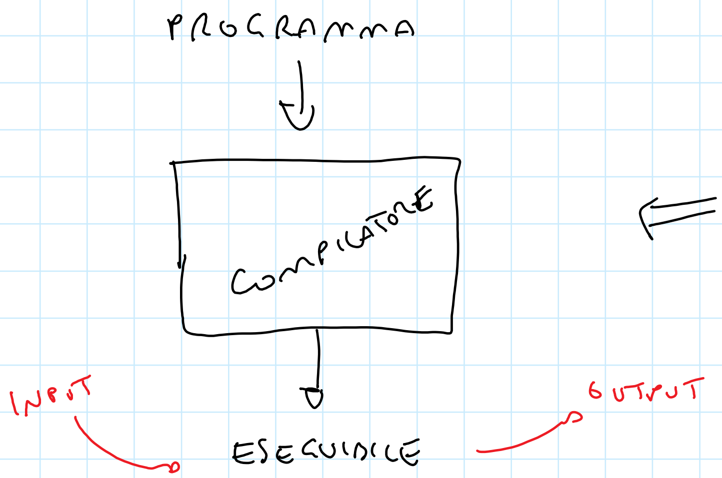


⟸ IL COMPILATORE DEVE CONTROLLARE CHE IL PROGRAMMA SIA SINTATTICAMENTE CORRETTO

while (x > 0) x = x + 1; ⟸ SINTATTICAMENTE CORRETTO (ANCHE SE POCO SENSATO ... NON TERMINA)

while x > 0 x = x + 1; ⟸ SINTATTICAMENTE NON CORRETTO (MANCANO LE PARENTESI ... IL COMPILATORE SE NE ACCORGE)

IL COMPILATORE AL SUO INTERNO DEVE AVERE UN COMPONENTE CHE, DATO UN PROGRAMMA, VERIFICA SE È SINTATTICAMENTE CORRETTO

TEORIA DEI LINGUAGGI FORMALI

(VEDERE DISPENSA SULLA PAGINA WEB)

LINGUAGGIO FORMALE ⟷ LINGUAGGIO NATURALE (ES. ITALIANO)

LINGUAGGIO FORMALE
- ↳ REGOLE GRAMMATICALI PRECISE
- ↳ NO AMBIGUITÀ

LINGUAGGIO NATURALE (ES. ITALIANO)
- ↳ REGOLE GRAMMATICALI CON TANTE ECCEZIONI
- ↳ AMBIGUITÀ

e.g. "Paolo guarda Mario con il binocolo"

↳ chi ha il binocolo?

POSSIAMO DEFINIRE UN LINGUAGGIO COME UN INSIEME

↳ LINGUAGGIO NATURALE (ES. ITALIANO) = INSIEME DI FRASI COSTRUITE UTILIZZANDO PAROLE PROVENIENTI DA UN DATO DIZIONARIO

↓

(INSIEME DI TUTTE LE PAROLE POSSIBILI)

↳ LINGUAGGIO FORMALE = INSIME DI STRINGHE COSTITUITE UTILIZZANDO SIMBOLI PROVENIENTI DA UN DATO ALFABETO

↓

INSIEME DI TUTTI I SIMBOLI

DIZIONARIO = { Paolo, Mario, mangia, beve, ... }

LINGUA ITALIANA = { Paolo mangia, Paolo beve, Mario mangia, ... }

] LING. NATURALE

ALFABETO = { a, b, c }

] LING.

ALFABETO = { a , b , c }

$L = \{ aa, ab, bcc, aba, \dots \}$ LING. FORMALE

COME SI DEFINISCE UN LINGUAGGIO FORMALE?

Un linguaggio può essere infinito:

$$L = \{ab, abb, acb, aa, \dots\}$$

NON VORREI USARE I PUNTINI

DUE MODI:

→ GRAMMATICHE
(approccio generativo)
definire delle regole grammaticali che consentono di generare tutte le stringhe del linguaggio

→ AUTOMI A STATI FINITI
(approccio riconoscitivo)
data una stringa mi dice se essa appartiene al linguaggio oppure no.

VEDREMO PRIMA GLI AUTOMI E POI
→ LE GRAMMATICHE

→ IL COMPILATORE AL SUO INTERNO UTILIZZA UN AUTOMA

DEFINIZIONI

Alfabeto insieme finito di simboli

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

sigma maiuscolo

$$\Sigma' = \{0, 1, 2, \dots, 9\}$$

$$\Sigma'' = \{a, \dots, z, 0, \dots, 9, =, <, >, !, ;, \dots\}$$

Stringa su un dato alfabeto Σ

è una sequenza finita di simboli di Σ

abc
abbbcc
bbca

$abc \neq bca$

LUNGHEZZA DI UNA STRINGA

è il numero di simboli che la compongono

abc è una stringa di lunghezza 3
abbc " " " " " 4
ba " " " " " 2

ε RAPPRESENTA UNA STRINGA DI LUNGHEZZA 0

EPSILON

ATTENZIONE $\varepsilon \notin \Sigma$

Dato un alfabeto Σ denotiamo con Σ^* ← STAR
l'insieme di tutte e sole le stringhe su Σ
di lunghezza finita.

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

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

$\varepsilon \in \Sigma^*$ qualunque sia Σ

NOTAZIONE

per semplicità scriveremo

$$a^n = \underbrace{aa \ldots a}_{n \text{ volte}}$$

$aabbb = a^2 b^3$

$abbaa = a b^2 a^2$

$a^1 = a$

$a^0 = \varepsilon$

useremo la stessa notazione anche su porzioni di stringhe

$$a^2 (ab)^3 = aaababab$$

( ) NON SONO SIMBOLI IN Σ
LE USO SOLO PER RAGGRUPPARE ab

$$aaabbb \neq (ab)^3 = ababab$$

$$\shortparallel$$

$$a^3 b^3$$

DEFINIZIONE DI LINGUAGGIO

Dato un alfabeto Σ, un linguaggio su Σ è un sottoinsieme di Σ^*

$$L \subseteq \Sigma^*$$

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

$L_1 = \{a, ab\}$ $L_1 \subseteq \Sigma^*$

$L_2 = \{a, aa, aba, bba, ab\}$

L_1 e L_2 SONO FINITI

$L_3 = \{a^n b^m \mid n, m > 0\}$

DESCRIVE IL LINGUAGGIO INFINITO DELLE STRINGHE FATTE DA UNA SEQUENZA DI a SEGUITA DA UNA SEQUENZA DI b

$aabbb \in L_3$ $n = 2$ $m = 3$

$aaabbbb \in L_3$

$bbaa \notin L_3$

$bbb \notin L_3$ $n = 0$

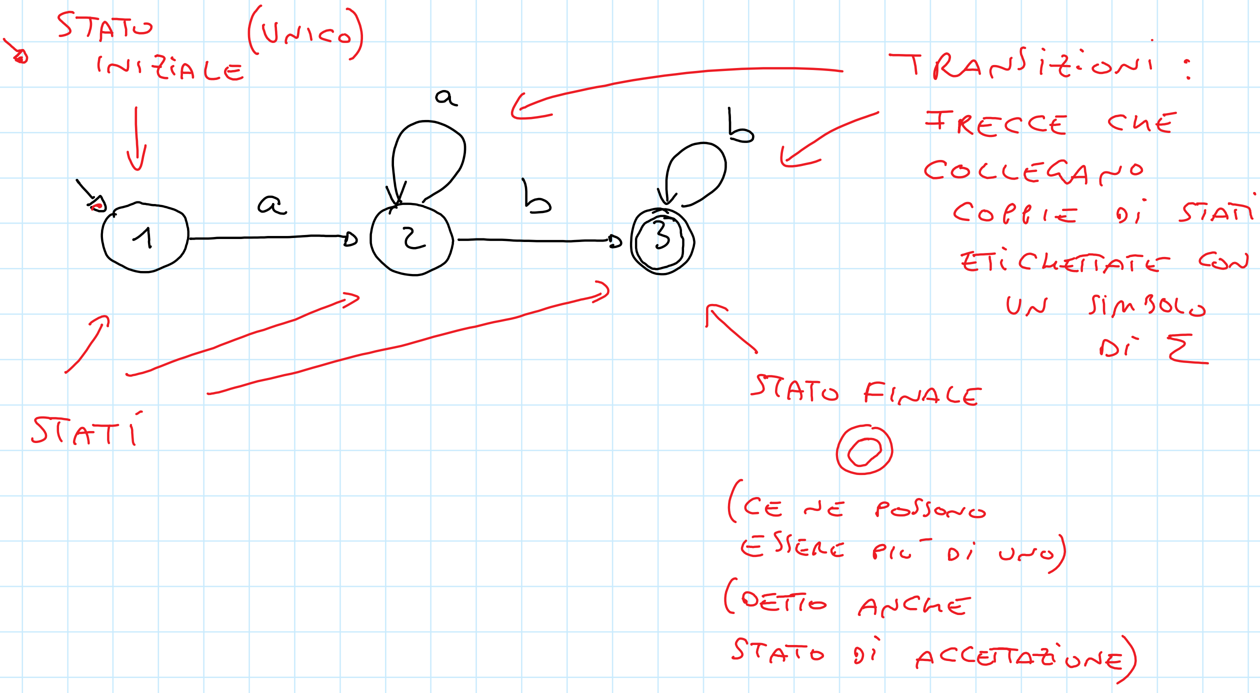
martedì 10 ottobre 2017 12:12

AUTOMI A STATI FINITI (ASF)

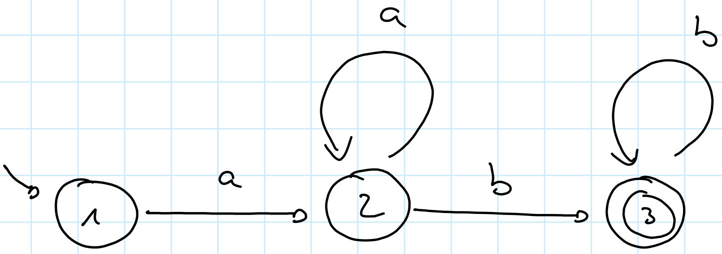
Strumento che, dato un alfabeto Σ e una stringa s su tale alfabeto $s \in \Sigma^*$, mi dice se s appartiene a un dato linguaggio $L \subseteq \Sigma^*$ oppure no.

→ RISPONDE ALLA DOMANDA $s \in L$?

GRAFICAMENTE (UN ESEMPIO) $\Sigma = \{a, b\}$



COME SI USA UN AUTOMA



data una stringa vediamo tramite l'automa se appartiene al linguaggio che l'automa definisce.

→ vediamo se l'automa ACCETTA/RICONOSCE la stringa

ESEMPIO: aaab appartiene al linguaggio?

VEDIAMO SE ESISTE UN PERCORSO DALLO STATO INIZIALE A QUELLO FINALE COSTITUITO DA TRANSIZIONI CHE HANNO COME ETICHETTE I SIMBOLI DELLA STRINGA PRESI IN ORDINE

$$1 \xrightarrow{a} 2 \xrightarrow{a} 2 \xrightarrow{a} 2 \xrightarrow{b} 3$$

3 È FINALE

$\Rightarrow aaab \in L$

ESEMPIO: aaa

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

$aaa \notin L$

POICHÉ 2 NON È

$1 \longrightarrow 2 \longrightarrow 2 \longrightarrow 2$

$\notin L$

POICHÉ 2 NON È FINALE

ESEMPIO: $aabb\underline{\underline{a}}$

$1 \xrightarrow{a} 2 \xrightarrow{a} 2 \xrightarrow{b} 3 \xrightarrow{b} 3 \cancel{\xrightarrow{a}}$

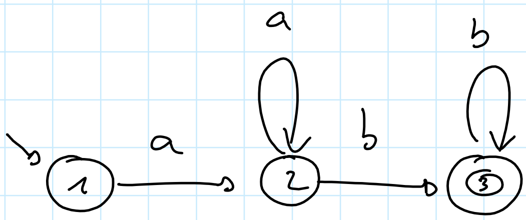
$aabba \notin L$

POICHÉ NON C'È UNA TRANSIZIONE CORRISPONDENTE ALL'ULTIMA a

IL LINGUAGGIO DEFINITO DA UN ASF

È L'INSIEME DI TUTTE E SOLE LE STRINGHE

CHE ESSO ACCETTA



che linguaggio definisce?

aaab $\in L$

aaabbbb $\in L$

ab $\in L$

$$L = \{ a^n b^m \mid n, m > 0 \}$$

DEFINIZIONE FORMALE DI ASF

Un automa a stati finiti A è una quintupla

è fatto di 5 componenti

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

delta minuscolo

Σ alfabeto

Q insieme finito di stati

S stato iniziale $S \in Q$

F insieme di stati finali $F \subseteq Q$ (di accettazione)

δ relazione di transizione

$$\delta \subseteq \underbrace{Q \times \Sigma \times Q}$$

× è prodotto cartesiano

$$Q \times \Sigma \times Q = \{ \langle q, s, q' \rangle \mid q, q' \in Q, \; s \in \Sigma \}$$

$$\langle q, S, q' \rangle \equiv (q) \xrightarrow{S} (q')$$

martedì 10 ottobre 2017 12:40

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

$$\begin{tikzcd} \arrow[r] & 0 \arrow[r, "a"] & 1 \arrow[r, "b"] & \circledcirc 2 \arrow[l, bend left, "a"] \end{tikzcd}$$

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

$Q = \{0, 1, 2\}$

$S = 0$

$F = \{2\}$

$\delta = \{ \langle 0, a, 1 \rangle, \langle 1, b, 2 \rangle, \langle 2, a, 1 \rangle \}$

QUAL È IL LINGUAGGIO DEFINITO?

$ab \in L$

$aba \notin L$

$abab \in L$

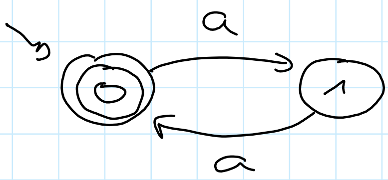
$abababab \in L$

$\varepsilon \notin L$ (RIMANE FERMO NELLO STATO INIZIALE CHE NON È FINALE)

$L = \{ (ab)^n \mid n > 0 \}$

martedì 10 ottobre 2017 12:46

$\Sigma = \{a\}$



$\varepsilon \in L$

$a \notin L$

$aa \in L$

$aaa \notin L$

$aaaa \in L$

$L = \{ a^n \mid n \text{ è pari} \}$

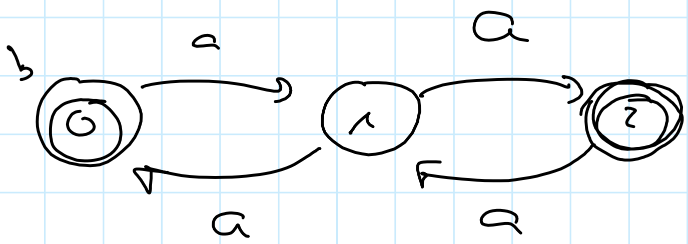
VI RICORDO CHE 0 È PARI

LO POSSO SCRIVERE ANCHE

$L = \{ a^n \mid n \% 2 = 0 \}$

$L = \{ (aa)^n \mid n \geq 0 \}$

$L = \{ (aa)^n \mid n \in \mathbb{N} \}$



$\Sigma = \{a\}$

ANCHE QUESTO DEFINISCE IL LINGUAGGIO

$$L = \{ a^n \mid n \text{ è pari} \}$$

POSSONO ESISTERE AUTOMI DIVERSI PER LO STESSO LINGUAGGIO

ESEMPIO AL CONTRARIO: Partiamo dal linguaggio e definiamo l'automa

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

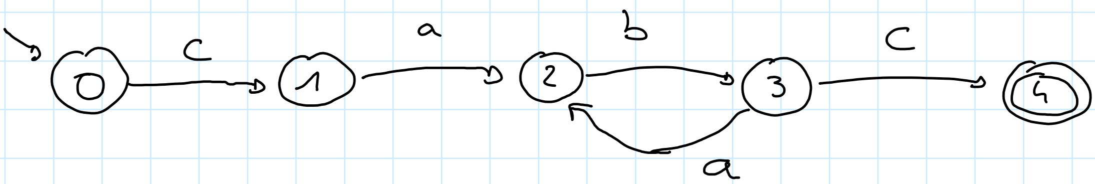
$$L = \{ c(ab)^n c \mid n > 0 \}$$

$cabc \in L$

$cababc \in L$

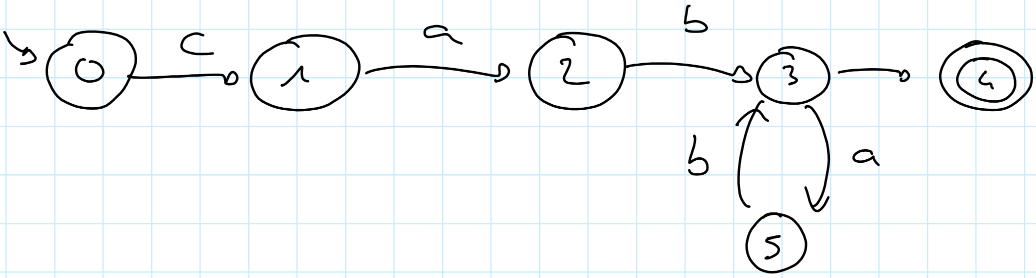
$cabababc \in L$

$\vdots$

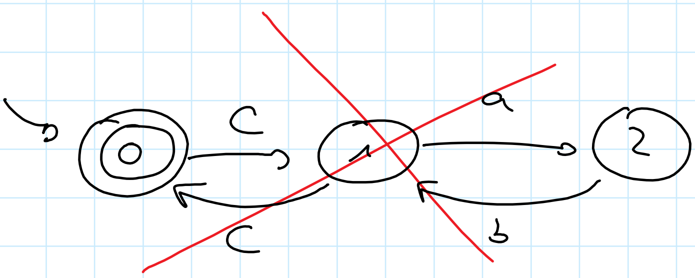


✓

ALTRE SOLUZIONI



✓



SBAGLIATO!

ACCETTA ANCHE ε