# Lezione 10 16/11/2023

## Automa a stati finiti

Ora stiamo affrontando il problema dello string matching esatto.

**INPUT**:

Un **testo T** di lunghezza n e un **pattern P** di lunghezza m, definiti su alfabeto $\Sigma$

**OUTPUT**:

Tutte le posizioni i di T tale che T[i, i+m-1] = P (occorrenze esatte di P in T)

Risolviamo questo problemi attraverso un automa a stati finiti. Un **automa** è un modello di calcolo che riconosce un linguaggio (insieme di stringhe che godono di una proprietà). Un **automa a stati finiti** è un automa che riconosce un linguaggio regolare.

Un **Automa a Stati Finiti** è formalmente una quintupla $(Q, \Sigma, \delta, q_0, F)$:

1. Q, insieme finito di stati
2. $\Sigma$, alfabeto in input
3. $\delta : Q \times \Sigma \rightarrow Q$, funzione di transizione
   $\delta(q,\sigma) \rightarrow$ stato di arrivo a partire da q dopo la lettura di $\sigma$
4. $q_0$, stato inziale
5. F (sottoinsieme di Q), insieme degli stati accettanti

**ESEMPIO**: ASF per riconoscere le stringhe di bit che terminano con 1

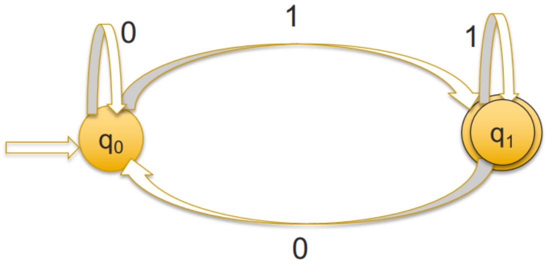
- $Q = \{q_0, q_1\}$
- $q_0$, stato iniziale
- $\Sigma = \{0,1\}$
- $F = \{q_1\}$
- Funzione di transizione $\delta : \{q_0, q_1\} \times \{0,1\} \rightarrow \{q_0, q_1\}$
  $\delta(q_0, 0) = q_0$
  $\delta(q_0, 1) = q_1$
  $\delta(q_1, 0) = q_0$
  $\delta(q_1, 1) = q_1$

## Diagramma di stato

**DIAGRAMMA DI STATO (struttura a grafo)**

- Insieme Q degli stati → vertici
- Esiste l'arco $(q_1,q_2)$ se almeno un simbolo $\sigma$ è tale per cui $\delta(q_1,\sigma) = q_2$
- L'arco $(q_1,q_2)$ viene etichettato dalla lista di simboli che permettono la transizione da $q_1$ a $q_2$
- $q_0$ è indicato tramite un arco entrante che non esce da uno stato
- Gli stati accettanti sono indicati da un doppio bordo

Esempio precedente:




| | 0 | 1 |
|---|---|---|
| $q_0$ | $q_0$ | $q_1$ |
| $q_1$ | $q_0$ | $q_1$ |

Useremo molto questa tabella, che sarà chiesta in esame al posto del diagramma.

Il **bordo** di una stringa X è più lungo prefisso proprio di X che occorre anche come suffisso di X. Esempi:

X = baaccbbaac
B(X) → baac

X = aaaccbbaac
B(X) → $\varepsilon$

X = abababa
B(X) → ababa

X = aaaaaaaa
B(X) → aaaaaaa

X = a
B(X) → $\varepsilon$

**Concatenazione di X con $\sigma$** → stringa $X\sigma$

ESEMPIO:

X = bba $\sigma$ = d

$X\sigma$ → bbad

# Automa a stati finiti per la ricerca esatta di un pattern

L'**Automa a Stati Finiti** per la ricerca esatta di un pattern P di lunghezza m definito su alfabeto $\Sigma$ è la quintupla $(Q, \Sigma, \delta, q_0, F)$ con:

- Q = {0,1,2,3,…,m}
- $\Sigma$, alfabeto di definizione di P
- $\delta : Q \times \Sigma \rightarrow Q$, funzione di transizione
- $q_0 = 0$, stato iniziale
- F = {m}, stato accettante

Inizia dallo stato 0, sostanzialmente prima del primo simbolo del pattern.

1. Costruzione dell'automa per il pattern P
   → Calcolo della funzione $\delta$ in tempo $\theta(m|\Sigma|)$

2. Uso dell'automa per riconoscere, in un testo T definito su alfabeto $\Sigma$, tutte le occorrenze esatte di P
   → Scansione del testo T in tempo $\theta(n)$

Nel primo step bisogna costruire l'automa tramite le sue funzioni di transizione.

Nel secondo passo bisogna dare il testo in mano all'automa per trovare le occorrenze esatte del pattern.

## Funzione di transizione

**Funzione di transizione $\delta$ per un pattern P di lunghezza m**

$\delta : Q \times \Sigma \rightarrow Q$ con $Q = \{0,1,2,3,\ldots,m\}$

Per ogni $(j,\sigma) \in Q \times \Sigma$
$\delta(j,\sigma) \rightarrow$ stato a cui si arriva da j attraverso $\sigma$

**DEFINIZIONE DI $\delta$**

$\delta(j,\sigma) = j+1$ se $j < m$ **and** $P[j+1] = \sigma$
$\delta(j,\sigma) = k$ se $j = m$ **or** $P[j+1] \neq \sigma$
$k = |B(P[1,j]\sigma)|$

$k \leq j$

Per definirla (anche in esame visto che la chiede) bisogna fare il preambolo iniziale, ovvero che si riferisce ad un pattern P di lunghezza m, con il numero di stati che dipende dalla lunghezza del pattern.

Chiamo j lo stato perché voglio riferirmi al fatto che gli stati sono delle posizioni sul pattern. Quindi si legge la coppia j.

La definizione significa che se sono in uno stato j (posizione nel pattern), leggo un simbolo sigma dal testo, allora l'automa passa allo stato successivo. Se non siamo alla fine (j<m) e il simbolo successivo del pattern è sigma, ovvero il simbolo che l'automa sta leggendo, allora andiamo nello stato j+1.

Altrimenti se siamo alla fine del pattern, oppure la posizione j+1 del pattern non contiene il simbolo sigma, allora andiamo nello stato k che è la lunghezza del bordo dato dal prefisso del pattern da 1 a j concatenato con sigma.

$\delta$ fornisce la transizione, attraverso il simbolo $\sigma$, dallo *stato* j al nuovo *stato* $\delta(j,\sigma)$

j < m **and** $P[j+1] = \sigma \Rightarrow \quad \delta(j,\sigma) = j+1$

Transizione dallo *stato* **j** allo *stato successivo* **j+1** attraverso il simbolo $\sigma$

P a c a c b a b b a a b a c
6

$\delta(6,b)$ = ?

P a c a c b a b b a a b a c
6 **7**

$\delta(6,b) = 7$

P a c a c b a b b a a b a c
0 **1**

$\delta(0,a) = 1$

a c a c b a c

P a c a c b a b b a a b a c
6

$\delta(6,c) = |B(\mathbf{P[1,6]c})| = |ac| = 2$

I primi 6 simboli del pattern con sigma accodato, ovvero *c* in questo caso. Prendiamo la lunghezza del bordo più lungo possibile che in questo caso è 2.

P a c a c b a b b a a b a c
0

$\delta(0,c) = |B(P[1,0]c)| = |B(\varepsilon c)| = |B(c)| = |\varepsilon| = 0$

Qui stesso ragionamento, ma il prefisso da 1 a 0 è il prefisso nullo.

a c a c b a b b a a b a c a

P a c a c b a b b a a b a c
3 13

$\delta(13,a) = |B(\mathbf{P[1,13]a})| = |aca| = 3$

1. Dallo stato **0** si arriva allo stato **0** per qualsiasi simbolo ≠ **P[1]**

2. Dallo stato **0** si arriva allo stato **1** attraverso il simbolo **P[1]**

Dallo stato **j=m** si arriva sempre a uno stato **k ≤ m**

→ dallo stato **m** si può giungere quindi di nuovo allo stato **m**

**ESEMPIO**:

P=aaaaaa (m=6)

$\delta(6,\mathbf{a}) = |B(\text{aaaaaa}\mathbf{a})| = |\text{aaaaaa}| = 6$

# Esercizio (simil esame)

P

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| a | c | a | c | b | a | c |

m = 7 $\Sigma = \{a,b,c\}$

- Q = {0,1,2,3,4,5,6,7}
- $\Sigma = \{a,b,c\}$
- $\delta : Q \times \Sigma \to Q$, funzione di transizione
- 0, stato iniziale
- 7, stato accettante

$$\delta(j,\sigma) = j+1 \qquad j \xrightarrow{\sigma} j+1$$

P

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| a | c | a | c | b | a | c |

m = 7 $\Sigma = \{a,b,c\}$

| δ(j,σ) | a | b | c |
|---|---|---|---|
| 0 | 1 | | |
| 1 | | | 2 |
| 2 | 3 | | |
| 3 | | | 4 |
| 4 | | 5 | |
| 5 | 6 | | |
| 6 | | | 7 |
| 7 | | | |

$\delta(6,c) = 7 \qquad 6 \xrightarrow{c} 7$

Le altre celle vanno riempite con le transizioni da j a k

$\delta(j,\sigma) = k \qquad j \xrightarrow{\sigma} k$

| δ(j,σ) | a | b | c |
|---|---|---|---|
| 0 | 1 | | |
| 1 | | | 2 |
| 2 | 3 | | |
| 3 | | 0 | 4 |
| 4 | 3 | 5 | |
| 5 | 6 | | |
| 6 | | | 7 |
| 7 | 3 | | |

alcune riempite a campione

P

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| a | c | a | c | b | a | c |

m = 7 $\Sigma = \{a,b,c\}$

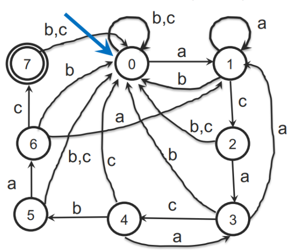
| δ(j,σ) | a | b | c |
|---|---|---|---|
| 0 | 1 | | |
| 1 | | | 2 |
| 2 | 3 | | |
| 3 | | 0 | 4 |
| 4 | | 5 | |
| 5 | 6 | | |
| 6 | | | 7 |
| 7 | | | |

$\delta(6,c) = 7 \qquad 6 \xrightarrow{c} 7$

$\delta(3,b) = 0 \qquad 3 \xrightarrow{b} 0$

$\delta(3,b) = |B(acab)| = 0$

| δ(j,σ) | a | b | c |
|---|---|---|---|
| 0 | 1 | 0 | 0 |
| 1 | 1 | 0 | 2 |
| 2 | 3 | 0 | 0 |
| 3 | 1 | 0 | 4 |
| 4 | 3 | 5 | 0 |
| 5 | 6 | 0 | 0 |
| 6 | 1 | 0 | 7 |
| 7 | 3 | 0 | 0 |



E se io aggiungo d all'alfabeto?

| δ(j,σ) | a | b | c | d |
|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 |
| 1 | 1 | 0 | 2 | 0 |
| 2 | 3 | 0 | 0 | 0 |
| 3 | 1 | 0 | 4 | 0 |
| 4 | 3 | 5 | 0 | 0 |
| 5 | 6 | 0 | 0 | 0 |
| 6 | 1 | 0 | 7 | 0 |
| 7 | 3 | 0 | 0 | 0 |

$\delta(j,\mathbf{d}) = |B(P[1,j])\mathbf{d}| = |\varepsilon| = 0$