CFG e PDA

Ogni linguaggio generato da una CFG può essere accettato da un PDA e viceversa.

Il PDA costruito a partire da una CFG è un modello di analizzatore sintattico che lavora non deterministicamente in tempo lineare nella lunghezza della parola input.

Da CFG a PDA: l'idea

Sia G = (T,V,S,P) una CFG.

Descriviamo informalmente come costruire un PDA M equivalente:

1. impila il marcatore di pila vuota $ e poi il simbolo iniziale della CFG, S.

2. ripeti

se in cima alla pila c'è una variabile A, non deterministicamente scegli una regola con parte sinistra A e rimpiazza A nella pila con la parte destra della regola scelta;

se in cima alla pila c'è un terminale che coincide con il simbolo in lettura allora esegui un pop dalla pila;

3. se in cima alla pila c'è $ vai nello stato finale

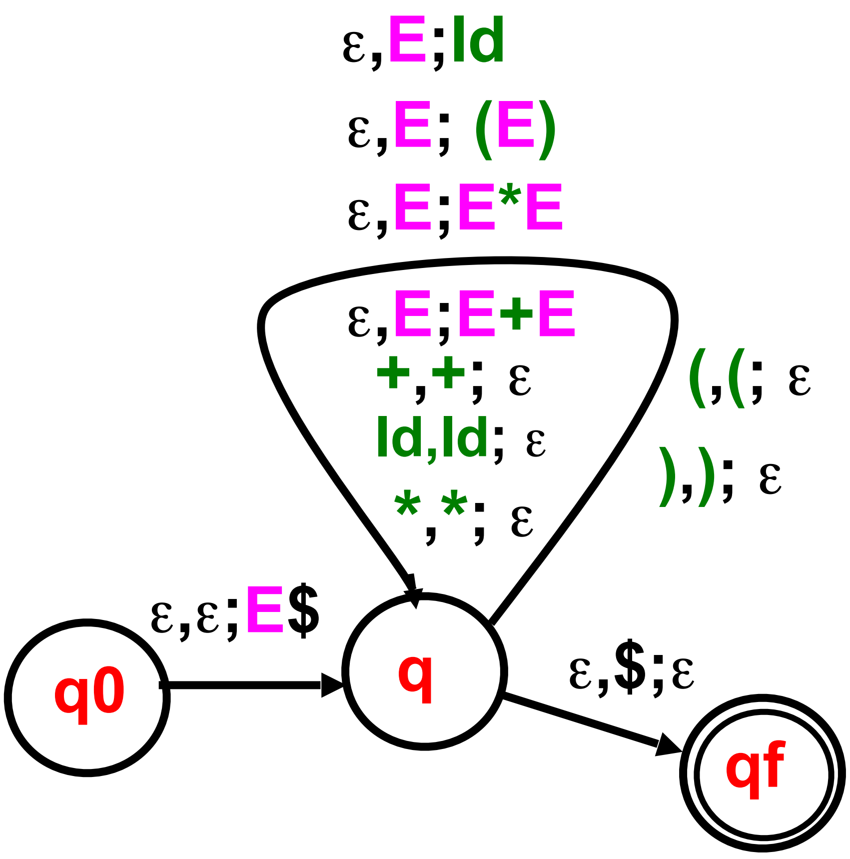
Esempio ESPRESSIONI ARITMETICHE

$E \rightarrow E+E$

$E \rightarrow E * E$

$E \rightarrow (E)$

$E \rightarrow Id$



Esempio ESPRESSIONI ARITMETICHE

Consideriamo una derivazione:

$$E \rightarrow E+E$$
$$E \rightarrow E*E$$
$$E \rightarrow (E)$$
$$E \rightarrow Id$$

$$E \Rightarrow E*E \Rightarrow (E)*E \Rightarrow (E+E)*E$$

$$\Rightarrow (id+E)*E \Rightarrow (Id+id)*E \Rightarrow (Id+Id)*Id$$

$$(q0, (Id+Id)*Id, \varepsilon) \vdash^* (q, (Id+Id)*Id, E\$) \vdash (q, (Id+Id)*Id, E*E\$)$$

$$\vdash (q, (Id+Id)*Id, (E)*E\$) \vdash (q, Id+Id)*Id, E)*E\$) \mid\!-$$

$$(q, Id+Id)*Id, E+E)*E\$) \vdash (q, Id+Id)*Id, Id+E)*E\$)$$

$$\vdash^* (q, Id)*Id, E)*E\$) \vdash (q, Id)*Id, Id)*E\$) \vdash^* (q, Id, E\$)$$

$$(q, Id, Id\$) \vdash (q, \varepsilon, \$)$$

Da CFG a PDA: la costruzione

Sia CFG G = (T,V,S,P), costruiamo un PDA equivalente
M = ({q0,q,qf}, T, V $\cup$ T $\cup$ {\$}, δ, q0,{qf}) dove

- δ(q0, ε, ε) = {(q, ε, S\$)} "inizializzazione della pila"
- δ(q, ε, A) contiene (q, γ) per ogni A $\rightarrow$ γ in P,
- δ(q, a, a) contiene (q, ε) per ogni a in T,
- δ(q, ε, \$) = {(qf, ε, ε)} "se l'input è terminato, accetta"

$\boxed{A \in V,\ \gamma \in (T \cup V)^*}$

Da CFG a PDA: la correttezza

Si deve dimostrare che

$$S \Rightarrow_G^* x \text{ se e solo se } (q,x,S\$) \mid\!-_M^* (q, \varepsilon, \$)$$

da cui $x \in L(G)$ se e solo se $x \in L(M)$.

Si dimostra una proprietà più generale
per induzione sulla lunghezza della derivazione

$$A \Rightarrow_G^* x\alpha \text{ se e solo se } (q,x,A) \mid\!-_M^* (q, \varepsilon, \alpha),$$

dove A è una qualunque variabile, x è di soli terminali o la parola vuota e α è una parola composta di variabili e terminali, che inizia con una variabile

Da CFG a PDA: la correttezza

$$A \Rightarrow_G^* x\alpha \iff (q,x,A) \vdash_M^* (q, \varepsilon, \alpha)$$

dove A è una qualunque variabile, x è di soli terminali o la parola vuota e α è una parola composta di variabili e terminali, che inizia con una variabile

$\Rightarrow$

Base: Se $A \Rightarrow_G x\alpha$ in un solo passo allora questo è

possibile se $A \to x\alpha$ è una produzione della grammatica. Per costruzione si ha che

$$(q,x,A) \vdash_M (q,x,x\alpha) \vdash_M^* (q, \varepsilon, \alpha),$$

perché x è di soli terminali.

Da CFG a PDA: la correttezza

$A \Rightarrow_G^* x\alpha$ se e solo se $(q,x,A) \vdash_M^* (q, \varepsilon, \alpha)$.

$\Rightarrow$

Passo induttivo:
Se $A \Rightarrow_G^i x\alpha$ allora facciamo vedere che $(q,x,A) \vdash_M^* (q, \varepsilon, \alpha)$.

Isolando l'ultimo passo di derivazione si ha
$A \Rightarrow_G^{i-1} yB\beta \Rightarrow_G yz\delta\beta$ dove $yz=x$ e $\delta\beta=\alpha$, con y e z di soli terminali o

la parola vuota e $\delta\beta$ di variabili o terminali, che inizia con una variabile, applicando la regola $B \rightarrow z\delta$.

Per ipotesi induttiva $A \Rightarrow_G^{i-1} yB\beta$ implica che $(q,y,A) \vdash_M^* (q,\varepsilon,B\beta)$ e,

poiché l'input ancora letto non influisce si ha anche

$$(q,yz,A) \vdash_M^* (q,z,B\beta).$$

Per costruzione si ha $(q,z,B\beta) \vdash_M (q,z,z\delta\beta) \vdash_M^* (q,\varepsilon,\delta\beta)$. Quindi

$$(q,yz,A) \vdash_M^* (q,\varepsilon,\delta\beta).$$

L'implicazione contraria si dimostra analogamente.