22. LENGUAJES

—**Definición 50 (Lenguajes):** $lenguaje_A(L) \stackrel{\Delta}{\Leftrightarrow} L \subseteq A^*$, si $A \neq \emptyset$.

Escolio: un lenguaje L sobre un alfabeto A es una colección de palabras sobre A. Como A^* denota el conjunto de todas las palabras sobre A, se sigue que un lenguaje L no es más que una subclase de A^* .

Ejemplo: Sea $A = \{a, b\}$. Algunos lenguajes sobre A son los siguientes:

- (i) $L_1 = \{a, ab, ab^2, ...\} = \{ab^n | n \ge 0\}$ (L_1 consta de todas las palabras que empiezan con una a seguidas de cero o más b. No incluye la palabra vacía).
- (ii) $L_2 = \{a^m b^n | m > 0 \land n > 0\}$ (L_2 consta de todas las palabras que empiezan con una o más a seguidas de una o más b. No incluye la palabra vacía).
- (iii) $L_3 = \{a^m b^m | m > 0\}$ (L_3 consta de todas las palabras que empiezan con una o más a seguidas por el mismo número de b. No incluye la palabra vacía).
- (iv) $L_4 = \{b^m a b^n | m \ge 0 \land n \ge 0\}$ (L_3 consta de todas las palabras que empiezan con una o más a seguidas por el mismo número de b. No incluye la palabra vacía).

—**Teorema 64:** $lenguaje_A(L) \Leftrightarrow L \in \mathcal{P}(A^*)$, $\operatorname{si} \mathcal{M}(A) \land A \neq \emptyset$. **Demostración:**

Si $\mathcal{M}(A) \land A \neq \emptyset$ entonces

$$L \in \mathcal{P}(A^*) \iff L \subseteq A^*$$
 por definiciones 12, 5(i)
 $\iff lenguaje_A(L)$ por definición 50

La tesis se sigue de la simetría de la equivalencia (corolario II.XXIII). ■

—Definición 51 (Operación sobre los Lenguajes: la Concatenación de Lenguajes): Si $\mathcal{M}(A) \land A \neq \emptyset$, entonces

- (i) $con(L \ M) \triangleq LM \triangleq \{x_{A,m}y_{A,n} | (\forall m,n)[x_{A,m} \in L \land y_{A,n} \in M]\}, \forall L,M \in \mathcal{P}(A^*),$ donde $con: \mathcal{P}^2(A^*) \longrightarrow \mathcal{P}(A^*)$
- (ii) $L^0 \triangleq {\lambda} = {\emptyset} = 1, \forall L \in \mathcal{P}(A^*)$
- (iii) $L^1 \triangleq L, \forall L \in \mathcal{P}(A^*)$
- (iv) $L^{n+1} \triangleq L^n L, \forall L \in \mathcal{P}(A^*), \forall n \geq 1$

Escolio: la anterior definición establece que si L y M son lenguajes sobre un alfabeto A, entonces la "concatenación" de L y M, que se denota por LM, es el nuevo lenguaje definido como sigue:

$$LM = \{uv \mid u \in L \land v \in V\}$$

Es decir, LM denota el conjunto de todas las palabras que provienen de la concatenación de una palabra de L con otra de M.

En general, a L^n se le llama usualmente la potencia de un lenguaje.

Ejemplo: suponga que

$$L_1 = \{a, b^2\}, L_2 = \{a^2, ab, b^3\}, L_3 = \{a^2, a^4, a^6, \ldots\}$$

Entonces:

$$L_1L_1 = \{a^2, ab^2, b^2a, b^4\}, L_1L_2 = \{a^3, a^2b, ab^3, b^2a^2, b^2ab, b^5\},$$

 $L_1L_3 = \{a^3, a^5, a^7, \dots, b^2a^2, b^2a^4, b^2a^6, \dots\}$

Además, ya que, L_2L_1 = { a^3 , aba, b^3a , a^2 b^2 , a b^3 , b^5 } se sigue que, en general

$$LM \neq ML$$

Sin embargo, ya que la concatenación de palabras es asociativa, resulta evidente que la concatenación de lenguajes es asociativa.

- —**Teorema 65:** Si $\mathcal{M}(A) \wedge A \neq \emptyset$,
 - (i) $oper_{\mathcal{P}(A^*)}(con)$
 - $(LM)N = L(MN), \forall L, M, N \in \mathcal{P}(A^*)$ (ii)
 - $L^0 \in \mathcal{P}(A^*)$ (iii)
 - $L^n = LL \dots L, n veces, \forall L \in \mathcal{P}(A^*).$ (iv)
- **—Teorema 66:** $Semigr(\mathcal{P}(A^*), con, \{\lambda\})$, $si \mathcal{M}(A) \land A \neq \emptyset$.

—Definición 52 (La Cerradura de Kleene):

- $\begin{array}{l} L^* \triangleq \bigcup_{k=0}^{\infty} L^k \triangleq \bigcup \{L^k | L \in \mathcal{P}(A^*) \wedge n \geq 0\}, \, \mathrm{si} \,\, \mathcal{M}(A) \wedge A \neq \emptyset. \\ L^+ \triangleq \bigcup_{k=1}^{\infty} L^k \triangleq \bigcup \{L^k | L \in \mathcal{P}(A^*) \wedge n \geq 1\}, \, \mathrm{si} \,\, \mathcal{M}(A) \wedge A \neq \emptyset. \end{array}$ (ii)

Escolio: La operación unaria L^* (que se lee "L estrella") de un lenguaje L, que se denomina cerradura de Kleene de L se define como la unión infinita:

$$L^* = \bigcup_{k=0}^{\infty} L^k = L^0 \cup L \cup L^2 \cup \dots$$

La definición de L^* coincide con la notación A^* , que consta de todas las palabras sobre A. L^+ se define como la unión de L^1, L^2, \ldots es decir, L^+ es lo mismo que L^* , aunque sin la palabra vacía λ.

23. EXPRESIONES REGULARES, LENGUAJES REGULARES

Escolio: En esta sección se definen una expresión regular r sobre A y un lenguaje L(r)sobre A en asociación con la expresión regular r. La expresión r y su lenguaje correspondiente L(r) se definen inductivamente como sigue.

—Definición 53 (Expresiones Regulares de un Alfabeto): Si $\mathcal{M}(A) \wedge A \neq \emptyset$, entonces

- (i) λ (la palabra vacía) y () (la expresión vacía) son expresiones regulares (de (el alfabeto) A).
- $a \in A$ es una expresión regular (de A). Es decir, toda letra en A es una (ii) expresión regular.
- (iii) Si r es una expresión regular entonces (r^*) es una expresión regular (de A).
- Si r_1 y r_2 son expresiones regulares entonces $(r_1 \lor r_2)$ es una expresión (iv) regular (de A).
- Si r_1 y r_2 son expresiones regulares entonces (r_1r_2) es una expresión regular (v)
- r es una expresión regular (de A) ssi se obtiene a partir de los incisos (i)–(v), (vi) anteriores.

Escolio: Observe que una expresión regular r es un tipo especial de palabra (cadena) que usa las letras de *A* y los cinco símbolos:

$$() * V \lambda$$

Se recalca que ningún otro símbolo se usa para las expresiones regulares.

—Definición 54 (Lenguaje Regular sobre una Expresión Regular de un Alfabeto): Si $\mathcal{M}(A) \land A \neq \emptyset$, entonces

- (i) $L(\lambda) \triangleq {\lambda} = {\emptyset} = 1, \forall L \in \mathcal{P}(A^*)$
- (ii) $L(()) \triangleq \emptyset = 0, \forall L \in \mathcal{P}(A^*)$
- (iii) $L(a) \triangleq \{a\}, \forall L \in \mathcal{P}(A^*), \forall a \in A$
- (iv) $L(r^*) \triangleq L(r)^*, \forall L \in \mathcal{P}(A^*)$, si r es una expresión regular de A.
- (v) $L(r_1 \lor r_2) \triangleq L(r_1) \cup L(r_2), \forall L \in \mathcal{P}(A^*)$, si $r_1 \lor r_2$ son expresiones regulares de A.
- (vi) $L(r_1r_2) \triangleq L(r_1)L(r_2), \forall L \in \mathcal{P}(A^*), \text{ si } r_1 \text{ y } r_2 \text{ son expresiones regulares de } A.$
- (vii) L es un lenguaje regular sobre r, una expresión regular, el cual denotaremos por L(r), ssi se obtiene a partir de alguno de los incisos (i)—(vi), anteriores.

Escolio: En la anterior definición $L(r)^*$ denota la cerradura de Kleene para L(r). Además, $L(r_1)L(r_2)$ denota la concatenación de lenguajes. Cuando es posible, en las expresiones regulares se omiten los paréntesis. Puesto que la concatenación de lenguajes y la unión de lenguajes son asociativas, es posible omitir muchos de los paréntesis. También, al adoptar la convención de que "*" tiene precedencia sobre la concatenación y ésta a su vez tiene precedencia sobre "V," es posible omitir otros paréntesis.

Ejemplo 1: Sea $A = \{a, b\}$. Cada una de las siguientes es una expresión r y su lenguaje correspondiente es L(r):

- a) Sea $r = a^*$. Entonces L(r) consta de todas las potencias de a incluso la palabra vacía λ .
- b) Sea $r = aa^*$. Entonces L(r) consta de todas las potencias positivas de a excepto la palabra vacía λ .
- c) Sea $r = a \vee b^*$. Entonces L(r) consta de a o de cualquier palabra en b; es decir,

$$L(r) = \{a, \lambda, b, b^2, \cdots\}.$$

- d) Sea $r=(a \vee b)^*.$ Observe que $L(a \vee b)=\{a\} \cup \{b\}=A;$ por tanto, $L(r)=A^*,$ todas las palabras sobre A.
- e) Sea $r = (a \lor b)*bb$. Entonces L(r) consta de la concatenación de cualquier palabra en A con bb; es decir, todas las palabras que terminan en b^2 .
- f) Sea $r = a \wedge b^*$. L(r) no existe puesto que r no es una expresión regular. (En este caso \wedge no es uno de los símbolos que se usan para expresiones regulares.)

Ejemplo 2: Considere los siguientes lenguajes sobre $A = \{a, b\}$.

a)
$$L_1 = \{a^m b^n \mid m > 0, n > 0\}; b) L_2 = \{b^m a b^n \mid m > 0, n > 0\}; c) L_3 = \{a^m b^m \mid m > 0\}.$$

Encontrar una expresión regular r sobre $A = \{a, b\}$ tal que $L_i = L(r)$ para $i \in \{1, 2, 3\}$.

- a) L_1 consta de aquellas palabras que empiezan con una o más a seguidas por una o más
- b. Por tanto, $r = aa^*bb^*$. Observe que r NO es única; por ejemplo, $r = a^*abb^*$ es otra solución.

- b) L_2 consta de todas las palabras que empiezan con una o más b seguidas por una sola a que luego es seguida por una o más b; es decir, todas las palabras que contienen exactamente una a que no es la primera o la última letra. Por tanto, r = bb*abb* es una solución.
- c) L_3 consta de todas las palabras que empiezan con una o más a seguidas por el mismo número de b. No existe ninguna expresión regular r tal que L_3 = L(r); es decir, L_3 no es un lenguaje regular. La demostración de este hecho se deja en la hoja de trabajo adjunta.