Linguaggi Formali e Traduttori

2.4 Espressioni regolari

Sommario

Automi

- approcci riconoscitivi per descrivere linguaggi regolari
- 3 varianti equivalenti: deterministici, non deterministici, con ε-transizioni

Espressioni regolari

- approccio generativo per descrivere linguaggi regolari

In questa lezione

1. Definiamo la sintassi ed il significato delle espressioni regolari
2. Enunciamo alcune leggi fondamentali delle espressioni regolari
3. Mostriamo che le espressioni regolari generano tutti e soli i linguaggi regolari

Sintassi delle espressioni regolari

Definizione

Le **espressioni regolari** su un alfabeto Σ (abbreviate **RE**, da **R**egular **E**xpressions) sono definite induttivamente come segue:

- $\emptyset$ ed ε sono espressioni regolari;
- se $\boldsymbol{a} \in \Sigma$, allora $\boldsymbol{a}$ è un'espressione regolare;
- se $\boldsymbol{E}$ ed $\boldsymbol{F}$ sono espressioni regolari, allora $\boldsymbol{E} + \boldsymbol{F}$ ed $\boldsymbol{EF}$ sono espressioni regolari;
- se $\boldsymbol{E}$ è un'espressione regolare, allora $\boldsymbol{E}^*$ è un'espressione regolare.

Convenzioni

- assumiamo la **precedenza** degli operatori $+ <$ concatenazione $<$ *
- usiamo le **parentesi** per disambiguare la struttura di un'espressione regolare

Esempi

- $\boldsymbol{ab + c = (ab) + c \neq a(b + c)}$
- $\boldsymbol{01^* = 0(1^*) \neq (01)^*}$
- $\boldsymbol{0 + 11^* = 0 + (1(1^*))}$

Significato di un'espressione regolare

Se E è un'espressione regolare, il **linguaggio generato** da E, denotato da $L(E)$, è definito per induzione sulla struttura di E come segue:

$$
\begin{array}{rcll}
L(\emptyset) & = & \emptyset & \text{linguaggio vuoto} \\
L(\varepsilon) & = & \{\varepsilon\} & \text{stringa vuota} \\
L(a) & = & \{a\} & \text{simbolo dell'alfabeto} \\
L(E+F) & = & L(E) \cup L(F) & \text{unione} \\
L(EF) & = & L(E)L(F) & \text{concatenazione} \\
L(E^*) & = & L(E)^* & \text{chiusura di Kleene}
\end{array}
$$

Diciamo che E ed F sono **equivalenti**, notazione $E = F$, se $L(E) = L(F)$.

Esercizio

Calcolare il linguaggio generato dalle espressioni regolari $(\mathbf{a} + \mathbf{b})^*$ e $(\mathbf{ab})^*$.

$\{ab\}^* \quad \{(ab)^n \mid n \in \mathbb{N}\}$

Proprietà delle espressioni regolari

Unione

- commutatività: $E + F = F + E$
- associatività: $E + (F + G) = (E + F) + G$
- idempotenza: $E + E = E$
- identità: $E + \emptyset = \emptyset + E = E$

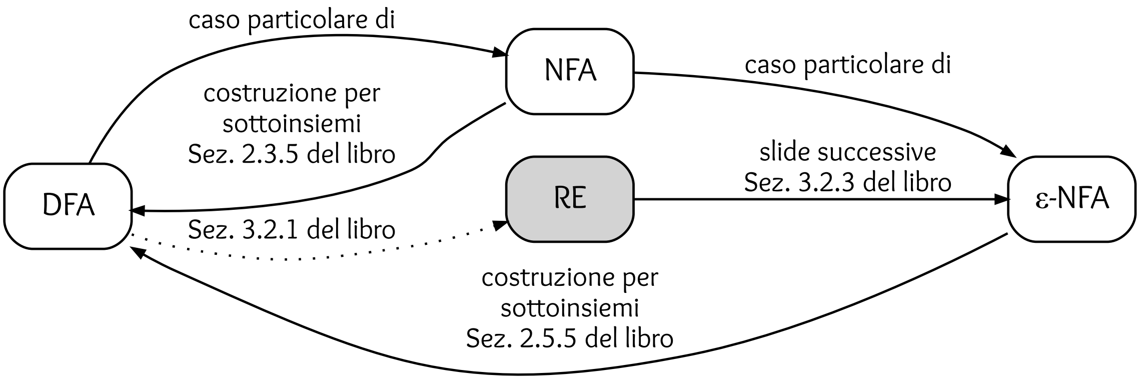
Concatenazione

- associatività: $E(FG) = (EF)G$
- identità: $E\varepsilon = \varepsilon E = E$
- assorbimento: $E\emptyset = \emptyset E = \emptyset$
- distributività sinistra della concatenazione sull'unione: $E(F + G) = EF + EG$
- distributività destra della concatenazione sull'unione: $(E + F)G = EG + FG$

Chiusura di Kleene

- idempotenza: $(E^*)^* = E^*$
- casi banali: $\varepsilon^* = \emptyset^* = \varepsilon$

Espressioni e linguaggi regolari

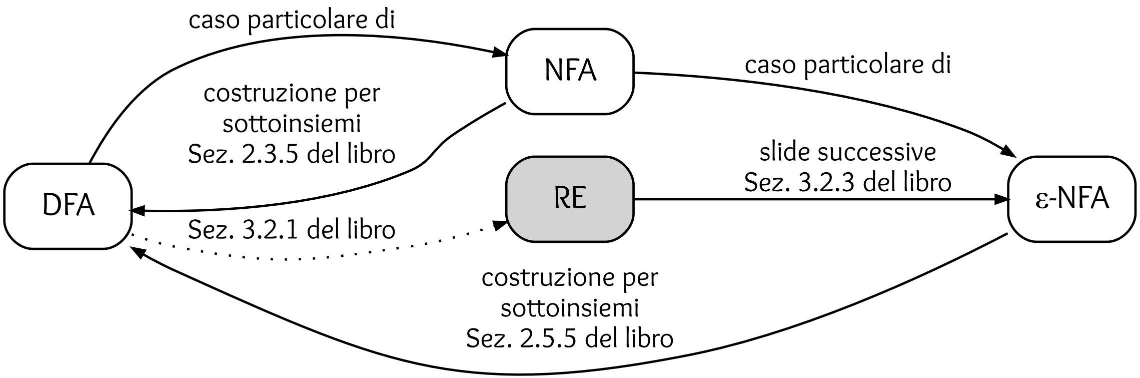


Conseguenza

DFA, NFA, **ε**-NFA ed espressioni regolari sono approcci diversi ma **equivalenti** di definire (riconoscere, generare) linguaggi regolari

$(abb+b)^*$

Espressioni e linguaggi regolari



Conseguenza

DFA, NFA, ε-NFA ed espressioni regolari sono approcci diversi ma **equivalenti** di definire (riconoscere, generare) linguaggi regolari

Teorema

Per ogni DFA A, esiste un'espressione regolare E tale che $L(A) = L(E)$.

Dimostrazione

Si veda la Sez. 3.2.1 del libro (lettura facoltativa)

Espressione regolare $\longrightarrow$ ε-NFA (1/4)

Teorema

Data un'espressione regolare E, esiste un ε-NFA A tale che $L(A) = L(E)$.

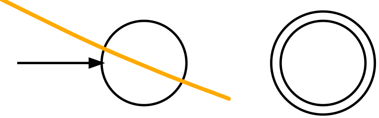
Dimostrazione

dim

Costruiamo A per induzione sulla struttura di E e per casi sulla sua forma, facendo in modo che l'ε-NFA ottenuto abbia sempre **esattamente uno stato finale** (quello più a destra nei diagrammi che seguono).

Espressione regolare $\longrightarrow$ ε-NFA (2/4)

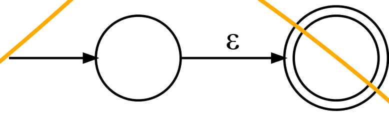
Caso $\emptyset$ (linguaggio vuoto)



Espressione regolare ⟶ ε-NFA (2/4)

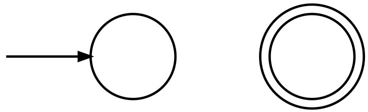
Caso $\emptyset$ (linguaggio vuoto)

Caso ε (linguaggio che contiene la sola stringa vuota)

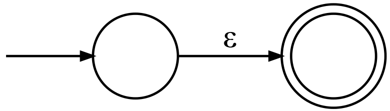


Espressione regolare $\longrightarrow$ ε-NFA (2/4)

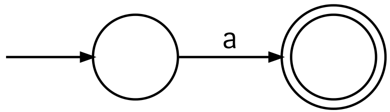
Caso $\emptyset$ (linguaggio vuoto)



Caso ε (linguaggio che contiene la sola stringa vuota)

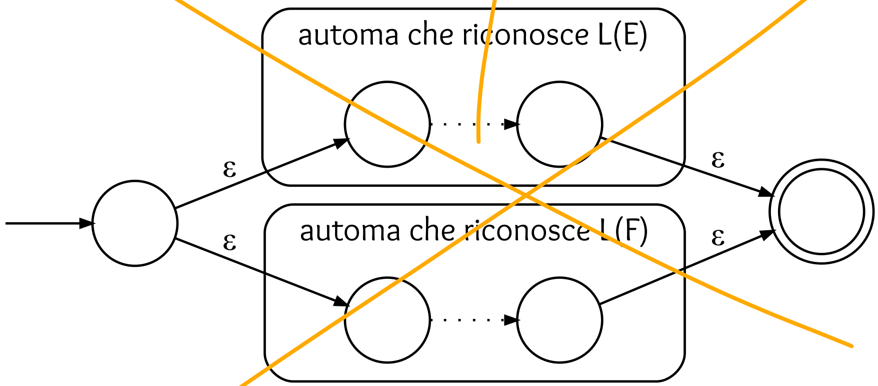


Caso $\boldsymbol{a}$ (linguaggio che contiene solo $\boldsymbol{a}$)



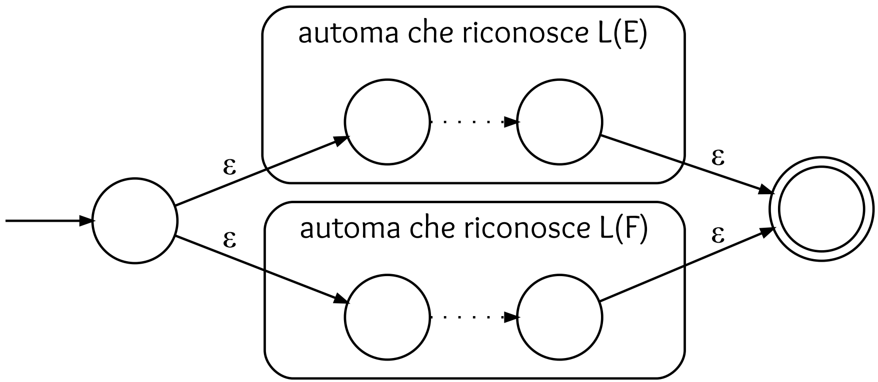
Espressione regolare ⟶ ε-NFA (3/4)

Caso $E + F$ (unione)

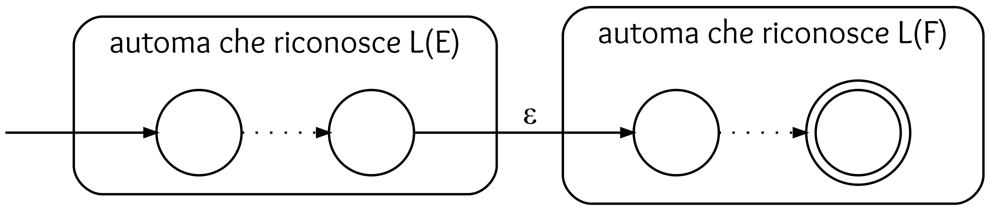


Espressione regolare ⟶ ε-NFA (3/4)

Caso $E + F$ (unione)

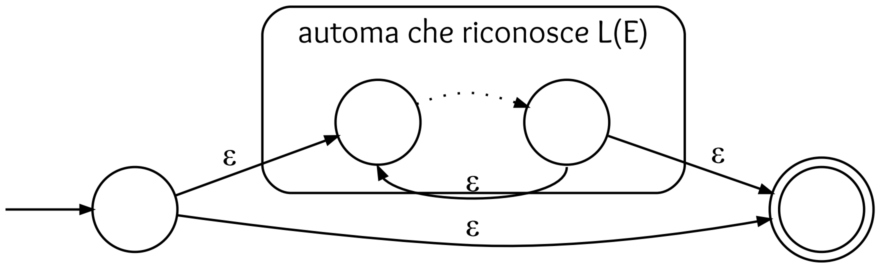


Caso EF (concatenazione)



Espressione regolare ⟶ ε-NFA (4/4)

Caso E^* (chiusura di Kleene)



Esempio: sequenze di a seguite da sequenze di b

$$
\begin{aligned}
L(\mathtt{a}^*\mathtt{b}^*) &= L(\mathtt{a}^*)L(\mathtt{b}^*) \\
&= L(\mathtt{a})^*L(\mathtt{b})^* \\
&= \{\mathtt{a}\}^*\{\mathtt{b}\}^* \\
&= \{\varepsilon, \mathtt{a}, \mathtt{aa}, \mathtt{aaa}, \ldots\}\{\varepsilon, \mathtt{b}, \mathtt{bb}, \mathtt{bbb}, \ldots\} \\
&= \{\varepsilon, \mathtt{a}, \mathtt{b}, \mathtt{aa}, \mathtt{ab}, \mathtt{bb}, \mathtt{aaa}, \mathtt{aab}, \mathtt{abb}, \mathtt{bbb}, \ldots\}
\end{aligned}
$$

Esempio: 0 oppure sequenze non vuote di 1

easy

$$
\begin{aligned}
L(0+11^*) &= L(0)\cup L(11^*)\\
&= \{0\}\cup L(1)L(1^*)\\
&= \{0\}\cup \{1\}L(1)^*\\
&= \{0\}\cup \{1\}\{1\}^*\\
&= \{0\}\cup \{1\}\{\varepsilon,1,11,111,\ldots\}\\
&= \{0\}\cup \{1,11,111,1111,\ldots\}
\end{aligned}
$$

Esempio: ogni a è seguita da bb

$$
\begin{aligned}
L((\mathtt{abb}+\mathtt{b})^*) &= L(\mathtt{abb}+\mathtt{b})^* \\
&= (L(\mathtt{abb}) \cup L(\mathtt{b}))^* \\
&= (L(\mathtt{a})L(\mathtt{b})L(\mathtt{b}) \cup L(\mathtt{b}))^* \\
&= (\{\mathtt{a}\}\{\mathtt{b}\}\{\mathtt{b}\} \cup \{\mathtt{b}\})^* \\
&= (\{\mathtt{abb}\} \cup \{\mathtt{b}\})^* \\
&= \{\mathtt{abb}, \mathtt{b}\}^*
\end{aligned}
$$

peasy

Esempio: esiste a seguita da bb

$(a+b)^* abb (a+b)^*$

$$
\begin{aligned}
L((\mathtt{a}+\mathtt{b})^*\mathtt{abb}(\mathtt{a}+\mathtt{b})^*) &= L((\mathtt{a}+\mathtt{b})^*)L(\mathtt{abb})L((\mathtt{a}+\mathtt{b})^*) \\
&= L(\mathtt{a}+\mathtt{b})^*L(\mathtt{a})L(\mathtt{b})L(\mathtt{b})L(\mathtt{a}+\mathtt{b})^* \\
&= (L(\mathtt{a})\cup L(\mathtt{b}))^*L(\mathtt{a})L(\mathtt{b})(L(\mathtt{a})\cup L(\mathtt{b}))^* \\
&= (\{\mathtt{a}\}\cup\{\mathtt{b}\})^*\{\mathtt{a}\}\{\mathtt{b}\}\{\mathtt{b}\}(\{\mathtt{a}\}\cup\{\mathtt{b}\})^* \\
&= \{\mathtt{a},\mathtt{b}\}^*\{\mathtt{abb}\}\{\mathtt{a},\mathtt{b}\}^* \\
&= \{\varepsilon,\mathtt{a},\mathtt{b},\mathtt{aa},\mathtt{ab},\mathtt{ba},\mathtt{bb},\ldots\}\{\mathtt{abb}\}\{\varepsilon,\mathtt{a},\mathtt{b},\mathtt{aa},\mathtt{ab},\mathtt{ba},\mathtt{bb},\ldots\}
\end{aligned}
$$

lemma

slide 15

Esercizi sulla definizione di espressioni regolari

Definire espressioni regolari che generino i seguenti linguaggi:

1. stringhe di a, b e c che iniziano con due a e finiscono con due b
2. stringhe di `0` e 1 la cui lunghezza è un multiplo di 3
3. stringhe di `0` e 1 con un numero pari di `0`
4. stringhe di a, b e c che **non contengono** la sottostringa ab
5. costanti numeriche binarie pari senza `0` inutili a sinistra (es. `0`, 1`0`, ma non `010` o 11)
6. costanti numeriche decimali con virgola facoltativa (es. 42, `.5`, 12`.`3, 12`.` ma non `.`)

Esercizi sulla conversione di espressione regolari

Convertire le seguenti espressioni regolari in ε-NFA e gli automi ottenuti in DFA:

1. $(a+b)^*$
2. $(ab)^*$
3. a^*b^*
4. $a^* + b^*$