TEORIA DEI LINGUAGGI FORMALI

lunedì 9 ottobre 2017 08:45

Linguaggi formali (linguaggi di programmazione) ⟷ Linguaggi naturali (italiano, inglese)

- Artificiali
- Non ambigui
- Semantica semplice e univoca

Definizione delle SINTASSI dei linguaggi

Come si costruiscono frasi (programmi) nei linguaggi formali che siano corretti dal punto di vista della forma

in C

parentesi

if $x > 0$ ~~then~~ $x = x + 1$;

if $(x > 0)$ $x = x + 1$;

formalmente corretto;

while $(x > 0)$ $x = x + 1$;

formalmente corretto (sintatticamente corretto)

semanticamente SCORRETTO

lunedì 9 ottobre 2017 08:45

Definizione delle SINTASSI dei linguaggi formali.

Come si definiscono programmi corretti dal punto di vista della forma.

Gerarchia di LINGUAGGI FORMALI

linguaggi più complessi di altri

⋮

Linguaggi dipendenti dal contesto

Linguaggi liberi dal contesto

Linguaggi regolari

Linguaggi di programmazione →

I linguaggi hanno strumenti per definire le loro sintassi che dipendono dalla loro complessità (livello nella gerarchia)

lunedì 9 ottobre 2017 08:45

Definizione di "linguaggio formale"

Alla base della definizione sintetica abbiamo un <u>ALFABETO</u>

Un alfabeto è un INSIEME FINITO di SIMBOLI.

ES:

$\Sigma_1 = \{a, b\}$

$\Sigma_2 = \{\downarrow, \uparrow, \rightarrow, \leftarrow\}$

$\Sigma_3 = \{0, +, -, *\}$

$\vdots$

Dato Σ (alfabeto)

Σ^* è l'insieme di tutte le <u>sequenze finite</u> di simboli di Σ (stringhe)

$\Sigma = \{a, b\}$

$\Sigma^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, aab, \dots\}$

sequenze vuote (stringa vuota)

Σ^* è l'insieme di tutte le possibili FRASI che si possono costruire a

partire da $\mathcal{E}$

Alfabeto dei linguaggi formali $\Longleftrightarrow$ Dizionario dei linguaggi naturali

Σ^* $\Longleftrightarrow$ tutte le possibili sequenze di parole di un dizionario (la la la mela)

Un linguaggio su Σ è un sottoinsieme di Σ^*

$\Sigma = \{a, b\}$

$L_0 = \{a, ab\}$

le sequenze a e ab sono corrette sintatticamente nel linguaggio L_0, tutte le altre no!

$L_1 = \{ab, aab, aaab, aaaab, \ldots\}$

le frasi corrette nel linguaggio L_1 sono tutte e sole quelle che hanno una sequenza di a seguita da una sola b.

$aaaaab \in L_1$? sì!

$abaab \notin L_1$ no

$L_1 = \{ ab, aab, aaab, \ldots \}$ def infinita

Non è una definizione INFORMATICA !!

tutte le definizioni informatiche devono avere senso per un esecutore (calcolatore)

Tutte le definizioni informatiche devono essere finite

---

nei linguaggi formali

$\underbrace{aa \ldots a}_{n}$ si può esprimere con a^n

$L_1 = \{ a^n b \mid n \in \mathbb{N}^+ \}$

$n \in \mathbb{N} \wedge n > 0$

L_2 è il linguaggio su $\Sigma = \{a, b\}$ in cui il numero delle a è pari > 0.

$L_2 = \{ aa, baa, aba, aab, abab, \ldots \}$

lunedì 9 ottobre 2017 08:45

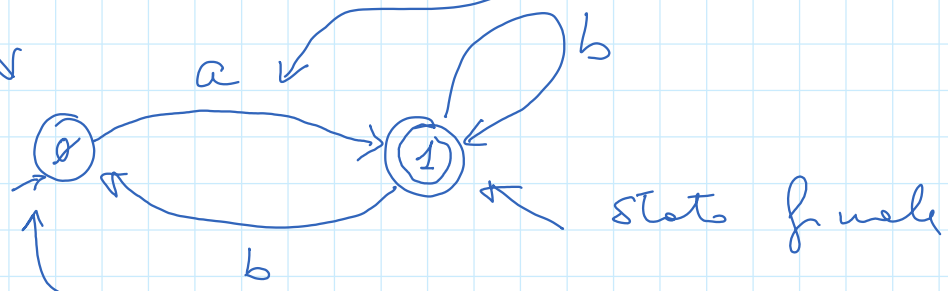
AUTOMA a STATI FINITI

Graficamente

L'insieme dei nodi è finito

Nodi (stati)

Archi orientati, etichettati



Nodo (stato) iniziale (uno solo)

Stati finali (o di accettazione)
(doppio cerchio)

Formalmente:

lunedì 9 ottobre 2017 08:45

Un ASF è una quintupla

$$A = (\Sigma, Q, S, F, \delta)$$

Σ è un alfabeto (insieme finito di simboli)

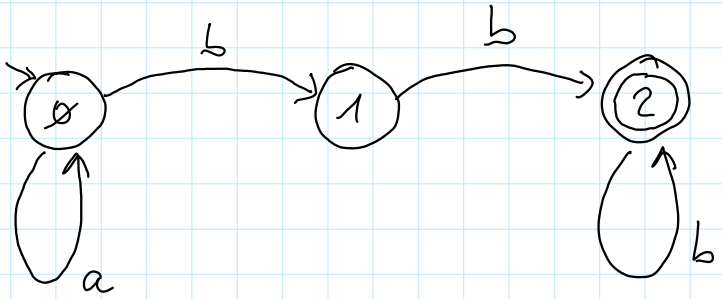
Q è un insieme finito di stati

$S \in Q$ è lo stato iniziale

$F \subseteq Q$ è l'insieme degli stati finali

$\delta \subseteq \underbrace{Q \times \Sigma \times Q}$ relazione di transizione (archi)

$$A = (\underbrace{\{a,b\}}_{\Sigma}, \underbrace{\{\emptyset,1,2\}}_{Q}, \underbrace{\emptyset}_{S}, \underbrace{\{2\}}_{F}, \underbrace{\{\langle \emptyset, a, \emptyset\rangle, \langle \emptyset, b, 1\rangle, \langle 1, b, 2\rangle, \langle 2, b, 2\rangle\}}_{\delta})$$



Come può un ASF definire la sintassi di un linguaggio?

ASF funziona come RICONOSCITORE delle stringhe che appartengono a un linguaggio.

Il linguaggio DEFINITO da un ASF è il linguaggio di tutte, e solamente, le stringhe riconosciute dall'ASF.

lunedì 9 ottobre 2017 08:45

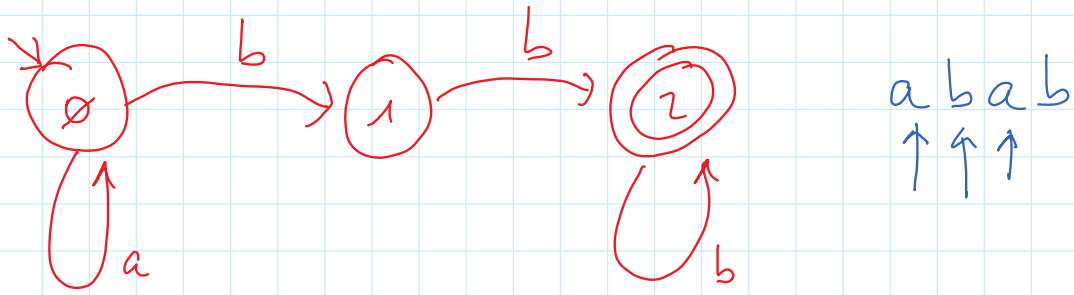
Come fa un ASF a riconoscere una stringa?

Partiamo dalla stringa da riconoscere

$a_0 a_1 a_2 \dots a_n \qquad a_i \in \Sigma$

e dallo stato iniziale dell'automa.

Per ogni simbolo della stringa, a partire dallo stato dell'automa, si esegue la transizione che ha come prime due componenti lo stato e il simbolo della stringa (se esiste)



Stato	simbolo	nuovo stato
0	a	0
0	b	1
1	a	

$\langle 0, a, 0 \rangle$

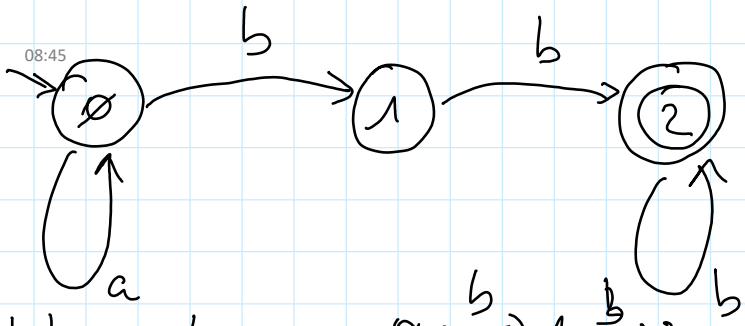
$\langle 0, b, 1 \rangle$

$\cancel{\langle 1, a, \rangle}$

Un automa riconosce una stringa $\in \Sigma^*$ se può eseguire una transizione per ogni simbolo della stringa, riesce a

scomporle tutte, e terminare in uno stato finale (o di accettazione)

lunedì 9 ottobre 2017 08:45



$bb \in L$ $\quad \emptyset \xrightarrow{b} 1 \xrightarrow{b} 2$

$abb \in L$ $\quad \emptyset \xrightarrow{a} \emptyset \xrightarrow{b} 1 \xrightarrow{b} 2$ $\quad 2 \in F$

$aab \notin L$ $\quad \emptyset \xrightarrow{a} \emptyset \xrightarrow{a} \emptyset \xrightarrow{b} 1$ $\quad 1 \notin F$

$aba \notin L$ $\quad \emptyset \xrightarrow{a} \emptyset \xrightarrow{b} 1 \not\xrightarrow{a}$

$aabbb \in L$ $\quad \emptyset \xrightarrow{a} \emptyset \xrightarrow{a} \emptyset \xrightarrow{b} 1 \xrightarrow{b} 2 \xrightarrow{b} 2$

$aabbba \notin L$

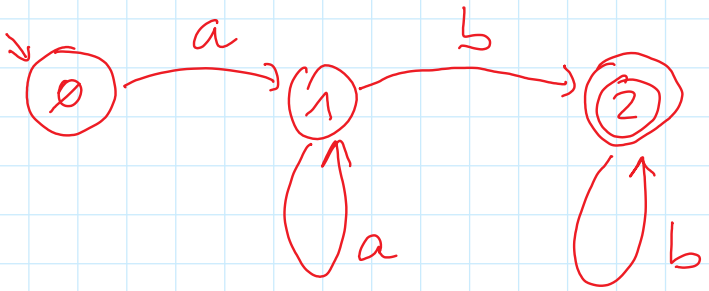
$$L = \{ a^n b^m \mid n \geq 0 \wedge m \geq 2 \}$$

$\Sigma = \{a, b\}$

lunedì 9 ottobre 2017 08:45

$L = \{ a^n b^m \mid n, m > \emptyset \}$

$L = \{ ab, aab, abb, aaab, \ldots \}$



$abbb \in L \quad \emptyset \xrightarrow{a} 1 \xrightarrow{b} 2 \xrightarrow{b} 2 \xrightarrow{b} 2 \qquad 2 \in F$

$bbb \notin L \quad \emptyset \not\xrightarrow{b}$

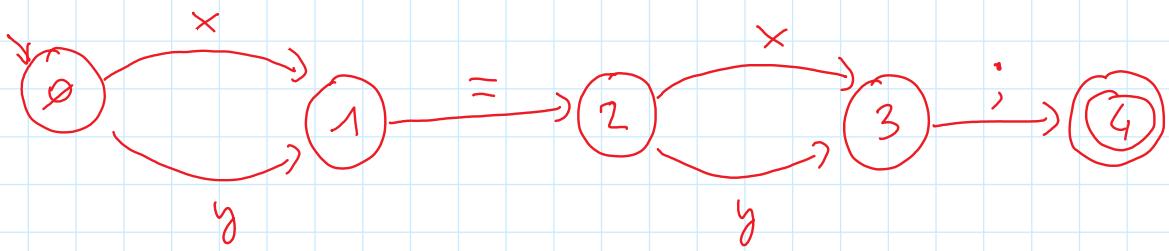
$aa \notin L \quad \emptyset \xrightarrow{a} 1 \xrightarrow{a} 1 \qquad 1 \notin F$

$aaba \notin L \quad \emptyset \xrightarrow{a} 1 \xrightarrow{a} 1 \xrightarrow{b} 2 \not\xrightarrow{a}$

lunedì 9 ottobre 2017 08:45

x, y $\quad x = y;$ $\quad y = y;$ $\quad x = x;$

$\Sigma = \{ x, y, =, ; \}$



$x == y; \quad \notin L$

$= x; \quad \notin L$

$x = x; \quad \in L \qquad 0 \xrightarrow{x} 1 \xrightarrow{=} 2 \xrightarrow{x} 3 \xrightarrow{;} 4 \in F$

L su $\Sigma = \{a, b\}$ le cui stringhe contengono un numero pari di $a > \emptyset$.

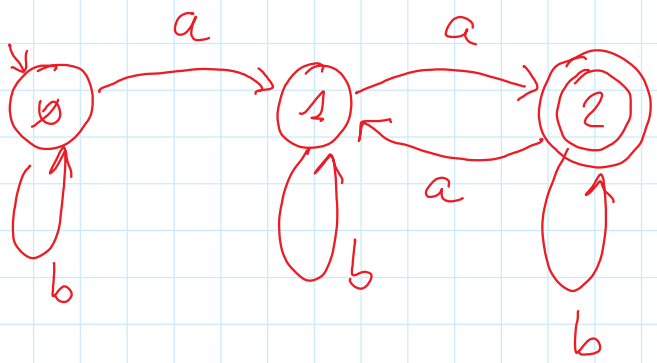
lunedì 9 ottobre 2017 08:45

aabababb $\in L$

bbaaa $\notin L$

bbabbaa $\notin L$

bbb $\notin L$



bababb

$\emptyset \xrightarrow{b} \emptyset \xrightarrow{a} 1 \xrightarrow{b} 1 \xrightarrow{a} 2 \xrightarrow{b} 2 \xrightarrow{b} 2$

lunedì 9 ottobre 2017 08:45

$$L = \{ \alpha\, a b\, \beta \mid \alpha, \beta \in \Sigma^* \}$$

$\varepsilon \in \Sigma^*$

$bb\ \underline{ab}\ aab \in L$

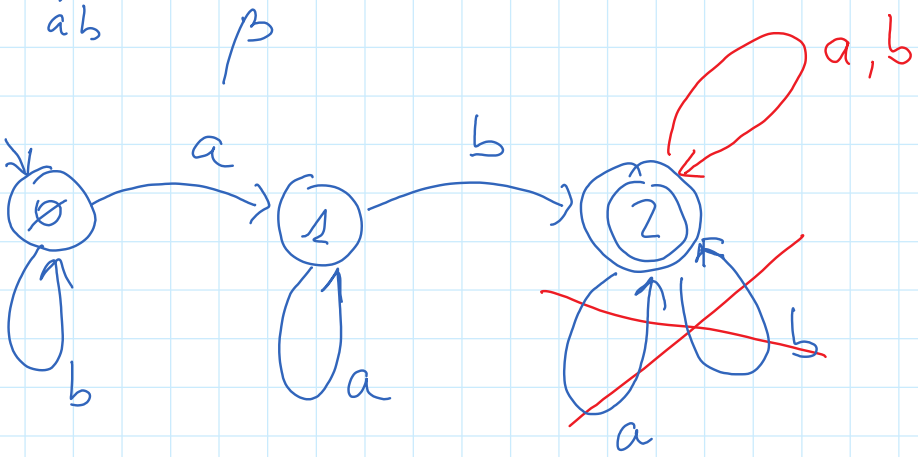
$\underline{ab}\ bbb \in L$

$bbb\ \underline{ab} \in L$

$ab \in L$

$\underbrace{ab}_{ab}\ \underbrace{abab}_{\beta} \in L$

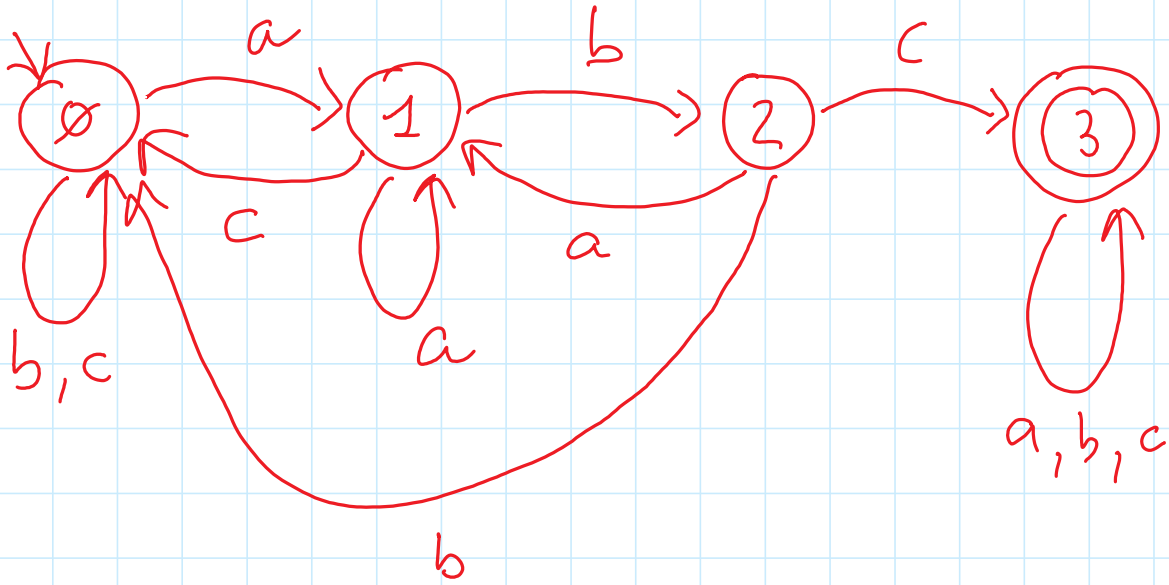
Tutte le stringhe che contengono la sequenza "ab"



lunedì 9 ottobre 2017 08:45

$$\Sigma = \{a, b, c\}$$

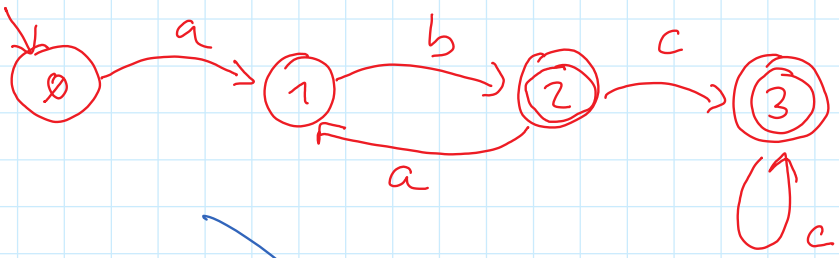
$$L = \{\alpha abc\beta \mid \alpha, \beta \in \Sigma^*\}$$



$\Sigma = \{a, b, c\}$

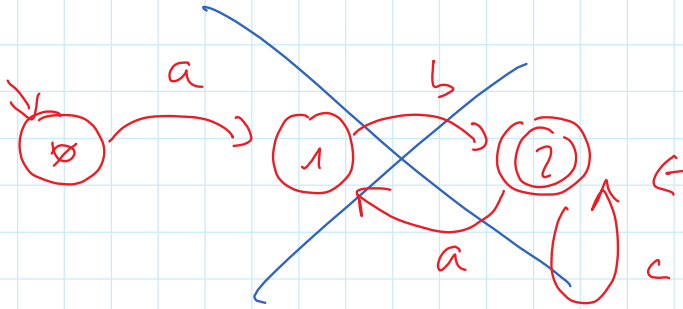
$$L = \{ (ab)^n c^m \mid n > 0 \wedge m \geq 0 \}$$

↳ l'intera sequenza "ab" deve essere ripetuta n volte



abbc $\notin L$

ababc $\in L$



← abcab $\notin L$