Teoría de Lenguajes

Clase Teórica 4

Expresiones Regulares y gramáticas regulares

Primer Cuatrimestre 2024

Bibliografía

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

En esta clase

- 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 $\mathcal{L}(r) = \mathcal{L}(M)$.

Sea L_1 un lenguaje definido sobre el alfabeto Σ_1, y sea L_2 un lenguaje definido sobre el alfabeto Σ_2. Definimos la concatenacion de L_1 y L_2 como el lenguaje

$$L_1 L_2 = \{xy : x \in L_1 \wedge y \in L_2\},$$

definido sobre el alfabeto $\Sigma_1 \cup \Sigma_2$.

La clausura de Kleene de un lenguaje $L \subseteq \Sigma^*$ es el lenguaje L^* que se define así:

- $L^0 = \{\lambda\}$
- $L^n = LL^{n-1}$ para $n \geq 1$
- $L^* = \bigcup_{n \geq 0} L^n$.

La clausura positiva de un lenguaje $L \subseteq \Sigma^*$ es el lenguaje L^+ definido por :

$$L^+ = \bigcup_{n \geq 1} L^n.$$

Por lo tanto $L^+ = LL^* = L^*L$, y que $L^* = L^+ \cup \{\lambda\}$.

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

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

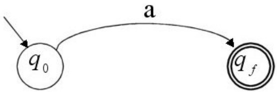
Teorema

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: $r = \emptyset$,

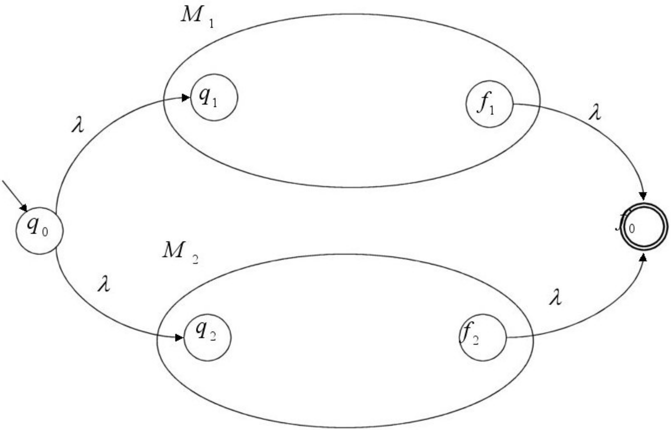
Caso base: $r = \lambda$,

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 $\mathcal{L}(M_2) = \mathcal{L}(r_2)$.

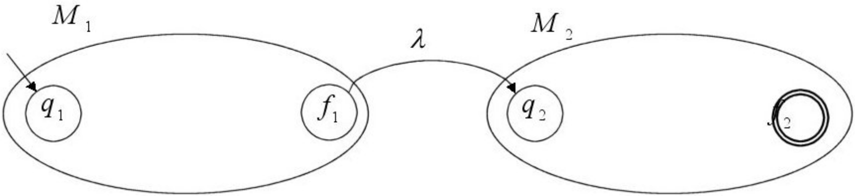
Caso $r = r_1 \mid r_2$. Por h.i. existen $M_1 = \langle Q_1, \Sigma_1, \delta_1, q_1, \{f_1\} \rangle$ y $M_2 = \langle Q_2, \Sigma_2, \delta_2, q_2, \{f_2\} \rangle$ tales que $\mathcal{L}(M_1) = \mathcal{L}(r_1)$ y $\mathcal{L}(M_2) = \mathcal{L}(r_2)$. Sea $M = \langle Q_1 \cup Q_1 \cup \{q_0, f_0\},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



- $\delta(q_0, \lambda) = \{q_1, q_2\}$
- $\delta(q, a) = \delta_1(q, a)$ 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 $r = r_1 r_2$.

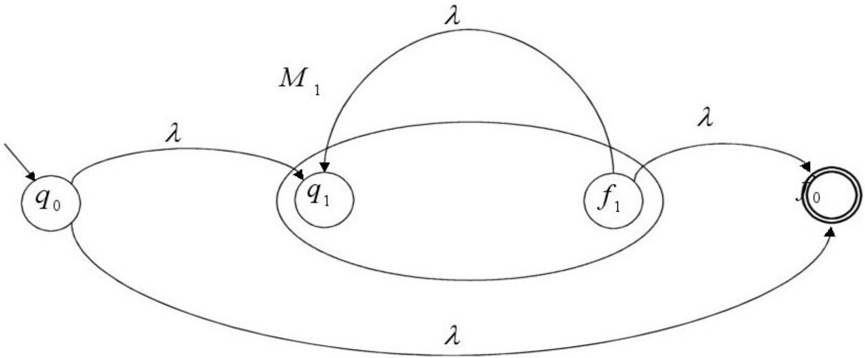
Por h.i. existen $M_1 = \langle Q_1, \Sigma_1, \delta_1, q_1, \{f_1\}\rangle$ y $M_2 = \langle Q_2, \Sigma_2, \delta_2, q_2, \{f_2\}\rangle$, tales que $\mathcal{L}(M_1) = \mathcal{L}(r_1)$ y $\mathcal{L}(M_2) = \mathcal{L}(r_2)$ respectivamente.
Entonces podemos construir el autómata
$M = \langle Q_1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



- $\delta(q, a) = \delta_1(q, a)$ 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

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 $Q = \{q_1, \ldots q_n\}$ 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 $\mathcal{L}(M) = r_{1j_1}^{n} \mid \ldots \mid r_{1j_m}^{n}$.