Definition (Definición Expresión regular)

- $\varnothing$ 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 $r \mid s$, rs, r^* y r^+ son expresiones regulares que denotan los conjuntos $R \cup S$, RS, R^* y R^+ respectivamente.

Notación

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

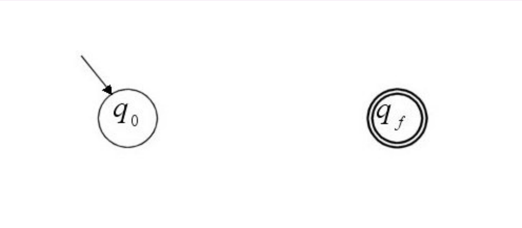
Example (Ejemplos)

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

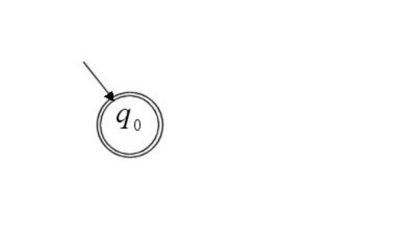
Theorem (Teorema)

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

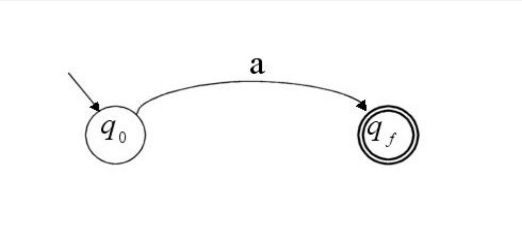
- caso base:
  - $r = \emptyset$,



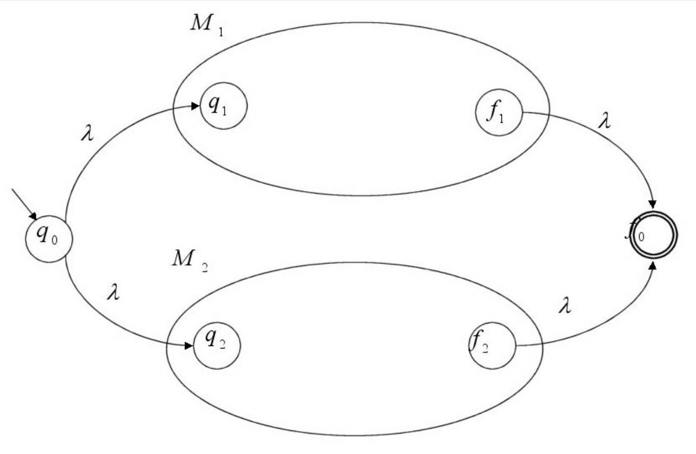
- caso base:
  - $r = \lambda$,



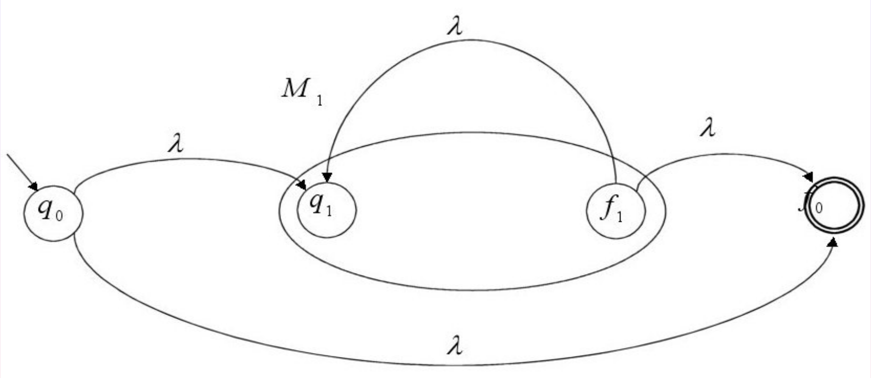
- caso base:
  - $r = a$.



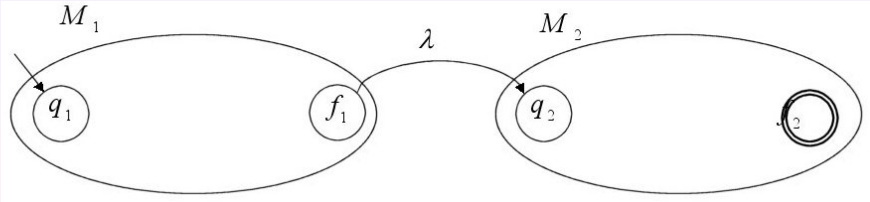
- inducción en la cantidad de operadores:
  - **Caso** $r = r_1 \mid r_2$**:** por h.i. existen $M_1 = \langle Q_1, \Sigma_1, \delta_1, q_1, \{f_1\} \rangle$ y $M_2 = \langle Q_2, \Sigma_2, \delta_2, q_2, \{f_2\} \rangle$, tales que $\mathcal{L}(M_1) = \mathcal{L}(r_1)$ y $\mathcal{L}(M_2) = \mathcal{L}(r_2)$ respectivamente. Entonces podemos construir el autómata $M = \langle Q_1 \cup Q_1 \cup \{q_0, f_0\}, \Sigma_1 \cup \Sigma_1, \delta, q_0, \{f_0\} \rangle$ con δ dada por:
    - $\delta(q_0, \lambda) = \{q_1, q_2\}$
    - $\delta(q, a) = \delta_1(q, a)$ 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 $a \in \Sigma_2 \cup \{\lambda\}$
    - $\delta(f_1, \lambda) = \delta(f_2, \lambda) = \{f_0\}$


M_1
q_1
f_1
λ
λ
q_0
f_0
M_2
λ
λ
q_2
f_2

- inducción en la cantidad de operadores:
  - **Caso** $r = r_1^*$**:** por h.i. existen $M_1 = \langle Q_1, \Sigma_1, \delta_1, q_1, \{f_1\}\rangle$, tal que $\mathcal{L}(M_1) = \mathcal{L}(r_1)$ Entonces podemos construir el autómata $M = \langle Q_1 \cup \{f_0, q_0\}, \Sigma_1, \delta, q_0, \{f_0\}\rangle$ con δ dada por:
  - $\delta(q, a) = \delta_1(q, a)$ para $q \in Q_1 - \{f_1\}$ y $a \in \Sigma_1 \cup \{\lambda\}$
  - $\delta(q_0, \lambda) = \delta(f_1, \lambda) = \{q_1, f_0\}$


λ
M_1
λ
λ
q_0
q_1
f_1
f_0
λ

- inducción en la cantidad de operadores:
  - **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
    $M = \langle Q_1 \cup Q_1, \Sigma_1 \cup \Sigma_1, \delta, q_1, \{f_2\}\rangle$ con δ dada por:
    - $\delta(q, a) = \delta_1(q, a)$ 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 $a \in \Sigma_2 \cup \{\lambda\}$



M_1
M_2
λ
q_1
f_1
q_2
f_2

Theorem (Teorema)

Dado un AFD $M = \langle \{q_1, \ldots, q_n\}, \Sigma, \delta, q_1, F \rangle$ *que acepta el lenguaje L,existe una expresión regular que denota el mismo lenguaje.*

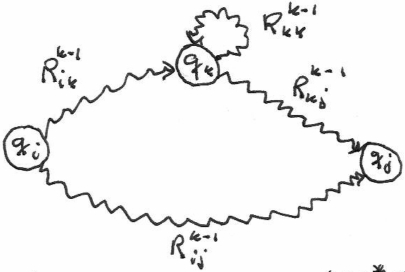
Demostración.

Denotemos con $R^k_{i,j}$ 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. Entonces, $R^n_{i,j}$ denota todos los caminos entre q_i y q_j.
Definamos $R^k_{i,j}$ en forma recursiva:

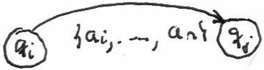
$$R^k_{i,j} = R^{k-1}_{i,k} \left(R^{k-1}_{kk} \right)^* R^{k-1}_{k,j} \cup R^{k-1}_{i,j} \text{ para } k \geq 1$$

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

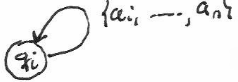
$\square$



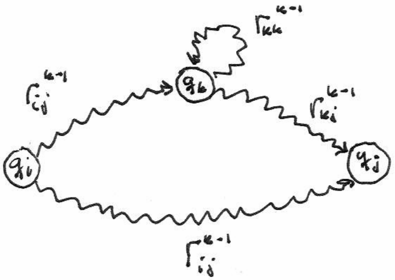
$$A_{ij}^{k} = R_{ij}^{k-1} \cup R_{ik}^{k-1} (R_{kk}^{k-1})^* R_{kj}^{k-1}$$



$i \neq j$



$i = j$



h.i.:

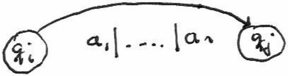
$\mathcal{L}(R_{ij}^{k-1}) = r_{ij}^{k-1}$

$\mathcal{L}(R_{ik}^{k-1}) = r_{ik}^{k-1}$

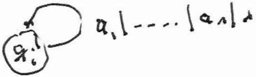
$\mathcal{L}(R_{kk}^{k-1}) = r_{kk}^{k-1}$

$\mathcal{L}(R_{kj}^{k-1}) = r_{kj}^{k-1}$

$$\mathcal{L}(R_{ij}^{k}) = r_{ij}^{k} = r_{ik}^{k-1} (r_{kk}^{k-1})^* r_{kj}^{k-1} \mid r_{ij}^{k-1}$$



$i \neq j$



$i = j$

Demostración (cont.)

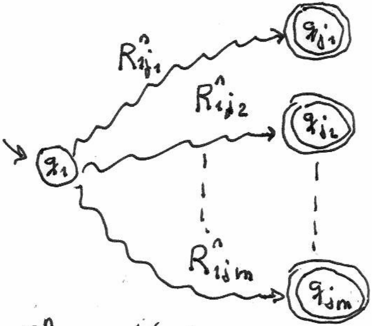
- Base: $k = 0$
  $R^0_{i,j}$ es el conjunto de cadenas de un solo caracter o λ. Por lo tanto, $r^0_{i,j}$ será igual a:
  - $a_1 \mid \ldots \mid a_p$,con $a_1, \ldots, a_p$ símbolos de Σ, si $\delta\left(q_{i,} a_s\right) = q_j$ para $s = 1, \ldots, p$ y $q_i \neq q_j$.
  - $a_1 \mid \ldots \mid a_p \mid \lambda$,con $a_1, \ldots, a_p$ símbolos de Σ, si idem anterior pero 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 idem anterior pero además $q_i = q_j$.

$\square$

Demostración (cont.)

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

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



$$R^n_{1j_i} = \mathcal{L}(r^n_{1j_i})$$
$$i = 1, \dots, m$$

$$F = \{q_{j_1}, \dots, q_{j_m}\}$$
$$|F| = m$$

$$\mathcal{L}(M) = \underbrace{\mathcal{L}(r^n_{1j_1})}_{R^n_{1j_1}} \cup \dots \cup \underbrace{\mathcal{L}(r^n_{1j_m})}_{R^n_{1j_m}}$$

$$= \mathcal{L}(\underbrace{r^n_{1j_1} | \dots | r^n_{1j_m}}_{\text{expresión regular del lenguaje aceptado por } M})$$

Demostración (cont.)

Por otro lado

$$\mathcal{L}(M) = \bigcup_{q_j \in F} R^n_{1j} = R^n_{1j_1} \cup \ldots \cup R^n_{1j_m},$$

con $F = \{q_{j_1}, \ldots, q_{j_m}\}$. De lo anterior, $\mathcal{L}\left(r^n_{1j_i}\right) = R^n_{1j_i}$ para $i = 1, \ldots, m$, por lo tanto,

$$\begin{aligned} \mathcal{L}(M) &= \mathcal{L}\left(r^n_{1j_1}\right) \cup \ldots \cup \mathcal{L}\left(r^n_{1j_m}\right) \\ &= \mathcal{L}\left(r^n_{1j_1} \mid \ldots \mid r^n_{1j_m}\right), \end{aligned}$$

entonces, el lenguaje $\mathcal{L}(M)$ es denotado por la expresión regular $r^n_{1j_1} \mid \ldots \mid r^n_{1j_m}$. □