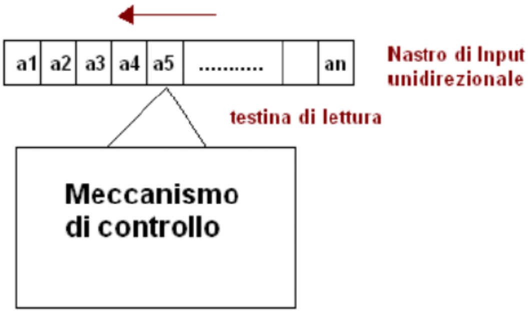
I metodi formali dell'Analisi Lessicale: Gli automi a stati finiti (FSA)

N.Fanizzi - V.Carofiglio

31 marzo 2016

Schema di un Automa a Stati Finiti



Automi a Stati Finiti

Dato un alfabeto X, un **automa a stati finiti** (FSA) è una quadrupla

$$M = (Q, \delta, q_0, F)$$

- X è l'*alfabeto d'ingresso*
- Q è un insieme finito e non vuoto di *stati*
- δ è la *funzione di transizione*:

$$\delta : Q \times X \longrightarrow Q$$

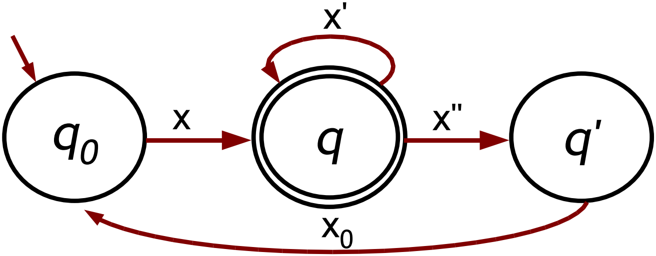
- q_0 è lo *stato iniziale*
- $F \subseteq Q$ è l'insieme degli *stati finali* o *d'accettazione*

δ funzione parziale: può essere indefinita per qualche coppia (q, x)
Si può ottenere una funzione δ totale considerando uno stato q_P (*pozzo*) dal quale non si raggiungano stati finali

Un FSA $M = (Q, \delta, q_0, F)$ è rappresentabile mediante:

▷ un grafo detto **diagramma di transizione** in cui:

- ogni stato $q \in Q$ è rappresentato da un cerchio con etichetta q
- lo stato iniziale q_0 ha un arco entrante libero
- per ogni $q \in Q$ ed ogni $x \in X$, se $q' = \delta(q, x)$ allora esiste un arco da q in q' etichettato con x



▷ una matrice **tavola di transizione** in cui:

- sulle righe: gli stati $q_i \in Q, \quad i = 1, \dots, m$
- sulle colonne: i simboli dell'alfabeto d'ingresso $x_j \in X$ $j = 1, \dots, n$
- in ogni casella: $q_i^j = \delta(q_i, x_j)$

δ	x_1	x_2	$\cdots$	x_n
$\rightarrow q_0$	q_0^1	q_0^2	$\cdots$	q_0^n
$*q_1$	q_1^1	q_1^2	$\cdots$	q_1^n
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
q_m	q_m^1	q_m^2	$\cdots$	q_m^n

La funzione δ^*

Dato un automa a stati finiti $M = (Q, \delta, q_0, F)$
Si definisce per induzione la funzione

$$\delta^* : \; Q \times X^* \longrightarrow Q$$

$$\forall w \in X^*, \forall q \in Q : \; \delta^*(q, w) = \begin{cases} q & \text{se } w = \lambda \\ \delta(\delta^*(q, w'), x) & \text{se } w = w'x \end{cases}$$

La funzione calcola lo stato di arrivo avendo in ingresso uno stato ed una parola sull'alfabeto X

Linguaggi accettati da FSA

- Una parola si dice **accettata** (o *riconosciuta*) da M se, partendo da q_0 e data la sequenza di ingresso w, M porta ad uno stato q finale

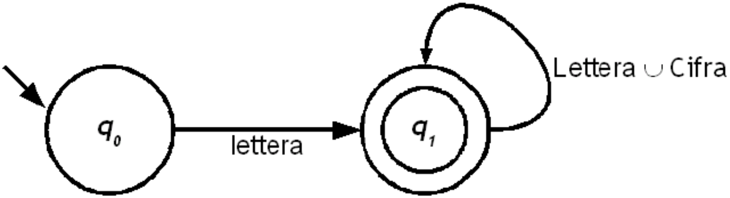
$$\delta^*(q_0, w) = q \in F$$

- Il **linguaggio accettato** (o *riconosciuto*) da M è dato dall'insieme:

$$\mathbf{T(M)} = \{w \in X^* \mid \delta^*(q_0, w) \in F\}$$

Gli identificatori C-like

$M = (\{q_0, q_1\}, \delta, q_0, F)$:



Equivalenza di FSA

Due **FSA** M_1 e M_2 si dicono **equivalenti** quando:

$$T(M_1) = T(M_2)$$

Un linguaggio L su un alfabeto X è un **linguaggio a stati finiti** (FSL) sse esiste un FSA con alfabeto di ingresso X tale che: $L = T(M)$
Risulta così definita la **Classe di Linguaggi a Stati Finiti**:
$\mathcal{L}_{FSL} = \{L \in \wp(X^*) \mid \exists M \; L = T(M)\}$

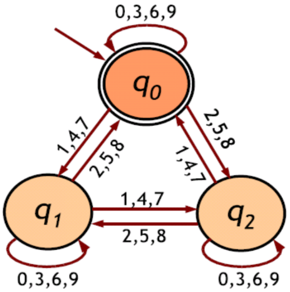
Esempio.

$$L = \{w \mid w \ mod \ 3 = 0\}$$

$D = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$
$M = (Q, \delta, q_0, F) \in FSA$
$Q = \{q_0, q_1, q_2\}$

- q_0 = classe resto 0
- q_1 = classe resto 1
- q_2 = classe resto 2

$F = \{q_0\}$



Proposizione. La classe $\mathcal{L}_{FSL}$ è chiusa rispetto al complemento
Dim. Sia $L \in \mathcal{L}_{FSL}$ su X.
Per definizione: $\exists M = (Q, \delta, q_0, F)$ FSA tale che $T(M) = L$.
Considerato $\overline{L} = X^* \setminus L$ e l'FSA $\overline{M} = (Q, \delta, q_0, Q \setminus F)$
risulta che $\overline{L} = T(\overline{M})$
Da dimostrare per induzione sulla lunghezza delle stringhe:

base Sia $w = \lambda$.
Se $\lambda \notin L$ allora $\delta^*(q_0, \lambda) \notin F$ per cui
$\delta^*(q_0, \lambda) \in Q \setminus F$ e quindi
$\lambda \in \overline{L} \cap T(\overline{M})$

passo Supponiamo di avere $w \in X^*$ tale che $|w| = n \in N$ e che $w = w'x$ con $|x| = 1$.
La parola w' deve essere supposta appartenere a $\overline{L} \cap T(\overline{M})$ per ipotesi di induzione, avendo lunghezza $n - 1$
Sia $q' = \delta^*(q_0, w')$ etc...

Esercizi.

- Costruire il FSA che accetta il linguaggio

$$L = \{a^n b^m | n \geq 2, \;\; m \geq 3\}$$

- Costruire il FSA che accetta il linguaggio su $X = \{a, b\}$ consistente di tutte le stringhe con non piú di 3 "a".
- Costruire il FSA per il linguaggio

$$L = \{a^n b^m \;|n + m \; un \; numero \; pari\}$$

- Costruire il FSA che accetta i linguaggi seguenti su $X = \{a, b\}$

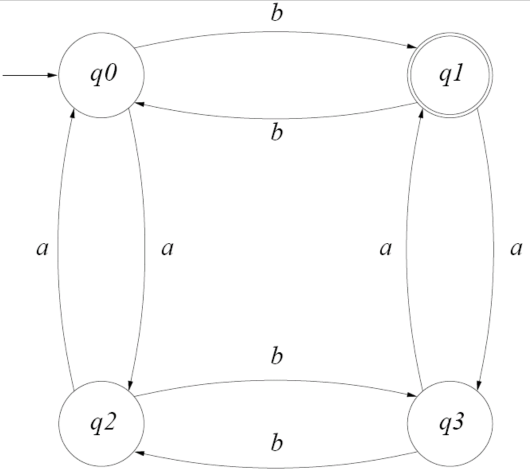
$$L = \{w : n_a(w) - n_b \; mod \; 3 \;\; = 1\}$$
$$L = \{w : n_a(w) - n_b \; mod \; 3 \;\; \neq 1\}$$

Esercizi

1. Costruire un FSA che accetti questo linguaggio:
   $L = \{w \in \{a, b\}^* \mid w \text{ ha un numero pari di } a \text{ e dispari di } b\}$
2. Costruire un FSA che accetti questo linguaggio:
   $L = \{w \in \{a, b\}^* \mid w \neq \alpha aa\beta,\ \alpha, \beta \in \{a, b\}^*\}$

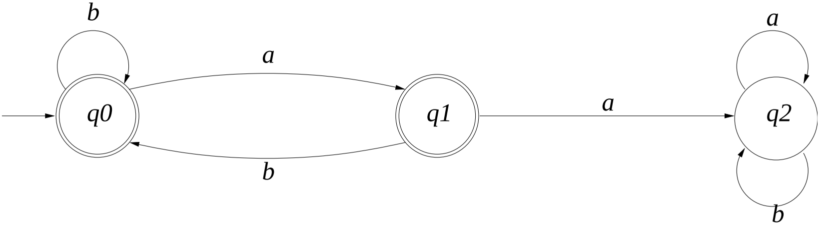
Esercizio 1. Costruire un FSA che accetti questo linguaggio:
$L = \{w \in \{a, b\}^* \mid w$ ha un numero pari di a e dispari di $b\}$
Soluzione: Sia $M = (Q, \delta, q_0, F) \in$ FSA

- $Q = \{q_0, q_1, q_2, q_3\}$ dove
  - q_0 stato per un numero pari di a e di b
  - q_1 stato per un numero pari di a e dispari di b
  - q_2 stato per un numero dispari di a e pari di b
  - q_3 stato per un numero dispari di a e di b
- funzione di transizione δ definita:
  - $\delta(q_0, a) = \delta(q_3, b) = q_2$
  - $\delta(q_0, b) = \delta(q_3, a) = q_1$
  - $\delta(q_1, a) = \delta(q_2, b) = q_3$
  - $\delta(q_1, b) = \delta(q_2, a) = q_0$
- q_0 stato iniziale
- $F = \{q_1\}$


q0
q1
q2
q3
b
b
b
b
a
a
a
a

Esercizio 2. Costruire un FSA che accetti questo linguaggio:
$L = \{w \in \{a, b\}^* \mid w \neq \alpha aa\beta,\ \alpha, \beta \in \{a, b\}^*\}$
Soluzione: Sia $M = (Q, \delta, q_0, F) \in$ FSA

- $Q = \{q_0, q_1, q_2\}$ dove
  - q_0 stato per parole non contenenti due o più a consecutive e terminanti con b
  - q_1 stato per parole non contenenti due o più a consecutive e terminanti con a
  - q_2 stato pozzo per parole contenenti due o più a consecutive
- funzione di transizione δ definita:
  - $\delta(q_0, a) = q_1$
  - $\delta(q_0, b) = \delta(q_1, b) = q_0$
  - $\delta(q_1, a) = \delta(q_2, a) = \delta(q_2, b) = q_2$
- q_0 stato iniziale
- $F = \{q_0, q_1\}$


b
a
a
a
q0
q1
q2
b
b

Automi Non Deterministici

Un **automa** a stati finiti **non deterministico** (NDA)

$$M = (Q, \delta, q_0, F)$$

- Q è un insieme finito e non vuoto di *stati*
- q_0 è lo *stato iniziale*
- $F \subseteq Q$ è l'insieme degli *stati finali* o *d'accettazione*
- δ è la *funzione di transizione*

$$\delta : Q \times X \longrightarrow \wp(Q)$$

Per ogni stato e simbolo dell'alfabeto X si ha ora un insieme di stati successivi possibili in cui transitare.
Si può definire anche in questo caso l'estensione di δ alle stringhe:

$$\delta^* : \ \wp(Q) \times X^* \longrightarrow \wp(Q)$$

$$\forall p \in \wp(Q)\, \forall w \in X^* : \ \delta^*(p, w) = \begin{cases} p & \text{se } w = \lambda \\ \bigcup_{q \in \delta^*(p,v)} \delta(q, x) & \text{se } w = vx \end{cases}$$

- Una parola si dice **accettata** (o *riconosciuta*) dal NDA M se, partendo da q_0 con una sequenza di ingresso w, M transita ad uno stato finale in almeno un cammino

$$\delta^*(\{q_0\}, w) \cap F \neq \emptyset$$

- Il **linguaggio accettato** (o *riconosciuto*) da M è dato dall'insieme:

$$\mathbf{T(M)} = \{w \in X^* \mid \delta^*(\{q_0\}, w) \cap F \neq \emptyset\}$$

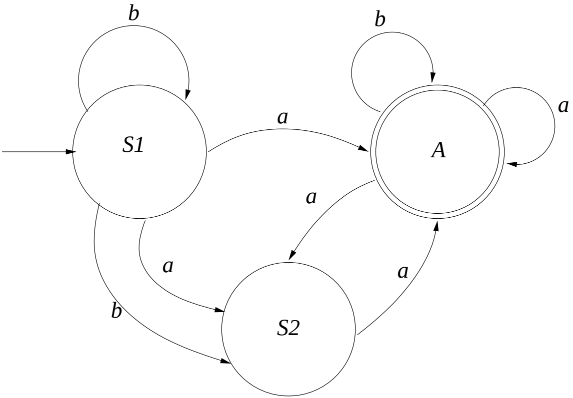
Due NDA sono **equivalenti** se generano lo stesso linguaggio.

Risulta così definita la seguente classe di linguaggi:

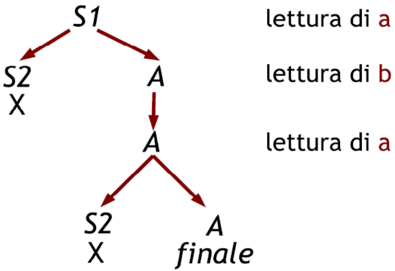
$$\mathcal{L}_{NDL} = \{L \in \wp(X^*) \mid \exists M \in \text{NDA } L = T(M)\}$$

Esempio.

Si consideri l'NDA: $(\{S_1, S_2, A\}, \delta, S_1, \{A\})$:



Si consideri la stringa $w = aba$.



Considerando la matrice di transizione:

δ	a	b
S_1	$\{A, S_2\}$	$\{S_1, S_2\}$
S_2	$\{A\}$	$\emptyset$
A	$\{A, S_2\}$	$\{A\}$

si può calcolare la stessa cosa nel seguente modo:

$\delta^*(\{S_1\}, aba) = \bigcup_{\underline{q \in \delta^*(\{S_1\}, ab)}} \delta(q, a)$

$\delta^*(\{S_1\}, ab) = \bigcup_{\underline{q' \in \delta^*(\{S_1\}, a)}} \delta(q', b)$

$\delta^*(\{S_1\}, a) = \bigcup_{\underline{q'' \in \delta^*(\{S_1\}, \lambda)}} \delta(q'', a)$

$\delta^*(\{S_1\}, \lambda) = \{S_1\}$

$\delta^*(\{S_1\}, a) = \delta(\{S_1\}, a) = \{A, S_2\}$

$\delta^*(\{S_1\}, ab) = \delta(A, b) \cup \delta(S_2, b) = \{A\} \cup \emptyset = \{A\}$

$\delta^*(\{S_1\}, aba) = \delta(A, a) = \{A, S_2\}$ e siccome $A \in F$:

$$\delta^*(\{S_1\}, aba) \cap F = \{A\} \neq \emptyset$$

Quindi M accetta w

Teorema. Le classi $\mathcal{L}_{FSL}$ e $\mathcal{L}_{NDL}$ sono equivalenti.

Dimostrazione.

$\underline{\mathcal{L}_{FSL} \subseteq \mathcal{L}_{NDL}}$

Si consideri $M_1 = (Q_1, \delta_1, q_1, F_1) \in \text{FSA}$

Definiamo l'NDA $M_2 = (Q_2, \delta_2, q_2, F_2)$ sullo stesso alfabeto di ingresso X, dove:

- $Q_2 = Q_1$
- $\delta_2 \colon Q_2 \times X \longrightarrow \wp(Q_2)$
  $\forall q \in Q_2 = Q_1 \ \forall x \in X \quad \delta_2(q, x) = \{\delta_1(q, x)\}$
- $q_2 = q_1$
- $F_2 = F_1$

Per induzione sulla lunghezza delle parole si dimostra che:

$T(M_2) = T(M_1)$

$\underline{\mathcal{L}_{NDL} \subseteq \mathcal{L}_{FSL}}$:
Sia $M = (Q, \delta, q_0, F) \in$ NDA
Algoritmo di costruzione dell'FSA equivalente $M' = (Q', \delta', q_0', F')$:

1. $Q' = \wp(Q)$
2. $q_0' = \{q_0\}$
3. $F' = \{p \subseteq Q \mid p \cap F \neq \emptyset\}$
4. $\delta' : Q' \times X \longrightarrow Q'$

$$\forall q' = \{q_1, q_2, \ldots, q_k\} \in Q' \quad \forall x \in X$$

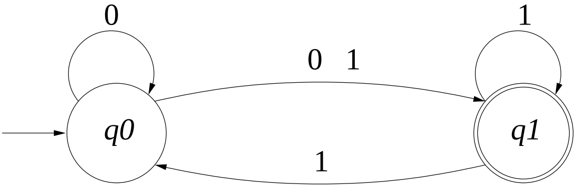
$$\delta'(q', x) = \delta'(\{q_1, q_2, \ldots, q_k\}, x) = \bigcup_{j=1}^{k} \delta(q_j, x) = \bigcup_{q \in q'} \delta(q, x)$$

Occorre ora dimostrare che $T(M) = T(M')$ per induzione sulla lunghezza della parola w

(base) $|w| = 0,\ w = \lambda$
$\lambda \in T(M') \Leftrightarrow \delta'^*(q'_0, \lambda) \in F'$
$\delta'^*(q'_0, \lambda) = \delta'^*(\{q_0\}, \lambda) = \{q_0\} = \delta(q_0, \lambda) = \delta^*(q_0, \lambda)$
ma $\{q_0\} \in F' \Rightarrow \delta^*(q_0, \lambda) \cap F \neq \emptyset$ quindi $\lambda \in T(M)$

(passo) Sia $w = va \in T(M')$ cioè $\delta'^*(q'_0, va) \cap F \neq \emptyset \qquad (1)$
Per ipotesi di induzione $\delta'^*(\{q_0\}, v) = \delta^*(\{q_0\}, v) \qquad (2)$
$\delta'^*(q'_0, va) = \delta'^*(\{q_0\}, va) = \delta'(\delta'^*(\{q_0\}, v), a) \overset{(2)}{=}$
$= \delta'(\delta^*(\{q_0\}, v), a) = \bigcup_{q' \in \delta^*(q_0, v)} \delta(q', a)$
Per definizione, $\delta^*(\{q_0\}, va) = \bigcup_{q' \in \delta^*(\{q_0\}, v)} \delta(q', a)$
Pertanto $\delta^*(\{q_0\}, va) = \delta'^*(q'_0, va)$
e quindi, tramite la (1) $\delta^*(\{q_0\}, va) \cap F \neq \emptyset$

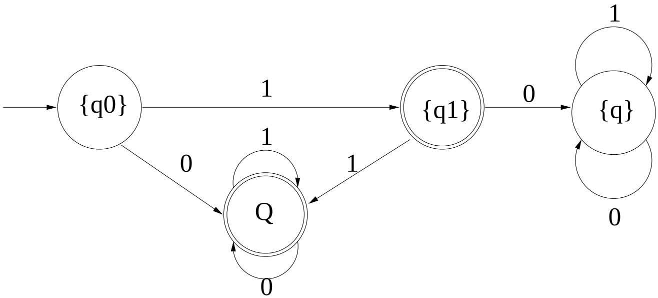
Esercizio Trasformare in FSA questo NDA:



Soluzione: Sia $M' = (Q', \delta', q'_0, F') \in$ FSA

- $Q' = \{\emptyset, \{q_0\}, \{q_1\}, Q\}$
- $\{q_0\}$ stato iniziale
- $F' = \{\{q_1\}, Q\}$
- funzione di transizione δ definita:

Esercizio



Con $\{q\}$ stato pozzo aggiunto per definire una funzione di transizione totale