UNIVERSITÀ
DEGLI STUDI
DI TORINO

Espressioni regolari e proprietà dei linguaggi regolari

a.a. 2018-2019

Espressioni regolari

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

Una espressione regolare è un modo *dichiarativo* per descrivere un linguaggio regolare (ne fornisce una descrizione algebrica).
Esempio: 01 + 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, che ricevono in input delle espressioni regolari che descrivono i pattern dei lessemi e generano automi 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

Un'espressione regolare (e.r.) r su un'alfabeto Σ è una espressione costruita partendo dai simboli Σ con l'aggiunta degli operatori . , +, *, ϕ , (,) che denota un linguaggio (insieme di stringhe).
Definiamo induttivamente le e.r.. Se r è in e.r. L(r) è il linguaggio denotato

espressione regolare	insieme denotato
Base	
$a \quad (a \in \Sigma)$	$L(a) = \{a\}$
ε	$L(\varepsilon) = \{\varepsilon\}$
ϕ	$L(\phi) = \phi$
Induttivo, siano R e S espressioni regolari:	
$R + S$	$L(R+S) = L(R) \cup L(S)$
$R.S$ (o solo RS)	$L(R.S) = L(R).L(S)$
R^*	$L(R^*) = L(R)^*$
(R)	$L((R)) = L(R)$

Operatori derivati: $R^+ = R R^*$ $\qquad R^h = \underbrace{R \ldots R}_{h}$

Espressioni regolari

I linguaggi denotati dalle e.r. si chiamano: **Linguaggi 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 $\Sigma \cup \{\varepsilon\}$.

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

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

Espressioni regolari

Esempi:

- ab + bca — denota l' insieme {ab, bca}
- ab* — denota l' insieme $\{ab^n \mid n \geq 0\}$
- $(ab)^+$ + bca — denota l' insieme $\{bca\} \cup \{(ab)^n \mid n > 0\}$
- (aa)* — denota l'insieme $\{a^{2n} \mid n \geq 0\}$
- (aa)*a — denota l'insieme $\{a^{2n+1} \mid n \geq 0\}$
- (a +b + (cc)*)* — denota l'insieme … (esercizio!)
- 1 (0 + 1)* + 0 — denota i numeri binari

Due espressioni regolari sono equivalenti se denotano lo stesso linguaggio.

Esempio

Linguaggio	Espressione regolare
1. $L = \{w \mid w \in \{a\}^* \;\&\; \lvert w \rvert = 2k\}$	$(aa)^*$
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.*	$(\mathrm{R}\ \mathrm{S})\ \mathrm{T}= R\ (S\ T)$
• $\Phi \cup L = L \cup \Phi = L$ 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
$L(M \cup N) = LM \cup LN$	$R(S + T) = RS + RT$
La concatenazione è *distributiva a sinistra* sull'unione.	
$(L \cup M).N = L.M \cup L.N$	$(R + S)T = RT + ST$
La concatenazione è *distributiva a destra* sull'unione.	
$L \cup L = L$	$R + R = R$
L'unione è *idempotente*.	
$(L^*)^* = L^*$	$(R^*)^* = R^*$
La chiusura è *idempotente*.	
$\Phi^* = \{\varepsilon\}$	$\Phi^* = \varepsilon$
$\{\varepsilon\}^* = \{\varepsilon\}$	$\varepsilon^* = \varepsilon$
$L.L^* = L^*.L = L^+$	$RR^* = R^*R = R^+$
$L^* = L^+ \cup \{\varepsilon\}$	$R^* = R^+ + \varepsilon$

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:

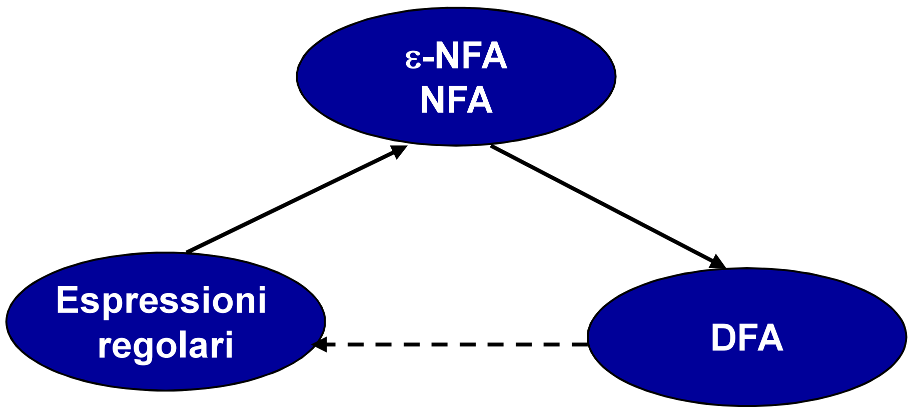
- ogni occorrenza del carattere *a* sia seguita immediatamente da almeno due occorrenze del carattere *b*.
- ogni occorrenza del carattere *a* sia seguita immediatamente da esattamente due occorrenze del carattere *b*.
- contengano almeno un'occorrenza delle sottostringhe *aa* e *bb*, nell'ordine.
- contengano almeno un'occorrenza delle sottostringhe *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.

Per mostrare che gli automi finiti sono i riconoscitori dei linguaggi denotati dalle espressioni regolari, si deve mostrare 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*)



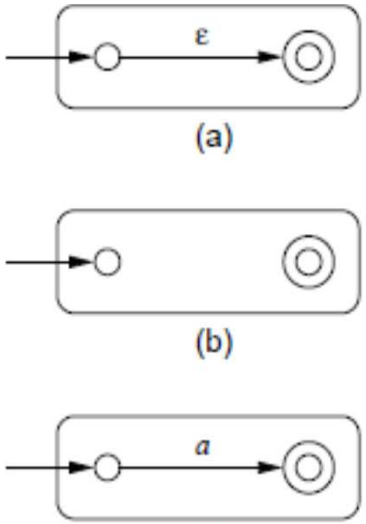
Vediamo solo la costruzione di un ε-NFA, a partire da un'espressione regolare, L'altra direzione (costruzione di un'espressione regolare a partire da un automa a stati finiti) richiede una prova un poco più complessa.

Da espressione regolare ad automa

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

Costruzione: Per induzione strutturale

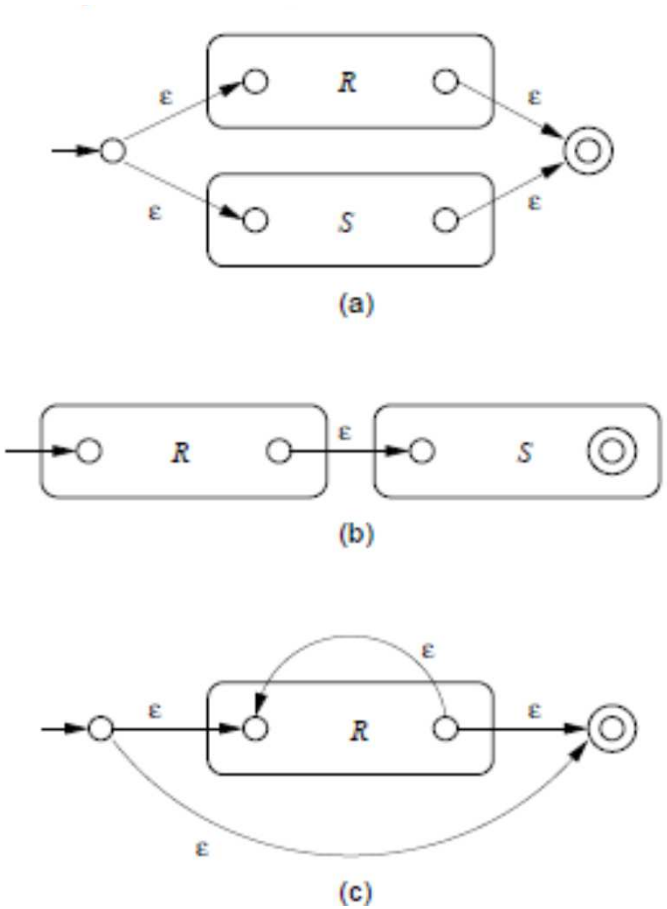
Base: Automa per ε, Φ e a.



(a) (b) (c)

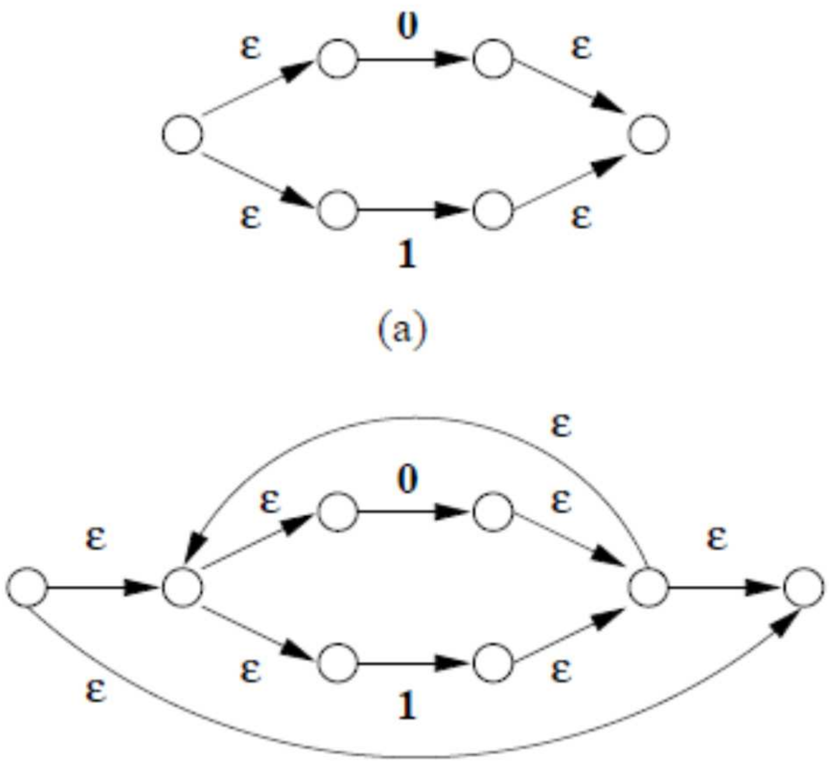
Da espressione regolare ad automa

Induzione: Automa per *R* + *S*, *RS* e *R*



Da espressione regolare ad automa: esempio

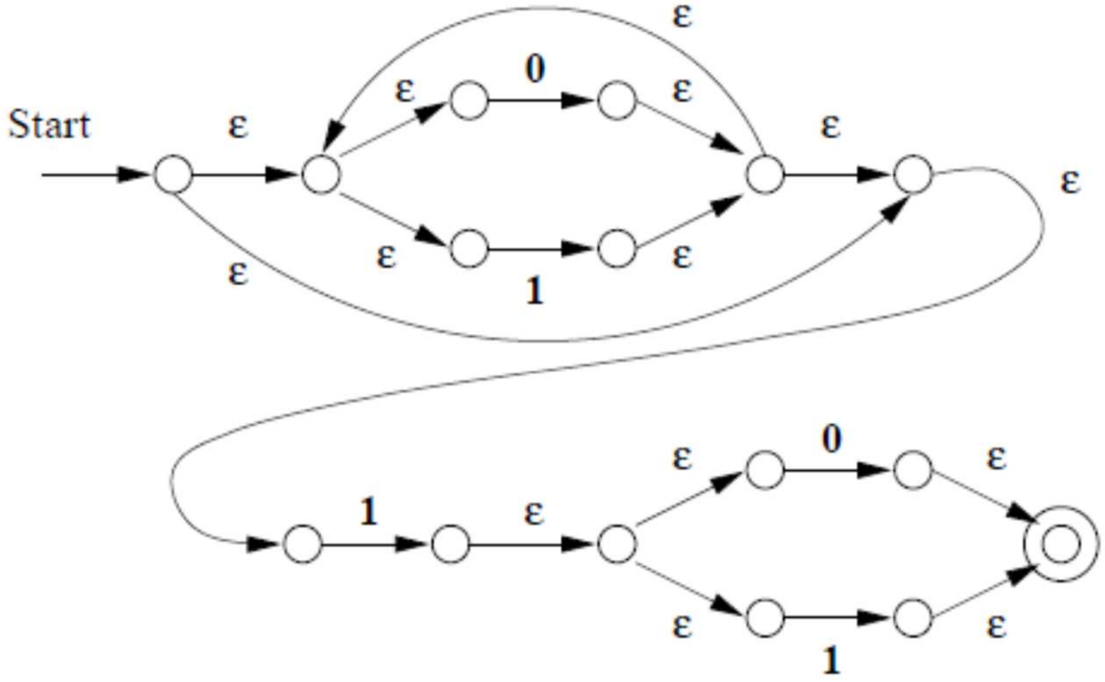
Trasformiamo (0 + 1)*1(0 + 1)



(a)

(b)

Da espressione regolare ad automa: esempio



(c)

Esercizi

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

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

Esercizi

- Fornire un'espressione regolare che denoti il linguaggio
  $L = \{w \mid w \in a^*b^* \text{ e } |w| = 2i+1,\ i \geq 0\}$.
  e costruire un automa che lo riconosca.

- Costruire automi che riconoscano rispettivamente l'intersezione, l'unione, la concatenazione, l'inversione, la differenza, il complemento e la chiusura dei linguaggi riconosciuti dai seguenti automi: