Tema 4: Aplicación de los autómatas: Lenguajes formales

3. Relacin entre los lenguajes regulares y los autmatas .

Definición. Sea S $= (S, E, \delta)$ un semiautómata con estado inicial e_1 y $E_1 \subseteq E$. Se llama **lenguaje representado por S respecto de** E_1 al conjunto

$$Ł(\mathrm{S}, E_1) = \{x \in \Omega_S | \hat{\delta}(e_1, x) \in E_1\}.$$

Es obvio que si $E_1 = \{e_{i1}, \ldots, e_{ir}\}$, entonces

$$Ł(\mathrm{S}, E_1) = \sum_{j=1}^{r} Ł(\mathrm{S}, e_{ij}).$$

Ejemplo. Consideramos el semiautómata S $= (S, E, \delta)$, donde $S = a_1, a_2$, $E = \{e_1, e_2, e_3\}$ y

$$\begin{array}{rcl} \delta : E \times S & \rightarrow & E \\ (e_1, a_1) & \mapsto & e_2 \\ (e_1, a_2) & \mapsto & e_3 \\ (e_2, a_1) & \mapsto & e_2 \\ (e_2, a_3) & \mapsto & e_2 \\ (e_3, a_1) & \mapsto & e_1 \\ (e_3, a_2) & \mapsto & e_2 \end{array}$$

Entonces,

$$Ł(\mathrm{S}, \{e_1\}) = \{(a_2 a_1)^i | i \in \mathbb{N} \cup \{0\}\}.$$

Observamos que el lenguaje del ejemplo anterior es un lenguaje regular, ya que es el lenguaje asociado a la expresión regular $(a_2 a_1)^*$. En lo que sigue vamos a estudiar la relación existente entre los lenguajes regulares y los lenguajes asociados a semiautómatas.

Definición. Sea S $= (S, E, \delta)$ un semiautómata y $e_i, e_j \in E$. Diremos que de e_i se **pasa** a e_j si existe $x \in \Omega_S$ tal que $\hat{\delta}(e_i, x) = e_j$. Si $x = s_1 \ldots s_r$, entonces a $\delta(e_i, s_1)$, $\hat{\delta}(e_i, s_1 s_2)$, ..., $\hat{\delta}(e_i, s_1 s_2 s_{r-1})$ se les llama **estados intermedios** del paso de e_i a e_j mediante x.

Definición. Sea S $= (S, E, \delta)$ un semiautómata tal que $E = \{e_1, \ldots, e_n\}$, y sean $i, j \in \{1, \ldots, n\}$ y $k \in \{0, 1, \ldots, n\}$. Se llama

$$Ł^k_{ij} = \{x \in \Omega_S | x = s_{i1} \ldots s_{ir}, \hat{\delta}(e_i, x) = e_j \text{ y } \hat{\delta}(e_i, s_{i1} s_{it}) \notin E - \{e_1, \ldots, e_k\}, t \in \{1, \ldots, r-1\}\}$$

entendiendo que si $k = 0$ la última condición significa que no hay estados intermedios en el paso de e_i a e_j. Esto es,

$$Ł^0_{ij} = \{x \in \Omega_S | \delta(e_i, x) = e_j \text{ y no hay estados intermedios}\}$$

y

$$Ł^n_{ij} = \{x \in \Omega_S | \delta(e_i, x) = e_j\}.$$

Un caso especial es cuando consideramos L^n_{1j} que coincide con Ł(S$, e_j$).

En lo que sigue vamos a probar que $Ł^k_{ij}$ es un lenguaje regular. En primer lugar relacionamos $Ł^k_{ij}$ con $Ł^{k-1}_{rs}$ para determinados índices r y s.

Lema 3.1. *Sea $S = (S, E, \delta)$ un semiautómata tal que $E = \{e_1, \ldots, e_n\}$, y sean $i, j, k \in \{1, \ldots, n\}$. Entonces,*

$$Ł^k_{ij} = Ł^{k-1}_{ij} + Ł^{k-1}_{ik} (Ł^{k-1}_{kk})^* Ł^{k-1}_{kj}.$$

Demostración. $L^k_{ij} \subseteq Ł^{k-1}_{ij} + Ł^{k-1}_{ik} (Ł^{k-1}_{kk})^* Ł^{k-1}_{kj}$. Sea $x \in L^k_{ij}$. Entonces, $\hat{\delta}(e_i, x) = e_j$ y los estados intermedios que aparecen son pertenecen al conjunto $\{e_1, \ldots, e_k\}$. Si e_k no es estado intermedio, entonces $x \in L^{k-1}_{ij}$. Si e_k es un estado intermedio, entonces podemos escribir $x = y_1 y_2 y_3$, donde $y_i \in \Omega_S$ y tales que

$$\begin{aligned}
\hat{\delta}(e_i, y_1) &= e_k && \text{y } e_k \text{ no es estado intermedio} \\
\hat{\delta}(e_k, y_2) &= e_k && \text{y los estados intermedios pertenecen a } \{e_1, \ldots e_k\} \\
\hat{\delta}(e_k, y_3) &= e_j && \text{y } e_k \text{ no es estado intermedio.}
\end{aligned}$$

Por tanto, $y_1 \in Ł^{k-1}_{ik}$ y $y_3 \in Ł^{k-1}_{kj}$. A su vez, descomponemos $y_2 = z_1 \ldots z_t$ de forma que

$$\hat{\delta}(e_k, z_r) = e_k \qquad \text{y } e_k \text{ no es estado intermedio para } r = 1, \ldots, t.$$

Entonces, $y_2 \in (Ł^{k-1}_{kk})^*$.

$Ł^{k-1}_{ij} + Ł^{k-1}_{ik} (Ł^{k-1}_{kk})^* Ł^{k-1}_{kj} \subseteq Ł^k_{ij}$. Sea $x \in Ł^{k-1}_{ij} + Ł^{k-1}_{ik} (Ł^{k-1}_{kk})^* Ł^{k-1}_{kj}$. Si $x \in Ł^{k-1}_{ij}$, entonces $x \in Ł^k_{ij}$ ya que $Ł^{k-1}_{ij} \subseteq Ł^k_{ij}$. Si $x \in Ł^{k-1}_{ik} (Ł^{k-1}_{kk})^* Ł^{k-1}_{kj}$, entonces podemos descomponer $x = y_1 y_2 y_3$, donde $y_1 \in Ł^{k-1}_{ik}$, $y_2 \in (Ł^{k-1}_{kk})^*$ y $y_3 \in Ł^k_{kj}$. Pero entonces,

$$\hat{\delta}(e_i, x) = \hat{\delta}(\hat{\delta}(\hat{\delta}(e_i, y_1), y_2), y_3) = \hat{\delta}(\hat{\delta}(e_k, y_2), y_3) = \hat{\delta}(e_k, y_3) = e_j$$

y los estados intermedios que aparecen cuando calculamos $\hat{\delta}(e_i, y_1)$ pertenecen al conjunto $\{e_1, \ldots, e_{k-1}\}$, $\hat{\delta}(e_k, y_2)$ pertenecen al conjunto $\{e_1, \ldots, e_k\}$ y $\hat{\delta}(e_k, y_3)$ pertenecen al conjunto $\{e_1, \ldots, e_{k-1}\}$. Por tanto, $x \in Ł^k_{ij}$. □

Proposición 3.2. *Sea $S = (S, E, \delta)$ un semiautómata tal que $E = \{e_1, \ldots, e_n\}$, y sean $i, j \in \{1, \ldots, n\}$ y $k \in \{0, \ldots, n\}$. Entonces, $Ł^k_{ij}$ es un lenguaje regular.*

Demostración. (Por inducción sobre k) Si $k = 0$, entonces $Ł^0_{ij} \subseteq S \cup \{\Lambda\}$ ó $Ł^0_{ij} = \emptyset$ y en cualquier caso es un lenguaje regular.

Supongamos que el resultado es cierto para $k < t$ y veamoslo para $k = t$. Entonces, por el lema anterior sabemos que

$$Ł^t_{ij} = Ł^{t-1}_{ij} + Ł^{t-1}_{it}(Ł^{t-1}_{tt})^* Ł^{t-1}_{tj}$$

y empleando inducción, tenemos que $Ł^{t-1}_{ij}$, $Ł^{t-1}_{it}$, $Ł^{t-1}_{tt}$ y $Ł^{t-1}_{tj}$ son lenguajes regulares. Como la suma, concatenación y clausura de lenguajes regulares es regular, se sigue que $Ł^t_{ij}$ es regular. □

Como Corolario de esta proposición, tenemos

Corolario 3.3. *Sea $S = (S, E, \delta)$ un semiautómata tal que $E = \{e_1, \ldots, e_n\}$. Entonces, $L(S, e_j$ es un lenguaje regular para todo $j \in \{1, \ldots, n\}$.*

Corolario 3.4. *Sea $S = (S, E, \delta)$ un semiautómata tal que $E = \{e_1, \ldots, e_n\}$ y $E_1 \subseteq E$. Entonces, $L(S, E_1)$ es un lenguaje regular.*

Ahora nos falta ver que a partir de un lenguaje regular se puede construir un semiautómata y un subconjunto E_1 de forma que el lenguaje asociado al semiautómata respecto de E_1. En primer lugar, vamos a dar un ejemplo del que extraeremos la estrategia para probar el caso general.

Ejemplo. Sea L el lenguaje regular cuya expresión regular viene dada por $a(a^* + b)^* + abb^*ab^*$, esto es,

$$L = |a(a^* + b)^* + abb^*ab^*|.$$

Paso 1. Transformamos la expresión regular que define L en otra en la que las letras que lo forman aparezcan subindicadas por la posición que ocupan

$$a(a^* + b)^* + abb^*ab^* \Rightarrow a_1(a_2^* + b_3)^* + a_4b_5b_6^*a_7b_8^*$$

A c_i le llamaremos **descendiente** de c, siendo c una de las letras que parece en la expresión regular. Al lenguaje que tiene por expresión regular a la nueva lo denotamos por L_1, esto es,

$$L_1 = |a_1(a_2^* + b_3)^* + a_4 b_5 b_6^* a_7 b_8^*|$$

Paso 2. Localizamos

(i) Pares de letras que aparezcan consecutivas en palabras de L_1. En nuestro caso,

$$\begin{aligned} P =& \{(a_1, a_2), (a_1, b_3), (a_2, a_2), (a_2, b_3), (b_3, a_2), (b_3, b_3), \\ & (b_4, a_5), (a_5, b_6), (b_6, b_6), (a_5, a_7), (b_6, a_7), (a_7, b_8), (b_8, b_8)\}. \end{aligned}$$

(ii) Letras de inicio de palabra. En nuestro caso,

$$I = \{a_1, b_4\}.$$

(iii) Letras de final de palabra. En nuestro caso,

$$F = \{a_1, a_2, b_3, a_7, b_8\}.$$

Paso 3. Se define el semiautómata S $= (S, E, \delta)$, donde S es el albabeto del lenguaje L y la función δ y E se definen como sigue: Dado $c \in S$,

$$\delta(e_1, c) = \{\text{descendientes de } c \text{ que satisfacen 2(ii) }\} \qquad \forall c \in S$$

$$\delta(e_i, c) = \{\text{descendientes de } c \text{ que siguen a alguna letra que aparece en } e_i\} \quad \forall i > 1.$$

En nuestro caso,

$$S = \{a, b\},$$

$$\begin{array}{ll}
\delta(e_1, a) = \{a_1\} = e_2 & \delta(e_1, b) = \{b_4\} = e_3 \\
\delta(e_2, a) = \{a_2\} = e_4 & \delta(e_2, b) = \{b_3\} = e_5 \\
\delta(e_3, a) = \{a_5\} = e_6 & \delta(e_3, b) = \emptyset = e_7 \\
\delta(e_4, a) = \{a_2\} = e_4 & \delta(e_4, b) = \{b_3\} = e_5 \\
\delta(e_5, a) = \{a_2\} = e_4 & \delta(e_5, b) = \{b_3\} = e_5 \\
\delta(e_6, a) = \{a_7\} = e_8 & \delta(e_6, b) = \{b_6\} = e_9 \\
\delta(e_7, a) = \emptyset = e_7 & \delta(e_7, b) = \emptyset = e_7 \\
\delta(e_8, a) = \emptyset = e_7 & \delta(e_7, b) = \{b_8\} = e_{10} \\
\delta(e_9, a) = \{a_7\} = e_8 & \delta(e_9, b) = \{b_6\} = e_9 \\
\delta(e_{10}, a) = \emptyset = e_7 & \delta(e_{10}, b) = \{b_8\} = e_{10}
\end{array}$$

y

$$E = \{e_1, \ldots, e_{10}\}.$$

Paso 4. Se toma como E_1 al subconjunto de estados de E tales que en su definición contienen al menos una letra final de palabra de L_1. En nuestro caso,

$$E_1 = \{e_2, e_4, e_5, e_8, e_{10}\}.$$

Siguiendo estos mismos pasos, se prueba

Proposición 3.5. *Sea L un lenguaje regular. Entonces, existe $S = (S, E, \delta)$ y $E_1 \subseteq E$ tal que*

$$L = Ł(S, E_1).$$