Esercizi di Informatica Teorica

Linguaggi regolari: espressioni regolari e grammatiche, proprietà decidibili e teorema di Myhill-Nerode

a cura di

Luca Cabibbo e Walter Didimo

Sommario

- espressioni regolari e grammatiche regolari
- proprietà decidibili dei linguaggi regolari
- teorema di Myhill-Nerode

notazioni sul livello degli esercizi: (*) facile, (**) non difficile
(***) media complessità, (****) difficile, (*****) quasi impossibile

Espressioni regolari e linguaggi regolari

teorema L è un linguaggio regolare $\Leftrightarrow$ L è definibile con una espressione regolare

- da una espressione regolare per L si ricava un ASFND applicando le proprietà di chiusura dei linguaggi regolari (dall'ASFND si può poi ricavare una grammatica regolare che genera L)

- da una grammatica regolare che genera L si ricava una espressione regolare risolvendo un sistema di equazioni lineari

Da grammatica ad espressione regolare

il sistema di equazioni lineari si ricava dalla grammatica sostituendo ogni insieme di produzioni del tipo:

$A \rightarrow a_1B_1 \mid a_2B_2 \mid .... \mid a_nB_n \mid b_1 \mid b_2 \mid .... \mid b_m$ al modo:

$A = a_1B_1 + a_2B_2 + ... + a_nB_n + b_1 + b_2 + ... + b_m$

dal sistema di equazioni lineari si ricava una espressione regolare applicando le due tecniche seguenti ripetutamente:

- sostituzione: si può sostituire un simbolo non terminale con una espressione equivalente (es. $A = aB + b$, $B = cA \Rightarrow A = acA + b$)
- eliminazione della ricursione: si può sostituire l'equazione

$A = \alpha_1A + \alpha_2A + ... + \alpha_nA + \beta_1 + \beta_2 + ... + \beta_m$ con l'equazione

$A = (\alpha_1 + \alpha_2 + ... + \alpha_n)^* (\beta_1 + \beta_2 + ... + \beta_m)$

Esercizi svolti: da grammatica a espressione regolare

Esercizio 1(**) ricavare una espressione regolare per il linguaggio generato dalla seguente grammatica regolare:

$S \rightarrow aA$	$S \rightarrow bC$
$A \rightarrow aA$	$A \rightarrow bC$
$C \rightarrow cC$	$C \rightarrow d$

Soluzione

si ricava il seguente sistema:

$S = aA + bC$

$A = aA + bC$

$C = cC + d$

Esercizi svolti: da grammatica a espressione regolare

si applicano le tecniche di sostituzione ed eliminazione della ricursione:

S = aA + bC		S = aA + bC		S= aA + bc*d
A = aA + bC	$\Rightarrow$	A = aA + bC	$\Rightarrow$	A = aA + bc*d
C = cC + d		C = c*d		

S = aA + bc*d	$\Rightarrow$	S = aa*bc*d + bc*d
A = a*bc*d		

dunque risulta: aa*bc*d + bc*d

che semplificata diventa: a*bc*d

Esercizi svolti: da grammatica a espressione regolare

Esercizio 2(**) ricavare una espressione regolare per il linguaggio generato dalla seguente grammatica regolare:

$S \to aX$
$X \to bY|a$
$Y \to bX$

Soluzione

$S = aX$	$S = aX$	$S = aX$	$S = a(bb)^*a$
$X = bY + a$	$X = bbX + a$	$X = (bb)^*a$	
$Y = bX$			

Esercizi svolti: da grammatica a espressione regolare

Esercizio 3(***) ricavare una espressione regolare per il linguaggio generato dalla seguente grammatica regolare:

$S \rightarrow aX|a$
$X \rightarrow bX|aY|\varepsilon$
$Y \rightarrow bY|aX$

Soluzione

$S = aX + a$
$X = bX + aY + \varepsilon$
$Y = bY + aX$

$S = aX + a$
$X = bX + aY + \varepsilon$
$Y = b^*aX$

$S = aX + a$
$X = bX + ab^*aX + \varepsilon$

Esercizi svolti: da grammatica a espressione regolare

$S = aX + a$
$X = bX + ab^*aX + \varepsilon$

$S = aX + a$
$X = (b + ab^*a)^*$

$S = a(b+ab^*a)^* + a$

che può essere semplificata al modo: $a(b+ab^*a)^*$

Esercizio 4(***) ricavare una espressione regolare per il linguaggio generato dalla seguente grammatica regolare:

$S \rightarrow bX$
$X \rightarrow aX|bX|aY|a$
$Y \rightarrow bY|b$

Esercizi svolti: da grammatica a espressione regolare

Soluzione

S = bX	S = bX
X = aX + bX + aY + a	X = aX + bX + aY + a
Y = bY + b	Y = b*b

S = bX	S = bX
X = aX + bX + ab*b + a	X = (a+b)*(ab*b + a)

S = b(a+b)*(ab*b + a)

che si semplifica al modo: b(a+b)*ab*

Esercizi da svolgere: da grammatica a espr. regolare

Esercizio 5(***) ricavare una espressione regolare per il linguaggio generato da ciascuna delle seguenti grammatiche regolari:

1) $S \rightarrow a|aA$

$A \rightarrow aA|bA|a|b$

2) $S \rightarrow aX$

$X \rightarrow aX|bX|b$

3) $S \rightarrow aB \mid aC$

$B \rightarrow bX \mid a$

$X \rightarrow bB$

$C \rightarrow cC \mid c$

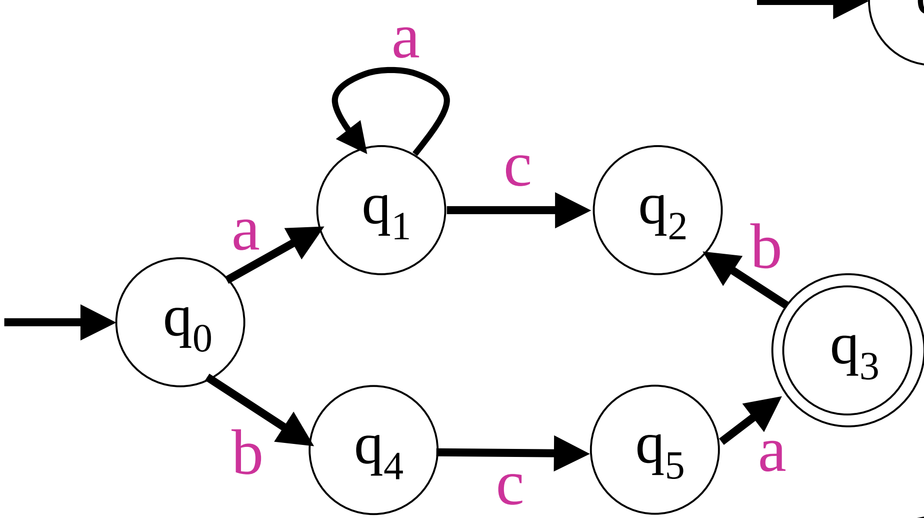
Proprietà decidibili dei linguaggi regolari

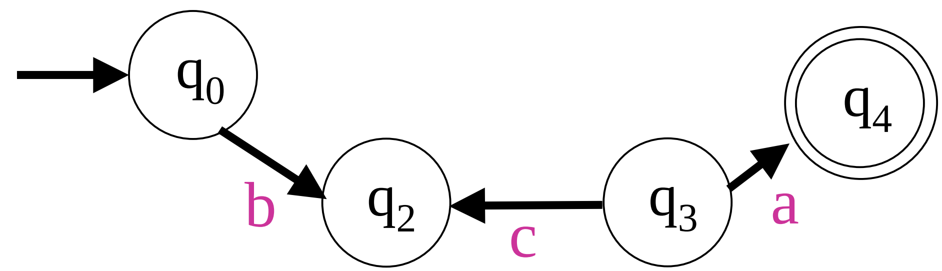
teorema è possibile decidere se un linguaggio regolare L è vuoto, finito o infinito

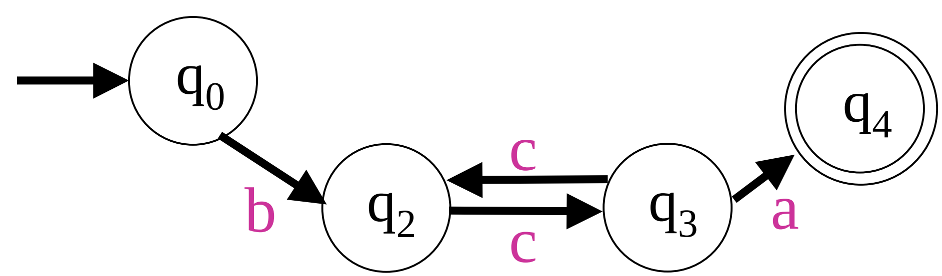
- è sufficiente studiare un ASF A che riconosce L: se n è il numero di stati di A, allora:
  - L è vuoto se e solo se A non accetta alcuna stringa di lunghezza minore di n
  - L è infinito se e solo se A accetta qualche stringa di lunghezza $k \in [n, 2n)$
  - altrimenti L è finito

Proprietà decidibili dei linguaggi regolari

Esercizio 6(*) dire se i linguaggi riconosciuti dai seguenti ASF sono vuoti, finiti o infiniti







Proprietà decidibili dei linguaggi regolari

teorema dati due linguaggi regolari L_1 ed L_2 è possibile decidere se:

- $L_1 \subseteq L_2$
- $L_1 = L_2$

infatti:

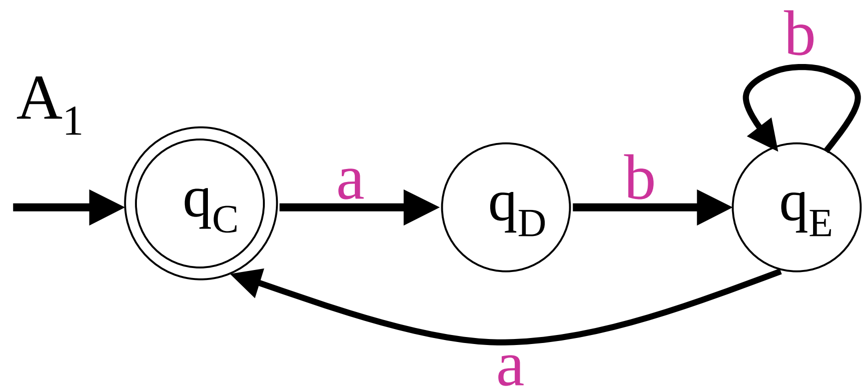
- $L_1 \subseteq L_2 \Leftrightarrow L_1 - L_2 = \varnothing$ $\quad$ $(L_1 - L_2 = c(c(L_1) \cup L_2)$
- $L_1 = L_2 \Leftrightarrow L_1 \subseteq L_2$ ed $L_2 \subseteq L_1$

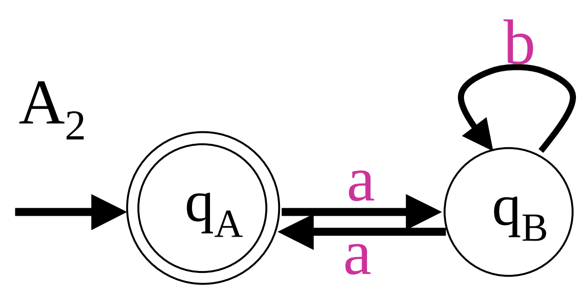
osservazione: $L_1 = L_2$ equivale anche a dire che

$$(L_1 \cap c(L_2)) \cup (L_2 \cap c(L_1)) = \varnothing$$

Proprietà decidibili dei linguaggi regolari

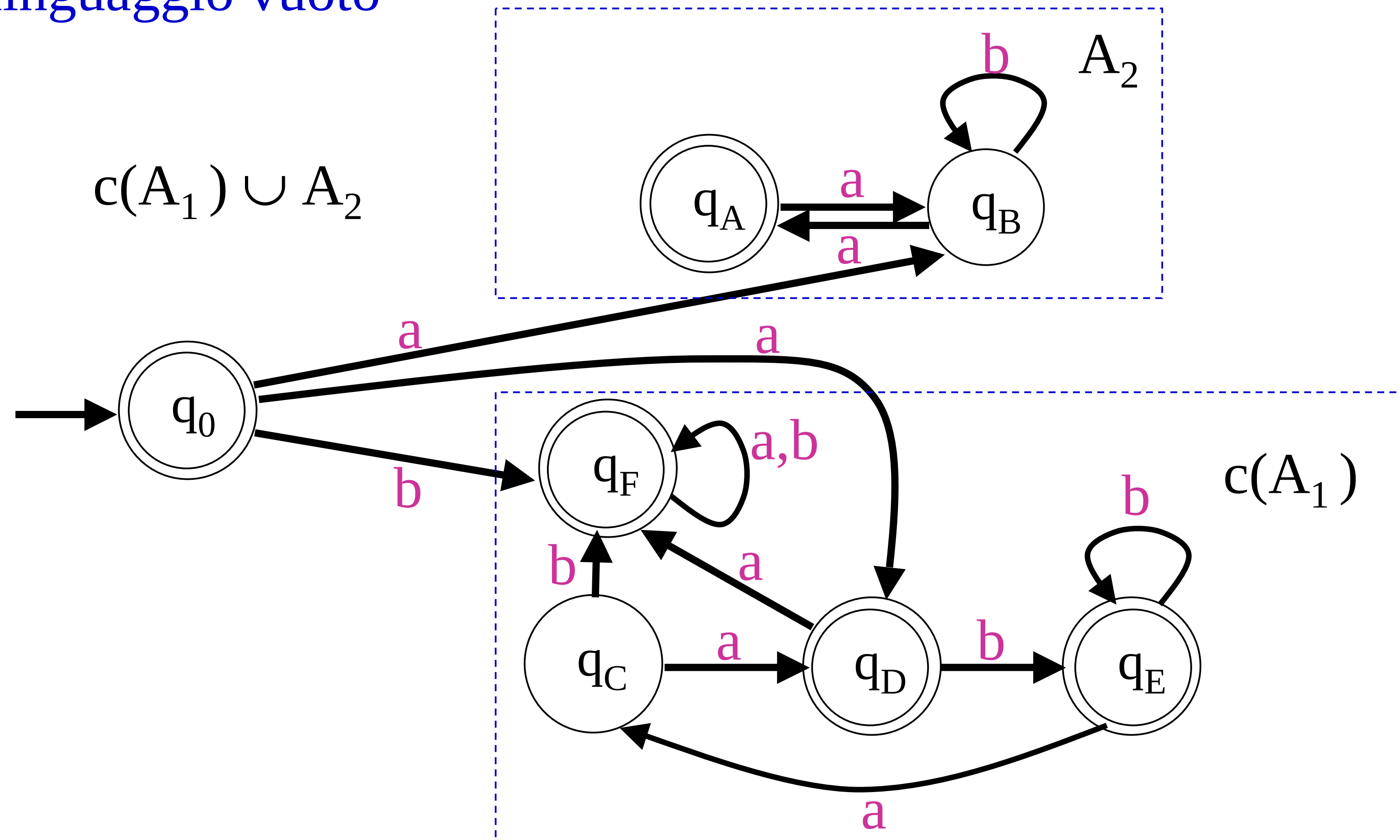
Esercizio 7(***) dimostrare formalmente che il linguaggio L_1 riconosciuto dall'ASF A_1 è contenuto nel linguaggio L_2 riconosciuto dall'ASF A_2.





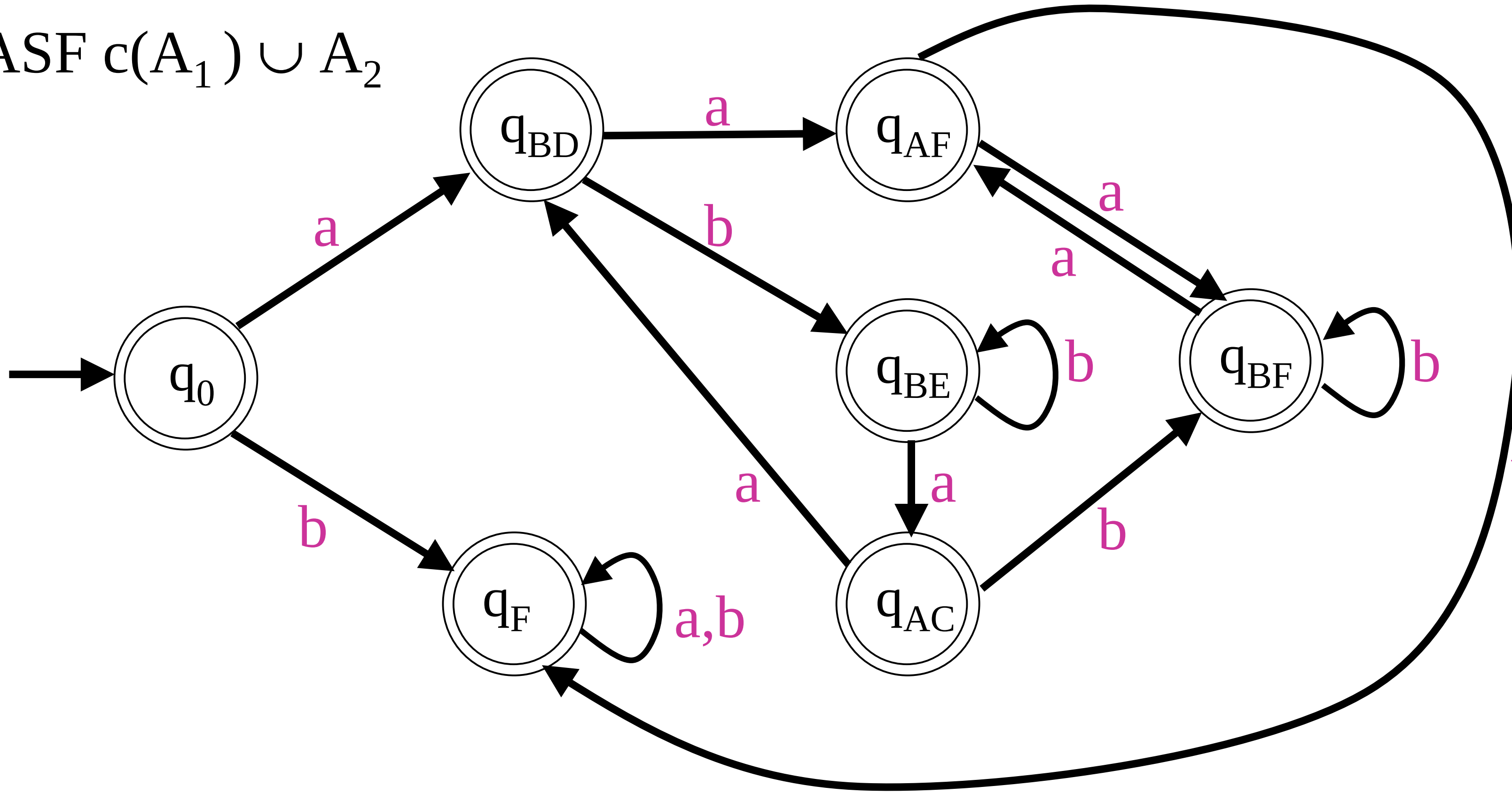
Proprietà decidibili dei linguaggi regolari

Soluzione dimostriamo che $A = A_1 - A_2$ è un automa che riconosce il linguaggio vuoto



Proprietà decidibili dei linguaggi regolari

ASF $c(A_1) \cup A_2$



quindi, il complementare di questo ASF non avrà stati finali, e dunque riconoscerà il linguaggio vuoto.

Teroma di Myhill-Nerode

<u>teorema</u> sia L un linguaggio sull'alfabeto Σ; sia data la seguente relazione di equivalenza su Σ^*:

$$xR_Ly \Leftrightarrow (\forall z \in \Sigma^* \ \ xz \in L \Leftrightarrow yz \in L)$$

R_L ha indice finito $\Leftrightarrow$ L è regolare

<u>osservazioni</u>:

• si ricordi che <u>l'indice</u> di R_L è il <u>numero delle sua classi di equivalenza</u>, cioè il numero di elementi dell'insieme quoziente R_L/Σ^*

• il teroma di Myhill-Nerode fornisce una <u>caratterizzazione</u> dei linguaggi regolari, e può quindi essere <u>usato per provare sia la regolarità che la non regolarità</u> di un linguaggio

Esercizi svolti sul teorema di Myhill-Nerode

Esercizio 8(**) determinare tutte le classi di equivalenza della relazione R_L per il linguaggio $L = a^*ba^*$.

Soluzione:

esistono tre distinte classi di equivalenza:

- $C_1=\{a^n: n \geq 0\}$ (nota: comprende anche ε)
- $C_2=\{a^n b a^m: n,m \geq 0\}$
- $C_3=\{w \in \{a,b\}^* : \text{non esiste z tale che wz} \in L\}$

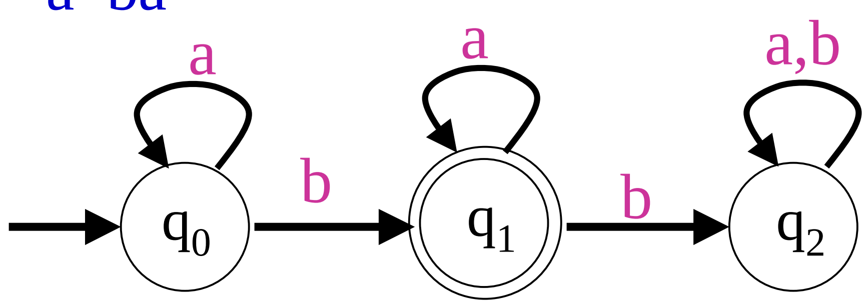
esercizio: mostrare qualche stringa di C_3

Esercizi svolti sul teorema di Myhill-Nerode

osservazione:

le classi di equivalenza di R_L rispetto ad un linguaggio regolare L sono associabili agli stati di un opportuno ASF (minimo) che riconosce L

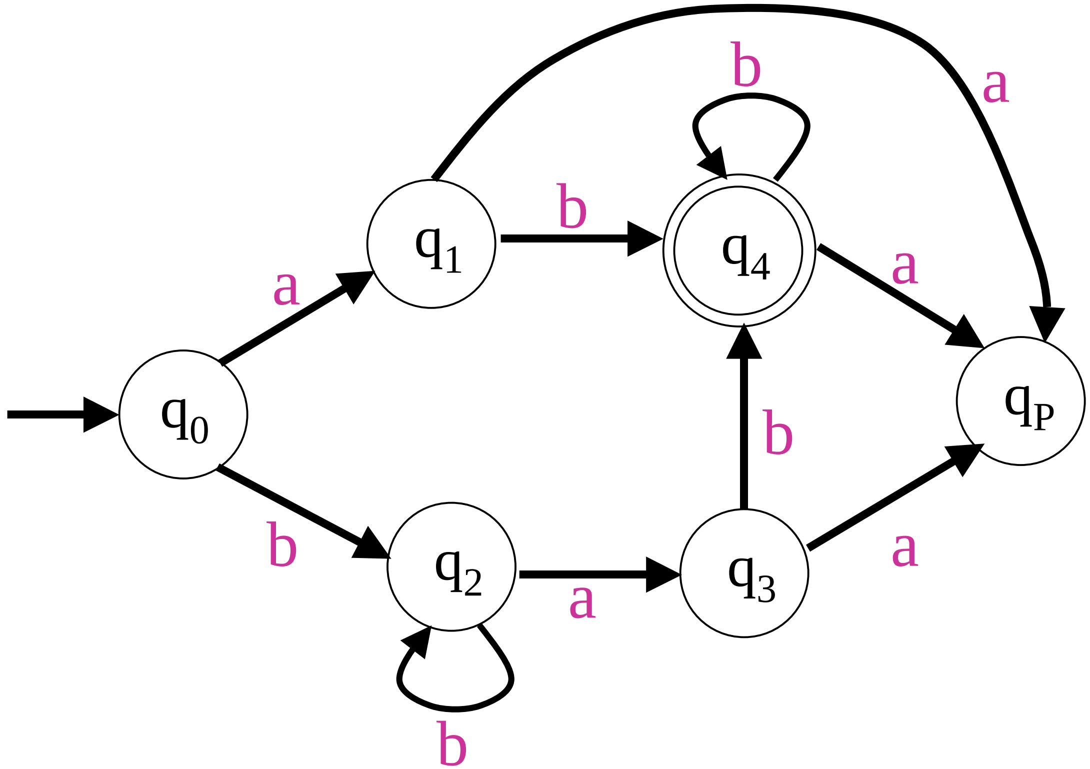
esempio per L = a*ba*



- $C_1=\{a^n: n \geq 0\} \leftrightarrow q_0$
- $C_2=\{a^n b a^m: n,m \geq 0\} \leftrightarrow q_1$
- $C_3=\{w \in \{a,b\}^* :$ non esiste z tale che $wz \in L\} \leftrightarrow q_2$

Esercizi svolti sul teorema di Myhill-Nerode

Esercizio 9(***) determinare tutte le classi di equivalenza della relazione R_L per il linguaggio L riconosciuto dal seguente ASF; qual'è l'indice di R_L?



Esercizi svolti sul teorema di Myhill-Nerode

Soluzione consideriamo la relazione di equivalenza $xR_My \Leftrightarrow \delta(q_0,x) = \delta(q_0,y)$; sappiamo che (vedi dimostrazione del teorema di Myhill-Nerode) se $xR_My \Rightarrow xR_Ly$, quindi R_M ha indice maggiore o uguale a quello di R_L (le classi di R_L sono ottenibili per unione di classi di R_M)

le classi di R_M si ottengono facilmente dall'ASF:

- $C_1=\{\varepsilon\} \leftrightarrow q_0$
- $C_2=\{a\} \leftrightarrow q_1$
- $C_3=\{bb^*\} \leftrightarrow q_2$
- $C_4=\{bb^*a\} \leftrightarrow q_3$
- $C_5=\{b^*abb^*\} \leftrightarrow q_4$ (nota che $C_5=L$)
- $C_6=\{w \in \{a,b\}^* : \text{non esiste } z \text{ tale che } wz \in L\} \leftrightarrow q_P$

Esercizi svolti sul teorema di Myhill-Nerode

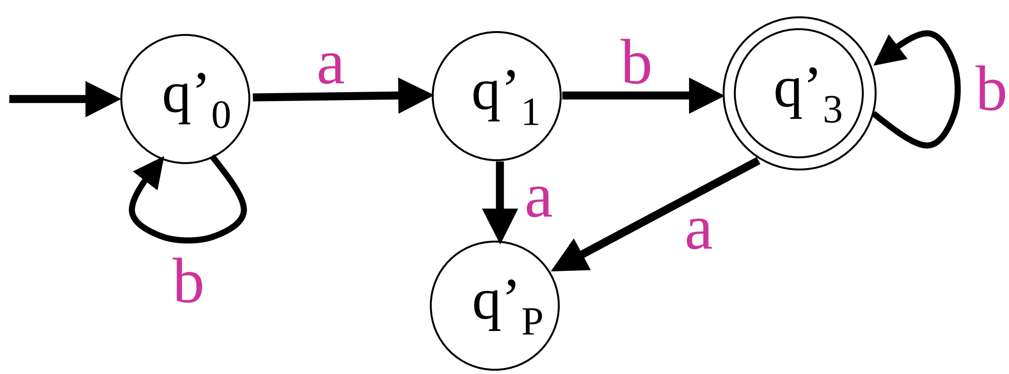
- $C_1=\{\varepsilon\} \leftrightarrow q_0$
- $C_2=\{a\} \leftrightarrow q_1$
- $C_3=\{bb^*\} \leftrightarrow q_2$
- $C_4=\{bb^*a\} \leftrightarrow q_3$
- $C_5=\{b^*abb^*\} \leftrightarrow q_4$ (nota che $C_5=L$)
- $C_6=\{w \in \{a,b\}^* :$ non esiste z tale che $wz \in L\} \leftrightarrow q_P$

per <u>ottenere le classi di equivalenza di R_L</u> si osserva che le classi C_2 e C_4 devono essere unite, in quanto $aR_L(bb^*a)$; inoltre risulta $\varepsilon R_L(bb^*)$, quindi anche C_1 e C_3 debbono essere unite; le classi di equivalenza di R_L sono dunque le seguenti:

Esercizi svolti sul teorema di Myhill-Nerode

- $C'_1 = \{b^*\} \leftrightarrow q'_0$ (unione di C_1 e C_3)
- $C'_2 = \{b^*a\} \leftrightarrow q'_1$ (unione di C_2 e C_4)
- $C'_3 = \{b^*abb^*\} \leftrightarrow q'_3$ (equivale a C_5)
- $C'_4 = \{w \in \{a,b\}^* :$ non esiste z tale che $wz \in L\} \leftrightarrow q'_P$ (equivale a C_6)

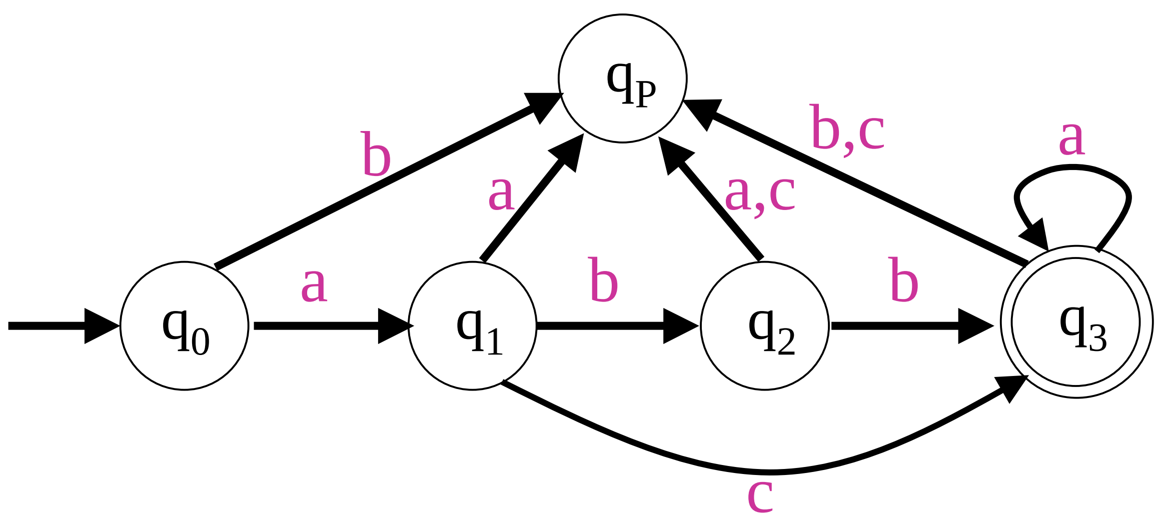
si può in effetti costruire un ASF (minimo) con soli 4 stati che riconosce L



Esercizio svolti sul teorema di Myhill-Nerode

Esercizio 10(***) determinare le classi di equivalenza della relazione R_L di Myhill-Nerode per il seguente linguaggio regolare:
$L = a(bb + c)a^*$.

Soluzione
consideriamo un ASF che riconosce L



Esercizio svolti sul teorema di Myhill-Nerode

le classi di R_M sono:

- $C_1=\{\varepsilon\} \leftrightarrow q_0$
- $C_2=\{a\} \leftrightarrow q_1$
- $C_3=\{ab\} \leftrightarrow q_2$
- $C_4=\{abba^*, aca^*\} \leftrightarrow q_3$
- $C_5 =\{w \in \{a,b\}^* :$ non esiste z tale che $wz \in L\} \leftrightarrow q_P$

d'altro canto, è facile osservare che non è possibile unire nessuna di queste classi nella relazione R_L (l'AFS ha il minimo numero di stati); quindi le classi di R_M <u>coincidono</u> con quelle di R_L.

Esercizio svolti sul teorema di Myhill-Nerode

Esercizio 11(***) dimostrare, utilizzando il teorema di Myhill-Nerode, che il linguaggio $L = \{a^n b^n : n \geq 0\}$ non è regolare; quali sono le classi di equivalenza della relazione R_L?

Soluzione

• la relazione R_L ha una classe di equivalenza $\{a^k\}$ distinta per ogni naturale k; infatti, comunque scelti $k > h$, risulta che la stringa $a^k b^k$ appartiene al linguaggio, mentre non vi appartiene la stringa $a^h b^k$; dunque, R_L ha sicuramente un numero infinito di classi di equivalenza, e pertanto L non è regolare.

• tutte le classi di equivalenza di R_L sono le seguenti:

Esercizio svolti sul teorema di Myhill-Nerode

- $\{\varepsilon\}$
- $\{a^k\}\ \forall k > 0$
- $\{a^k b^h\}\ \forall k, h > 0$
- $\{w \in \{a,b\}^*$: non esiste z tale che $wz \in L\}$

Esercizio 12(****) dato il linguaggio $L = \{a^n b^m c^{n+m}: n,m \geq 1\}$, determinare tutte le classi di equivalenza della relazione R_L.

Soluzione

osservazioni preliminare: le stringhe "aaabb", "aabbb", "abbbb", "aaaab" appartengono tutte alla stessa classe di equivalenza;

Esercizi svolti sul teorema di Myhill-Nerode

più in generale:

- per ogni $k > 1$ le stringhe del tipo $x = a^n b^m : n,m \geq 1$ ed $n+m=k$ appartengono alla stessa classe di equivalenza, infatti $xz \in L \Leftrightarrow z = b^h c^{k+h}$ ($h \geq 0$); quindi per ogni $k > 1$
$B_k = \{a^n b^m : n,m \geq 1 \text{ ed } n+m=k\}$ è una classe di equivalenza distinta;
- ragionando analogamente a sopra, per ogni $k > 0$ le stringhe del tipo $x = a^n b^m c^h : (n+m) - h = k$ ed $n,m,h \geq 1$, appartengono alla stessa classe di equivalenza, infatti $xz \in L \Leftrightarrow z = c^k$; quindi per ogni $k > 0$
$C_k = \{a^n b^m c^h : (n+m) - h = k \text{ ed } n,m,h \geq 1\}$ è una classe di equivalenza distinta;
- le altre classi di equivalenza sono:
$A_k = \{a^k\}$ per ogni $k \geq 0$ (notare che $A_0=\{\varepsilon\}$) e la classe
$D = \{w \in \{a,b,c\}^* : \text{non esiste z tale che } wz \in L\}$

Esercizi da svolgere sul teorema di Myhill-Nerode

Esercizio 13(***) dato il linguaggio L = ba*(bb)*a, determinare tutte le classi di equivalenza della relazione R_L.

Esercizio 14(***) dimostrare, utilizzando il teorema di Myhill-Nerode, che il linguaggio L = $\{a^n b^m c^n : n,m \geq 0\}$ non è regolare; determinare inoltre tutte le classi di equivalenza della relazione R_L.

Esercizio 15(****) dato il linguaggio L = $\{a^n b^m c^{n+m} : n,m \geq 0\}$, determinare tutte le classi di equivalenza della relazione R_L. (attenzione: in questo caso possono anche mancare delle a o delle b}