Espressioni regolari

Leonardo Bizzoni

October 21, 2023

Le espressioni regolari vengono usate per **denotare** un linguaggio regolare.

La definizione di espressione regolare viene definita ricorsivamente:

- $\epsilon, \emptyset$ sono espressioni regolari che denotano:
  - $L(\epsilon) = \{\epsilon\}$
  - $L(\emptyset) = \emptyset$
- $a \in \Sigma$ allora $\mathbf{a}$ è un'espressione regolare che denota $L(\mathbf{a}) = \{a\}$
- una variabile $L \subseteq \Sigma^*$ che rappresenta un linguaggio è un'espressione regolare
- Se E è un'espressione regolare, allora anche (E) è un'espressione regolare che denota lo stesso linguaggio
- Se E, F sono espressioni regolari, allora $E+F$ è un'espressione regolare che denota $L(E+F) = L(E) \cup L(F)$
- Se E, F sono espressioni regolari, allora $E \cdot F$ è un'espressione regolare che denota $L(EF) = L(E) \cdot L(F)$
- Se E è un'espressione regolare, allora E^* è un'espressione regolare che denota $L(E^*) = (L(E))^*$

1 Precedenza degli operatori

1. Chiusura di Kleene
2. Concatenazione $\cdot$ che:
   - è associativa $E(FG) = (EF)G$

- non è commutativa $EG \neq GE$

3. Unione $+$ che:
   - è associativa $E + (F + G) = (E + F) + G$
   - è commutativa $E + F = F + E$

2 Proprietà

- $\emptyset$ è annichilatore per la concatenazione $\emptyset L = L\emptyset = \emptyset$.
- Distributività sinistra della concatenazione sull'unione:

$$L(M + N) = LM + LN$$

- Distributività destra della concatenazione sull'unione:

$$(M + N)L = ML + NL$$

3 Esempi

$\Sigma = \{0, 1\}$ il linguaggio denotato dalla regex $\mathbf{01}$ è $L(\mathbf{01}) = L(\mathbf{0}) \cdot L(\mathbf{1}) = \{0\} \cdot \{1\} = \{01\}$.

$\Sigma = \{0, 1\}$ scrivere una regex che denoti le stringhe in $(01)^*$ ovvero il linguaggio formato da stringhe di linghezza arbitraria di 0 e 1 alternati.

$L((\mathbf{01})^*) = (L(\mathbf{0}) \cdot L(\mathbf{1}))^* = \{01\}^*$ denota tutti i linguaggi che:

- sono di linghezza pari
- iniziano con 0
- finiscono con 1

$L((\mathbf{10})^*) = (L(\mathbf{1}) \cdot L(\mathbf{0}))^* = \{10\}^*$ denota tutti i linguaggi che:

- sono di linghezza pari
- iniziano con 1
- finiscono con 0

$L((\mathbf{10})^* + (\mathbf{01})^*)$ è l'unione delle 2 regex precendeti.
$L(\mathbf{0}(\mathbf{10})^*) = L(\mathbf{0}) \cdot L((\mathbf{10})^*) = \{0\} \cdot \{10\}^*$ denota tutti i linguaggi che:

- sono di linghezza dispari
- iniziano e finiscono con 0

$L(\mathbf{1}(\mathbf{01})^*) = L(\mathbf{1}) \cdot L((\mathbf{01})^*) = \{1\} \cdot \{01\}^*$ denota tutti i linguaggi che:

- sono di linghezza dispari
- iniziano e finiscono con 1

L'espressione regolare che denota il linguaggio richiesto è quindi l'unione di tutte queste regex: $(\mathbf{01})^* + (\mathbf{10})^* + \mathbf{0}(\mathbf{10})^* + \mathbf{1}(\mathbf{01})^* = (\epsilon + \mathbf{1})(\mathbf{01})^* + (\epsilon + \mathbf{0})(\mathbf{10})^*$.

Regex $= (\mathbf{0} + \mathbf{1})^*\mathbf{0}^*(\mathbf{01})^*$ denota la stringa 001? Si perchè:

- da $(\mathbf{0} + \mathbf{1})^*$ prendo ϵ
- da $\mathbf{0}^*$ prendo 0
- da $\mathbf{01}^*$ prendo 01

Regex $= (\mathbf{0} + \mathbf{1})^*\mathbf{0}^*(\mathbf{01})^*$ denota la stringa 1001? Si perchè:

- da $(\mathbf{0} + \mathbf{1})^*$ prendo 1
- da $\mathbf{0}^*$ prendo 0
- da $\mathbf{01}^*$ prendo 01

Regex $= (\mathbf{0} + \mathbf{1})^*\mathbf{0}^*(\mathbf{01})^*$ denota la stringa 0101? Si perchè:

- da $(\mathbf{0} + \mathbf{1})^*$ prendo ϵ
- da $\mathbf{0}^*$ prendo ϵ
- da $\mathbf{01}^*$ prendo per 2 volte 01

Regex $= (\mathbf{0} + \mathbf{1})^*\mathbf{0}^*(\mathbf{01})^*$ denota la stringa 0? Si perchè:

- da $(\mathbf{0} + \mathbf{1})^*$ prendo 0
- da $\mathbf{0}^*$ prendo ϵ

- da $\mathbf{01}^*$ prendo ϵ

Regex = $(\mathbf{0}+\mathbf{1})^*\mathbf{0}^*(\mathbf{01})^*$ denota la stringa 10? Si perchè:

- da $(\mathbf{0}+\mathbf{1})^*$ prendo 10
- da $\mathbf{0}^*$ prendo ϵ
- da $\mathbf{01}^*$ prendo ϵ

Qual'è la differenza tra le regex $(\mathbf{0}+\mathbf{1})^*$ e $(\mathbf{0}^*+\mathbf{1}^*)$?

$L((\mathbf{0}+\mathbf{1})^*) = \{0,1\}^*$ denota tutte le stringhe binarie.

$L((\mathbf{0}^*+\mathbf{1}^*)) = L(\mathbf{0}^*) \cup L(\mathbf{1}^*) = \{0\}^* \cup \{1\}^*$ denota le sequenze, possibilmente nulle, di 0 seguiti da 1.