Teoría de Lenguajes

Clase Teórica 3

Expresiones Regulares

Primer cuartimestre 2016

Material compilado por el Profesor Julio Jacobo, a lo largo de distintas ediciones de la materia Teoría de Lenguajes en el Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires.

Bibliografía: Capítulo 3, *Introduction to Automata Theory, Languages and Computation*, J. Hopcroft, R. Motwani, J. Ullman, Second Edition, Addison Wesley, 2001.

En esta teórica:

- Definición de expresión regular
- Teorema: Para cada expresion regular r hay un AFND-λ M con un solo estado final y sin transiciones de salida tal que $\mathcal{L}(r) = \mathcal{L}(M)$.
- Teorema: Para cada AFD M hay una expresion regular r tal que $\mathcal{L}(r) = \mathcal{L}(M)$.
- Teorema: Para cada gramatica regular G existe un AFND M tal que $\mathcal{L}(G) = \mathcal{L}(M)$.
- Teorema: Para cada AFD M existe una gramática regular G tal que $\mathcal{L}(G) = \mathcal{L}(M)$.

Definición
Dado un alfabeto Σ, una expresión regular denota un lenguaje sobre Σ:

- $\varnothing$ es una expresión regular que denota el conjunto vacío $\emptyset$.
- λ es una expresión regular que denota el conjunto $\{\lambda\}$.
- para cada $a \in \Sigma$, a es es una expresión regular que denota el conjunto $\{a\}$.
- si r y s denotan los lenguajes R y S entonces

$r \mid s$	denota	$R \cup S$,
rs	denota	RS,
r^*	denota	R^*, y
r^+	denota	R^+.

Notación
Lo anterior puede escribirse como : $\mathcal{L}(r) = R$, $\mathcal{L}(s) = S$, $\mathcal{L}(r \mid s) = R \cup S$, $\mathcal{L}(rs) = RS$, $\mathcal{L}(r^*) = R^*$ y $\mathcal{L}(r^+) = R^+$.

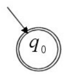
Ejemplo 1.

- 00
- $(0 \mid 1)^*$
- $(0 \mid 1)^* \, 00 \, (0 \mid 1)^*$
- $(1 \mid 10)^*$
- $(0 \mid \lambda) \, (1 \mid 10)^*$

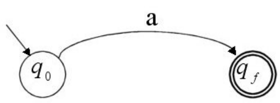
Teorema 2. *Dada una expresión regular r, existe un AFND-λ M con un solo estado final y sin transiciones a partir del mismo tal que $\mathcal{L}(M) = \mathcal{L}(r)$.*

Demostración. Caso base: $r = \emptyset$,

Caso base: $r = \lambda$,



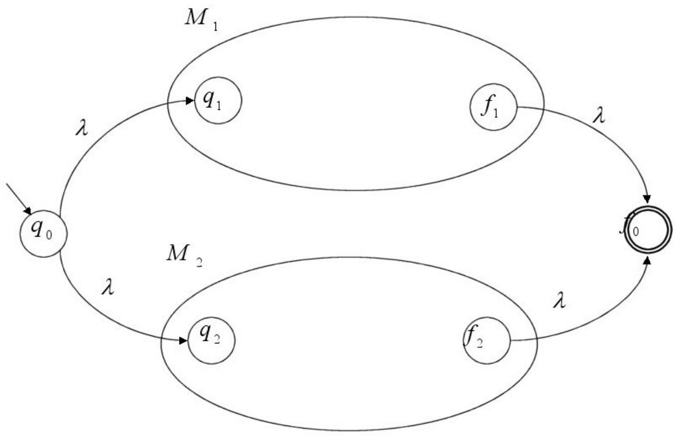
Caso base: $r = a$.



Caso inductivo: supongamos la expresión regular es $r_1|r_2, r_1r_2, r_1^*$, ó r_2^+ y asumimos que vale la propiedad para r_1 y para r_2.

Es decir, tanto para r_1 como para r_2 existen AFND-λ M_1 y M_2 con un solo estado final y sin transiciones a partir del mismo, tal que $\mathcal{L}(M_1) = \mathcal{L}(r_1)$ y $\mathcal{L}(M_2) = \mathcal{L}(r_2)$.

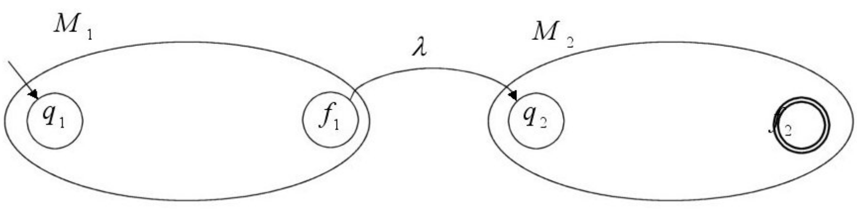
Caso $r = r_1 \mid r_2$. Por h.i. existen $M_1 = \langle Q_1, \Sigma_1, \delta_1, q_1, \{f_1\}\rangle$ y $M_2 = \langle Q_2, \Sigma_2, \delta_2, q_2, \{f_2\}\rangle$ tales que $\mathcal{L}(M_1) = \mathcal{L}(r_1)$ y $\mathcal{L}(M_2) = \mathcal{L}(r_2)$.Sea $M = \langle Q_1 \cup Q_1 \cup \{q_0, f_0\}, \Sigma_1 \cup \Sigma_1, \delta, q_0, \{f_0\}\rangle$



- $\delta(q_0, \lambda) = \{q_1, q_2\}$
- $\delta(q, a) = \delta_1(q, a)$ para $q \in Q_1 - \{f_1\}$ y $a \in \Sigma_1 \cup \{\lambda\}$
- $\delta(q, a) = \delta_2(q, a)$ para $q \in Q_2 - \{f_2\}$ y $a \in \Sigma_2 \cup \{\lambda\}$
- $\delta(f_1, \lambda) = \delta(f_2, \lambda) = \{f_0\}$.

Caso $r = r_1r_2$.

Por h.i. existen $M_1 = \langle Q_1, \Sigma_1, \delta_1, q_1, \{f_1\}\rangle$ y $M_2 = \langle Q_2, \Sigma_2, \delta_2, q_2, \{f_2\}\rangle$, tales que $\mathcal{L}(M_1) = \mathcal{L}(r_1)$ y $\mathcal{L}(M_2) = \mathcal{L}(r_2)$ respectivamente. Entonces podemos construir el autómata $M = \langle Q_1 \cup Q_2, \Sigma_1 \cup \Sigma_2, \delta, q_1, \{f_2\}\rangle$

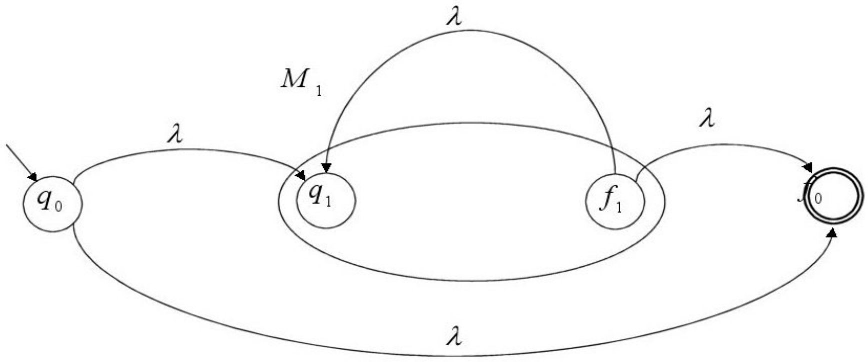


- $\delta(q, a) = \delta_1(q, a)$ para $q \in Q_1 - \{f_1\}$ y $a \in \Sigma_1 \cup \{\lambda\}$

- $\delta(f_1, \lambda) = \{q_2\}$
- $\delta(q, a) = \delta_2(q, a)$ para $q \in Q_2 - \{f_2\}$ y $a \in \Sigma_2 \cup \{\lambda\}$

Caso $r = r_1^*$.

Por h.i. existe $M_1 = \langle Q_1, \Sigma_1, \delta_1, q_1, \{f_1\}\rangle$, tal que $\mathcal{L}(M_1) = \mathcal{L}(r_1)$. Entonces podemos construir el autómata $M = \langle Q_1 \cup \{f_0, q_0\}, \Sigma_1, \delta, q_0, \{f_0\}\rangle$



- $\delta(q, a) = \delta_1(q, a)$ para $q \in Q_1 - \{f_1\}$ y $a \in \Sigma_1 \cup \{\lambda\}$
- $\delta(q_0, \lambda) = \delta(f_1, \lambda) = \{q_1, f_0\}$.

Caso $r = r_1^+$.

Dado que $r_1^+ = r_1 r_1^*$, queda demostrado por los casos anteriores.

Indicar Verdadero o Falso, justificar

1. $r|s = s|r$
2. $(r^*)^* = r^*$
3. $\emptyset^* = \lambda$
4. $(r|s)^* = r^*|s^*$
5. $(rs|r)^*r = r(sr|r)^*$

Teorema 3. *Dado un AFD $M = \langle \{q_1, \ldots, q_n\}, \Sigma, \delta, q_1, F\rangle$ existe una expresión regular r tal que $\mathcal{L}(M) = \mathcal{L}(r)$.*

Demostración. Denotemos con $R_{i,j}^k$ el conjunto de cadenas de Σ^* que llevan al autómata M desde el estado q_i al estado q_j pasando por estados cuyo índice es, a lo sumo, k. Definamos $R_{i,j}^k$ en forma recursiva:

$$R_{i,j}^k = R_{i,k}^{k-1}\left(R_{kk}^{k-1}\right)^* R_{k,j}^{k-1} \cup R_{i,j}^{k-1} \text{ para } k \geq 1$$

$$R_{i,j}^0 = \begin{cases} \{a : \delta(q_i, a) = q_j\} \ , a \in \Sigma & \text{si } i \neq j \\ \{a : \delta(q_i, a) = q_j\} \cup \{\lambda\} \ , a \in \Sigma & \text{si } i = j \end{cases}$$

Caso base: $k = 0$. Debemos dar r_{ij}^0, tal que $\mathcal{L}\left(r_{ij}^0\right) = R_{ij}^0$.

$R_{i,j}^0$ es el conjunto de cadenas de un solo caracter o λ.

Por lo tanto, $r_{i,j}^0$ es:

- $a_1 \mid \ldots \mid a_p$,con $a_1, \ldots, a_p$ símbolos de Σ, si $\delta(q_i, a_s) = q_j$ para $s = 1, \ldots, p$ y $q_i \neq q_j$.
- $a_1 \mid \ldots \mid a_p \mid \lambda$,con $a_1, \ldots, a_p$ símbolos de Σ, si $\delta(q_i, a_s) = q_j$ para $s = 1, \ldots, p$ y además $q_i = q_j$.
- $\varnothing$, si no existe ningún a_i que una q_i y q_j y $q_i \neq q_j$.
- λ, si no existe ningún a_i que una q_i y q_j y además $q_i = q_j$.

Caso inductivo. Por hipótesis inductiva, tenemos:
$\mathcal{L}(r_{ik}^{k-1}) = R_{ik}^{k-1}$, $\mathcal{L}(r_{kk}^{k-1}) = R_{kk}^{k-1}$, $\mathcal{L}(r_{kj}^{k-1}) = R_{kj}^{k-1}$ y$\mathcal{L}(r_{ij}^{k-1}) = R_{ij}^{k-1}$.
Definimos $r_{ij}^{k} = r_{ik}^{k-1}\left(r_{kk}^{k-1}\right)^* r_{kj}^{k-1} \mid r_{ij}^{k-1}$ y verificamos que

$$\begin{aligned}
\mathcal{L}\left(r_{ij}^{k}\right) &= \mathcal{L}\left(r_{ik}^{k-1}\left(r_{kk}^{k-1}\right)^* r_{kj}^{k-1} \mid r_{ij}^{k-1}\right) \\
&= \mathcal{L}\left(r_{ik}^{k-1}\left(r_{kk}^{k-1}\right)^* r_{kj}^{k-1}\right) \cup \mathcal{L}\left(r_{ij}^{k-1}\right) \\
&= \mathcal{L}\left(r_{ik}^{k-1}\right)\mathcal{L}\left(r_{kk}^{k-1}\right)^* \mathcal{L}\left(r_{kj}^{k-1}\right) \cup \mathcal{L}\left(r_{ij}^{k-1}\right) \\
&= R_{ik}^{k-1}\left(R_{kk}^{k-1}\right)^* R_{kj}^{k-1} \cup R_{ij}^{k-1} \\
&= R_{ij}^{k}.
\end{aligned}$$

Entonces, como $Q = \{q_1, \dots q_n\}$ y q_1 es el estado inicial de M,

$$\begin{aligned}
\mathcal{L}(M) &= R_{1j_1}^{n} \cup \ldots \cup R_{1j_m}^{n}, \text{ con } F = \{q_{j_1}, \ldots, q_{j_m}\} \\
&= \mathcal{L}(r_{1j_1}^{n}) \cup \ldots \cup \mathcal{L}(r_{1j_m}^{n}) \\
&= \mathcal{L}(r_{1j_1}^{n} \mid \ldots \mid r_{1j_m}^{n})
\end{aligned}$$

Concluimos $\mathcal{L}(M) = r_{1j_1}^{n} \mid \ldots \mid r_{1j_m}^{n}$.

Teorema 4. *Dada una gramática regular $G = \langle V_n, V_T, P, S\rangle$ existe un AFND $M = \langle Q, \Sigma, \delta, q_0, F\rangle$ tal que $\mathcal{L}(G) = \mathcal{L}(M)$.*

Demostración. Definamos M de la siguiente manera:

- $Q = V_N \cup \{q_f\}$, para mayor claridad, llamaremos q_A al estado correspondiente al no terminal A
- $\Sigma = V_T$
- $q_0 = q_S$
- $q_B \in \delta(q_A, a) \Leftrightarrow A \to aB \in P$
- $q_f \in \delta(q_A, a) \Leftrightarrow A \to a \in P$
- $q_A \in F \Leftrightarrow A \to \lambda \in P$
- $q_f \in F$.

Asumamos Lema: Para todo $w \in V_T^*$, si $A \stackrel{*}{\Rightarrow} wB$ entonces $q_B \in \delta(q_A, w)$.

$$\begin{aligned}
wa \in \mathcal{L}(G) \Leftrightarrow & S \stackrel{*}{\Rightarrow} wa \\
\Leftrightarrow & \left(\exists A \in V_N, S \stackrel{*}{\Rightarrow} wA \wedge A \to a \in P\right) \vee \\
& \left(\exists B \in V_N, S \stackrel{*}{\Rightarrow} waB \wedge B \to \lambda \in P\right) \text{ (únicas dos formas)} \\
\Leftrightarrow & (\exists q_A \in Q, q_A \in \delta(q_S, w) \wedge q_f \in \delta(q_A, a)) \vee \\
& (\exists q_B \in Q, q_B \in \delta(q_S, wa) \wedge q_B \in F) \text{ (por el Lema)} \\
\Leftrightarrow & q_f \in \delta(q_S, wa) \vee (\exists q_B \in Q, q_B \in \delta(q_S, wa) \wedge q_B \in F) \\
\Leftrightarrow & wa \in \mathcal{L}(M).
\end{aligned}$$

$$\begin{aligned}
\lambda \in \mathcal{L}(G) \Leftrightarrow & S \stackrel{*}{\Rightarrow} \lambda \\
\Leftrightarrow & S \to \lambda \in P \\
\Leftrightarrow & q_S \in F \\
\Leftrightarrow & \lambda \in \mathcal{L}(M).
\end{aligned}$$

Lema: Para todo $w \in V_T^*$, Si $A \stackrel{*}{\Rightarrow} wB$ entonces $q_B \in \delta(q_A, w)$.

Demostración. Por inducción en la longitud de w.

Caso base $|w| = 0$, es decir $w = \lambda$. Como $A \overset{*}{\Rightarrow} A$ y $q_A \in \delta(q_A, \lambda)$,

$$A \overset{*}{\Rightarrow} A \Leftrightarrow q_A \in \delta(q_A, \lambda).$$

Caso $|w| = n + 1, n \geq 0$, es decir, $w = \alpha a$ con $\alpha = n$. Asumamos h.i. para longitud n, es decir, vale para α.

$$\begin{aligned}
A \overset{*}{\Rightarrow} \alpha a B \quad & \Leftrightarrow \exists C \in V_N, A \overset{*}{\Rightarrow} \alpha C \wedge C \to aB \in P \\
& \Leftrightarrow \ \exists q_C \in Q, q_C \in \delta(q_A, \alpha) \wedge q_B \in \delta(q_C, a) \ \text{ por h.i.} \\
& \Leftrightarrow q_B \in \delta(\delta(q_A, \alpha), a) \\
& \Leftrightarrow q_B \in \delta(q_A, \alpha a)
\end{aligned}$$

Teorema 5. *Dado un AFD $M = \langle Q, \Sigma, \delta, q_0, F\rangle$ existe una gramática regular $G = \langle V_n, V_T, P, S\rangle$ tal que $\mathcal{L}(G) = \mathcal{L}(M)$.*

Demostración. Debemos definir gramática G.
$V_N = Q$, (llamaremos A_p al no terminal correspondiente a $p \in Q$).

$$\begin{aligned}
V_T &= \Sigma \\
S &= A_{q_0} \\
A_p \to aA_q \in P \Leftrightarrow &\ \delta(p, a) = q \\
A_p \to a \in P \Leftrightarrow &\ \delta(p, a) = q \in F \\
S \to \lambda \in P \Leftrightarrow &\ q_0 \in F
\end{aligned}$$

Asumamos Lema: $\delta(p, w) = q$ si y solo si $A_p \overset{*}{\Rightarrow} wA_q$.

$$\begin{aligned}
wa \in \mathcal{L}(M) &\Leftrightarrow \delta(q_0, wa) \in F \\
&\Leftrightarrow \exists p \in Q, \delta(q_0, w) = p \wedge \delta(p, a) \in F \\
&\Leftrightarrow \exists A_p, A_{q_0} \overset{*}{\Rightarrow} wA_p \wedge A_p \to a \in P \\
&\Leftrightarrow A_{q_0} \overset{*}{\Rightarrow} wa \\
&\Leftrightarrow wa \in \mathcal{L}(G) \\
\lambda \in \mathcal{L}(M) &\Leftrightarrow \ q_0 \in F \\
&\Leftrightarrow S \to \lambda \in P \\
&\Leftrightarrow S \overset{*}{\Rightarrow} \lambda \\
&\Leftrightarrow \lambda \in \mathcal{L}(G).
\end{aligned}$$

Lema: $\delta(p, w) = q$ si y solo si $A_p \overset{*}{\Rightarrow} wA_q$.

Demostración. Por inducción en la longitud de w.
Para $w = \lambda$, es cierto que $\delta(p, \lambda) = p$ y que $A_p \overset{*}{\Rightarrow} A_p$, por lo tanto $\delta(p, \lambda) = p \Leftrightarrow A_p \overset{*}{\Rightarrow} A_p$,
Asumamos h.i. vale para α de longitud n, con $n \geq 0$, y veamos que vale para para $w = \alpha a$.

$$\begin{aligned}
\delta(p, \alpha a) = q &\Leftrightarrow \exists r \in Q, \delta(p, \alpha) = r \wedge \delta(r, a) = q \\
&\Leftrightarrow \exists A_r, A_p \overset{*}{\Rightarrow} \alpha A_r \wedge A_r \to aA_q \in P \text{ por h.i.} \\
&\Leftrightarrow A_p \overset{*}{\Rightarrow} \alpha a A_q.
\end{aligned}$$