Tema 5: Expressions Regulars

U.D. Computació

DSIC - UPV

Índex

Definicions

- Inductivament, una expressió regular sobre Σ es defineix com:
  - $\emptyset$ denota el llenguatge buit
  - λ denota el llenguatge $\{\lambda\}$
  - $\forall a \in \Sigma$, a denota el llenguatge $\{a\}$
  - Si r i s són expressions regulars que denoten L_r i L_s:
    - (r) denota el llenguatge L_r
    - $r + s$ denota el llenguatge $L_r \cup L_s$
    - rs denota el llenguatge $L_r L_s$
    - $(r)^*$ denota el llenguatge L_r^*
  - Només són expressions regulars les construïdes d'aquesta forma

Propietats

- Siguen α, β y γ expressiones regulars
  1. $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$
  2. $\alpha(\beta\gamma) = (\alpha\beta)\gamma$
  3. $\alpha + \beta = \beta + \alpha$
  4. $\alpha(\beta + \gamma) = (\alpha\beta) + (\alpha\gamma)$
  5. $(\alpha + \beta)\gamma = (\alpha\gamma) + (\beta\gamma)$
  6. $\alpha\lambda = \lambda\alpha = \alpha$
  7. $\alpha + \emptyset = \emptyset + \alpha = \alpha$
  8. $\alpha\emptyset = \emptyset\alpha = \emptyset$
  9. $\lambda^* = \lambda$
  10. $\emptyset^* = \lambda$
  11. $\alpha^* = \lambda + \alpha\alpha^*$
  12. $(\alpha^* + \beta^*)^* = (\alpha^*\beta^*)^* = (\alpha + \beta)^*$
  13. $(\alpha\beta)^*\alpha = \alpha(\beta\alpha)^*$
  14. $(\alpha^*\beta)^*\alpha^* = (\alpha + \beta)^*$
  15. $(\alpha^*\beta)^* = (\alpha + \beta)^*\beta + \lambda$

Construccions

- Homomorfisme
- Revers

Construccions

- Homomorfisme
  Donada una expressió regular α i un homomorfisme $h : \Sigma_\alpha \to \Delta^*$, per a obtenir una expressió regular per al llenguatge $h(L(\alpha))$, només cal substituir cada símbol a de α per $h(a)$

  Per exemple, considerant $\alpha = a(bb^* + (aa)^*)^* b$ i l'homomorfisme: $h(a) = 0$ i $h(b) = 11$, l'expressió regular per a $h(L(\alpha))$ seria:

$$0(11(11)^* + (00)^*)^* 11$$

Construccions

- Revers
  Donada una expressió regular α, per a obtenir una expressió regular α^r tal que $L(\alpha^r) = (L(\alpha))^r$, apliquem recursivament las regles següents:
  - Si $\alpha = \emptyset$, $\alpha = \lambda$ o $\alpha = a \in \Sigma$, aleshores $\alpha^r = \alpha$
  - Si $\alpha = \beta + \gamma$, aleshores $\alpha^r = \beta^r + \gamma^r$
  - Si $\alpha = \beta\gamma$, aleshores $\alpha^r = \gamma^r\beta^r$
  - Si $\alpha = \beta^*$, aleshores $\alpha^r = (\beta^r)^*$

  Per exemple, considerant $\alpha = a(b(a+b)^* + (bba)^*)^* b$, l'expressió regular per a $(L(\alpha))^r$ seria:

$$\alpha^r = b((a+b)^* b + (abb)^*)^* a$$

Síntesi d'AFs a partir d'ERs

- Algorisme de Brzozowski
- Autòmat de Posició
- Autòmat Follow

Càlcul de Derivades

- Regles per al càlcul de les derivades
  - Respecte de símbols ($a, b \in \Sigma, \quad r, s$ E.R.)
    1. $a^{-1}\emptyset = \emptyset$
    2. $a^{-1}\lambda = \emptyset$
    3. $a^{-1}b = \begin{cases} \emptyset & \text{si } a \neq b \\ \lambda & \text{si } a = b \end{cases}$
    4. $a^{-1}(r + s) = a^{-1}r + a^{-1}s$
    5. $a^{-1}(rs) = \begin{cases} (a^{-1}r)s & \text{si } \lambda \notin r \\ (a^{-1}r)s + a^{-1}s & \text{si } \lambda \in r \end{cases}$
    6. $a^{-1}r^* = (a^{-1}r)r^*$
  - Respecte de cadenes ($a \in \Sigma, \quad x \in \Sigma^*$)
    1. $\lambda^{-1}r = r$
    2. $(xa)^{-1}r = a^{-1}(x^{-1}r)$

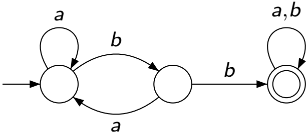
Algorisme de Brzozowski

Entrada: α expressió regular sobre Σ
Eixida: AFD mínim per a $L(\alpha)$
Mètode:

$Q = \{\alpha\}$; $q_0 = \alpha$; $F = \emptyset$; $\delta = \emptyset$;
if $\lambda \in L(\alpha)$ **then**
  $F = F \cup \{\alpha\}$
end if
$actius = \{\alpha\}$
while $actius \neq \{\}$ **do**
  $\beta = First(actius)$
  $actius = Rest(actius)$
  for all $a \in \Sigma$ **do**
    $\beta' = a^{-1}\beta$
    if $\nexists r \in Q \ : \ L(r) = L(\beta')$ **then**
      $Q = Q \cup \{\beta'\}$
      $\delta = \delta \cup \{(\beta, a, \beta')\}$
      $actius = actius \cup \{\beta'\}$
      if $\lambda \in L(\beta')$ **then**
        $F = F \cup \{\beta'\}$
      end if
    end if
  end for
end while
Return $(Q, \Sigma, \delta, q_0, F)$
Fi Mètode

Algorisme de Brzozowski. Exemple

- Considerem $\alpha = (a+b)^*bb(a+b)^*$:
  - $q_0 = \alpha = (a+b)^*bb(a+b)^*$; $\lambda \notin L(q_0)$ per tant $F = \emptyset$
  - $a^{-1}q_0 = q_0$
    $b^{-1}q_0 = (a+b)^*bb(a+b)^* + b(a+b)^* = q_1$; $\lambda \notin L(q_1)$ per tant $F = \emptyset$.
  - $a^{-1}q_1 = q_0$
    $b^{-1}q_1 = (a+b)^*bb(a+b)^* + b(a+b)^* + (a+b)^* = (a+b)^* = q_2$; $\lambda \in L(q_2)$ per tant $F = \{q_2\}$.
  - $a^{-1}q_2 = b^{-1}q_2 = q_2$



Autòmat de posició

- Autòmat local. Llenguatge Local
- Expressió regular linealitzada
- AFD per a una expressió regular linealitzada
- Autòmat de posició

Autòmat local. Llenguatge Local

- El AFD $A = (Q, \Sigma, \delta, q_0, F)$ és *local* si i només si per a qualsevol $a \in \Sigma$ el conjunt $\{\delta(q, a) \ : \ q \in Q\}$ conté com a molt un element
- Si a més a més no existeix cap arc que arribe a q_0, l'autòmat és *local estàndard*
- Un llenguatge és local si i només si és reconegut per un autòmat local estàndard

Expressió regular linealitzada

- Siga α una expressió regular i siga n el nombre de símbols en α sense comptar parèntesi ni símbols d'operació. L'expressió linealitzada de α (denotada per $\overline{\alpha}$) s'obté afegint un subíndex $j \in \{1, \ldots, n\}$ a cada símbol de α indicant la seua posició.
  p.e.: Siga

$$\alpha = (a + b)(a^* + ba^* + b^*)^*$$

la versió linealitzada és

$$\overline{\alpha} = (a_1 + b_2)(a_3^* + b_4 a_5^* + b_6^*)^*$$

Expressió regular linealitzada

- Si Σ_α i $\Sigma_{\overline{\alpha}}$ són els alfabets de α i $\overline{\alpha}$ respectivament, i $h : \Sigma_{\overline{\alpha}}^* \to \Sigma_\alpha^*$ és un homomorfisme que borra els subíndex, aleshores:

$$h(L(\overline{\alpha})) = L(\alpha)$$

- Per tant, es pot obtenir un autòmat finit per a $L(\alpha)$ construint un autòmat per a $L(\overline{\alpha})$ i posteriorment eliminant els subíndex d'aquest autòmat (*autòmat de posició*)

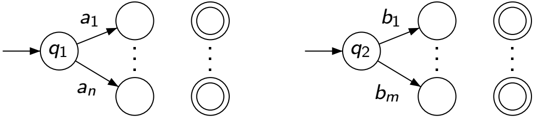
AFD per a una expressió regular linealitzada

- Tota expressió regular linealitzada denota un llenguatge local (reconegut per un AF local estàndard)
  Es pot vore per inducció sobre l'estructura de les expressions regulars.
- Casos base:

$\emptyset$ λ a

AFD per a una expressió regular linealitzada

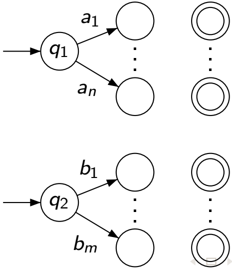
- Expressions compostes:
  Siguen $\overline{\alpha}$ i $\overline{\beta}$ expressions regulars linealitzadas, i siguen $A(\overline{\alpha}) = (Q_1, \Sigma_1, \delta_1, q_1, F_1)$ i $A(\overline{\beta}) = (Q_2, \Sigma_2, \delta_2, q_2, F_2)$, amb $\Sigma_1 \cap \Sigma_2 = \emptyset$, autòmats locals que accepten $L(\overline{\alpha})$ i $L(\overline{\beta})$ respectivament:



$(\overline{\alpha})$ $(\overline{\beta})$

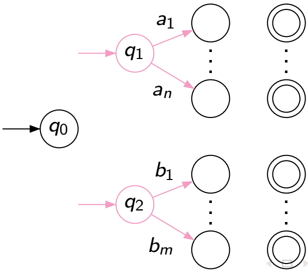
AFD per a una expressió regular linealitzada

- Unió $(\overline{\alpha} + \overline{\beta})$:
  - $Q = (Q_1 - \{q_1\}) \cup (Q_2 - \{q_2\}) \cup \{q_0\}$, $q_0 \notin Q_1 \cup Q_2$.
  - $\delta = \{(q, a, q') \in \delta_1 \cup \delta_2 : q \notin \{q_1, q_2\}\} \cup \{(q_0, a, q) : (q_1, a, q) \in \delta_1 \vee (q_2, a, q) \in \delta_2\}$,
  - $F = \begin{cases} F_1 \cup F_2 & \text{si } q_1 \notin F_1 \wedge q_2 \notin F_2 \\ (F_1 - \{q_1\}) \cup (F_2 - \{q_2\}) \cup \{q_0\} & \text{en qualsevol altre cas.} \end{cases}$



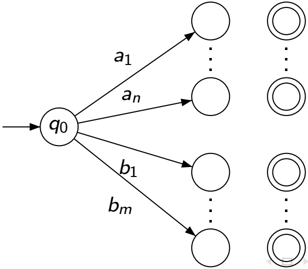
AFD per a una expressió regular linealitzada

- Unió $(\overline{\alpha} + \overline{\beta})$:
  - $Q = (Q_1 - \{q_1\}) \cup (Q_2 - \{q_2\}) \cup \{q_0\}$, $q_0 \notin Q_1 \cup Q_2$.
  - $\delta = \{(q, a, q') \in \delta_1 \cup \delta_2 : q \notin \{q_1, q_2\}\} \cup \{(q_0, a, q) : (q_1, a, q) \in \delta_1 \vee (q_2, a, q) \in \delta_2\}$,
  - $F = \begin{cases} F_1 \cup F_2 & \text{si } q_1 \notin F_1 \wedge q_2 \notin F_2 \\ (F_1 - \{q_1\}) \cup (F_2 - \{q_2\}) \cup \{q_0\} & \text{en qualsevol altre cas.} \end{cases}$



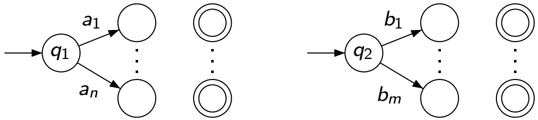
AFD per a una expressió regular linealitzada

- Unió $(\overline{\alpha} + \overline{\beta})$:
  - $Q = (Q_1 - \{q_1\}) \cup (Q_2 - \{q_2\}) \cup \{q_0\}$, $q_0 \notin Q_1 \cup Q_2$.
  - $\delta = \{(q, a, q') \in \delta_1 \cup \delta_2 : q \notin \{q_1, q_2\}\} \cup \{(q_0, a, q) : (q_1, a, q) \in \delta_1 \vee (q_2, a, q) \in \delta_2\}$,
  - $F = \begin{cases} F_1 \cup F_2 & \text{si } q_1 \notin F_1 \wedge q_2 \notin F_2 \\ (F_1 - \{q_1\}) \cup (F_2 - \{q_2\}) \cup \{q_0\} & \text{en qualsevol altre cas.} \end{cases}$



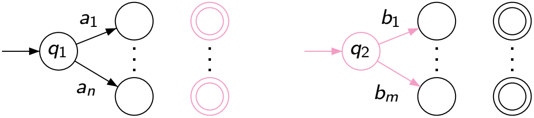
AFD per a una expressió regular linealitzada

- Producte $(\overline{\alpha} \cdot \overline{\beta})$ $(q_2 \notin F_2)$:
  - $Q = (Q_1 \cup Q_2) - \{q_2\})$,
  - $\delta = \delta_1 \cup \{(q, a, q') \in \delta_2 : q \neq q_2\} \cup \{(q, a, q') : q \in F_1 \wedge (q_2, a, q') \in \delta_2\}$,
  - $q_0 = q_1$
  - $F = F_2$



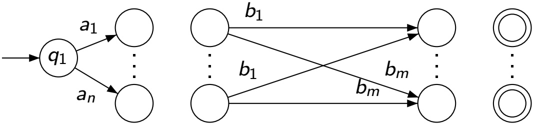
AFD per a una expressió regular linealitzada

- Producte $(\overline{\alpha} \cdot \overline{\beta})$ $(q_2 \notin F_2)$:
  - $Q = (Q_1 \cup Q_2) - \{q_2\})$,
  - $\delta = \delta_1 \cup \{(q, a, q') \in \delta_2 : q \neq q_2\} \cup \{(q, a, q') : q \in F_1 \wedge (q_2, a, q') \in \delta_2\}$,
  - $q_0 = q_1$
  - $F = F_2$



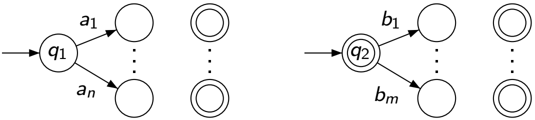
AFD per a una expressió regular linealitzada

- Producte $(\overline{\alpha} \cdot \overline{\beta})$ $(q_2 \notin F_2)$:
  - $Q = (Q_1 \cup Q_2) - \{q_2\})$,
  - $\delta = \delta_1 \cup \{(q, a, q') \in \delta_2 : q \neq q_2\} \cup \{(q, a, q') : q \in F_1 \wedge (q_2, a, q') \in \delta_2\}$,
  - $q_0 = q_1$
  - $F = F_2$



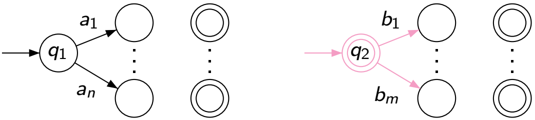
AFD per a una expressió regular linealitzada

- Producte $(\overline{\alpha} \cdot \overline{\beta})$ $(q_2 \in F_2)$:
  - $Q = (Q_1 \cup Q_2) - \{q_2\}$),
  - $\delta = \delta_1 \cup \{(q, a, q') \in \delta_2 : q \neq q_2\} \cup \{(q, a, q') : q \in F_1 \wedge (q_2, a, q') \in \delta_2\}$,
  - $q_0 = q_1$
  - $F = F_1 \cup (F_2 - \{q_2\})$



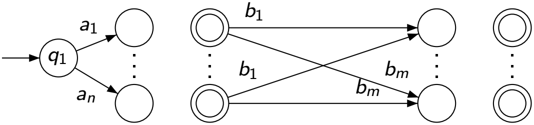
AFD per a una expressió regular linealitzada

- Producte $(\overline{\alpha} \cdot \overline{\beta})$ $(q_2 \in F_2)$:
  - $Q = (Q_1 \cup Q_2) - \{q_2\}$,
  - $\delta = \delta_1 \cup \{(q, a, q') \in \delta_2 : q \neq q_2\} \cup \{(q, a, q') : q \in F_1 \wedge (q_2, a, q') \in \delta_2\}$,
  - $q_0 = q_1$
  - $F = F_1 \cup (F_2 - \{q_2\})$



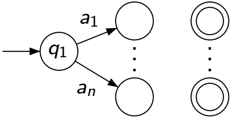
AFD per a una expressió regular linealitzada

- Producte $(\overline{\alpha} \cdot \overline{\beta})$ $(q_2 \in F_2)$:
  - $Q = (Q_1 \cup Q_2) - \{q_2\})$,
  - $\delta = \delta_1 \cup \{(q, a, q') \in \delta_2 : q \neq q_2\} \cup \{(q, a, q') : q \in F_1 \wedge (q_2, a, q') \in \delta_2\}$,
  - $q_0 = q_1$
  - $F = F_1 \cup (F_2 - \{q_2\})$



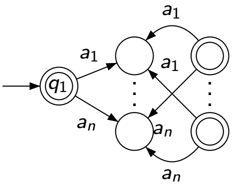
AFD per a una expressió regular linealitzada

- Clausura ($\overline{\alpha}^*$):
  - $\delta' = \delta \cup \{(q, a, q') : q \in F \wedge (q_0, a, q') \in \delta\}$
  - $F = F_1 \cup \{q_1\})$



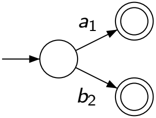
AFD per a una expressió regular linealitzada

- Clausura ($\overline{\alpha}^*$):
  - $\delta' = \delta \cup \{(q, a, q') : q \in F \land (q_0, a, q') \in \delta\}$
  - $F = F_1 \cup \{q_1\})$

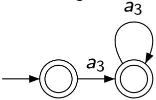


AFD per a una expressió regular linealitzada. Exemple

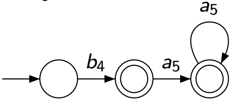
Siga $\alpha = (a + b)(a^* + ba^* + b^*)^*$.
Aleshores $\overline{\alpha} = (a_1 + b_2)(a_3^* + b_4 a_5^* + b_6^*)^*$



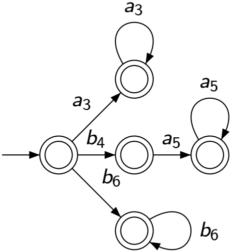
$a_1 + b_2$



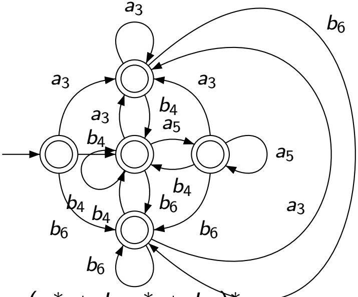
a_3^*



$b_4 a_5^*$



$a_3^* + b_4 a_5^* + b_6$



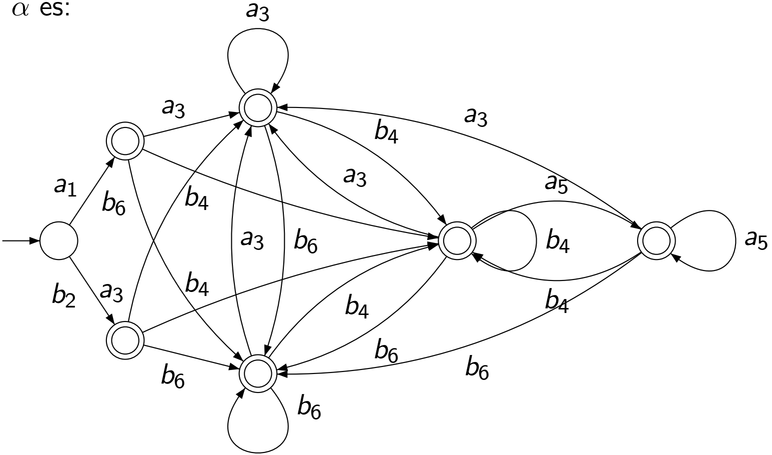
$(a_3^* + b_4 a_5^* + b_6)^*$

Autòmat de posició. Algorisme

Entrada: α expressió regular sobre Σ
Eixida: AF per a $L(\alpha)$
Mètode:
Obtenir $\overline{\alpha}$ versión linearitzada de α
Obtenir A un Autòmat local estàndard per a $\overline{\alpha}$
$A_{pos} = h(A)$, on h és un homomorfisme de borrat dels subíndex.
Return A_{pos}
Fi Método

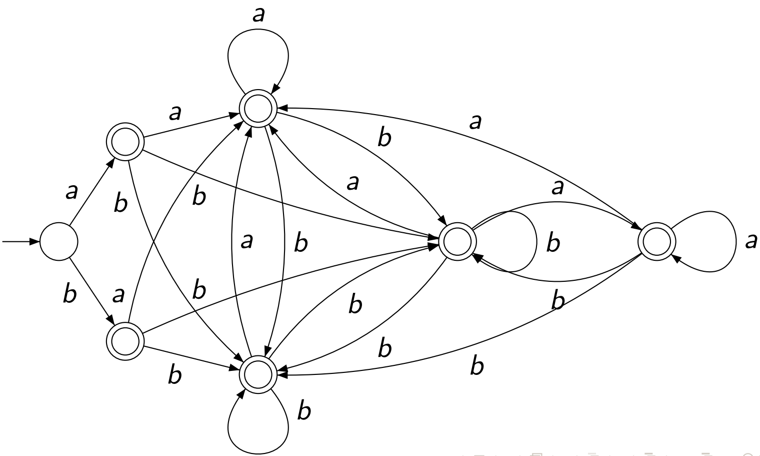
Autòmat de posició. Exemple

Siga $\alpha = (a + b)(a^* + ba^* + b^*)^*$ i la seua versió linealitzada $\overline{\alpha} = (a_1 + b_2)(a_3^* + b_4 a_5^* + b_6^*)^*$, l'autòmat local estàndard per a $\overline{\alpha}$ és:



Autòmat de posició. Exemple

i l'autòmat de posició per a $\alpha = (a+b)(a^* + ba^* + b^*)^*$ és:



Autòmat Follow

- Relació follow
- Autòmat follow

Autòmat Follow

- L'*autòmat follow* d'una expressió regular α és l'autòmat quocient de l'autòmat de posició per la relació següent:

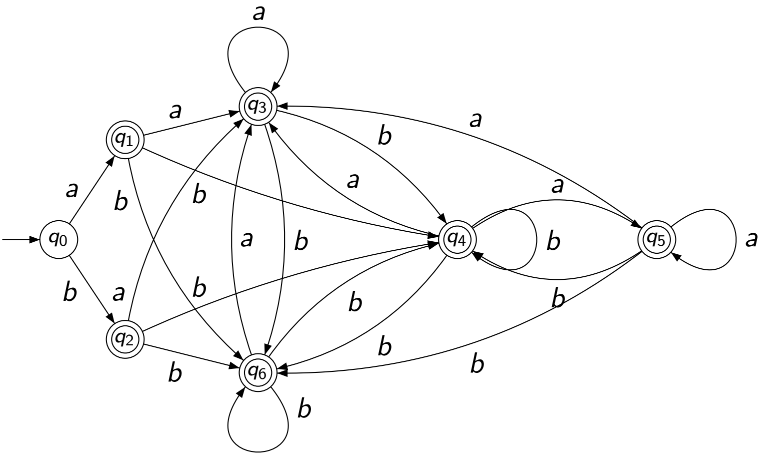
$$p \equiv_f q \Leftrightarrow \begin{cases} p, q \in F \ o \ be \ p, q \in Q - F \\ follow(p) = follow(q) \end{cases}$$

on $follow(p) = \{q \in Q \ : \ \exists a \in \Sigma, \delta(p, a) = q\}$

- L'autòmat quocient resultant és una reducció parcial de l'autòmat de posició.

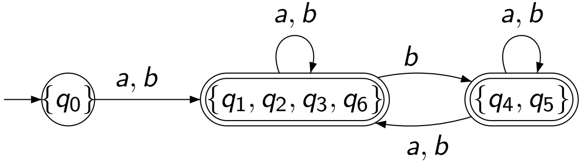
Autòmat Follow

Recordem l'autòmat de posició per a $\alpha = (a+b)(a^* + ba^* + b^*)^*$:



Autòmat Follow

Les classes d'equivalència són: $\{q_0\}, \{q_1, q_2, q_3, q_6\}, \{q_4, q_5\}$, amb la qual cosa l'autòmat follow per a α queda:



Anàlisi d'autòmats finits

- Sistemes d'equacions en expressions regulars
- Lema d'Arden
- Anàlisi d'autòmats finits

Sistemes d'equacions en expressions regulars. Lema d'Arden

- Equació en expressions regulars: Equació lineal on variables i coeficients prenen la forma d'expressions regulars.

$$X = rX + s$$

- Lema d'Arden: Siga $X = rX + s$ una equació en expressions regulars. $X = r^*s$ és una solució per a l'equació. És única si $\lambda \notin r$
  - demostrem que r^*s és solució:

$$rX + s \underset{X=r^*s}{=} rr^*s + s = (rr^* + \lambda)s \underset{rr^*+\lambda=r^*}{=} r^*s$$

  - Si $\lambda \in r$ existeixen infinites solucions: $\forall t \subseteq \Sigma^*$, $r^*(s + t)$ és solució:

$$\begin{aligned} X =& rX + s = rr^*(s + t) + s = rr^*s + rr^*t + s = \\ =& (rr^* + \lambda)s + rr^*t \underset{rr^*+\lambda=r^*}{=} r^*s + r^*t \underset{X=r^*(s+t)}{=} X \end{aligned}$$

Sistemes d'equacions en expressions regulars

- Donat un sistema d'equacions en expressions regulars:

$$\begin{cases} X_1 = r_{11}X_1 + r_{12}X_2 + & \ldots + r_{1n}X_n + s_1 \\ X_2 = r_{21}X_1 + r_{22}X_2 + & \ldots + r_{2n}X_n + s_2 \\ & \ldots \\ X_n = r_{n1}X_1 + r_{n2}X_2 + & \ldots + r_{nn}X_n + s_n \end{cases}$$

la resolució ve de l'aplicació del mètode de Gauss utilitzant el Lemma d'Arden per a reduir.

Anàlisi d'AFs. Algorisme

1: **Entrada:** Autòmat finit $A = (Q, \Sigma, \delta, q_1, F)$ amb $Q = \{q_1, q_2, \ldots, q_n\}$
2: **Eixida:** Expressió regular per a $L(A)$
3: **Mètode:**
4: Per cada estat q_i introduir una variable X_i
5: Si $q_i \in F$ aleshores en la part dreta de la i-èssima equació afegir el terme λ
6: Si $q_j \in \delta(q_i, a)$ aleshores en la parte dreta de la i-èssima equació afegir el terme aX_j, amb $a \in \Sigma \cup \{\lambda\}$
7: Resodre el sistema d'equacions en expressions regulars utilitzant el Lema d'Arden per a reduir
8: Tornar l'expressió regular associada a l'estat inicial
9: **Fi Mètode**