STRUMENTI FORMALI PER LA BIOINFORMATICA

Espressioni Regolari

Introdurremo e utilizzeremo vari formalismi per la rappresentazione di linguaggi.

Introdurremo e utilizzeremo vari formalismi per la rappresentazione di linguaggi.

Un primo strumento, di natura algebrica, è costituito dalle *espressioni regolari*.

Introdurremo e utilizzeremo vari formalismi per la rappresentazione di linguaggi.

Un primo strumento, di natura algebrica, è costituito dalle *espressioni regolari*.

Esso consente di descrivere tutti i linguaggi appartenenti a un'importante classe di linguaggi, i *linguaggi regolari*.

Introdurremo e utilizzeremo vari formalismi per la rappresentazione di linguaggi.

Un primo strumento, di natura algebrica, è costituito dalle *espressioni regolari*.

Esso consente di descrivere tutti i linguaggi appartenenti a un'importante classe di linguaggi, i *linguaggi regolari*.

Le espressioni regolari offrono notevoli vantaggi di scrittura, e possono essere manipolate in base ad alcune loro proprietà algebriche senza modificare i linguaggi che rappresentano.

Le espressioni regolari sono un utile strumento nel progetto dei compilatori per i linguaggi di programmazione. Gli oggetti elementari in un linguaggio di programmazione, chiamati *token*, come i nomi delle variabili e le costanti, possono essere descritti con espressioni regolari.

Le espressioni regolari sono un utile strumento nel progetto dei compilatori per i linguaggi di programmazione. Gli oggetti elementari in un linguaggio di programmazione, chiamati *token*, come i nomi delle variabili e le costanti, possono essere descritti con espressioni regolari.

Una volta che la sintassi di un linguaggio di programmazione è stata descritta usando espressioni regolari per i suoi token, dei sistemi automatici possono generare l'analizzatore lessicale, la parte di un compilatore che elabora inizialmente il programma input.

Definizione ricorsiva delle espressioni regolari su un alfabeto Σ:

Definizione ricorsiva delle espressioni regolari su un alfabeto Σ:

PASSO BASE:

Per ogni $a \in \Sigma$, a è un'espressione regolare;
ϵ è un'espressione regolare;
$\emptyset$ è un'espressione regolare.

Definizione ricorsiva delle espressioni regolari su un alfabeto Σ:

PASSO BASE:
Per ogni $a \in \Sigma$, a è un'espressione regolare;
ϵ è un'espressione regolare;
$\emptyset$ è un'espressione regolare.

PASSO RICORSIVO: Se E_1, E_2 sono espressioni regolari, allora
(E_1) è un'espressione regolare;
$(E_1 + E_2)$ è un'espressione regolare;
$(E_1 E_2)$ è un'espressione regolare;
(E_1^*) è un'espressione regolare.

Nota. Nelle espressioni regolari, alcuni autori usano $+$ per denotare l'operatore $\cup$.

Un'espressione regolare E è appunto un'espressione, non un linguaggio.
Quando vogliamo riferirci al linguaggio denotato da E, useremo la notazione $L(E)$.
È consuetudine fare riferimento a E quando effettivamente si intende $L(E)$.

Definizione ricorsiva dei linguaggi rappresentati dalle espressioni regolari su un alfabeto Σ.

Data un'espressione regolare E, indicheremo con $L(E)$ il linguaggio che essa rappresenta, definito come segue

Definizione ricorsiva dei linguaggi rappresentati dalle espressioni regolari su un alfabeto Σ.

Data un'espressione regolare E, indicheremo con $L(E)$ il linguaggio che essa rappresenta, definito come segue

PASSO BASE:

Per ogni $a \in \Sigma$, $L(a) = \{a\}$;
$L(\epsilon) = \{\epsilon\}$;
$L(\emptyset) = \emptyset$.

Definizione ricorsiva dei linguaggi rappresentati dalle espressioni regolari su un alfabeto Σ.

Data un'espressione regolare E, indicheremo con $L(E)$ il linguaggio che essa rappresenta, definito come segue

PASSO BASE:

Per ogni $a \in \Sigma$, $L(a) = \{a\}$;
$L(\epsilon) = \{\epsilon\}$;
$L(\emptyset) = \emptyset$.

PASSO RICORSIVO: Se E_1, E_2 sono espressioni regolari, allora

$L((E_1)) = L(E_1)$;
$L(E_1 + E_2) = L(E_1) \cup L(E_2)$;
$L(E_1 E_2) = L(E_1)L(E_2)$;
$L(E_1^*) = L(E_1)^*$.

Regole di precedenza degli operatori

Operatore	precedenza
$*$	1
$\cdot$	2
$+$	3

Regole di precedenza degli operatori

Operatore	precedenza
$*$	1
$\cdot$	2
$+$	3

L'espressione regolare $ab^* + b$ corrisponde a $(a(b)^*) + b$ ed è diversa da $(ab)^* + b$.

Regole di precedenza degli operatori

Operatore	precedenza
$*$	1
$\cdot$	2
$+$	3

L'espressione regolare $ab^* + b$ corrisponde a $(a(b)^*) + b$ ed è diversa da $(ab)^* + b$.

L'espressione regolare $ab^* + b$ è anche diversa da $a(b^* + b)$.

Esempi

L'espressione regolare $(a + (b(cd)))$ definita sull'alfabeto $\Sigma = \{a, b, c, d\}$ può essere sostituita da $a + bcd$.

L'espressione regolare $(a + b)^* a$ rappresenta il linguaggio

L'espressione regolare $(a+b)^*a$ rappresenta il linguaggio

$$\begin{aligned}
L((a+b)^*a) &= L((a+b)^*)L(a) \\
&= (L(a+b))^*L(a) \\
&= (L(a)\cup L(b))^*L(a) \\
&= (\{a\}\cup\{b\})^*\{a\} \\
&= \{a,b\}^*\{a\} \\
&= \{w \mid w \in \{a,b\}^+, w \text{ termina con } a\}
\end{aligned}$$

Esempi

Negli esempi seguenti, assumiamo che l'alfabeto Σ sia $\{0, 1\}$.

Negli esempi seguenti, assumiamo che l'alfabeto Σ sia $\{0, 1\}$.

- $L(0^*10^*) = \{w \mid w \text{ contiene un solo } 1\}$

Esempi

Negli esempi seguenti, assumiamo che l'alfabeto Σ sia $\{0, 1\}$.

- $L(0^*10^*) = \{w \mid w \text{ contiene un solo } 1\}$
- $L(\Sigma^*1\Sigma^*) = \{w \mid w \text{ ha almeno un } 1\}$

Negli esempi seguenti, assumiamo che l'alfabeto Σ sia $\{0, 1\}$.

- $L(0^*10^*) = \{w \mid w \text{ contiene un solo } 1\}$
- $L(\Sigma^*1\Sigma^*) = \{w \mid w \text{ ha almeno un } 1\}$
- $L(\Sigma^*001\Sigma^*) = \{w \mid w$ contiene la stringa 001 come sottostringa$\}$

Esempi

Negli esempi seguenti, assumiamo che l'alfabeto Σ sia $\{0, 1\}$.

- $L(0^*10^*) = \{w \mid w \text{ contiene un solo } 1\}$
- $L(\Sigma^*1\Sigma^*) = \{w \mid w \text{ ha almeno un } 1\}$
- $L(\Sigma^*001\Sigma^*) = \{w \mid w$ contiene la stringa 001 come sottostringa$\}$
- $L(1^*(011^*)^*) = \{w \mid \text{ogni } 0 \text{ in } w \text{ è seguito da almeno un } 1\}$

Negli esempi seguenti, assumiamo che l'alfabeto Σ sia $\{0, 1\}$.

- $L(0^*10^*) = \{w \mid w \text{ contiene un solo } 1\}$
- $L(\Sigma^*1\Sigma^*) = \{w \mid w \text{ ha almeno un } 1\}$
- $L(\Sigma^*001\Sigma^*) = \{w \mid w \text{ contiene la stringa } 001 \text{ come sottostringa}\}$
- $L(1^*(011^*)^*) = \{w \mid \text{ogni } 0 \text{ in } w \text{ è seguito da almeno un } 1\}$
- $L((\Sigma\Sigma)^*) = \{w \mid w \text{ è una stringa di lunghezza pari}\}$

Negli esempi seguenti, assumiamo che l'alfabeto Σ sia $\{0, 1\}$.

Esempi

Negli esempi seguenti, assumiamo che l'alfabeto Σ sia $\{0, 1\}$.

- $L((\Sigma\Sigma\Sigma)^*) = \{w \mid w$ la lunghezza di w è un multiplo di $3\}$

Esempi

Negli esempi seguenti, assumiamo che l'alfabeto Σ sia $\{0, 1\}$.

- $L((\Sigma\Sigma\Sigma)^*) = \{w \mid w \text{ la lunghezza di } w \text{ è un multiplo di } 3\}$
- $L(01 + 10) = \{01, 10\}$

Esempi

Negli esempi seguenti, assumiamo che l'alfabeto Σ sia $\{0, 1\}$.

- $L((\Sigma\Sigma\Sigma)^*) = \{w \mid w$ la lunghezza di w è un multiplo di $3\}$
- $L(01 + 10) = \{01, 10\}$
- $L(0\Sigma^*0 + 1\Sigma^*1 + 0 + 1) = \{w \mid w$ inizia e termina con lo stesso simbolo$\}$

Negli esempi seguenti, assumiamo che l'alfabeto Σ sia $\{0, 1\}$.

- $L((\Sigma\Sigma\Sigma)^*) = \{w \mid w$ la lunghezza di w è un multiplo di $3\}$
- $L(01 + 10) = \{01, 10\}$
- $L(0\Sigma^*0 + 1\Sigma^*1 + 0 + 1) = \{w \mid w$ inizia e termina con lo stesso simbolo$\}$
- $(0 + \epsilon)1^* = 01^* + 1^*$

Negli esempi seguenti, assumiamo che l'alfabeto Σ sia $\{0, 1\}$.

- $L((\Sigma\Sigma\Sigma)^*) = \{w \mid w$ la lunghezza di w è un multiplo di $3\}$
- $L(01 + 10) = \{01, 10\}$
- $L(0\Sigma^*0 + 1\Sigma^*1 + 0 + 1) = \{w \mid w$ inizia e termina con lo stesso simbolo$\}$
- $(0 + \epsilon)1^* = 01^* + 1^*$
- $L((0 + \epsilon)(1 + \epsilon)) = \{\epsilon, 0, 1, 01\}$

Esercizio.
Determinare l'espressione regolare che, sull'alfabeto $\{a, b\}$, definisce l'insieme delle stringhe il cui terzultimo carattere è una b.

Esercizio.
Determinare l'espressione regolare che, sull'alfabeto $\{a, b\}$, definisce l'insieme delle stringhe il cui terzultimo carattere è una b.

Esercizio.
Determinare il linguaggio definito dall'espressione regolare

$$a^*((aa)^*b + (bb)^*a)b^*$$

Fornire espressioni regolari che generino i seguenti linguaggi sull'alfabeto $\{0, 1\}$:

Fornire espressioni regolari che generino i seguenti linguaggi sull'alfabeto $\{0, 1\}$:

- $\{w \mid w$ inizia con un 1 e termina con uno 0$\}$

Fornire espressioni regolari che generino i seguenti linguaggi sull'alfabeto $\{0, 1\}$:

- $\{w \mid w \text{ inizia con un 1 e termina con uno 0}\}$
- $\{w \mid w \text{ contiene almeno tre simboli uguali a 1}\}$

Fornire espressioni regolari che generino i seguenti linguaggi sull'alfabeto $\{0, 1\}$:

- $\{w \mid w$ inizia con un 1 e termina con uno $0\}$
- $\{w \mid w$ contiene almeno tre simboli uguali a $1\}$
- $\{w \mid w$ contiene la sottostringa 0101 (cioè, $w = x0101y$ per qualche x e y)$\}$

Fornire espressioni regolari che generino i seguenti linguaggi sull'alfabeto $\{0, 1\}$:

- $\{w \mid w$ inizia con un 1 e termina con uno 0$\}$
- $\{w \mid w$ contiene almeno tre simboli uguali a 1$\}$
- $\{w \mid w$ contiene la sottostringa 0101 (cioè, $w = x0101y$ per qualche x e y)$\}$
- $\{w \mid w$ ha lunghezza almeno 3 e il suo terzo simbolo è uno 0$\}$

Fornire espressioni regolari che generino i seguenti linguaggi sull'alfabeto $\{0, 1\}$:

- $\{w \mid w$ inizia con un 1 e termina con uno $0\}$
- $\{w \mid w$ contiene almeno tre simboli uguali a $1\}$
- $\{w \mid w$ contiene la sottostringa 0101 (cioè, $w = x0101y$ per qualche x e y)$\}$
- $\{w \mid w$ ha lunghezza almeno 3 e il suo terzo simbolo è uno $0\}$
- $\{w \mid w$ inizia con uno 0 e ha lunghezza dispari oppure inizia con un 1 e ha lunghezza pari$\}$

Fornire espressioni regolari che generino i seguenti linguaggi sull'alfabeto $\{0, 1\}$:

Fornire espressioni regolari che generino i seguenti linguaggi sull'alfabeto $\{0, 1\}$:

- $\{w \mid$ la lunghezza di w è al più $5\}$

Fornire espressioni regolari che generino i seguenti linguaggi sull'alfabeto $\{0, 1\}$:

- $\{w \mid$ la lunghezza di w è al più 5$\}$
- $\{w \mid w$ è una stringa diversa da 11 e 111$\}$

Fornire espressioni regolari che generino i seguenti linguaggi sull'alfabeto $\{0, 1\}$:

- $\{w \mid$ la lunghezza di w è al più 5$\}$
- $\{w \mid w$ è una stringa diversa da 11 e 111$\}$
- $\{w \mid$ in ogni posizione dispari di w c'è il simbolo 1$\}$

Fornire espressioni regolari che generino i seguenti linguaggi sull'alfabeto $\{0, 1\}$:

- $\{w \mid$ la lunghezza di w è al più 5$\}$
- $\{w \mid w$ è una stringa diversa da 11 e 111$\}$
- $\{w \mid$ in ogni posizione dispari di w c'è il simbolo 1$\}$
- $\{w \mid w$ contiene almeno due simboli uguali a 0 e al più un 1$\}$

Fornire espressioni regolari che generino i seguenti linguaggi sull'alfabeto $\{0, 1\}$:

Fornire espressioni regolari che generino i seguenti linguaggi sull'alfabeto $\{0, 1\}$:

- $\{\epsilon, 0\}$

Fornire espressioni regolari che generino i seguenti linguaggi sull'alfabeto $\{0, 1\}$:

- $\{\epsilon, 0\}$
- L'insieme vuoto

Fornire espressioni regolari che generino i seguenti linguaggi sull'alfabeto $\{0, 1\}$:

- $\{\epsilon, 0\}$
- L'insieme vuoto
- Tutte le stringhe eccetto la stringa vuota.

Per ciascuna delle seguenti espressioni regolari, fornire due stringhe che sono elementi dei corrispondenti linguaggi e due stringhe che *non* lo sono - un totale di quattro stringhe per ogni linguaggio. Assumere che l'alfabeto sia $\Sigma = \{a, b\}$ per ognuno di essi.

Per ciascuna delle seguenti espressioni regolari, fornire due stringhe che sono elementi dei corrispondenti linguaggi e due stringhe che *non* lo sono - un totale di quattro stringhe per ogni linguaggio. Assumere che l'alfabeto sia $\Sigma = \{a, b\}$ per ognuno di essi.

- $a^* b^*$

Per ciascuna delle seguenti espressioni regolari, fornire due stringhe che sono elementi dei corrispondenti linguaggi e due stringhe che *non* lo sono - un totale di quattro stringhe per ogni linguaggio. Assumere che l'alfabeto sia $\Sigma = \{a, b\}$ per ognuno di essi.

- a^*b^*
- $a(ba)^*b$

Per ciascuna delle seguenti espressioni regolari, fornire due stringhe che sono elementi dei corrispondenti linguaggi e due stringhe che *non* lo sono - un totale di quattro stringhe per ogni linguaggio. Assumere che l'alfabeto sia $\Sigma = \{a, b\}$ per ognuno di essi.

- $a^* b^*$
- $a(ba)^* b$
- $a^* + b^*$

Per ciascuna delle seguenti espressioni regolari, fornire due stringhe che sono elementi dei corrispondenti linguaggi e due stringhe che *non* lo sono - un totale di quattro stringhe per ogni linguaggio. Assumere che l'alfabeto sia $\Sigma = \{a, b\}$ per ognuno di essi.

- $a^* b^*$
- $a(ba)^* b$
- $a^* + b^*$
- $(aaa)^*$

Per ciascuna delle seguenti espressioni regolari, fornire due stringhe che sono elementi dei corrispondenti linguaggi e due stringhe che *non* lo sono—un totale di quattro stringhe per ogni linguaggio. Assumere che l'alfabeto sia $\Sigma = \{a, b\}$ per ognuno di essi.

Per ciascuna delle seguenti espressioni regolari, fornire due stringhe che sono elementi dei corrispondenti linguaggi e due stringhe che *non* lo sono—un totale di quattro stringhe per ogni linguaggio. Assumere che l'alfabeto sia $\Sigma = \{a, b\}$ per ognuno di essi.

- $\Sigma^* a \Sigma^* b \Sigma^* a \Sigma^*$

Per ciascuna delle seguenti espressioni regolari, fornire due stringhe che sono elementi dei corrispondenti linguaggi e due stringhe che *non* lo sono—un totale di quattro stringhe per ogni linguaggio. Assumere che l'alfabeto sia $\Sigma = \{a, b\}$ per ognuno di essi.

- $\Sigma^* a \Sigma^* b \Sigma^* a \Sigma^*$
- $aba + bab$

Per ciascuna delle seguenti espressioni regolari, fornire due stringhe che sono elementi dei corrispondenti linguaggi e due stringhe che *non* lo sono—un totale di quattro stringhe per ogni linguaggio. Assumere che l'alfabeto sia $\Sigma = \{a, b\}$ per ognuno di essi.

- $\Sigma^* a \Sigma^* b \Sigma^* a \Sigma^*$
- $aba + bab$
- $(\epsilon + a)b$

Per ciascuna delle seguenti espressioni regolari, fornire due stringhe che sono elementi dei corrispondenti linguaggi e due stringhe che *non* lo sono—un totale di quattro stringhe per ogni linguaggio. Assumere che l'alfabeto sia $\Sigma = \{a, b\}$ per ognuno di essi.

- $\Sigma^* a \Sigma^* b \Sigma^* a \Sigma^*$
- $aba + bab$
- $(\epsilon + a)b$
- $(a + ba + bb)\Sigma^*$