13068. Teoria d'Autòmats i Llenguatges Formals

3. Propietats dels llenguatges regulars

Francesc J. Ferri
Dept. d'Informàtica. Universitat de València

27 de novembre de 2002

Regular?

Com saber si un llenguatge és regular?

Es regular el resultat de transformar un L. regular?

Quin és el millor autòmat donat un L. regular?

Comptar, no comptar ...

Intuïtivament, no són regulars aquells llenguatges les cadenes dels quals requereixen algun tipus de "comptatge".

$$\{a^n b^n\}, \quad \{ww^{-1}\}, \quad \{w : |w|_a = |w|_b\}$$

Però, no sempre:

$$\{a^n | n \text{ primer}\}, \{a^{n^2} | n \geq 0\}$$

Seria més correcte dir que les cadenes dels llenguatges regulars requereixen un comptatge finit (o utilitzen una quantitat finita de memòria; la dels estats del corresponent autòmat).

Quan un llenguatge és regular?

Si trobem un autòmat finit, una gramàtica regular o una E.R. (anomenem açò descripció regular) equivalent el llenguatge és regular.

(Atenció: el problema de la verificació no és trivial)

Però, i si no en trobem cap? És perque no existeix o perque no sabem?

Com podem, doncs caracteritzar els llenguatges regulars front als que no ho són?

Condicions necessàries i suficients

L regular $\iff$ $\exists$ descripció regular per a L.

Aquesta és una condició necessària i suficient per a la regularitat. Serveix tant per comprovar que si com que no.

L'únic problema és que és fàcil aplicarla en un sentit (existència) i molt més complicada en l'altre (no existència).

Caldria trobar una condició necessària que ens permetera comprovar la no regularitat. És a dir,

L regular $\implies$ condició.

Equivalentment, no condició $\implies$ L no regular!

El lema de bombament, informalment

Si L és regular, aleshores ...??

- existeix $A : L = L(A)$.
- tota cadena de L arriba a un final de A.
- si una cadena de L és més llarga que el nombre d'estats, ha de passar dues vegades (o més) per algun estat.
- si una cadena ha passat dues vegades per algun estat, aleshores ha recorregut un **bucle**.
- la subcadena que correspon a recórrer un bucle es pot eliminar o repetir indefinidament i la cadena seguirà arribant al mateix estat final.

[El lema de bombament, informalment]

Si L és regular, aleshores:

> tota cadena suficientment llarga ha de contindre una subcadena que es puga **bombar**.

bombar una subcadena: elevar-la a n, $n \geq 0$.

Però aquest enunciat informal, per a usar en privat, quasi clandestí

... no té validesa "legal"!

Lema de bombament. Enunciat

Si $L \in \mathcal{L}_R$ aleshores

$\exists n \in \mathbb{N}$ de manera que $\forall z \in L$, $|z| \geq n$ es compleix que

- $\exists z = uvw$ (factorització de z) : $|uv| \leq n \wedge |v| \geq 1$
- $uv^i w \in L$, $\forall i \geq 0$

DEMOSTRACIÓ:

-n és el nombre d'estats mínim que calen per acceptar L.

-per això la subcadena "bombable" està a una distància del principi menor que n.

Exemple. L regular

- $0^\star 1^\star$ és regular.

  Per a cadenes de longitut major o igual que 1, sempre és possible trobar una subcadena (el primer símbol) que es pot bombar.

  $(0)^i 0$, $(1)^i 1111$, $(0)^i 0011$, ...

- $100(0+1)^\star$ és regular.

  ara la n hauria de ser 4 i el símbol que es podria bombar el quart.

Exemple. L no regular

- $L = \{0^n 1^n | n \geq 0\}$ no és regular.

  Si L NO compleix el lema de bombament, aleshores ho podrem assegurar.

No complir el lema: $\exists$ almenys una cadena major que n que no conté cap subcadena vàlida ($|uv| \leq n$) que es puga bombar.

És a dir, que almenys per a UN valor de i, $uv^i w \notin L$!

Què és n? n és un valor que suposarem existeix i tot estarà expressat en funció seua. La conclusió al final serà que n no pot existir.

Demostració ($\{0^n1^n|n \geq 0\}$ no regular)

- Suposem que n existeix.
- Siga $z = 0^n1^n$. Ara n és un valor concret i z una cadena concreta (en funció de n).
- Les úniques subcadenes vàlides ($|uv| \leq n$) estan formades per 1 o més zeros.
- Per exemple, per a $i = 0$, $uv^0w \notin L$ per a qualsevol factorització vàlida de z.
- És a dir, siga quin siga el valor de n, tenim UNA cadena z, que no pot contindre subcadenes vàlides bombables. Conclusió: L no pot ser regular.

Aplicacions del lema de bombament

Donat una autòmat A es pot decidir:

- si $L(A) = \emptyset$

$$\{w \in \Sigma^\star : |w| < |Q|\}$$

- si $|L(A)| < \infty$

$$\{w \in \Sigma^\star : |w| < 2|Q|\}$$

Propietats de clausura

Sota quines operacions $\oplus$ és tancada la classe $\mathcal{L}_R$?

Si L_1, L_2 regulars,
$L_1 \oplus L_2$ regular?

També es pot plantejar per a operacions unàries o n-àries en general.

$\mathcal{L}_R$ és tancada respecte de $\cup, \cdot, ^\star$

Trivial a partir de la definició d'expressió regular

$\mathcal{L}_R$ també és tancada respecte de la **complementació**

Si L regular, existeix $A = \langle Q, \Sigma, \delta, q_0, F \rangle$ determinista i complet amb $L = L(A)$.

Aleshores $A' = \langle Q, \Sigma, \delta, q_0, Q - F \rangle$ accepta $\overline{L}$.

Són dos autòmats equivalents?

Es pot decidir si dos autòmats A_1 i A_2 accepten o no el mateix llenguatge. Només cal construir el llenguatge regular

$$L_3 = \left[L(A_1) \cap \overline{L(A_2)}\right] \cup \left[\overline{L(A_1)} \cap L(A_2)\right]$$

Aleshores es compleix que $L(A_1) = L(A_2)$ si i només si $L_3 = \emptyset$.

$\mathcal{L}_R$ és tancada respecte de $\cap$, substitucions i homomorfismes

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$$

Si $L \subseteq \Sigma^\star$ regular i $f(a) \subseteq \Delta^\star$ regular $\forall a \in \Sigma$, existeixen expressions regulars α_L i $\alpha_a \forall a$.

α_L està definida sobre Σ. Si se substitueix tot símbol a per α_a s'obté una expressió regular sobre Δ que és equivalent a $f(L)$.

$\mathcal{L}_R$ tancada respecte homomorfismes inversos

$h^{-1}(L) = \{w | h(w) \in L\}$

Si L és regular, hi ha un autòmat A tal que $L = L(A)$.

Per acceptar una cadena $w \in h^{-1}(L(A))$ hi ha que comprovar que $h(w)$ és acceptada per A.

Si $w = a_1 a_2 \cdots a_k$, aleshores $h(w) = h(a_1)h(a_2) \cdots h(a_k)$.

Cal definir un nou autòmat A' on, per cada símbol a_i s'ha de conseguir l'efecte de passar-li $h(a_i)$ a A.

És a dir, $\delta'(q, a) = \delta(q, h(a)), \quad \forall a \in \Sigma$

Homomorfisme invers d'un L regular

Per exemple, $h : \{0, 1, 2\} \longrightarrow \{a, b\}$

Partim d'un autòmat A sobre $\{a, b\}$ la taula del qual és

	a	b
Q	$\delta(q, a)$	$\delta(q, b)$

La taula de l'autòmat corresponent a $h^{-1}(L(A))$ serà

	0	1	2
Q	$\delta(q, h(0))$	$\delta(q, h(1))$	$\delta(q, h(2))$

$\mathcal{L}_R$ és tancada respecte de quocients

Si L és regular, aleshores L/L_0 és regular.
(independentment de si ho és L_0!)

El motiu és que L/L_0 està format per prefixs de cadenes de L de forma que el sufix corresponent està en L_0.

Tot prefix de cadenes de L arriba a algun estat del corresponent autòmat A. Per ser L regular, si un prefix és acceptat, tota cadena que arribe al mateix estat ho serà també.

L'autòmat que accepta L/L_0 és el mateix A però canviant els estats finals per

$$F' = \{q \in Q | \exists y \in L_0 : \delta(q, y) \in F\}$$

Relacions binàries d'equivalència

- aRb, es diu que a està **relacionat** amb b.
- Si R és reflexiva, simètrica i transitiva aleshores és una **relació binària d'equivalència**.

  per exemple: aRb si i només si $|a| = |b|$
- Una **classe d'equivalència** de R està formada per tots els elements relacionats entre ells.
- Les classes d'equivalència de R formen una **partició** del conjunt original, A, que s'anomena **conjunt quocient**, i s'escriu A/R.
- Si el conjunt quocient és finit, es diu que la relació és d'**índex finit**.

Congruències

Siga un conjunt A amb una operació interna $\oplus$ i una relació R.

• R s'anomena **congruència dreta** en A respecte de $\oplus$ si per a tot parell d'elements de A, x, y es compleix que

$$xRz \implies (x \oplus z)R(y \oplus z), \quad \forall z \in A$$

si $xRz \implies (z \oplus x)R(z \oplus y)$ s'anomena **congruència esquerra**.

També es pot dir que R és **invariant** per la dreta o l'esquerra respecte de $\oplus$.

Congruència associada a un autòmat

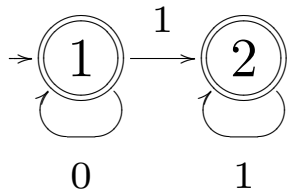
Donat un autòmat determinista $A = \langle Q, \Sigma, \delta, q_0, F \rangle$, es pot definir la següent congruència dreta sobre $\Sigma^\star$:

$$x \mathbb{R}_A y \iff \delta(q_0, x) = \delta(q_0, y)$$

Dues cadenes estan relacionades si arriben al mateix estat dins de A (o si no arriben a cap estat).

Les classes d'equivalència són els conjunts R^n_{1i}. Hi haurà tantes com estats (més una si l'autòmat no és complet).

Exemple $\mathbb{R}_A$



Hi ha tres classes d'equivalència. El conjunt quocient serà

$$\Sigma^\star/\mathbb{R}_A = \{0^\star,\ 0^\star 11^\star,\ 0^\star 11^\star 0(0+1)^\star\}$$

Congruència associada a L

Siga L un llenguatge (no necessàriament regular). Podem definir la següent congruència dreta sobre $\Sigma^\star$:

$x \mathbb{R}_L y$ si i només si ,

$$xz \in L \iff yz \in L, \quad \forall z \in \Sigma^\star$$

Dues cadenes estan relacionades si afegint el mateix al final, s'obté el mateix resultat (quant a estar o no en L).

Quantes i quines són les classes d'equivalència.

Exemple $\mathbb{R}_L$

Siga L format per cadenes de longitut parella.

$\mathbb{R}_L$ tindrà dues classes d'equivalència: parelles i senars.

$$\Sigma^\star/\mathbb{R}_L = \{L, \overline{L}\}$$

Siga $L = \{00\}$, Què val $\Sigma^\star/\mathbb{R}_{\{00\}}$?

$$\Sigma^\star/\mathbb{R}_{\{00\}} = \{\{\varepsilon\}, \{0\}, \{00\}, \overline{\{\varepsilon, 0, 00\}}\}$$

Altre exemple $\mathbb{R}_L$

Siga $L = \{0^n 1^n | n \geq 0\}$. Què val $\Sigma^\star/\mathbb{R}_L$?

$$\begin{array}{l} \varepsilon \not\mathbb{R}_L 0 \ (z = 1) \\ 0 \not\mathbb{R}_L 00 \ (z = 1) \\ \varepsilon \not\mathbb{R}_0 0 \ (z = 11) \\ \vdots \\ 0^k \not\mathbb{R}_L 0^m \ (z = 1^k) \ \underline{\texttt{si}} \ \ k \neq m! \end{array}$$

Hi ha infinites classes d'equivalència.

exercici: calcular $\Sigma^\star/\mathbb{R}_L$

Propietats de $\mathbb{R}_L$

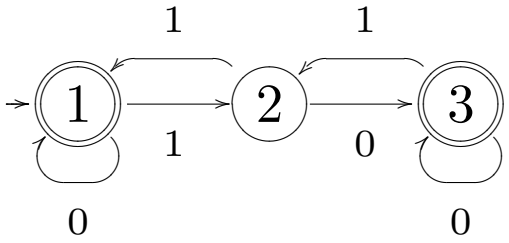
Si L és regular i A l'accepta,

> dues cadenes que arriben al mateix estat de A, els passarà el mateix independentment del que s'afegesca al final.

Teorema: $x\mathbb{R}_A y \implies x\mathbb{R}_L y$

Es compleix al reves?

Contraexemple



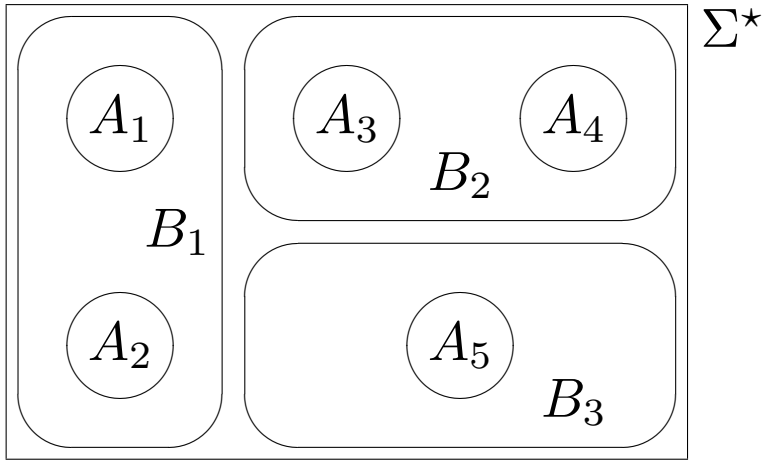
$\mathbb{R}_A$ indueix 3 classes, però ...

des dels estats 1 i 3 s'arriba al mateix quan afegim cadenes a la dreta!

$\mathbb{R}_L$ induex 2 classes: les que arriben a 1 o a 3, i les que arriben a 2.

$$x \mathbb{R}_L y \not\Rightarrow x \mathbb{R}_A y$$

$\mathbb{R}_A$ és un refinament de $\mathbb{R}_{L(A)}$



Es compleix que

$$|\Sigma^\star/\mathbb{R}_{L(A)}| \leq |\Sigma^\star/\mathbb{R}_A|, \quad \forall A = \langle Q, \Sigma, \delta, q_0, F\rangle$$

Teorema de Myhill-Nerode

Si L és regular $\implies$ $|\Sigma^\star/\mathbb{R}_L| < \infty$.

$$|\Sigma^\star/\mathbb{R}_L| \leq |\Sigma^\star/\mathbb{R}_A| = |Q| < \infty$$

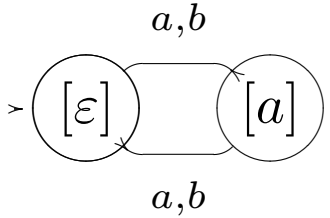
Si $|\Sigma^\star/\mathbb{R}_L| < \infty$ $\implies$ L és regular.

Es pot definir $A_L = \langle Q, \Sigma, \delta, q_0, F \rangle$

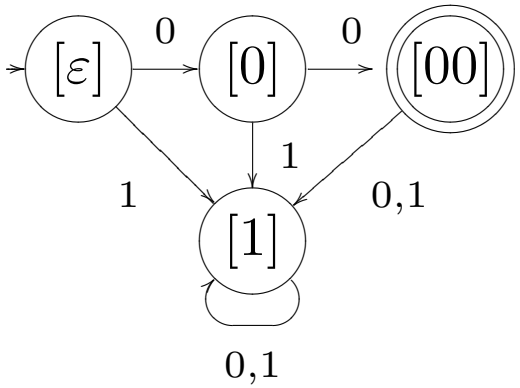
1. $Q = \Sigma^\star/\mathbb{R}_L$ (un estat per cada classe d'equivalència).
2. $q_0 = [\varepsilon]$ (la classe d'equivalència que conté ε).
3. $F = \{[x] | x \in L\}$ (aquelles classes d'equivalència que contenen cadenes de L).
4. $\delta([x], a) = [xa], \forall a \in \Sigma$.

Exemple. A_L

L cadenes parelles,



$L = \{00\}$,



Cororl·lari: A_L és l'autòmat mínim

Suposem que hi haguera un A amb menys estats que A_L. Aleshores,

$$|\Sigma^\star/\mathbb{R}_{A_L}| = |\Sigma^\star/\mathbb{R}_{L(A)}| > |\Sigma^\star/\mathbb{R}_A|$$

Impossible!

No pot haver autòmats més menuts que A_L

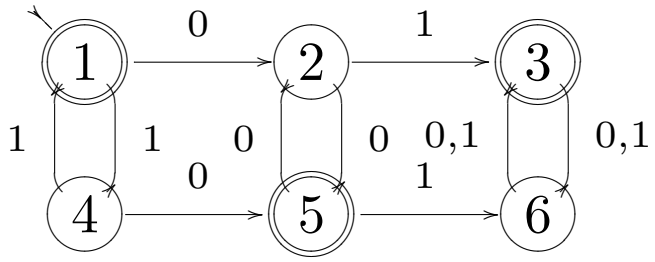
Aplicacions del T. de Myhill-Nerode

- Tenim una condició necessària i **suficient** per a la regularitat de qualsevol llenguatge.

- Podem "calcular" l'autòmat determinista mínim que accepta un llenguatge regular. Com?

> Partint d'un autòmat qualsevol i la corresponent $\mathbb{R}_A$, es poden **aglomerar** els seus blocs fins que es trobe $\mathbb{R}_L$.

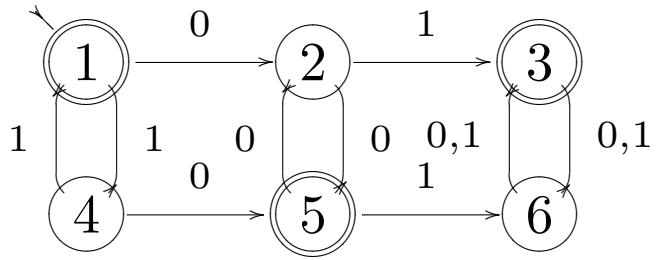
Autòmat mínim ($\mathbb{R}_L$)



	0	1
→ 1	2	4
2	5	3
3	6	6
4	5	1
5	2	6
6	3	3

[Autòmat mínim ($\mathbb{R}_L$)]

Hi ha 6 classes d'equivalència:



estat 1: $[\varepsilon]_{\mathbb{R}_A} = (11)^\star$

estat 4: $[1]_{\mathbb{R}_A} = 1(11)^\star$

estat 2: $[0]_{\mathbb{R}_A} = ([\varepsilon]_{\mathbb{R}_A} \cdot 0 + [1]_{\mathbb{R}_A} \cdot 00) \cdot (00)^\star$

estat 5: $[10]_{\mathbb{R}_A} = ([1]_{\mathbb{R}_A} \cdot 0 + [\varepsilon]_{\mathbb{R}_A} \cdot 00) \cdot (00)^\star$

estat 3: $[01]_{\mathbb{R}_A} = ([0]_{\mathbb{R}_A} \cdot 1 + [10]_{\mathbb{R}_A} \cdot 1(0+1)) \cdot ((0+1)^2)^\star$

estat 6: $[101]_{\mathbb{R}_A} = ([10]_{\mathbb{R}_A} \cdot 1 + [0]_{\mathbb{R}_A} \cdot 1(0+1)) \cdot ((0+1)^2)^\star$

Aglomerar?

Per cada parell de classes de $\mathbb{R}_A$, cal preguntarse:

$$[x]_{\mathbb{R}_A} \cup [y]_{\mathbb{R}_A} \subseteq [x]_{\mathbb{R}_L}$$

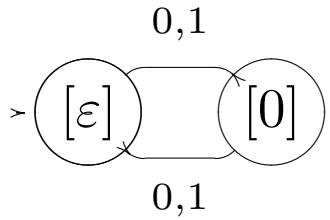
i això serà deveres si i només si

$$x \mathbb{R}_L y$$

Per tant, nomes cal considerar una cadena concreta per cada classe d'equivalència de $\mathbb{R}_A$, i comprovar si estan o no relacionades segons $\mathbb{R}_L$.

[Aglomerar?]

En l'exemple:



- $\varepsilon \mathbb{R}_L 10$.
- $\varepsilon \mathbb{R}_L 01$
- $1 \mathbb{R}_L 0$.
- $1 \mathbb{R}_L 101$.

Per tant,

$$[\varepsilon]_{\mathbb{R}_L} = [\varepsilon]_{\mathbb{R}_A} \cup [10]_{\mathbb{R}_A} \cup [01]_{\mathbb{R}_A}$$
$$[1]_{\mathbb{R}_L} = [1]_{\mathbb{R}_A} \cup [0]_{\mathbb{R}_A} \cup [101]_{\mathbb{R}_A}$$

Minimització d'autòmats finits

- L'objectiu és trobar un mètode més senzill.
- Basat directament en els estats i no en congruències sobre $\Sigma^\star$.
- Definirem una relació d'equivalència entre estats.

Estats equivalents

Siga $A = \langle Q, \Sigma, \delta, q_0, F \rangle$.

> Dos estats són equivalents si trobar-se en un o en l'altre és irrelevant respecte de l'acceptació o no de la cadena que s'està analitzant.

En altres paraules: les les configuracions (p, x) i (q, x) han de ser equivalents en el sentit de què

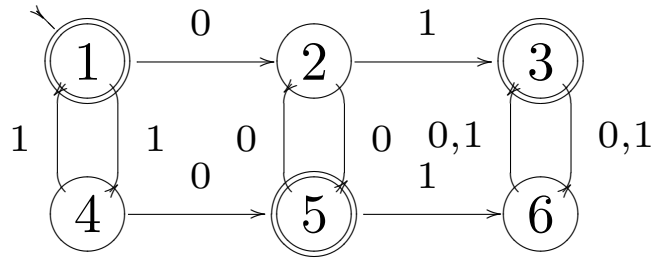
si $(p, x) \vdash^\star (r, \varepsilon)$ i $(q, x) \vdash^\star (s, \varepsilon)$, aleshores

$$r \in F \iff s \in F$$

Més formalment

Dos estats p i q de $A = \langle Q, \Sigma, \delta, q_0, F \rangle$ són equivalents ($p \equiv q$) si i només si

$$\delta(p, x) \in F \iff \delta(q, x) \in F, \quad \forall x \in \Sigma^\star$$



$$1 \equiv 3 \equiv 5, \quad 2 \equiv 4 \equiv 6$$

Estats equivalents?

- Però, com podem demostrar en un cas concret que $p \equiv q$?
- Caldria en general fer una demostració per inducció per cada parell d'estats.
- Cal trobar una manera més pràctica de calcular les equivalències

Estat equivalents (però no tant)

Dos estats p i q de $A = \langle Q, \Sigma, \delta, q_0, F \rangle$ són n-equivalents ($p \equiv_n q$) si i només si

$$\delta(p, x) \in F \iff \delta(q, x) \in F, \quad \forall x \in \Sigma^\star : |x| \leq n$$

Propietats: (la condició implica un nombre finit de cadenes)

- $Q/\equiv_0 = \{F, Q - F\}$ (trivial!)
- $p \equiv_{n+1} q \implies p \equiv_n q, \quad \forall n \geq 0$

  Si anem augmentant la n obtenim successius refinaments de $\equiv_0$.

  Però, com obtenim un refinament a partir de l'anterior??

Obtenció de les n-equivalències

Teorema: es pot calcular una equivalència a partir de l'anterior.

$$\left.\begin{array}{l} p \equiv_n q \\ \delta(p,a) \equiv_n \delta(q,a),\ \forall a \in \Sigma \end{array}\right\} \iff p \equiv_{n+1} q$$

Val. I per a què volem aquests successius refinaments?

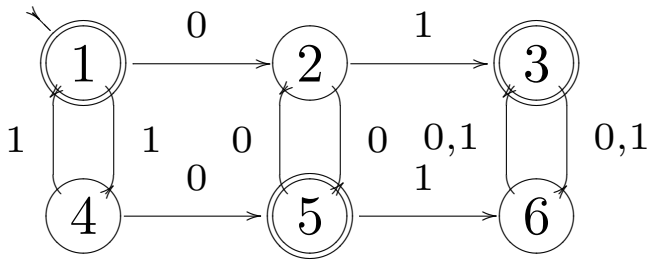
Per que tendeixen a $\equiv$.

$$\equiv_n = \equiv_{n+1} \implies \equiv_n = \equiv$$

Mètode per calcular l'AFD mínim

- Primer es calcula $Q/\equiv_0$: finals i no finals.
- A continuació es calcula $Q/\equiv_i$, $i = 1, 2, \ldots$ fins que per a dos valors consecutius el conjunt quocient no canvie.
- En eixe moment, tenim els estats que són equivalents.

Exemple:



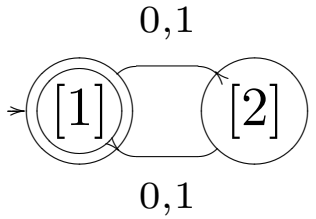
$$Q/\equiv_0 = \{\{1,3,5\},\{2,4,6\}\}$$

$$Q/\equiv_1 = \{\{1,3,5\},\{2,4,6\}\} = Q/\equiv$$

Autòmat associat a $\equiv$

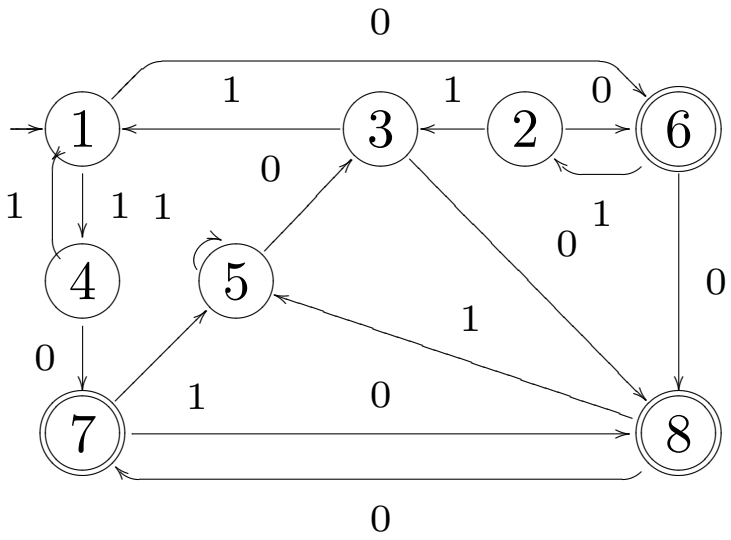
Una vegada obtinguda $\equiv$ se li pot associar el següent autòmat que serà el mínim: (en realitat es fa exactament el mateix que quan eliminavem estats idèntics)

$A_{\equiv} = \langle Q', \Sigma, \delta', q_0', F' \rangle$ on



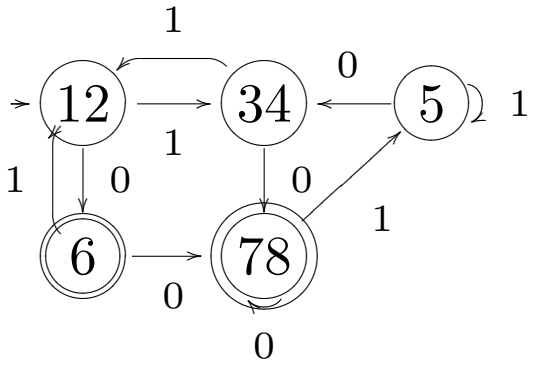
- $Q' = Q/\equiv$ (un estat per cada classe)
- $q_0' = [q_0]$ (la classe que conté q_0.)
- $F' = F/\equiv$ (les classes que continguen estats finals.)
- $\delta'([q], a) = [\delta(q, a)]$, $\forall a \in \Sigma$ i $\forall [q] \in Q'$.

Exemple



	0	1	$\equiv_0$	$\equiv_1$	$\equiv_2$	$\equiv_3$	$\equiv_4$
$\rightarrow$ 1	6	4	♠	♠	♠	♠	♠
2	6	3	♠	♠	♠	♠	♠
3	8	1	♠	♠	♠	♢	♢
4	7	1	♠	♠	♠	♢	♢
5	3	5	♠				
(6)	8	2	⊙	⊙			
(7)	8	5	⊙	⊙	⊙	⊙	⊙
(8)	7	5	⊙	⊙	⊙	⊙	⊙

Exemple: autòmat mínim



Altra representació

Marcar els parells d'estats **no** relacionats:

	0	1
→ 1	6	4
2	6	3
3	8	1
4	7	1
5	3	5
(6)	8	2
(7)	8	5
(8)	7	5

1								
2								
3	T	T						
4	T	T						
5	Y	Y	Y	Y				
6	X	X	X	X	X			
7	X	X	X	X	X	Z		
8	X	X	X	X	X	Z		
	1	2	3	4	5	6	7	8