UNIVERSITÀ
DEGLI STUDI
DI TORINO

Espressioni regolari

a.a. 2018-2019

Riferimenti bibliografici

Automi, Linguaggi e Calcolabilità
J. E. Hopcroft, R. Motwani, J. D. Ullman.

Espressioni regolari [cap. 3]

3.1 Espressioni regolari.

3.4 Proprietà algebriche per le espressioni regolari.

3.2 Automi a stati finiti ed espressioni regolari.

3.2.1 Solo enunciato del teorema 3.4 (senza dimostrazione).

3.2.3 Conversione di espressioni regolari in automi

Espressioni regolari

Un FA (NFA o DFA) è una macchina che riconosce linguaggi regolari.

Una espressione regolare è un modo dichiarativo per descrivere le parole (stringhe) che fanno parte di un linguaggio regolare (ne fornisce una descrizione algebrica).

Esempio: 01 + 10 (che descrive l'insieme $\{01\} \cup \{10\}$)

Le espressioni regolari sono usate come linguaggi di input per molti sistemi che trattano stringhe, ad esempio:

- comandi per la ricerca di stringhe in un browser o in programmi per la composizione di testi.
- strumenti per l'analisi lessicale come (Lex (Lexical analyzer generator) e Flex (Fast Lex)) di UNIX, ricevono in input delle espressioni regolari che descrivono i pattern dei lessemi e generano DFA per riconoscerli e costruire i token.

Espressioni regolari

Le espressioni regolari su un alfabeto Σ formano un'algebra.
Tutte le algebre sono definite a partire da espressioni elementari (costanti e variabili) e permettono di costruire altre espressioni applicando un insieme di operatori a espressioni elementari o a espressioni già costruite.
Per la loro definizione è importante ricordare tre operazioni che sono state definite sui linguaggi:

- *Unione*:

  $L \cup M = \{w \mid w \in L \text{ oppure } w \in M\}$

- *Concatenazione*:

  $L.M = \{w \mid w = xy,\ x \in L,\ y \in M\}$

- *Chiusura di Kleene*:

  $L = \bigcup_{i=0}^{\infty} L^i$

Dove L^i è la potenza i-esima di L: $L^0 = \{\varepsilon\}$, $L^i = L.L^{i-1}$

Espressioni regolari

Definizione

Base (espressioni elementari):

- Φ, ε sono espressioni regolari.
  $L(\Phi) = \{\ \}$ e $L(\varepsilon) = \{\varepsilon\}$.
- Se $a \in \Sigma$, allora *a* è un'espressione regolare.
  $L(a) = \{a\}$.

Induzione:

- Se R è un'espressione regolare, allora (R) è un'espressione regolare. $L((R)) = L(R)$.
- Se R e S sono espressioni regolari, allora R + S è un'espressione regolare. $L(R + S) = L(R) \cup L(S)$.
- Se R e S sono espressioni regolari, allora R.S (oppure RS) è un'espressione regolare. $L(R.S) = L(R).L(S)$.
- Se R è un'espressione regolare, allora R* è un'espressione regolare. $L(R^*) = (L(R))^*$.

Espressioni regolari

Un linguaggio regolare su un alfabeto Σ è un linguaggio che può essere espresso mediante concatenazione, unione e chiusura di Kleene a partire dai simboli di Σ.

Precedenza degli operatori

1. Chiusura (*)
2. Concatenazione (.)
3. Unione (+)

Esempio: 01 + 1 è abbreviazione per (01) + 1

01* + 1 è abbreviazione per (0(1)*) + 1

$0 + 1^*10 + 1 \neq (0 + 1)^*1(0 + 1)$

Esempio

Linguaggio	Espressione regolare
1. $L = \{w \mid w \in \{a\}^* \ \& \ \|w\| = 2k\}$	$(aa)^*$
2. $L = \{w \mid w \in \{a,b\}^* \ \& \ n_a(w) = 2k\}$	$b^*(ab^*ab^*)^*$
3. $L = \{w \mid w \in \{a,b,c\}^* \ \& \ n_a(w) = 2k\}$	?
4. $L = \{w \mid w \in \{0,1\}^*$ & 0 e 1 alternati$\}$	$(01)^* + (10)^* + 0(10)^* + 1(01)^*$ oppure $(\varepsilon + 1)(01)^*(\varepsilon + 0)$

Leggi algebriche

Linguaggi	**Espressioni regolari**
• $L \cup M – M \cup L$ L’unione è *commutativa.*	$R + S = S + R$
• $(L \cup M) \cup N = L \cup (M \cup N)$ L’unione è *associativa.*	$(R + S) + T = R + (S + T)$
• $(L.M).N = L.(M.N)$ La concatenazione è *associativa.*	$(RS)T = R(ST)$
• $\Phi \cup L = L \cup \Phi = L$ Φ è *l’elemento neutro* per l’unione.	$\Phi + R = R + \Phi = R$
• $\{\varepsilon\}.L = L.\{\varepsilon\} = L$ $\{\varepsilon\}$ è *l’elemento neutro* per la concatenazione.	$\varepsilon R = R \varepsilon = R$
• $\Phi.L = L.\Phi = \Phi$ Φ è *l’elemento zero* per la concatenazione.	$\Phi R = R \Phi = \Phi$

Leggi algebriche

Linguaggi	Espressioni regolari
• $L(M \cup N) = LM \cup LN$ La concatenazione è *distributiva a sinistra* sull'unione.	$R(S + T) = RS + RT$
• $(L \cup M).N = L.N \cup M.N$ La concatenazione è *distributiva a destra* sull'unione.	$(R + S)T = RT + ST$
• $L \cup L = L$ L'unione è *idempotente.*	$R + R = R$
• $(L^*)^* = L^*$ La chiusura è *idempotente*.	$(R^*)^* = R^*$
• $\Phi^* = \{\varepsilon\}$	$\Phi^* = \varepsilon$
• $\{\varepsilon\}^* = \{\varepsilon\}$	$\varepsilon^* = \varepsilon$
• $L.L^* = L^*.L = L^+$	$RR^* = R^*R = R^+$
• $L^* = L^+ \cup \{\varepsilon\}$	$R^* = R^+ + \varepsilon$

Esercizi

Per ognuna delle seguenti espressioni regolari scrivere almeno tre stringhe appartenenti al linguaggio denotato

- a^*bc^*
- $(a + b)^*abb$
- $(ab + ac)^*a$
- $a(da + a)^*b$
- $(ab^* + c)^*a$
- $(0 + 1\ (11)^*0)^*\ (11)^* + \varepsilon$
- $(a + b)^*aa\ (a + b)^*$
- $(00 + 11)^*((01 + 10)(00 + 11)^*\ (01 + 10)\ (00 + 11)^*)^*$

Esercizi

Determinare la cardinalità dei linguaggi denotati dalle seguenti espressioni regolari:

- $ab(c + a) + a(b + bc)$
- $\varepsilon + \Phi^*$
- $\Phi^* + a$
- a^*b

Dire se le seguenti uguaglianze sono identità e, in caso affermativo dimostrarle, in caso negativo confutarle con un esempio.

- $L - \{\varepsilon\} = L$
- $L^* - \{\varepsilon\} = L^+$
- $L - \Phi = L$
- $L \subseteq L^m, (m \geq 0)$
- $a(b + \Phi)ab = a(b + \varepsilon\varepsilon)ab$
- $(R + S)^* = (RS)^*$
- $(R + S)^* = (R + S + RS)^*$

Esercizi

Fornire espressioni regolari che denotino tutte e sole le stringhe sull'alfabeto {*a*, *b*} che soddisfano una delle seguenti condizioni:

- ogni occorrenza del carattere *a* sia seguita immediatamente da almeno due occorrenze del carattere *b*.
- ogni occorrenza del carattere *a* sia seguita immediatamente da esattamente due occorrenze del carattere *b*.
- contengano almeno un'occorrenza delle sottostrighe *aa* e *bb*, nell'ordine.
- contengano almeno un'occorrenza delle sottostrighe *aa* e *bb*, in un ordine qualunque.
- inizino con almeno due *a* e terminino con almeno due *b*.
- contengano un numero pari di *a* e un numero pari di *b*.
- contengano un numero di *a* divisibile per 3.

Espressioni regolari ed automi finiti

Abbiamo visto che ε-NFA, NFA e DFA sono equivalenti, nel senso che riconoscono gli stessi linguaggi: i linguaggi regolari.

Per dimostrare che gli automi finiti sono i riconoscitori dei linguaggi denotati dalle espressioni regolari, si deve dimostrare che:

1. Per ogni automa finito A si può costruire un'espressione regolare *R* tale che L(*R*) = L(A)
2. Per ogni espressione regolare *R* esiste un automa finito A tale che L(A) = L(*R*)

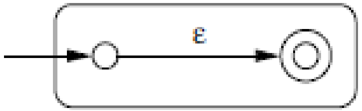
Vediamo solo la costruzione di un automa a stati finiti, un ε-NFA, a partire da un'espressione regolare, mentre non costruiamo un'espressione regolare che denoti il linguaggio riconosciuto da un automa a stati finiti.

Da espressione regolare ad automa

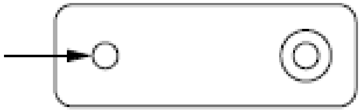
Teorema
Per ogni espressione regolare *R* possiamo costruire un ε-NFA A, tale che L(A) = L(*R*).

Costruzione per induzione strutturale.

Base: Automa per ε, Φ e *a*.



(a)

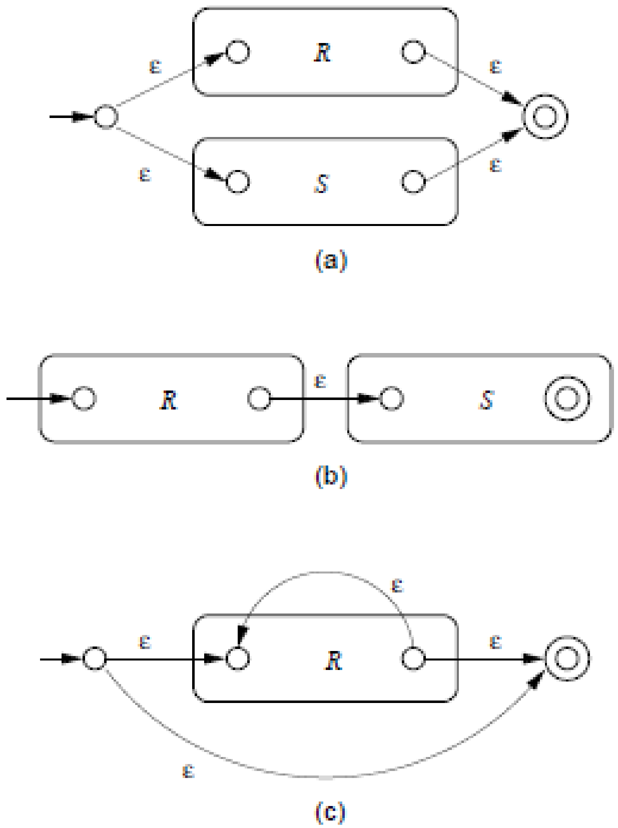


(b)

(c)

Da espressione regolare ad automa

<u>Induzione</u>: Automa per *R* + *S*, *RS* e *R*

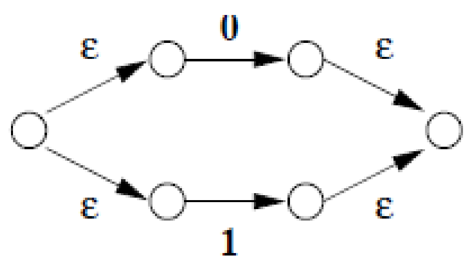


Da espressione regolare ad automa: esempio

Applichiamo l’algoritmo all’espressione regolare

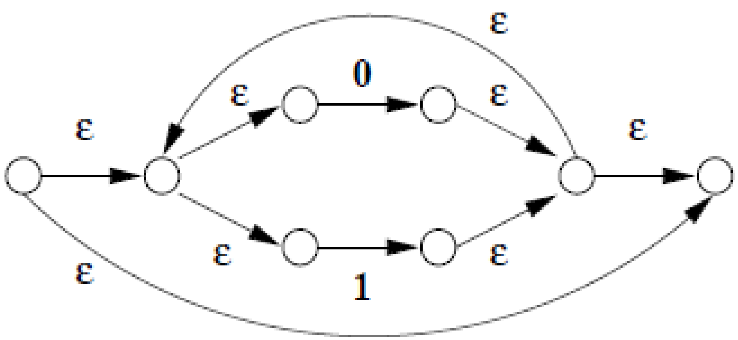
(0 + 1)*1(0 + 1)

0 + 1



(a)

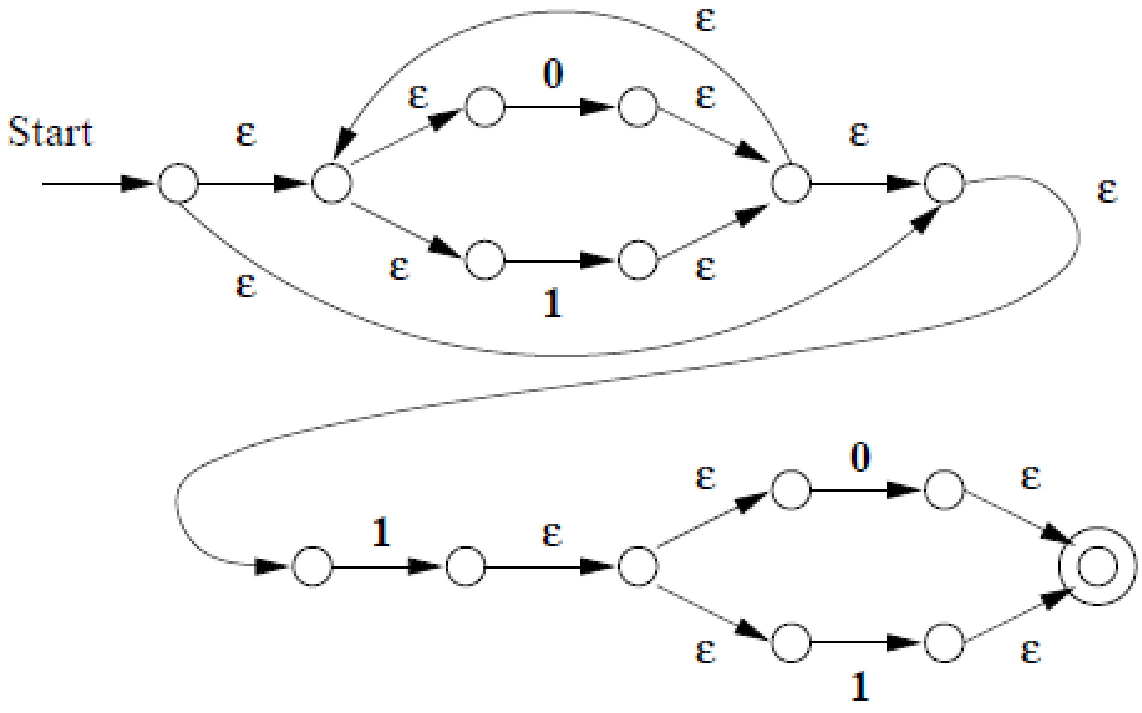
(0 + 1)*



(b)

Da espressione regolare ad automa: esempio

$(0 + 1)^*1(0 + 1)$



(c)

Esercizi

Per ognuna delle seguenti espressioni regolari, costruire un automa che riconosca il linguaggio denotato:

- a^*bc^*
- $(a + b)^*abb$
- $(ab + ac)^*a$
- $a(da + a)^*b$
- $(ab^* + c)^*a$
- $(0 + 1\,(11)^*0)^*\,(11)^* + \varepsilon$
- $(a + b)^*aa\,(a + b)^*$
- $(00 + 11)^*((01 + 10)(00 + 11)^*\,(01 + 10)\,(00 + 11)^*)^*$