



UNIVERSITÀ DEGLI STUDI  
DI GENOVA

UNIVERSITÀ DEGLI STUDI DI GENOVA

# Teoria degli Automi e Calcolabilità

*Lorenzo Vaccarecci*

# Capitolo 0: Indice

<b>1</b>	<b>Preliminari</b>	<b>2</b>
1.1	Alfabeti, stringhe, linguaggi . . . . .	2
<b>2</b>	<b>Automi a stati finiti</b>	<b>4</b>
2.1	Linguaggi regolari . . . . .	4
2.1.1	Automi a stati finiti . . . . .	4
2.2	Minimizzazione di DFA . . . . .	10
2.2.1	Algoritmo . . . . .	11
2.2.2	Espressioni regolari . . . . .	13
2.3	Proprietà dei linguaggi regolari . . . . .	13
2.3.1	Pumping Lemma . . . . .	13

# Capitolo 1: Preliminari

## 1.1 Alfabeti, stringhe, linguaggi

### Definizione 1.1.1: Alfabeto

Insieme finito **non vuoto** di oggetti detti *simboli*.

### Definizione 1.1.2: Stringa

Una stringa  $u$  su un alfabeto  $\Sigma$  è una funzione totale da  $[1, n]$  in  $\Sigma$ , per qualche  $n \in \mathbb{N}$ .  $n$  si dice *lunghezza* di  $u$  e si indica con  $|u|$ .

- $[1, n] \rightarrow$  sono le posizioni dei simboli all'interno della stringa
- Per **funzione totale** intendiamo che per ogni posizione nell'intervallo  $[1, n]$  deve avere un simbolo corrispondente

Da ora in poi:

- $\sigma \rightarrow$  simboli generici
- $u, v, w \rightarrow$  stringhe generiche
- $\Lambda$  o  $\varepsilon \rightarrow$  stringa vuota con lunghezza zero ( $|\Lambda| = 0$  o  $|\varepsilon| = 0$ )

### Definizione 1.1.3: Linguaggio $L$

E' un insieme di stringhe su  $\Sigma$ , ossia un sottoinsieme di  $\Sigma^*$  **infinito e numerabile**.

- **Infinito**: contiene un numero illimitato di elementi
- **Numerabile**: esiste una funzione iniettiva da  $\Sigma^*$  all'insieme dei numeri naturali  $\mathbb{N}$

L'insieme  $\emptyset$  è un linguaggio che non contiene alcun elemento (neanche la stringa vuota). L'insieme  $\{\varepsilon\}$  è composto solo da una stringa di lunghezza 0. *Esempio*:

- $\Sigma = \{a, b\}$
- $L = \Sigma^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, aab, \dots\}$

Un linguaggio può non contenere  $\varepsilon$ .

### Definizione 1.1.4: Operazioni su stringhe (concatenazione)

Se  $u$  e  $v$  sono stringhe di lunghezza  $n$  ed  $m$  rispettivamente, allora  $u \cdot v$  è la stringa di lunghezza  $n + m$ , definita da

$$(u \cdot v)(k) = \begin{cases} u(k) & \text{se } 1 \leq k \leq n \\ v(k - n) & \text{se } n < k \leq n + m \end{cases}$$

Questa operazione è associativa.

### Definizione 1.1.5: Operazioni su linguaggi (concatenazione)

Se  $L$  e  $L'$  sono linguaggi,  $L \cdot L' = \{u \cdot v \mid u \in L, v \in L'\}$ . Se si esegue:

$$\begin{aligned} L \cdot \{\varepsilon\} &= L & L \cdot \emptyset &= \emptyset \\ L^0 &= \{\varepsilon\} & L^{n+1} &= L \cdot L^n \quad \text{con } n \geq 0 \end{aligned}$$

- **Chiusura di Kleene**  $L^*$ :  $L^* = \cup_{n \geq 0} L^n \rightarrow$  collezione di tutte le sequenze possibili di elementi di  $L$ , inclusa la stringa vuota ( $L^0$ ). *Esempio:*

- $\Sigma = \{a, b\}$
- $L = \{a\}$
- $L^* = \{\varepsilon, a, aa, aaa, \dots, a^n\}$

- **Chiusura positiva**  $L^+$ :  $L^+ = \cup_{n > 0} L^n \rightarrow$  differisce da  $L^*$  solo per l'esclusione della stringa vuota, a meno che  $L$  stesso contenga  $\varepsilon$ . *Esempio:*

- $L^+ = \{a, aa, aaa, \dots, a^n\}$

Questa operazione è associativa.

# Capitolo 2: Automi a stati finiti

## 2.1 Linguaggi regolari

### 2.1.1 Automi a stati finiti

#### Definizione 2.1.1: Automa a stato finito deterministico (DFA)

E' una quintupla  $\mathcal{M} = \langle Q, \Sigma, \delta, q_0, F \rangle$  dove:

- $Q$  è un insieme finito di stati
- $\Sigma$  è un alfabeto di input
- $\delta : Q \times \Sigma \rightarrow Q$  è una funzione totale detta **funzione di transizione** ovvero stabilisce come l'automa si muove da uno stato all'altro in base al simbolo che legge dall'input
- $q_0 \in Q$  è lo stato iniziale
- $F \subseteq Q$  è l'insieme degli stati finali

La definizione del sistema di transizione associato a un DFA  $\mathcal{M}$  è:

- **Configurazioni:** sono rappresentate da una coppia  $\langle q, u \rangle$ :
  - $q \in Q$  è lo stato corrente in cui si trova l'automa
  - $u \in \Sigma^*$  è la parte della stringa di input che deve ancora essere letta
- **Relazione di Riduzione (o Transizione):** la relazione di riduzione, indicata con  $\rightarrow$ , definisce come l'automa si muove da una configurazione all'altra:

$$\langle q, \sigma u \rangle \rightarrow \langle q', u \rangle \text{ se } \delta(q, \sigma) = q'$$

- Significa che l'automa è nello stato  $q$  e il prossimo simbolo da leggere è  $\sigma$  (con  $u$  come resto della stringa) e  $\delta$  porta dallo stato  $q$  allo stato  $q'$  leggendo  $\sigma$ , allora l'automa si sposta nello stato  $q'$  e il simbolo  $\sigma$  viene "consumato" dall'input
- Questa relazione di riduzione è **deterministica** (per ogni configurazione, c'è al più una transizione possibile) e **terminante** (ad ogni passo viene consumato un simbolo, quindi la computazione termina sempre)
- **Configurazioni di Arresto (Halting Configurations):** Le configurazioni di arresto sono quelle in cui non ci sono più simboli da leggere, quindi nella forma  $\langle q, \varepsilon \rangle$ . Un DFA non si blocca mai prima di aver letto tutto l'input, dato che la sua funzione di transizione  $\delta$  è totale.
- **Direttive di Input/Output:** Queste definiscono come l'input viene processato e come viene determinato il risultato finale:
  - **Input** ( $f_{IN}(u)$ ): data la stringa di input  $u$ , la configurazione iniziale è  $f_{IN}(u) = \langle q_0, u \rangle$

- **Output** ( $f_{OUT}(\langle q, u \rangle)$ ): il risultato di una computazione viene estratto dalla configurazione finale  $\langle q, u \rangle$  nel modo seguente:

$$f_{OUT}(\langle q, u \rangle) = \begin{cases} \text{True} & q \in F, u = \varepsilon \Rightarrow \text{tutta la stringa è stata letta} \\ \text{False} & \text{altrimenti} \end{cases}$$

- **Linguaggio accettato:** Il linguaggio  $L(\mathcal{M})$  riconosciuto da un DFA  $\mathcal{M}$  è l'insieme di tutte le stringhe  $u$  tali per cui, partendo dalla configurazione iniziale  $\langle q_0, u \rangle$ , l'automa raggiunge una configurazione  $\langle q, \varepsilon \rangle$ . Formalmente:

$$L(\mathcal{M}) = \{u \mid \langle q_0, u \rangle \rightarrow^* \langle q, \varepsilon \rangle, \text{ per qualche } q \in F\}$$

Esiste un modo equivalente per definire il linguaggio accettato, che utilizza una funzione di transizione estesa  $\hat{\delta} : Q \times \Sigma^* \rightarrow Q$ :

- $\hat{\delta}(q, \varepsilon) = q$  (leggendo la stringa vuota, si rimane nello stato corrente)
- $\hat{\delta}(q, u\sigma) = \delta(\hat{\delta}(q, u), \sigma)$  (per leggere una stringa  $u$  seguita da un simbolo  $\sigma$ , si calcola prima lo stato raggiunto dopo aver letto  $u$ , e da quello stato si applica la funzione  $\delta$  per leggere  $\sigma$ )

Una stringa  $u$  è accettata se  $\hat{\delta}(q_0, u) \in F$ . Il linguaggio accettato (riconosciuto) da  $\mathcal{M}$  è quindi:

$$L(\mathcal{M}) = \{u \mid \hat{\delta}(q_0, u) \in F\}$$

I **linguaggi regolari** sono quelli accettati da qualche DFA.

### Esercizio

*Proviamo che  $\emptyset$ ,  $\{\varepsilon\}$  e  $\Sigma^*$  sono insiemi regolari.*

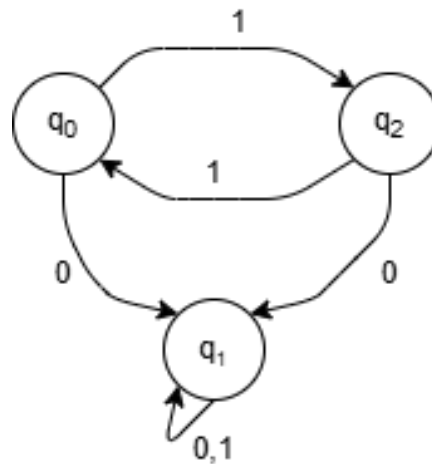
- $\emptyset$ : Per essere accettato abbiamo bisogno che  $\mathcal{M} = \langle Q, \Sigma, \delta, q_0, F = \emptyset \rangle$  perchè  $\emptyset \in F$  e perchè non viene accettato da quelle  $\mathcal{M} \mid F \neq \emptyset$
- $\{\varepsilon\}$ :  $F = \{q_0\}$  per definizione, quindi:
  - $Q = \{q_0, q_1\}$
  - $\Sigma$  qualunque ad esempio  $\{a\}$
  - $\delta$ 
    - \*  $\delta(q_0, a) = q_1$
    - \*  $\delta(q_1, a) = q_1$

Quindi  $\{\varepsilon\}$  è accettato perchè la DFA ha come solo stato finale  $q_0$  e se dessimo in input la stringa "aa" porterebbe allo stato  $q_1$  non finale e rimarrebbe bloccato lì ma  $q_1 \notin F$  quindi viene rifiutata.

- $\Sigma^*$ : Possiamo costruire una  $\mathcal{M}$  molto semplice
  - $Q = \{q_0\}$
  - $F = \{q_0\}$
  - $\delta$ 
    - \*  $\delta(q_0, \sigma) = q_0$

In questo modo qualsiasi sia l'alfabeto, l'automa consuma l'input e quando la stringa è vuota viene accettata.

Un DFA può essere rappresentato come un grafo orientato etichettato detto **grafo di transizione**



Oppure dando una matrice di transizione

	0	1
$\rightarrow q_0$	$q_1$	$q_2$
$*q_1$	$q_1$	$q_1$
$q_2$	$q_1$	$q_0$

Dove lo stato iniziale è indicato con  $\rightarrow$  e lo stato finale con  $*$ .

### Definizione 2.1.2: Automa a stato finito non deterministico (NFA)

È una quintupla  $\mathcal{M} = \langle Q, \Sigma, \delta, q_0, F \rangle$  dove  $\delta : Q \times \Sigma \rightarrow \wp(Q)$ .

$\wp(Q)$  mappa un **insieme di stati**, cioè l'insieme di tutti i sottoinsiemi possibili di  $Q$ . Ad esempio se abbiamo  $Q = \{q_0, q_1, q_2\}$ ,  $\wp(Q) = \{\{q_0\}, \{q_1\}, \{q_2\}, \{q_0, q_1\}, \dots, \{q_0, q_1, q_2\}\}$

Analogamente ai DFA, un NFA può essere rappresentato in due modi principali:

- **Grafo di transizione:** rimane concettualmente la stessa dei DFA
- **Tabella di transizione:** ogni casella della tabella (corrispondente a una coppia stato/simbolo) può contenere un insieme di stati

	0	1
$\rightarrow q_0$	$\{q_0, q_1\}$	$q_0$
$q_1$	$\emptyset$	$q_2$
$*q_2$	$q_2$	$q_2$

Il comportamento di un NFA è descritto da un sistema di transizione:

- **Configurazioni:** Sono della forma  $\langle q, u \rangle$ , dove  $q \in Q$  è lo stato corrente e  $u \in \Sigma^*$  è la stringa ancora da leggere
- **Relazione di riduzione:** È definita come:  $\langle q, \sigma u \rangle \rightarrow \langle q', u \rangle$  se  $q' \in \delta(q, \sigma)$ . **Non è deterministica** ma comunque **terminante**, poichè a ogni passo viene consumato un simbolo di input (a meno che non si blocchi)
- **Configurazioni di arresto:** Un NFA può bloccarsi in due modi:
  1. Quando ha letto tutta la stringa:  $\langle q, \epsilon \rangle$
  2. Quando si trova in una configurazione  $\langle q, \sigma u \rangle$  ma non esistono transizioni possibili per il simbolo  $\sigma$  dallo stato  $q$  (cioè  $\delta(q, \sigma) = \emptyset$ )

- **Direttive di Input/Output:** Le direttive di input/output sono simili a quelle dei DFA:

- **Input:**  $f_{IN}(u) = \langle q_0, u \rangle$
- **Output:**  $f_{OUT} = \begin{cases} \langle q, u \rangle = \text{True} & \text{se } q \in F, u = \varepsilon \\ \text{False} & \text{altrimenti} \end{cases}$

- **Linguaggio Accettato:** Il linguaggio  $L(\mathcal{M})$  accettato (o riconosciuto) da un NFA  $\mathcal{M}$  è l'insieme delle stringhe  $u$  per cui esiste almeno una computazione che, partendo dalla configurazione iniziale  $\langle q_0, u \rangle$ , raggiunge una configurazione  $\langle q, \varepsilon \rangle$  dove  $q$  è uno stato finale. Questo è un punto chiave del non determinismo: basta una computazione accettante tra tutte quelle possibili.

Un modo equivalente per definire il linguaggio accettato usa la funzione estesa  $\hat{\delta} : Q \times \Sigma^* \rightarrow \wp(Q)$ :

- $\hat{\delta}(q, \varepsilon) = q$
- $\hat{\delta}(q, u\sigma) = \bigcup_{q' \in \hat{\delta}(q, u)} \delta(q', \sigma)$

Con questa definizione, una stringa  $u$  è accettata se e solo se esiste uno stato  $q$  nell'insieme  $\hat{\delta}(q_0, u)$  che sia uno stato finale ( $q \in F$ ). In altre parole,  $L(\mathcal{M}) = \{u \mid \hat{\delta}(q_0, u) \cap F \neq \emptyset\}$

### Teorema 2.1.1: Rabin-Scott

Sia  $\mathcal{M} = \langle Q, \Sigma, \delta_N, q_0, F_N \rangle$  un NFA. Allora esiste un DFA  $\mathcal{M}_D$  tale che  $L(\mathcal{M}_D) = L(\mathcal{M})$

**Prova:** Costruiamo  $\mathcal{M}_D$  come la quintupla  $\langle \wp(Q), \Sigma, \delta_D, \{q_0\}, F_D \rangle$ . L'idea centrale della costruzione è che ogni stato del nuovo DFA  $\mathcal{M}_D$  corrisponde a un insieme di stati dell'NFA originale.

- $Q_D = \wp(Q_N)$ : questo significa che ogni stato del DFA è un sottoinsieme degli stati dell'NFA. Se l'NFA ha  $|Q_N|$  stati, il DFA risultante può avere fino a  $2^{|Q_N|}$  stati, sebbene non tutti siano necessariamente raggiungibili.
- $\delta_D$ : per ogni stato  $q_D \in \wp(Q_N)$  (che è un insieme di stati dell'NFA) e per ogni simbolo  $\sigma \in \Sigma$ , la transizione  $\delta_D(q_D, \sigma)$  porta a un nuovo stato del DFA che è l'unione di tutti gli stati raggiungibili nell'NFA da qualsiasi stato in  $q_D$  leggendo  $\sigma$ . Formalmente:

$$\delta_D(q_D, \sigma) = \bigcup_{q \in q_D} \delta_N(q, \sigma)$$

- $F_D$ : Uno stato  $q_D \subseteq Q_N$  del DFA è uno stato finale se e solo se contiene almeno uno stato finale dell'NFA. Formalmente:

$$F_D = \{q_D \subseteq Q_N \mid q_D \cap F_N \neq \emptyset\}$$



## Esempio

- **NFA Originale** ( $\mathcal{M}_N$ ):  $\mathcal{M} = (\{q_0, q_1, q_2\}, \{0, 1\}, \delta_N, q_0, q_2)$  con

$$\delta_N =$$

	0	1
$\rightarrow q_0$	$q_0$	$\{q_0, q_1\}$
$q_1$	$\emptyset$	$q_2$
$*q_2$	$q_2$	$q_2$

- **DFA Derivante** ( $\mathcal{M}_D$ ):  $\mathcal{M}_D = (Q_D, \{0, 1\}, \delta_D, q_0, F_D)$

- $Q_D = \{\emptyset, \{q_0\}, \{q_1\}, \{q_2\}, \{q_0, q_1\}, \{q_0, q_2\}, \{q_1, q_2\}, \{q_0, q_1, q_2\}\}$
- $F_D = \{\{q_2\}, \{q_0, q_2\}, \{q_1, q_2\}, \{q_0, q_1, q_2\}\}$  (tutti gli stati che contengono lo stato finale della NFA)
- $\delta_D$ :
  1. **Da**  $\{q_0\}$ :
    - \*  $\delta_D(q_0, 0) = \delta_N(q_0, 0) = q_0$
    - \*  $\delta_D(q_0, 1) = \delta_N(q_0, 1) = \{q_0, q_1\}$
  2. **Da**  $\{q_0, q_1\}$ :
    - \*  $\delta_D(\{q_0, q_1\}, 0) = \delta_N(q_0, 0) \cup \delta_N(q_1, 0) = \{q_0\} \cup \emptyset = q_0$
    - \*  $\delta_D(\{q_0, q_1\}, 1) = \delta_N(q_0, 1) \cup \delta_N(q_1, 1) = \{q_0, q_1\} \cup \{q_2\} = \{q_0, q_1, q_2\}$
  3. **Da**  $\{q_0, q_1, q_2\}$ :
    - \*  $\delta_D(\{q_0, q_1, q_2\}, 0) = \delta_N(q_0, 0) \cup \delta_N(q_1, 0) \cup \delta_N(q_2, 0) = \{q_0\} \cup \emptyset \cup \{q_2\} = \{q_0, q_2\}$
    - \*  $\delta_D(\{q_0, q_1, q_2\}, 1) = \delta_N(q_0, 1) \cup \delta_N(q_1, 1) \cup \delta_N(q_2, 1) = \{q_0, q_1\} \cup \{q_2\} \cup \{q_2\} = \{q_0, q_1, q_2\}$
  4. **Da**  $\{q_0, q_2\}$ :
    - \*  $\delta_D(\{q_0, q_2\}, 0) = \delta_N(q_0, 0) \cup \delta_N(q_2, 0) = \{q_0\} \cup \{q_2\} = \{q_0, q_2\}$
    - \*  $\delta_D(\{q_0, q_2\}, 1) = \delta_N(q_0, 1) \cup \delta_N(q_2, 1) = \{q_0, q_1\} \cup \{q_2\} = \{q_0, q_1, q_2\}$

Quindi

$$\delta_D =$$

	0	1
$\rightarrow q_0$	$q_0$	$\{q_0, q_1\}$
$\{q_0, q_1\}$	$q_0$	$\{q_0, q_1, q_2\}$
$*\{q_0, q_1, q_2\}$	$\{q_0, q_2\}$	$\{q_0, q_1, q_2\}$
$*\{q_0, q_2\}$	$\{q_0, q_2\}$	$\{q_0, q_1, q_2\}$

Abbiamo costruito la tabella includendo unicamente gli stati raggiungibili a partire da  $\{q_0\}$ . Secondo la definizione di  $\delta_D$ , non è necessario calcolare le transizioni per stati come  $\{q_1\}$ ,  $\{q_1, q_2\}$  e  $\{q_2\}$ , dato che non vengono mai raggiunti.

## Dimostrazione per induzione

### Base

- Per la definizione della funzione di transizione estesa in un DFA,  $\hat{\delta}_D(q_0, \varepsilon) = q_0$
- Per la definizione della funzione di transizione estesa in un NFA,  $\hat{\delta}_N(q_0, \varepsilon) = q_0$

Quindi, la condizione è soddisfatta per la stringa vuota:  $\hat{\delta}_D(q_0, \varepsilon) = \hat{\delta}_N(q_0, \varepsilon)$ .

### Passo (Stringa $u\sigma$ )

- Assumiamo, per ipotesi induttiva, che per una generica stringa  $u$ , valga  $\hat{\delta}_D(q_0, u) = \hat{\delta}_N(q_0, u)$ , vogliamo dimostrare che la stessa uguaglianza vale per la stringa  $u\sigma$
- Appliciamo la definizione di  $\hat{\delta}_D$  a  $u\sigma$ :  $\hat{\delta}_D(q_0, u\sigma) = \delta_D(\hat{\delta}_D(q_0, u), \sigma)$ , possiamo sostituire  $\hat{\delta}_D(q_0, u)$  con  $\hat{\delta}_N(q_0, u)$ :  $= \delta_D(\hat{\delta}_N(q_0, u), \sigma)$
- Ora applichiamo la definizione di  $\delta_D$ : è l'unione delle transizioni  $\delta_N(q', \sigma)$  per tutti gli stati  $q'$  nell'insieme  $\hat{\delta}_N(q_0, u) := \bigcup_{q' \in \hat{\delta}_N(q_0, u)} \delta_N(q', \sigma)$ . Questa è precisamente la definizione di  $\hat{\delta}_N(q_0, u\sigma)$

Quindi,  $\hat{\delta}_D(q_0, u\sigma) = \hat{\delta}_N(q_0, u\sigma)$ .

Avendo dimostrato per induzione che  $\hat{\delta}_D(q_0, u) = \hat{\delta}_N(q_0, u)$  per ogni stringa  $u \in \Sigma^*$ , possiamo concludere sull'accettazione del linguaggio:

- Una stringa  $u$  è accettata dal DFA  $\mathcal{M}_D$  se e solo se  $\hat{\delta}_D(q_0, u) \in F_D$
- Per definizione di  $F_D$ , ciò significa che  $\hat{\delta}_D(q_0, u)$  deve contenere almeno uno stato finale di  $\mathcal{M}_N$ , ovvero  $\hat{\delta}_D(q_0, u) \cap F_N \neq \emptyset$
- Poiché  $\hat{\delta}_D(q_0, u) = \hat{\delta}_N(q_0, u)$ , questo è equivalente a  $\hat{\delta}_N(q_0, u) \cap F_N \neq \emptyset$
- E quest'ultima è la condizione di accettazione di una stringa per un NFA  $\mathcal{M}_N$

Pertanto,  $u$  è accettata da  $\mathcal{M}_D$  se e solo se è accettata da  $\mathcal{M}_N$ , il che significa che  $L(\mathcal{M}_D) = L(\mathcal{M}_N)$ .

### Definizione 2.1.3: $\varepsilon$ -NFA o NFA con transizioni silenti

E' una quintupla  $M = \langle Q, \Sigma, \delta, q_0, F \rangle$  dove la funzione di transizione, che ha la forma

$$\delta : Q \times (\Sigma \cup \{\varepsilon\}) \rightarrow \wp(Q)$$

La peculiarità di questa definizione è l'inclusione di  $\varepsilon$  nell'alfabeto di input della funzione di transizione, il che consente transizioni tra stati senza consumare alcun simbolo di input (transizioni silenti). Questo permette, in molti casi, di costruire automi più semplici e leggibili

Il comportamento di un  $\varepsilon$ -NFA è descritto da un sistema di transizione con le seguenti caratteristiche:

- **Configurazioni:** Sono della forma  $\langle q, u \rangle$ , dove  $q \in Q$  è lo stato corrente e  $u \in \Sigma^*$  è la stringa ancora da leggere
- **Relazione di riduzione:** È definita in due modi:
  - $\langle q, \sigma u \rangle \rightarrow \langle q', u \rangle$  se  $q' \in \delta(q, \sigma)$  (transizione standard, consumando un simbolo)
  - $\langle q, u \rangle \rightarrow \langle q', u \rangle$  se  $q' \in \delta(q, \varepsilon)$  (transizione silente, senza consumare simboli)

La relazione di riduzione è non deterministica, anche per via della possibilità di scegliere tra leggere o non leggere un simbolo. È generalmente non terminante, ma le computazioni infinite sono dovute solo a cicli di transizioni  $\varepsilon$  dallo stesso stato, che possono essere eliminati.

- **Configurazione di arresto:** Oltre alle configurazioni della forma  $\langle q, \varepsilon \rangle$ , includono anche quelle in cui l'automa si blocca perché non ci sono transizioni possibili (né con simboli di input, né silenti) da un dato stato con il simbolo corrente
- **Linguaggio accettato:** Le direttive di input/output sono le stesse degli NFA. Una stringa  $u$  è accettata se esiste almeno una computazione che, partendo dalla configurazione iniziale  $\langle q_0, u \rangle$ , raggiunge una configurazione  $\langle q, \varepsilon \rangle$  dove  $q$  è uno stato finale. Un modo equivalente per definire il linguaggio accettato utilizza la funzione di transizione estesa  $\hat{\delta} : Q \times \Sigma^* \rightarrow \wp(Q)$ , che si basa sul concetto di  $\varepsilon$ -closure:

–  $\varepsilon$ -closure( $q$ ): È l'insieme di tutti gli stati raggiungibili da  $q$  utilizzando zero o più transizioni  $\varepsilon$ . Formalmente, include  $q$  stesso, tutti gli stati raggiungibili da  $q$  tramite  $\delta(q, \varepsilon)$ , e ricorsivamente tutti gli stati raggiungibili da questi ultimi tramite ulteriori transizioni  $\varepsilon$

– **Funzione di transizione estesa  $\hat{\delta}$ :**

$$* \hat{\delta}(q, \varepsilon) = \varepsilon\text{-closure}(q)$$

$$* \hat{\delta}(q, u\sigma) = \bigcup_{q' \in \hat{\delta}(q, u)} \varepsilon\text{-closure}(\delta(q', \sigma))$$

Una stringa  $u$  è accettata se e solo se l'insieme di stati raggiungibili dopo aver letto  $u$  (cioè  $\hat{\delta}(q_0, u)$ ) contiene almeno uno stato finale ( $\hat{\delta}(q_0, u) \cap F \neq \emptyset$ )

### Esempio

$$\mathcal{M}_N = (\{q_0, q_1, q_2\}, \{a, b\}, \delta_N, q_0, F_N)$$

$$\delta_N =$$

	a	b	$\varepsilon$
$\rightarrow q_0$	...	...	$q_1$
$q_1$	...	...	$q_2$
$*q_2$	...	...	$q_1$

Calcolo le  $\varepsilon$ -closure:

- $\varepsilon\text{-closure}(q_0) = \{q_0, q_1, q_2\}$
- $\varepsilon\text{-closure}(q_1) = \{q_1, q_2\}$
- $\varepsilon\text{-closure}(q_2) = \{q_2, q_1\}$

## 2.2 Minimizzazione di DFA

La minimizzazione di un DFA mira a trovare un automa equivalente, che accetti lo stesso linguaggio, ma con il numero minimo di stati.

- **Stati indistinguibili** ( $q \sim q'$ ): sono considerati indistinguibili (o equivalenti) se il linguaggio accettato a partire da  $q$  è identico al linguaggio accettato a partire da  $q'$ . Ciò significa che per qualsiasi stringa  $u \in \Sigma^*$ , se  $\hat{\delta}(q, u) \in F$  allora anche  $\hat{\delta}(q', u) \in F$  e viceversa. Questa relazione di indistinguibilità è una relazione di equivalenza.
- **Classe di equivalenza** ( $[q]_{\sim}$ ): è l'insieme di tutti gli elementi di  $Q$  che sono in relazione con  $q$  ( $\forall x \in Q, x \sim q$ )
- **Quoziente di  $Q$**  ( $Q/\sim$ ): è l'insieme i cui elementi sono le classi di equivalenza stesse,  $Q/\sim = \{[q]_{\sim} \mid q \in Q\}$

Due stati indistinguibili possono intuitivamente essere trasformati in un unico stato.

Quindi:

$$\mathcal{M}' = \langle Q/\sim, \Sigma, \delta', [q_0]_{\sim}, F' \rangle$$

### 2.2.1 Algoritmo

1. Suddividiamo gli stati in due classi di equivalenza: quelli finali e quelli non finali
2. A ogni passo consideriamo le classi di equivalenza ottenute al passo precedente
  - (a) Per ognuna di esse immaginiamo che tutti gli stati all'interno leggano un  $\sigma$
  - (b) Se tutti gli stati finiscono in altri stati della stessa classe di equivalenza, allora questi stati sono ancora considerati simili per quel  $\sigma$
  - (c) Altrimenti quei due stati non sono più indistinguibili. A questo punto la classe di equivalenza viene suddivisa in sottogruppi più piccoli, separando gli stati che si comportano in modo diverso
3. Torniamo al punto (2.) e ripetiamo il processo, l'algoritmo si ferma solo quando, dopo aver controllato tutti i gruppi con tutti i simboli, non c'è più nulla da dividere

```
ind = {insieme di (q,q') indistinguibili};
old_ind = null;
while(ind != old_ind) {
    old_ind = ind;
    ind = null;
    for each (q,q') in old_ind {
        ok = true;
        for simb in alfabeto {
            if(!indistinguibili(delta(q,simb),delta(q',simb))){
                ok = false;
                break;
            }
        }
        if (ok) {
            ind = ind + (q,q');
        }
    }
}
```

## Esempio

Consideriamo il seguente DFA:

	a	b
$\rightarrow q_0$	$q_0$	$q_1$
$q_1$	$q_1$	$q_2$
$q_2$	$q_2$	$q_3$
$*q_3$	$q_3$	$q_3$

1.  $\sim = \{\{q_3\}, \{q_0, q_1, q_2\}\}$

2. **Prima iterazione:**

(a) Consideriamo  $\{q_0, q_1, q_2\}$

(b) Leggiamo 'a':

i.  $\delta(q_0, a) = q_0 \in \{q_0, q_1, q_2\}$

ii.  $\delta(q_1, a) = q_1 \in \{q_0, q_1, q_2\}$

iii.  $\delta(q_2, a) = q_2 \in \{q_0, q_1, q_2\}$

(c) Leggiamo 'b':

i.  $\delta(q_0, b) = q_1 \in \{q_0, q_1, q_2\}$

ii.  $\delta(q_1, b) = q_2 \in \{q_0, q_1, q_2\}$

iii.  $\delta(q_2, b) = q_3 \notin \{q_0, q_1, q_2\}$

(d) Quindi  $\sim = \{\{q_3\}, \{q_0, q_1\}, \{q_2\}\}$

3. **Seconda iterazione:**

(a) Consideriamo  $\{q_0, q_1\}$

(b) Leggiamo 'a':

i.  $\delta(q_0, a) = q_0 \in \{q_0, q_1\}$

ii.  $\delta(q_1, a) = q_1 \in \{q_0, q_1\}$

(c) Leggiamo 'b':

i.  $\delta(q_0, b) = q_1 \in \{q_0, q_1\}$

ii.  $\delta(q_1, b) = q_2 \notin \{q_0, q_1\}$

(d) Quindi  $\sim = \{\{q_3\}, \{q_0\}, \{q_1\}, \{q_2\}\}$

Quindi il DFA era già minimo.

## 2.2.2 Espressioni regolari

### Definizione 2.2.1

Un'espressione regolare è una stringa che descrive schematicamente un insieme di stringhe. I linguaggi denotati dalle RE sono definiti induttivamente:

- $\emptyset$  è una RE che denota il linguaggio vuoto
- $\varepsilon$  è una RE che denota il linguaggio contenente solo la stringa vuota  $\{\varepsilon\}$
- Per ogni simbolo  $\sigma \in \Sigma$ ,  $\sigma$  è una RE che denota il linguaggio  $\{\sigma\}$
- Se  $r_1$  è un'espressione regolare che denota il linguaggio  $L_1$  e  $r_2$  denota  $L_2$ , allora:
  - $r_1 + r_2$  (o  $r_1 \mid r_2$ ) denota l'unione dei linguaggi  $L_1 \cup L_2$
  - $r_1 r_2$  denota la concatenazione dei linguaggi  $L_1 \cdot L_2$
- Se  $r$  è un'espressione regolare che denota  $L$ ,  $r^*$  denota la chiusura di Kleene  $L^*$

## 2.3 Proprietà dei linguaggi regolari

- **Unione:** Se  $L$  e  $L'$  sono linguaggi regolari, anche la loro unione  $(L \cup L')$  è un linguaggio regolare. Questa proprietà è "ovvia per definizione" in relazione alle espressioni regolari.
- **Concatenazione:** Se  $L$  e  $L'$  sono linguaggi regolari, anche la loro concatenazione  $(L \cdot L')$  è un linguaggio regolare. Anche questa è "ovvia per definizione".
- **Chiusura di Kleene:** Se  $L$  è un linguaggio regolare, anche la sua chiusura di Kleene  $(L^*)$  è un linguaggio regolare. Questa proprietà è anch'essa "ovvia per definizione".
- **Complementazione:** Se  $L$  è un linguaggio regolare su un alfabeto  $\Sigma$ , allora anche il suo complementare  $(\bar{L} = \Sigma^* \setminus L)$  è regolare. Questo si dimostra prendendo un DFA che riconosce  $L$  e scambiando gli stati finali con quelli non finali.
- **Intersezione:** Se  $L$  e  $L'$  sono linguaggi regolari, allora anche la loro intersezione  $(L \cap L')$  è un linguaggio regolare. La chiusura rispetto all'intersezione si deduce facilmente dalla chiusura per complementazione, dato che  $L \cap L' = \overline{\bar{L} \cup \bar{L}'}$ .

### 2.3.1 Pumping Lemma

#### Teorema 2.3.1: Pumping Lemma

Sia  $L$  un linguaggio regolare. Allora esiste una costante  $n \in \mathbb{N}$  (detta "lunghezza di pompaggio" o "numero di stati del DFA") tale che, per ogni  $z \in L$  con  $|z| \geq n$ , possiamo decomporre  $z$  come  $uvw$  in modo che:

- $|uv| \geq n$  (la porzione "pompabile"  $v$  deve trovarsi entro i primi  $n$  simboli della stringa)
- $|v| > 0$  (la porzione "pompabile" non deve essere vuota)
- per ogni  $i \geq 0$  si ha che  $uv^i w \in L$  ( $v$  può essere ripetuta  $i$  volte, inclusa l'eliminazione  $i = 0$ , e la stringa rimarrà nel linguaggio)

## Dimostrazione

- Assumiamo  $L$  regolare  $\Rightarrow$  per il Pumping Lemma, esiste una costante  $n$ .
- Si sceglie una stringa  $z = \sigma_1\sigma_2\ldots\sigma_{|z|} \in L$  la cui lunghezza  $|z| \geq n$ .
- Dato che la stringa è accettata dall'automaa, abbiamo  $q_0 \xrightarrow{\sigma_1} q_1 \xrightarrow{\sigma_2} \ldots \xrightarrow{\sigma_{|z|}} q_{|z|} = q_F \in F$
- Dato che il DFA ha solo  $n$  stati in totale, necessariamente si ha che almeno uno stato  $\bar{q}$  viene attraversato due volte nei primi  $n$  passi, ossia  $\bar{q} = q_i = q_j$  per qualche  $0 \leq i < j \leq n$

- **Decomposizione di  $z$ :**

- $u = \sigma_1 \ldots \sigma_i$
- $v = \sigma_{i+1} \ldots \sigma_j$
- $w = \sigma_{j+1} \ldots \sigma_{|z|}$

si ha  $q_0 \xrightarrow{u} \bar{q} \xrightarrow{v} \bar{q} \xrightarrow{w} q_F$ , quindi  $q_0 \xrightarrow{u} \bar{q} \xrightarrow{v^i} \bar{q} \xrightarrow{w} q_F$  perchè  $\bar{q} \xrightarrow{v} \bar{q}$

Il pumping lemma viene tipicamente utilizzato per provare che un linguaggio  $L$  **non** è regolare. A tale scopo:

- Si sceglie un numero arbitrario  $n$
- Si trova una stringa  $z \in L$ , con lunghezza  $|z| \geq n$
- Si dimostra che, per ogni modo in cui si può dividere  $z$  in tre parti  $uvw$  (dove  $|uv| \leq n$  e  $|v| > 0$ ), se si ripete la parte  $v$  (o la si elimina), la stringa che si ottiene **non** appartiene più al linguaggio

Questo contraddice la definizione di linguaggio regolare, provando così che il linguaggio non lo è.

## Esempio

Possiamo dimostrare, utilizzando il pumping lemma, che il linguaggio  $\{0^n 1^n\}$  non è regolare. Infatti, preso  $n$  arbitrario, consideriamo la stringa  $0^n 1^n$ . Decomponendo tale stringa in tre parti  $u, v, w$  tali che la lunghezza delle prime due sia  $\leq n$  e la seconda sia non vuota, si ha chiaramente  $u = 0^a$ ,  $v = 0^b$  e  $w = 0^c 1^n$  con  $a + b + c = n$ ,  $b > 0$ . Allora, per esempio per  $i = 0$ , si ha che  $uv^0w = 0^a 0^c 1^n$  non appartiene al linguaggio in quanto  $a + c < n$ .