STRUMENTI FORMALI PER LA BIOINFORMATICA

Grammatiche e Linguaggi Context-free
Parte 2

Ringraziamenti

Le figure sono prese dai libri (o dispense):

P. Degano, *Fondamenti di Informatica: Calcolabilità e Complessità*, Dispense, 2019.

J. Hopcroft, R. Motwani, J. Ullman , *Automi, Linguaggi e Calcolabilità*, Addison Wesley Pearson Education Italia s.r.l, Terza Edizione, 2009.

Michael Sipser, *Introduzione alla teoria della Computazione*, Apogeo Education, Maggioli Editore, 2016 (traduzione italiana di Introduction to the Theory of Computation, 3rd Edition).

Definizione

Sia $G = (V, T, P, S)$ *una grammatica context free e sia* $\alpha \in (V \cup T)^*$.

Definizione

Sia $G = (V, T, P, S)$ *una grammatica context free e sia* $\alpha \in (V \cup T)^*$.

- *La stringa* α *è una* **forma sentenziale** *se* $S \stackrel{*}{\Rightarrow} \alpha$

Definizione

Sia $G = (V, T, P, S)$ *una grammatica context free e sia* $\alpha \in (V \cup T)^*$.

- *La stringa* α *è una* **forma sentenziale** *se* $S \stackrel{*}{\Rightarrow} \alpha$
- *La stringa* α *è una* **forma sentenziale sinistra** *se* $S \underset{lm}{\stackrel{*}{\Rightarrow}} \alpha$

Definizione

Sia $G = (V, T, P, S)$ *una grammatica context free e sia* $\alpha \in (V \cup T)^*$.

- *La stringa* α *è una* **forma sentenziale** *se* $S \overset{*}{\Rightarrow} \alpha$
- *La stringa* α *è una* **forma sentenziale sinistra** *se* $S \underset{lm}{\overset{*}{\Rightarrow}} \alpha$
- *La stringa* α *è una* **forma sentenziale destra** *se* $S \underset{rm}{\overset{*}{\Rightarrow}} \alpha$

Negli esempi seguenti ci riferiremo alla grammatica $G_E = (\{E, I\}, T, P, E)$ per espressioni aritmetiche dove

$$T = \{+, *, (,), a, b, 0, 1\}$$

e

$$\begin{aligned} P &= \{E \to E + E \mid E * E \mid (E) \mid I, \\ &\quad I \to a \mid b \mid Ia \mid Ib \mid I0 \mid I1\} \end{aligned}$$

- Nella grammatica G_E delle espressioni aritmetiche, $E * (I + E)$ è una forma sentenziale:

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

Forme sentenziali - Esempi

- Nella grammatica G_E delle espressioni aritmetiche, $E * (I + E)$ è una forma sentenziale:

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

- $a * E$ è una forma sentenziale sinistra:

$$E \underset{lm}{\Rightarrow} E * E \underset{lm}{\Rightarrow} I * E \underset{lm}{\Rightarrow} a * E$$

Forme sentenziali - Esempi

- Nella grammatica G_E delle espressioni aritmetiche, $E * (I + E)$ è una forma sentenziale:

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

- $a * E$ è una forma sentenziale sinistra:

$$E \underset{lm}{\Rightarrow} E * E \underset{lm}{\Rightarrow} I * E \underset{lm}{\Rightarrow} a * E$$

- $E * (E + E)$ è una forma sentenziale destra:

$$E \underset{rm}{\Rightarrow} E * E \underset{rm}{\Rightarrow} E * (E) \underset{rm}{\Rightarrow} E * (E + E)$$

Sia $G = (V, T, P, S)$ una grammatica context free, sia $A \in V$.

Sia $G = (V, T, P, S)$ una grammatica context free, sia $A \in V$.

Il **linguaggio della variabile** A è il linguaggio $L(G_A)$ generato dalla grammatica $G_A = (V, T, P, A)$.

Abbiamo visto come in alcune grammatiche una stringa di terminali possa essere generata in diversi modi a partire dal simbolo iniziale, anche fissando a priori un criterio per la riscrittura cioè usando derivazioni a sinistra.

Abbiamo visto come in alcune grammatiche una stringa di terminali possa essere generata in diversi modi a partire dal simbolo iniziale, anche fissando a priori un criterio per la riscrittura cioè usando derivazioni a sinistra.

Ad esempio $a + a * a$ ha due derivazioni a sinistra in G_E:

$$E \underset{lm}{\Rightarrow} E * E \underset{lm}{\Rightarrow} E + E * E \underset{lm}{\Rightarrow} I + E * E \underset{lm}{\Rightarrow} a + E * E \underset{lm}{\Rightarrow}$$
$$a + I * E \underset{lm}{\Rightarrow} a + a * E \underset{lm}{\Rightarrow} a + a * I \underset{lm}{\Rightarrow} a + a * a$$

$$E \underset{lm}{\Rightarrow} E + E \underset{lm}{\Rightarrow} I + E \underset{lm}{\Rightarrow} a + E \underset{lm}{\Rightarrow} a + E * E \underset{lm}{\Rightarrow}$$
$$a + I * E \underset{lm}{\Rightarrow} a + a * E \underset{lm}{\Rightarrow} a + a * I \underset{lm}{\Rightarrow} a + a * a$$

Abbiamo osservato che questa grammatica non sembra particolarmente adatta a tradurre l'intuizione che abbiamo rispetto a come vanno eseguiti i programmi.

Abbiamo osservato che questa grammatica non sembra particolarmente adatta a tradurre l'intuizione che abbiamo rispetto a come vanno eseguiti i programmi.

Per vederlo con più chiarezza, ricorriamo a una rappresentazione delle derivazioni che prende il nome di albero sintattico (parse tree, albero di derivazione)

Il parse tree o albero sintattico, è una rappresentazione per le derivazioni mediante un albero. Ed è una rappresentazione che mostra chiaramente come i simboli di una stringa di terminali sono raggruppati in sottoparole, ognuna delle quali appartiene al linguaggio di una delle variabili della grammatica.

In particolare, quando usata in un compilatore, l'albero sintattico è la struttura dati che rappresenta il programma sorgente nel processo di traduzione in codice eseguibile.

Alberi sintattici

Vedremo le relazioni tra alberi sintattici e derivazioni.

In alcune grammatiche una stringa del linguaggio può avere più di un albero sintattico. Questo è legato al problema dell'ambiguità nella grammatica che, come abbiamo osservato, rende la grammatica poco adatta a un linguaggio di programmazione.

Si assume nota la terminologia sugli alberi (nodi, foglie, radice, figli ordinati da sinistra...).

Definizione

Sia $G = (V, T, P, S)$ una grammatica context free. Gli alberi sintattici (o parse tree) di G sono gli alberi che soddisfano le seguenti condizioni.

Definizione

Sia $G = (V, T, P, S)$ una grammatica context free. Gli alberi sintattici (o parse tree) di G sono gli alberi che soddisfano le seguenti condizioni.

1. *Ciascun nodo interno è etichettato da una variabile in V.*

Definizione

Sia $G = (V, T, P, S)$ una grammatica context free. Gli alberi sintattici (o parse tree) di G sono gli alberi che soddisfano le seguenti condizioni.

1. *Ciascun nodo interno è etichettato da una variabile in V.*
2. *Ogni foglia è etichettata da una variabile o da un terminale o da ϵ. Se una foglia ha come etichetta ϵ, deve essere l'unico figlio del suo genitore.*

Definizione

Sia $G = (V, T, P, S)$ una grammatica context free. Gli alberi sintattici (o parse tree) di G sono gli alberi che soddisfano le seguenti condizioni.

1. *Ciascun nodo interno è etichettato da una variabile in V.*
2. *Ogni foglia è etichettata da una variabile o da un terminale o da ϵ. Se una foglia ha come etichetta ϵ, deve essere l'unico figlio del suo genitore.*
3. *Se un nodo interno è etichettato A e i suoi figli sono etichettati, a partire da sinistra,*

$$X_1, X_2, \ldots, X_k$$

allora $A \to X_1X_2, \cdots X_k$ è una produzione in P. Se esiste i, $1 \leq i \leq k$, tale che $X_i = \epsilon$, allora $k = 1$ e $A \to \epsilon \in P$ (per la precedente condizione).

Esempio 1.
Un albero sintattico per la grammatica $G_E = (\{E, I\}, T, P, E)$ per espressioni aritmetiche dove

$$T = \{+, *, (,), a, b, 0, 1\}$$

e

$$\begin{aligned} P &= \{E \to E + E \mid E * E \mid (E) \mid I, \\ &\qquad I \to a \mid b \mid Ia \mid Ib \mid I0 \mid I1\} \end{aligned}$$

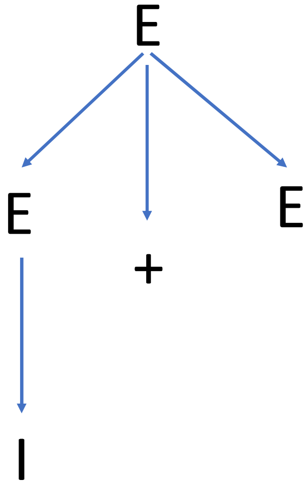


Figura: Esempio 1.

Esempio 2.
Un albero sintattico per la grammatica delle parole palindrome binarie $G_{pal} = (\{S\}, \{0, 1\}, P, S)$ e

$$P = \{S \rightarrow \epsilon \mid 0 \mid 1 \mid 0S0 \mid 1S1\}$$

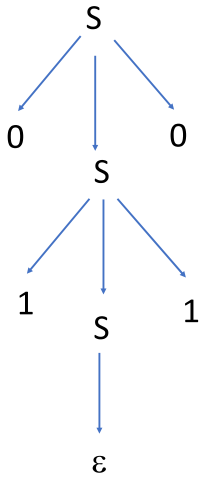


Figura: Esempio 2.

Il prodotto di un albero sintattico

Definizione

Il prodotto (o frontiera) di un albero sintattico è la parola che si ottiene concatenando le etichette delle foglie a partire da sinistra.

Definizione

Il prodotto (o frontiera) di un albero sintattico è la parola che si ottiene concatenando le etichette delle foglie a partire da sinistra.

- Di particolare importanza sono gli alberi sintattici che soddisfano le due seguenti condizioni.

Definizione

Il *prodotto* *(o frontiera) di un albero sintattico è la parola che si ottiene concatenando le etichette delle foglie a partire da sinistra.*

- Di particolare importanza sono gli alberi sintattici che soddisfano le due seguenti condizioni.
  1. Il prodotto è una stringa sull'alfabeto dei terminali, cioè ogni foglia ha come etichetta un terminale o ϵ.

Il prodotto di un albero sintattico

Definizione

Il prodotto (o frontiera) di un albero sintattico è la parola che si ottiene concatenando le etichette delle foglie a partire da sinistra.

- Di particolare importanza sono gli alberi sintattici che soddisfano le due seguenti condizioni.
  1. Il prodotto è una stringa sull'alfabeto dei terminali, cioè ogni foglia ha come etichetta un terminale o ϵ.
  2. La radice ha come etichetta il simbolo iniziale.

Definizione

Il prodotto (o frontiera) di un albero sintattico è la parola che si ottiene concatenando le etichette delle foglie a partire da sinistra.

- Di particolare importanza sono gli alberi sintattici che soddisfano le due seguenti condizioni.
  1. Il prodotto è una stringa sull'alfabeto dei terminali, cioè ogni foglia ha come etichetta un terminale o ϵ.
  2. La radice ha come etichetta il simbolo iniziale.
- Questi alberi sono importanti perché vedremo che una parola è nel linguaggio di una grammatica se e solo se è il prodotto di un albero di questo tipo.

Esempio 3
Un albero sintattico per la grammatica $G_E = (\{E, I\}, T, P, E)$ per espressioni aritmetiche dove

$$T = \{+, *, (,), a, b, 0, 1\}$$

e

$$\begin{aligned} P &= \{E \to E + E \mid E * E \mid (E) \mid I, \\ &\quad\; I \to a \mid b \mid Ia \mid Ib \mid I0 \mid I1\} \end{aligned}$$

Il prodotto $a * (a + b00)$ è una stringa sull'alfabeto dei terminali.
La radice ha come etichetta il simbolo iniziale della grammatica.

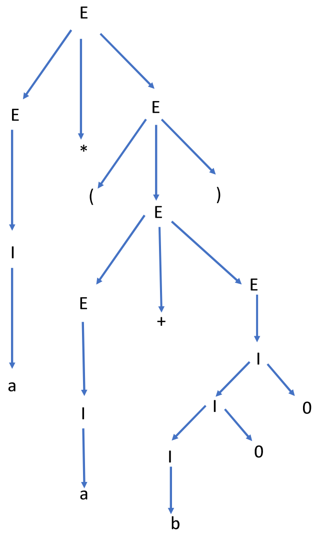


Figura: Esempio 3.

Esempio 3
Abbiamo mostrato una derivazione della stringa $a * (a + b00)$ in G_E.

$$E \Rightarrow E * E \Rightarrow I * E \Rightarrow a * E \Rightarrow a * (E) \Rightarrow$$
$$a * (E + E) \Rightarrow a * (I + E) \Rightarrow a * (a + E) \Rightarrow a * (a + I) \Rightarrow$$
$$a * (a + I0) \Rightarrow a * (a + I00) \Rightarrow a * (a + b00)$$

Vedremo che questo particolare parse tree è una rappresentazione della derivazione.

Teorema

Sia $G = (V, T, P, S)$ una grammatica context free, sia $A \in V$ e $w \in T^*$. Le condizioni seguenti sono equivalenti.

Teorema

Sia $G = (V, T, P, S)$ una grammatica context free, sia $A \in V$ e $w \in T^*$. Le condizioni seguenti sono equivalenti.

1. *Esiste un albero sintattico con radice etichettata da A e con prodotto w*

Teorema

Sia $G = (V, T, P, S)$ una grammatica context free, sia $A \in V$ e $w \in T^*$. Le condizioni seguenti sono equivalenti.

1. *Esiste un albero sintattico con radice etichettata da A e con prodotto w*
2. $A \overset{*}{\underset{lm}{\Rightarrow}} w$

Teorema

Sia $G = (V, T, P, S)$ *una grammatica context free, sia* $A \in V$ *e* $w \in T^*$. *Le condizioni seguenti sono equivalenti.*

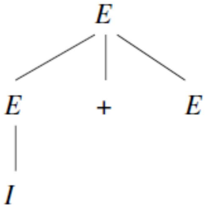
1. *Esiste un albero sintattico con radice etichettata da* A *e con prodotto* w
2. $A \underset{lm}{\overset{*}{\Rightarrow}} w$
3. $A \underset{rm}{\overset{*}{\Rightarrow}} w$

Teorema

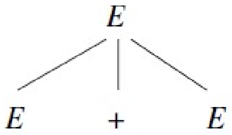
Sia $G = (V, T, P, S)$ una grammatica context free, sia $A \in V$ e $w \in T^*$. Le condizioni seguenti sono equivalenti.

1. *Esiste un albero sintattico con radice etichettata da A e con prodotto w*
2. $A \underset{lm}{\overset{*}{\Rightarrow}} w$
3. $A \underset{rm}{\overset{*}{\Rightarrow}} w$
4. $A \overset{*}{\Rightarrow} w$

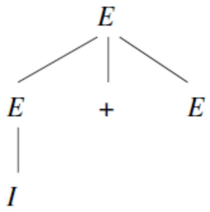
L'albero sintattico che corrisponde a $E \stackrel{*}{\Rightarrow} I + E$ nella grammatica delle espressioni aritmetiche.



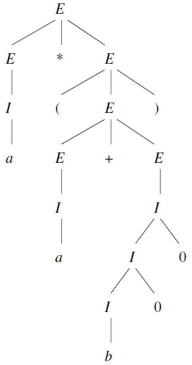
Nel primo passo di $E \Rightarrow E + E \Rightarrow I + E$ usiamo la produzione $E \rightarrow E + E$. Questo è l'albero corrispondente.



Nel secondo passo di $E \Rightarrow E + E \Rightarrow I + E$ usiamo la produzione $E \rightarrow I$. Questo è l'albero corrispondente.

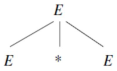


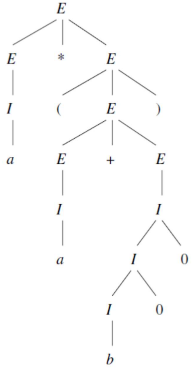
L'albero sintattico che corrisponde a $E \overset{*}{\Rightarrow} a * (a + b00)$ nella grammatica delle espressioni aritmetiche.



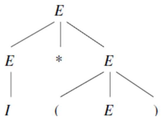
Alberi sintattici e derivazioni

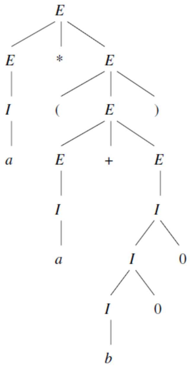
Nel primo passo di $E \Rightarrow E * E \overset{*}{\Rightarrow} a * (a + b00)$ usiamo la produzione $E \rightarrow E * E$. L'albero sulla sinistra è l'albero corrispondente.





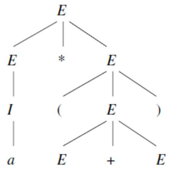
Nel secondo passo di $E \Rightarrow E * E \Rightarrow I * E \Rightarrow I * (E) \overset{*}{\Rightarrow} a * (a + b00)$ usiamo la produzione $E \rightarrow I$ e nel terzo $E \rightarrow (E)$. L'albero sulla sinistra è l'albero corrispondente.

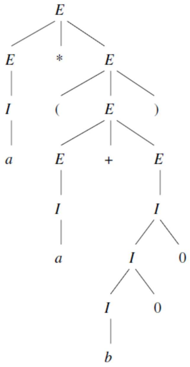




Alberi sintattici e derivazioni

L'albero sulla sinistra è l'albero corrispondente all'applicazione delle produzioni $I \to a$ e $E \to E + E$.





$G = (V, T, P, S)$ grammatica context free, $A \in V$ e $w \in T^*$.

1. Esiste un albero sintattico con radice etichettata da A e con prodotto w

$G = (V, T, P, S)$ grammatica context free, $A \in V$ e $w \in T^*$.

1. Esiste un albero sintattico con radice etichettata da A e con prodotto w
2. $A \overset{*}{\underset{lm}{\Rightarrow}} w$

$G = (V, T, P, S)$ grammatica context free, $A \in V$ e $w \in T^*$.

1. Esiste un albero sintattico con radice etichettata da A e con prodotto w
2. $A \underset{lm}{\overset{*}{\Rightarrow}} w$
3. $A \underset{rm}{\overset{*}{\Rightarrow}} w$

$G = (V, T, P, S)$ grammatica context free, $A \in V$ e $w \in T^*$.

1. Esiste un albero sintattico con radice etichettata da A e con prodotto w
2. $A \underset{lm}{\overset{*}{\Rightarrow}} w$
3. $A \underset{rm}{\overset{*}{\Rightarrow}} w$
4. $A \overset{*}{\Rightarrow} w$

$G = (V, T, P, S)$ grammatica context free, $A \in V$ e $w \in T^*$.

1. Esiste un albero sintattico con radice etichettata da A e con prodotto w
2. $A \underset{lm}{\overset{*}{\Rightarrow}} w$
3. $A \underset{rm}{\overset{*}{\Rightarrow}} w$
4. $A \overset{*}{\Rightarrow} w$

   Proveremo che (1) implica (2)

$G = (V, T, P, S)$ grammatica context free, $A \in V$ e $w \in T^*$.

1. Esiste un albero sintattico con radice etichettata da A e con prodotto w
2. $A \underset{lm}{\overset{*}{\Rightarrow}} w$
3. $A \underset{rm}{\overset{*}{\Rightarrow}} w$
4. $A \overset{*}{\Rightarrow} w$

   Proveremo che (1) implica (2)

   e che (1) implica (3)

$G = (V, T, P, S)$ grammatica context free, $A \in V$ e $w \in T^*$.

1. Esiste un albero sintattico con radice etichettata da A e con prodotto w
2. $A \underset{lm}{\overset{*}{\Rightarrow}} w$
3. $A \underset{rm}{\overset{*}{\Rightarrow}} w$
4. $A \overset{*}{\Rightarrow} w$

   Proveremo che (1) implica (2)

   e che (1) implica (3)

   Poi che (4) implica (1).

$G = (V, T, P, S)$ grammatica context free, $A \in V$ e $w \in T^*$.

1. Esiste un albero sintattico con radice etichettata da A e con prodotto w
2. $A \underset{lm}{\overset{*}{\Rightarrow}} w$
3. $A \underset{rm}{\overset{*}{\Rightarrow}} w$
4. $A \overset{*}{\Rightarrow} w$

   Proveremo che (1) implica (2)

   e che (1) implica (3)

   Poi che (4) implica (1).

   La dimostrazione è terminata. Infatti poiché è evidente che (2) implica (4), avremo provato che le condizioni (1), (2), (4) sono equivalenti.

$G = (V, T, P, S)$ grammatica context free, $A \in V$ e $w \in T^*$.

1. Esiste un albero sintattico con radice etichettata da A e con prodotto w
2. $A \underset{lm}{\overset{*}{\Rightarrow}} w$
3. $A \underset{rm}{\overset{*}{\Rightarrow}} w$
4. $A \overset{*}{\Rightarrow} w$

   Proveremo che (1) implica (2)

   e che (1) implica (3)

   Poi che (4) implica (1).

   La dimostrazione è terminata. Infatti poiché è evidente che (2) implica (4), avremo provato che le condizioni (1), (2), (4) sono equivalenti.

   Analogamente, poiché è evidente che (3) implica (4), avremo provato che le condizioni (1), (3), (4) sono equivalenti.

$G = (V, T, P, S)$ grammatica context free, $A \in V$ e $w \in T^*$.

1. Esiste un albero sintattico con radice etichettata da A e con prodotto w
2. $A \overset{*}{\underset{lm}{\Rightarrow}} w$
3. $A \overset{*}{\underset{rm}{\Rightarrow}} w$
4. $A \overset{*}{\Rightarrow} w$

Proveremo che (1) implica (2)

e che (1) implica (3)

Poi che (4) implica (1).

La dimostrazione è terminata. Infatti poiché è evidente che (2) implica (4), avremo provato che le condizioni (1), (2), (4) sono equivalenti.

Analogamente, poiché è evidente che (3) implica (4), avremo provato che le condizioni (1), (3), (4) sono equivalenti.

Quindi le quattro condizioni sono equivalenti.

Nelle dimostrazioni vengono utilizzati due risultati intermedi.

Ricordiamo la definizione di derivazione diretta: se $A \to \gamma$ è una produzione in G allora $\alpha A \beta \underset{G}{\Rightarrow} \alpha\gamma\beta$, per ogni α, γ, con α, γ stringhe di variabili e terminali.

Il primo risultato che si utilizza è una estensione di tale definizione: Se $A \stackrel{*}{\Rightarrow} \gamma$, questa derivazione si può inserire in un contesto, cioè $\alpha A \beta \stackrel{*}{\Rightarrow} \alpha\gamma\beta$. Con un ipotesi aggiuntiva su α (risp. β) lo stesso risultato vale per le derivazioni a sinistra (risp. a destra).

Chiameremo questa proprietà la **proprietà dell'inserzione**.

Consideriamo la grammatica context-free G definita dalle produzioni

$$S \to aS \mid Sb \mid a \mid b$$

Consideriamo la grammatica context-free G definita dalle produzioni

$$S \to aS \mid Sb \mid a \mid b$$

Risulta $S \Rightarrow aS \Rightarrow aaS$ e $S \Rightarrow aS \Rightarrow aSb$, quindi

$$S \stackrel{*}{\Rightarrow} aaS, \quad S \stackrel{*}{\Rightarrow} aSb$$

Proprietà dell'inserzione - Esempi

Consideriamo la grammatica context-free G definita dalle produzioni

$$S \to aS \mid Sb \mid a \mid b$$

Risulta $S \Rightarrow aS \Rightarrow aaS$ e $S \Rightarrow aS \Rightarrow aSb$, quindi

$$S \stackrel{*}{\Rightarrow} aaS, \quad S \stackrel{*}{\Rightarrow} aSb$$

Per la proprietà dell'inserzione, possiamo "inserire la seconda derivazione nella prima". Precisamente, possiamo sostituire alla variabile S nella stringa a destra della prima derivazione la stringa a destra della seconda derivazione e ottenere una nuova derivazione:

$$S \stackrel{*}{\Rightarrow} aaS \stackrel{*}{\Rightarrow} aaaSb$$

Consideriamo la grammatica context-free G definita dalle produzioni

$$S \to aSbS \mid bSaS \mid \epsilon$$

Proprietà dell'inserzione - Esempi

Consideriamo la grammatica context-free G definita dalle produzioni

$$S \to aSbS \mid bSaS \mid \epsilon$$

Risulta $S \Rightarrow aSbS \Rightarrow aaSbSbS$ e $S \Rightarrow bSaS \Rightarrow baSbSaS$, quindi

$$S \overset{*}{\Rightarrow} aaSbSbS, \quad S \overset{*}{\Rightarrow} baSbSaS$$

Consideriamo la grammatica context-free G definita dalle produzioni

$$S \to aSbS \mid bSaS \mid \epsilon$$

Risulta $S \Rightarrow aSbS \Rightarrow aaSbSbS$ e $S \Rightarrow bSaS \Rightarrow baSbSaS$, quindi

$$S \overset{*}{\Rightarrow} aaSbSbS, \quad S \overset{*}{\Rightarrow} baSbSaS$$

Per la proprietà dell'inserzione, possiamo "inserire la seconda derivazione nella prima". Precisamente, possiamo sostituire a una delle tre occorrenze della variabile S nella stringa a destra della prima derivazione la stringa a destra della seconda derivazione e ottenere una nuova derivazione. Se scegliamo la prima occorrenza otteniamo come risultato

$$S \overset{*}{\Rightarrow} aaSbSbS \overset{*}{\Rightarrow} aabaSbSaSbSbS$$

Proposizione

[proprietà dell'inserzione]. *Sia* $G = (V, T, P, S)$ *una grammatica context free. Per ogni* $A \in V$, $\alpha, \gamma_1, \gamma_2 \in (V \cup T)^*$,

Proposizione

[proprietà dell'inserzione]. *Sia* $G = (V, T, P, S)$ *una grammatica context free. Per ogni* $A \in V$, $\alpha, \gamma_1, \gamma_2 \in (V \cup T)^*$,

- *se* $A \overset{*}{\underset{G}{\Rightarrow}} \alpha$ *allora* $\gamma_1 A \gamma_2 \overset{*}{\underset{G}{\Rightarrow}} \gamma_1 \alpha \gamma_2$;

Proposizione

[proprietà dell'inserzione]. *Sia* $G = (V, T, P, S)$ *una grammatica context free. Per ogni* $A \in V$, $\alpha, \gamma_1, \gamma_2 \in (V \cup T)^*$,

- *se* $A \overset{*}{\underset{G}{\Rightarrow}} \alpha$ *allora* $\gamma_1 A \gamma_2 \overset{*}{\underset{G}{\Rightarrow}} \gamma_1 \alpha \gamma_2$*;*
- *se* $A \overset{*}{\underset{lm}{\Rightarrow}} \alpha$ *e* $\gamma_1 \in T^*$ *allora* $\gamma_1 A \gamma_2 \overset{*}{\underset{lm}{\Rightarrow}} \gamma_1 \alpha \gamma_2$

Proposizione

[proprietà dell'inserzione]. *Sia $G = (V, T, P, S)$ una grammatica context free. Per ogni $A \in V$, $\alpha, \gamma_1, \gamma_2 \in (V \cup T)^*$,*

- *se* $A \underset{G}{\overset{*}{\Rightarrow}} \alpha$ *allora* $\gamma_1 A \gamma_2 \underset{G}{\overset{*}{\Rightarrow}} \gamma_1 \alpha \gamma_2$;
- *se* $A \underset{lm}{\overset{*}{\Rightarrow}} \alpha$ *e* $\gamma_1 \in T^*$ *allora* $\gamma_1 A \gamma_2 \underset{lm}{\overset{*}{\Rightarrow}} \gamma_1 \alpha \gamma_2$
- *se* $A \underset{rm}{\overset{*}{\Rightarrow}} \alpha$ *e* $\gamma_2 \in T^*$ *allora* $\gamma_1 A \gamma_2 \underset{rm}{\overset{*}{\Rightarrow}} \gamma_1 \alpha \gamma_2$

Teorema

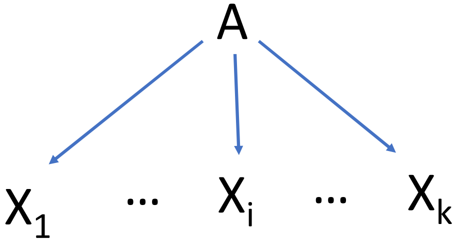
Sia $G = (V, T, P, S)$ *una grammatica context free.*

Se esiste un albero sintattico con radice etichettata da una variabile A *e con prodotto* w*, dove* $w \in T^*$*, allora esiste una derivazione a sinistra* $A \underset{lm}{\overset{*}{\Rightarrow}} w$ *nella grammatica* G*.*

(Dall'albero sintattico alla derivazione a sinistra)

(Dall'albero sintattico alla derivazione a sinistra)

La prova è per induzione strutturale sull'altezza dell'albero.



$$X_1 \dots X_k = w$$

Figura: Passo Base.

(Dall'albero sintattico alla derivazione a sinistra)

(Dall'albero sintattico alla derivazione a sinistra)

La prova è per induzione strutturale sull'altezza dell'albero.

(Dall'albero sintattico alla derivazione a sinistra)

La prova è per induzione strutturale sull'altezza dell'albero.

Passo Base. L'altezza minima che può avere un albero sintattico con radice in V e un prodotto che sia una stringa di terminali è 1.

(Dall'albero sintattico alla derivazione a sinistra)

La prova è per induzione strutturale sull'altezza dell'albero.

Passo Base. L'altezza minima che può avere un albero sintattico con radice in V e un prodotto che sia una stringa di terminali è 1.

Se $\mathcal{T}$ è un albero sintattico di altezza 1, con radice etichettata da una variabile A e con prodotto w, dove $w \in T^*$, allora l'albero è formato dalla radice A che ha come figli, da sinistra, etichettati $X_1, X_2, \ldots, X_k$ e $X_1 X_2 \cdots X_k = w$.

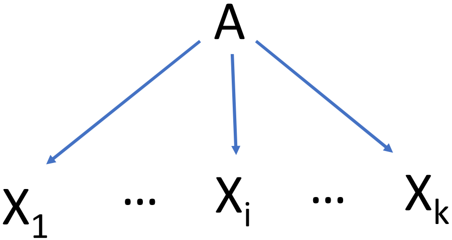
(Dall'albero sintattico alla derivazione a sinistra)

La prova è per induzione strutturale sull'altezza dell'albero.

Passo Base. L'altezza minima che può avere un albero sintattico con radice in V e un prodotto che sia una stringa di terminali è 1.

Se $\mathcal{T}$ è un albero sintattico di altezza 1, con radice etichettata da una variabile A e con prodotto w, dove $w \in T^*$, allora l'albero è formato dalla radice A che ha come figli, da sinistra, etichettati $X_1, X_2, \ldots, X_k$ e $X_1 X_2 \cdots X_k = w$.

Per definizione di albero sintattico, $A \to X_1 X_2 \cdots X_k$ è una produzione in G e quindi $A \underset{lm}{\overset{*}{\Rightarrow}} X_1 X_2 \cdots X_k = w$ nella grammatica G.



$$X_1 \dots X_k = w$$

Figura: Passo Base.

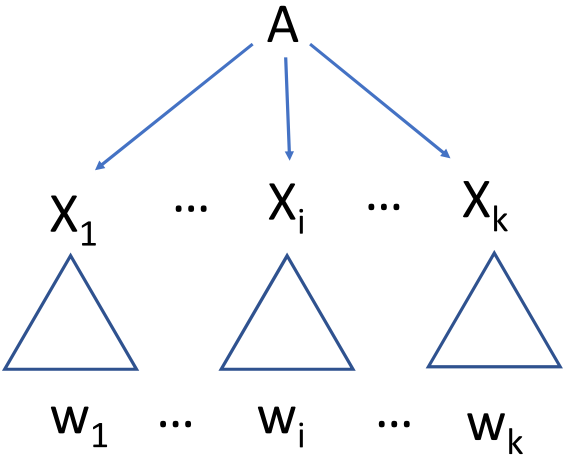


Figura: Passo Ricorsivo.

(Dall'albero sintattico alla derivazione a sinistra)

(Dall'albero sintattico alla derivazione a sinistra)

Passo Ricorsivo. Sia $\mathcal{T}$ un albero sintattico di altezza $n > 1$, con radice etichettata da una variabile A e con prodotto w, dove $w \in T^*$. La radice avrà figli, da sinistra, etichettati $X_1, X_2, \ldots, X_k$, con $X_1, X_2, \ldots, X_k \in V \cup T$.

(Dall'albero sintattico alla derivazione a sinistra)

Passo Ricorsivo. Sia $\mathcal{T}$ un albero sintattico di altezza $n > 1$, con radice etichettata da una variabile A e con prodotto w, dove $w \in T^*$. La radice avrà figli, da sinistra, etichettati $X_1, X_2, \ldots, X_k$, con $X_1, X_2, \ldots, X_k \in V \cup T$.

Se $X_i \in T$, poniamo $X_i = w_i$. Se $X_i \in V$ allora X_i è la radice di un sottoalbero (sintattico) $\mathcal{T}_i$ di $\mathcal{T}$ di altezza $n - 1 \geq 1$ (che soddisfa le ipotesi). Chiamiamo w_i il prodotto di $\mathcal{T}_i$.

(Dall'albero sintattico alla derivazione a sinistra)

Passo Ricorsivo. Sia $\mathcal{T}$ un albero sintattico di altezza $n > 1$, con radice etichettata da una variabile A e con prodotto w, dove $w \in T^*$. La radice avrà figli, da sinistra, etichettati $X_1, X_2, \ldots, X_k$, con $X_1, X_2, \ldots, X_k \in V \cup T$.

Se $X_i \in T$, poniamo $X_i = w_i$. Se $X_i \in V$ allora X_i è la radice di un sottoalbero (sintattico) $\mathcal{T}_i$ di $\mathcal{T}$ di altezza $n - 1 \geq 1$ (che soddisfa le ipotesi). Chiamiamo w_i il prodotto di $\mathcal{T}_i$.

Osserviamo che $w_1 w_2 \cdots w_k = w$

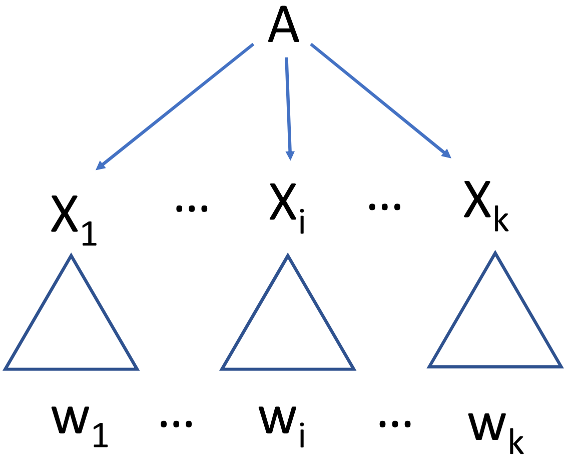


Figura: Passo Ricorsivo.

(Dall'albero sintattico alla derivazione a sinistra)

(Dall'albero sintattico alla derivazione a sinistra)

Passo Ricorsivo. Sia $\mathcal{T}$ un albero sintattico di altezza $n > 1$, con radice etichettata da una variabile A e con prodotto w, dove $w \in T^*$. La radice avrà figli, da sinistra, etichettati $X_1, X_2, \ldots, X_k$, con $X_1, X_2, \ldots, X_k \in V \cup T$.

(Dall'albero sintattico alla derivazione a sinistra)

Passo Ricorsivo. Sia $\mathcal{T}$ un albero sintattico di altezza $n > 1$, con radice etichettata da una variabile A e con prodotto w, dove $w \in T^*$. La radice avrà figli, da sinistra, etichettati $X_1, X_2, \ldots, X_k$, con $X_1, X_2, \ldots, X_k \in V \cup T$.

Se $X_i \in T$, poniamo $X_i = w_i$. Se $X_i \in V$ allora X_i è la radice di un sottoalbero (sintattico) $\mathcal{T}_i$ di $\mathcal{T}$ di altezza $n - 1 \geq 1$ (che soddisfa le ipotesi). Chiamiamo w_i il prodotto di $\mathcal{T}_i$.

(Dall'albero sintattico alla derivazione a sinistra)

Passo Ricorsivo. Sia $\mathcal{T}$ un albero sintattico di altezza $n > 1$, con radice etichettata da una variabile A e con prodotto w, dove $w \in T^*$. La radice avrà figli, da sinistra, etichettati $X_1, X_2, \ldots, X_k$, con $X_1, X_2, \ldots, X_k \in V \cup T$.

Se $X_i \in T$, poniamo $X_i = w_i$. Se $X_i \in V$ allora X_i è la radice di un sottoalbero (sintattico) $\mathcal{T}_i$ di $\mathcal{T}$ di altezza $n - 1 \geq 1$ (che soddisfa le ipotesi). Chiamiamo w_i il prodotto di $\mathcal{T}_i$.

Osserviamo che:

(Dall'albero sintattico alla derivazione a sinistra)

Passo Ricorsivo. Sia $\mathcal{T}$ un albero sintattico di altezza $n > 1$, con radice etichettata da una variabile A e con prodotto w, dove $w \in T^*$. La radice avrà figli, da sinistra, etichettati $X_1, X_2, \ldots, X_k$, con $X_1, X_2, \ldots, X_k \in V \cup T$.

Se $X_i \in T$, poniamo $X_i = w_i$. Se $X_i \in V$ allora X_i è la radice di un sottoalbero (sintattico) $\mathcal{T}_i$ di $\mathcal{T}$ di altezza $n - 1 \geq 1$ (che soddisfa le ipotesi). Chiamiamo w_i il prodotto di $\mathcal{T}_i$.

Osserviamo che:

1. $A \underset{lm}{\overset{*}{\Rightarrow}} X_1 X_2 \cdots X_k$ (per definizione di albero sintattico)

(Dall'albero sintattico alla derivazione a sinistra)

Passo Ricorsivo. Sia $\mathcal{T}$ un albero sintattico di altezza $n > 1$, con radice etichettata da una variabile A e con prodotto w, dove $w \in T^*$. La radice avrà figli, da sinistra, etichettati $X_1, X_2, \ldots, X_k$, con $X_1, X_2, \ldots, X_k \in V \cup T$.

Se $X_i \in T$, poniamo $X_i = w_i$. Se $X_i \in V$ allora X_i è la radice di un sottoalbero (sintattico) $\mathcal{T}_i$ di $\mathcal{T}$ di altezza $n - 1 \geq 1$ (che soddisfa le ipotesi). Chiamiamo w_i il prodotto di $\mathcal{T}_i$.

Osserviamo che:

1. $A \underset{lm}{\overset{*}{\Rightarrow}} X_1 X_2 \cdots X_k$ (per definizione di albero sintattico)
2. Se $X_i \in V$ allora $X_i \underset{lm}{\overset{*}{\Rightarrow}} w_i$ (per ipotesi induttiva)

(Dall'albero sintattico alla derivazione a sinistra)

Passo Ricorsivo. Sia $\mathcal{T}$ un albero sintattico di altezza $n > 1$, con radice etichettata da una variabile A e con prodotto w, dove $w \in T^*$. La radice avrà figli, da sinistra, etichettati $X_1, X_2, \ldots, X_k$, con $X_1, X_2, \ldots, X_k \in V \cup T$.

Se $X_i \in T$, poniamo $X_i = w_i$. Se $X_i \in V$ allora X_i è la radice di un sottoalbero (sintattico) $\mathcal{T}_i$ di $\mathcal{T}$ di altezza $n - 1 \geq 1$ (che soddisfa le ipotesi). Chiamiamo w_i il prodotto di $\mathcal{T}_i$.

Osserviamo che:

1. $A \underset{lm}{\overset{*}{\Rightarrow}} X_1 X_2 \cdots X_k$ (per definizione di albero sintattico)
2. Se $X_i \in V$ allora $X_i \underset{lm}{\overset{*}{\Rightarrow}} w_i$ (per ipotesi induttiva)
3. $w_1 w_2 \cdots w_k = w$

Proviamo, per induzione su i, che per $i = 0, \ldots, k$,

$$A \underset{lm}{\overset{*}{\Rightarrow}} w_1 w_2 \cdots w_i X_{i+1} X_{i+2} \cdots X_k$$

Proviamo, per induzione su i, che per $i = 0, \ldots, k$,

$$A \underset{lm}{\overset{*}{\Rightarrow}} w_1 w_2 \cdots w_i X_{i+1} X_{i+2} \cdots X_k$$

Passo Base. $i = 0$ e sappiamo che $A \underset{lm}{\overset{*}{\Rightarrow}} X_1 X_2 \cdots X_k$.

Proviamo, per induzione su i, che per $i = 0, \ldots, k$,

$$A \underset{lm}{\overset{*}{\Rightarrow}} w_1 w_2 \cdots w_i X_{i+1} X_{i+2} \cdots X_k$$

Passo Base. $i = 0$ e sappiamo che $A \underset{lm}{\overset{*}{\Rightarrow}} X_1 X_2 \cdots X_k$.

Passo Induttivo. Supponiamo che

$$A \underset{lm}{\overset{*}{\Rightarrow}} w_1 w_2 \cdots w_{i-1} X_i X_{i+2} \cdots X_k$$

Proviamo, per induzione su i, che per $i = 0, \ldots, k$,

$$A \underset{lm}{\overset{*}{\Rightarrow}} w_1 w_2 \cdots w_i X_{i+1} X_{i+2} \cdots X_k$$

Passo Base. $i = 0$ e sappiamo che $A \underset{lm}{\overset{*}{\Rightarrow}} X_1 X_2 \cdots X_k$.

Passo Induttivo. Supponiamo che

$$A \underset{lm}{\overset{*}{\Rightarrow}} w_1 w_2 \cdots w_{i-1} X_i X_{i+2} \cdots X_k$$

Se $X_i = w_i$, allora

$$A \underset{lm}{\overset{*}{\Rightarrow}} w_1 w_2 \cdots w_i X_{i+1} X_{i+2} \cdots X_k$$

Proviamo, per induzione su i, che per $i = 0, \ldots, k$,

$$A \underset{lm}{\overset{*}{\Rightarrow}} w_1 w_2 \cdots w_i X_{i+1} X_{i+2} \cdots X_k$$

Passo Base. $i = 0$ e sappiamo che $A \underset{lm}{\overset{*}{\Rightarrow}} X_1 X_2 \cdots X_k$.

Passo Induttivo. Supponiamo che

$$A \underset{lm}{\overset{*}{\Rightarrow}} w_1 w_2 \cdots w_{i-1} X_i X_{i+2} \cdots X_k$$

Se $X_i = w_i$, allora

$$A \underset{lm}{\overset{*}{\Rightarrow}} w_1 w_2 \cdots w_i X_{i+1} X_{i+2} \cdots X_k$$

Altrimenti $X_i \underset{lm}{\overset{*}{\Rightarrow}} w_i$ e grazie alla **proprietà dell'inserzione**

$$A \underset{lm}{\overset{*}{\Rightarrow}} w_1 w_2 \cdots w_i X_{i+1} X_{i+2} \cdots X_k$$

Proviamo, per induzione su i, che per $i = 0, \ldots, k$,

$$A \overset{*}{\underset{lm}{\Rightarrow}} w_1 w_2 \cdots w_i X_{i+1} X_{i+2} \cdots X_k$$

Passo Base. $i = 0$ e sappiamo che $A \overset{*}{\underset{lm}{\Rightarrow}} X_1 X_2 \cdots X_k$.

Passo Induttivo. Supponiamo che

$$A \overset{*}{\underset{lm}{\Rightarrow}} w_1 w_2 \cdots w_{i-1} X_i X_{i+2} \cdots X_k$$

Se $X_i = w_i$, allora

$$A \overset{*}{\underset{lm}{\Rightarrow}} w_1 w_2 \cdots w_i X_{i+1} X_{i+2} \cdots X_k$$

Altrimenti $X_i \overset{*}{\underset{lm}{\Rightarrow}} w_i$ e grazie alla **proprietà dell'inserzione**

$$A \overset{*}{\underset{lm}{\Rightarrow}} w_1 w_2 \cdots w_i X_{i+1} X_{i+2} \cdots X_k$$

Per $i = k$ otteniamo

$$A \overset{*}{\underset{lm}{\Rightarrow}} w_1 w_2 \cdots w_k = w$$

Teorema

Sia $G = (V, T, P, S)$ *una grammatica context free.*

Se esiste un albero sintattico con radice etichettata da una variabile A *e con prodotto* w*, dove* $w \in T^*$*, allora esiste una derivazione a destra* $A \underset{rm}{\overset{*}{\Rightarrow}} w$ *nella grammatica* G*.*

La dimostrazione di questo teorema è molto simile alla precedente.

Adesso vogliamo provare un enunciato che è quasi l'inverso: se esiste una derivazione $A \stackrel{*}{\Rightarrow} w$ nella grammatica G, dove $A \in V$ e $w \in T^*$, allora w è il prodotto di un albero sintattico con radice etichettata A.

Ci serve una seconda proprietà delle grammatiche context-free. Questa proprietà afferma che se una stringa α di variabili e terminali può essere derivata da $X_1 X_2 \cdots X_n$, dove ogni X_i è una variabile o un terminale, esiste una fattorizzazione $\alpha = \alpha_1 \alpha_2 \cdots \alpha_n$ di α tale che ogni X_i deriva la corrispondente stringa α_i.

Chiameremo questa proprietà la **proprietà della fattorizzazione**.

Consideriamo la grammatica context-free G definita dalle produzioni

$$S \rightarrow aSbS \mid bSaS \mid \epsilon$$

Consideriamo la grammatica context-free G definita dalle produzioni

$$S \to aSbS \mid bSaS \mid \epsilon$$

Consideriamo

$$aSbS \overset{*}{\Rightarrow} aabb$$

Proprietà della fattorizzazione - Esempi

Consideriamo la grammatica context-free G definita dalle produzioni

$$S \to aSbS \mid bSaS \mid \epsilon$$

Consideriamo

$$aSbS \stackrel{*}{\Rightarrow} aabb$$

Per la proprietà della fattorizzazione, esistono stringhe x, y tali che

$$axby = aabb, \quad S \stackrel{*}{\Rightarrow} x, \quad S \stackrel{*}{\Rightarrow} y$$

Proprietà della fattorizzazione - Esempi

Consideriamo la grammatica context-free G definita dalle produzioni

$$S \to aSbS \mid bSaS \mid \epsilon$$

Consideriamo

$$aSbS \stackrel{*}{\Rightarrow} aabb$$

Per la proprietà della fattorizzazione, esistono stringhe x, y tali che

$$axby = aabb, \quad S \stackrel{*}{\Rightarrow} x, \quad S \stackrel{*}{\Rightarrow} y$$

Soluzione: $x = ab$, $y = \epsilon$.

Proprietà della fattorizzazione - Esempi

Consideriamo la grammatica context-free G definita dalle produzioni

$$S \to aSbS \mid bSaS \mid \epsilon$$

Consideriamo

$$aSbS \stackrel{*}{\Rightarrow} aabb$$

Per la proprietà della fattorizzazione, esistono stringhe x, y tali che

$$axby = aabb, \quad S \stackrel{*}{\Rightarrow} x, \quad S \stackrel{*}{\Rightarrow} y$$

Soluzione: $x = ab$, $y = \epsilon$.

Infatti

$$axby = aabb, \quad S \stackrel{*}{\Rightarrow} ab, \quad S \stackrel{*}{\Rightarrow} \epsilon$$

Proprietà della fattorizzazione - Esempi

Consideriamo la grammatica context-free G di simbolo iniziale R e produzioni

$$
\begin{aligned}
R &\rightarrow XRX \mid S, \quad S \rightarrow aTb \mid bTa \\
T &\rightarrow XTX \mid X \mid \epsilon, \quad X \rightarrow a \mid b
\end{aligned}
$$

Consideriamo la grammatica context-free G di simbolo iniziale R e produzioni

$$\begin{aligned} R &\rightarrow XRX \mid S, \quad S \rightarrow aTb \mid bTa \\ T &\rightarrow XTX \mid X \mid \epsilon, \quad X \rightarrow a \mid b \end{aligned}$$

Consideriamo

$$XXRXX \stackrel{*}{\Rightarrow} abbaab$$

Proprietà della fattorizzazione - Esempi

Consideriamo la grammatica context-free G di simbolo iniziale R e produzioni

$$\begin{aligned} R &\rightarrow XRX \mid S, \quad S \rightarrow aTb \mid bTa \\ T &\rightarrow XTX \mid X \mid \epsilon, \quad X \rightarrow a \mid b \end{aligned}$$

Consideriamo

$$XXRXX \overset{*}{\Rightarrow} abbaab$$

Per la proprietà della fattorizzazione, esistono x_1, x_2, x_3, x_4, x_5 tali che

$$abbaab = x_1x_2x_3x_4x_5, \quad X \overset{*}{\Rightarrow} x_1, \quad X \overset{*}{\Rightarrow} x_2,$$
$$R \overset{*}{\Rightarrow} x_3, \quad X \overset{*}{\Rightarrow} x_4, \quad X \overset{*}{\Rightarrow} x_5$$

Proprietà della fattorizzazione - Esempi

Consideriamo la grammatica context-free G di simbolo iniziale R e produzioni

$$\begin{array}{rcl} R & \rightarrow & XRX \mid S, \quad S \rightarrow aTb \mid bTa \\ T & \rightarrow & XTX \mid X \mid \epsilon, \quad X \rightarrow a \mid b \end{array}$$

Consideriamo

$$XXRXX \stackrel{*}{\Rightarrow} abbaab$$

Per la proprietà della fattorizzazione, esistono x_1, x_2, x_3, x_4, x_5 tali che

$$abbaab = x_1x_2x_3x_4x_5, \quad X \stackrel{*}{\Rightarrow} x_1, \quad X \stackrel{*}{\Rightarrow} x_2,$$

$$R \stackrel{*}{\Rightarrow} x_3, \quad X \stackrel{*}{\Rightarrow} x_4, \quad X \stackrel{*}{\Rightarrow} x_5$$

Soluzione: $x_1 = x_4 = a$, $x_2 = x_5 = b$, $x_3 = ba$.

Proprietà della fattorizzazione - Esempi

Consideriamo la grammatica context-free G di simbolo iniziale R e produzioni

$$\begin{aligned} R &\rightarrow XRX \mid S, \quad S \rightarrow aTb \mid bTa \\ T &\rightarrow XTX \mid X \mid \epsilon, \quad X \rightarrow a \mid b \end{aligned}$$

Consideriamo

$$XXRXX \stackrel{*}{\Rightarrow} abbaab$$

Per la proprietà della fattorizzazione, esistono x_1, x_2, x_3, x_4, x_5 tali che

$$abbaab = x_1x_2x_3x_4x_5, \quad X \stackrel{*}{\Rightarrow} x_1, \quad X \stackrel{*}{\Rightarrow} x_2,$$

$$R \stackrel{*}{\Rightarrow} x_3, \quad X \stackrel{*}{\Rightarrow} x_4, \quad X \stackrel{*}{\Rightarrow} x_5$$

Soluzione: $x_1 = x_4 = a$, $x_2 = x_5 = b$, $x_3 = ba$.
Infatti

$$abbaab = x_1x_2x_3x_4x_5, \quad X \stackrel{*}{\Rightarrow} a, \quad X \stackrel{*}{\Rightarrow} b, \quad R \stackrel{*}{\Rightarrow} ba$$

- Consideriamo la grammatica context-free G definita dalle produzioni
$$S \to aS \mid Sb \mid a \mid b$$

- Consideriamo la grammatica context-free G definita dalle produzioni

$$S \to aS \mid Sb \mid a \mid b$$

Supponiamo che per una stringa w si abbia $S \Rightarrow aS \stackrel{*}{\Rightarrow} w$.

Proprietà della fattorizzazione - Esempi

- Consideriamo la grammatica context-free G definita dalle produzioni

  $$S \rightarrow aS \mid Sb \mid a \mid b$$

  Supponiamo che per una stringa w si abbia $S \Rightarrow aS \overset{*}{\Rightarrow} w$.

  Per la proprietà della fattorizzazione, esiste una stringa x tale che $w = ax$ e $S \overset{*}{\Rightarrow} x$.

Proprietà della fattorizzazione - Esempi

- Consideriamo la grammatica context-free G definita dalle produzioni

  $$S \to aS \mid Sb \mid a \mid b$$

  Supponiamo che per una stringa w si abbia $S \Rightarrow aS \stackrel{*}{\Rightarrow} w$.

  Per la proprietà della fattorizzazione, esiste una stringa x tale che $w = ax$ e $S \stackrel{*}{\Rightarrow} x$.

- Consideriamo la grammatica context-free G definita dalle produzioni

  $$S \to aSbS \mid bSaS \mid \epsilon$$

Proprietà della fattorizzazione - Esempi

- Consideriamo la grammatica context-free G definita dalle produzioni
$$S \to aS \mid Sb \mid a \mid b$$
Supponiamo che per una stringa w si abbia $S \Rightarrow aS \stackrel{*}{\Rightarrow} w$.
Per la proprietà della fattorizzazione, esiste una stringa x tale che $w = ax$ e $S \stackrel{*}{\Rightarrow} x$.
- Consideriamo la grammatica context-free G definita dalle produzioni
$$S \to aSbS \mid bSaS \mid \epsilon$$
Supponiamo che per una stringa w si abbia $S \Rightarrow aSbS \stackrel{*}{\Rightarrow} w$

Proprietà della fattorizzazione - Esempi

- Consideriamo la grammatica context-free G definita dalle produzioni
$$S \to aS \mid Sb \mid a \mid b$$
Supponiamo che per una stringa w si abbia $S \Rightarrow aS \overset{*}{\Rightarrow} w$.
Per la proprietà della fattorizzazione, esiste una stringa x tale che $w = ax$ e $S \overset{*}{\Rightarrow} x$.
- Consideriamo la grammatica context-free G definita dalle produzioni
$$S \to aSbS \mid bSaS \mid \epsilon$$
Supponiamo che per una stringa w si abbia $S \Rightarrow aSbS \overset{*}{\Rightarrow} w$
Per la proprietà della fattorizzazione, esistono stringhe x, y tali che $w = axby$, con $S \overset{*}{\Rightarrow} x$, $S \overset{*}{\Rightarrow} y$.

Proposizione

[proprietà della fattorizzazione]. *Sia* $G = (V, T, P, S)$ *una grammatica context free. Se*

$$X_1 X_2 \cdots X_n \overset{*}{\underset{G}{\Rightarrow}} \alpha$$

con $X_1, X_2, \ldots, X_n \in V \cup T$, $n \geq 1$, $\alpha \in (V \cup T)^*$, *allora esistono* $\alpha_1, \alpha_2, \ldots, \alpha_n \in (V \cup T)^*$ *tali che*

1. $\alpha = \alpha_1 \alpha_2 \cdots \alpha_n$

Proposizione

[proprietà della fattorizzazione]. *Sia* $G = (V, T, P, S)$ *una grammatica context free. Se*

$$X_1 X_2 \cdots X_n \overset{*}{\underset{G}{\Rightarrow}} \alpha$$

con $X_1, X_2, \ldots, X_n \in V \cup T$, $n \geq 1$, $\alpha \in (V \cup T)^*$, *allora esistono* $\alpha_1, \alpha_2, \ldots, \alpha_n \in (V \cup T)^*$ *tali che*

1. $\alpha = \alpha_1 \alpha_2 \cdots \alpha_n$
2. $X_i \overset{*}{\underset{G}{\Rightarrow}} \alpha_i$, *per* $i = 1, 2, \ldots, n$

Proposizione

[proprietà della fattorizzazione]. *Sia* $G = (V, T, P, S)$ *una grammatica context free. Se*

$$X_1 X_2 \cdots X_n \underset{G}{\overset{*}{\Rightarrow}} \alpha$$

con $X_1, X_2, \ldots, X_n \in V \cup T$, $n \geq 1$, $\alpha \in (V \cup T)^*$, *allora esistono* $\alpha_1, \alpha_2, \ldots, \alpha_n \in (V \cup T)^*$ *tali che*

1. $\alpha = \alpha_1 \alpha_2 \cdots \alpha_n$
2. $X_i \underset{G}{\overset{*}{\Rightarrow}} \alpha_i$, *per* $i = 1, 2, \ldots, n$
3. *Il numero dei passi (numero di derivazioni dirette) in* $X_1 X_2 \cdots X_n \underset{G}{\overset{*}{\Rightarrow}} \alpha$ *è la somma del numero dei passi in* $X_i \underset{G}{\overset{*}{\Rightarrow}} \alpha_i$, *per* $i = 1, 2, \ldots, n$

Proposizione

[proprietà della fattorizzazione]. *Sia $G = (V, T, P, S)$ una grammatica context free. Se*

$$X_1 X_2 \cdots X_n \overset{*}{\underset{G}{\Rightarrow}} \alpha$$

con $X_1, X_2, \ldots, X_n \in V \cup T$, $n \geq 1$, $\alpha \in (V \cup T)^*$, allora esistono $\alpha_1, \alpha_2, \ldots, \alpha_n \in (V \cup T)^*$ tali che

1. $\alpha = \alpha_1 \alpha_2 \cdots \alpha_n$
2. $X_i \overset{*}{\underset{G}{\Rightarrow}} \alpha_i$, *per* $i = 1, 2, \ldots, n$
3. *Il numero dei passi (numero di derivazioni dirette) in* $X_1 X_2 \cdots X_n \overset{*}{\underset{G}{\Rightarrow}} \alpha$ *è la somma del numero dei passi in* $X_i \overset{*}{\underset{G}{\Rightarrow}} \alpha_i$, *per* $i = 1, 2, \ldots, n$
4. *Se* $X_i \in T$ *allora* $X_i = \alpha_i$.

Teorema

Sia $G = (V, T, P, S)$ *una grammatica context free.*

Se esiste una derivazione $A \stackrel{*}{\Rightarrow} w$ *nella grammatica* G*, dove* $A \in V$ *e* $w \in T^*$*, allora esiste un albero sintattico con radice etichettata* A *e con prodotto* w*.*

(Dalla derivazione all'albero sintattico)

(Dalla derivazione all'albero sintattico)

La prova è per induzione sul numero dei passi (numero di derivazioni dirette) nella derivazione di w da A.

(Dalla derivazione all'albero sintattico)

La prova è per induzione sul numero dei passi (numero di derivazioni dirette) nella derivazione di w da A.

Passo Base. Il numero minimo di passi che può avere una derivazione $A \stackrel{*}{\Rightarrow} w$ di una stringa w di terminali da una variabile A è 1.

(Dalla derivazione all'albero sintattico)

La prova è per induzione sul numero dei passi (numero di derivazioni dirette) nella derivazione di w da A.

Passo Base. Il numero minimo di passi che può avere una derivazione $A \stackrel{*}{\Rightarrow} w$ di una stringa w di terminali da una variabile A è 1.

Quindi $A \Rightarrow w$ e allora $A \rightarrow w$ è una produzione in G.

(Dalla derivazione all'albero sintattico)

La prova è per induzione sul numero dei passi (numero di derivazioni dirette) nella derivazione di w da A.

Passo Base. Il numero minimo di passi che può avere una derivazione $A \stackrel{*}{\Rightarrow} w$ di una stringa w di terminali da una variabile A è 1.

Quindi $A \Rightarrow w$ e allora $A \rightarrow w$ è una produzione in G.

Siccome $w \in T^*$, allora $w = \epsilon$ oppure $w = w_1 w_2 \cdots w_k$, con $w_j \in T$, $1 \leq j \leq k$.

(Dalla derivazione all'albero sintattico)

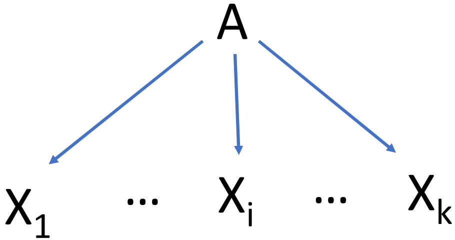
La prova è per induzione sul numero dei passi (numero di derivazioni dirette) nella derivazione di w da A.

Passo Base. Il numero minimo di passi che può avere una derivazione $A \stackrel{*}{\Rightarrow} w$ di una stringa w di terminali da una variabile A è 1.

Quindi $A \Rightarrow w$ e allora $A \rightarrow w$ è una produzione in G.

Siccome $w \in T^*$, allora $w = \epsilon$ oppure $w = w_1 w_2 \cdots w_k$, con $w_j \in T$, $1 \leq j \leq k$.

L'albero $\mathcal{T}$ di altezza 1, con radice etichettata da A e unico figlio etichettato ϵ se $w = \epsilon$ oppure figli, da sinistra, etichettati $w_1, w_2, \ldots, w_k$ se $w \neq \epsilon$, è per definizione un albero sintattico di radice A e prodotto w.



$$X_1 \ldots X_k = w$$

Figura: Passo Base.

(Dalla derivazione all'albero sintattico)

(Dalla derivazione all'albero sintattico)

Passo Ricorsivo. Supponiamo che $A \stackrel{*}{\Rightarrow} w$ in un numero di passi $n > 1$.

(Dalla derivazione all'albero sintattico)

Passo Ricorsivo. Supponiamo che $A \overset{*}{\Rightarrow} w$ in un numero di passi $n > 1$.

Allora esistono $X_1, X_2, \ldots, X_k \in V \cup T$ tali che

$$A \Rightarrow X_1 X_2 \cdots X_k \overset{*}{\Rightarrow} w$$

Osserviamo che:

(Dalla derivazione all'albero sintattico)

Passo Ricorsivo. Supponiamo che $A \overset{*}{\Rightarrow} w$ in un numero di passi $n > 1$.

Allora esistono $X_1, X_2, \ldots, X_k \in V \cup T$ tali che

$$A \Rightarrow X_1 X_2 \cdots X_k \overset{*}{\Rightarrow} w$$

Osserviamo che:

1. $A \to X_1 X_2 \cdots X_k$ è una produzione in G (per definizione di derivazione diretta)

(Dalla derivazione all'albero sintattico)

Passo Ricorsivo. Supponiamo che $A \stackrel{*}{\Rightarrow} w$ in un numero di passi $n > 1$.

Allora esistono $X_1, X_2, \ldots, X_k \in V \cup T$ tali che

$$A \Rightarrow X_1 X_2 \cdots X_k \stackrel{*}{\Rightarrow} w$$

Osserviamo che:

1. $A \to X_1 X_2 \cdots X_k$ è una produzione in G (per definizione di derivazione diretta)
2. Per la **proprietà della fattorizzazione**, esiste una fattorizzazione $w = w_1 w_2 \cdots w_k$ di w, con $w_j \in T^*$, $1 \leq j \leq k$ tale che $X_j \stackrel{*}{\Rightarrow} w_j$

(Dalla derivazione all'albero sintattico)

Passo Ricorsivo. Supponiamo che $A \overset{*}{\Rightarrow} w$ in un numero di passi $n > 1$.

Allora esistono $X_1, X_2, \ldots, X_k \in V \cup T$ tali che

$$A \Rightarrow X_1 X_2 \cdots X_k \overset{*}{\Rightarrow} w$$

Osserviamo che:

1. $A \rightarrow X_1 X_2 \cdots X_k$ è una produzione in G (per definizione di derivazione diretta)
2. Per la **proprietà della fattorizzazione**, esiste una fattorizzazione $w = w_1 w_2 \cdots w_k$ di w, con $w_j \in T^*$, $1 \leq j \leq k$ tale che $X_j \overset{*}{\Rightarrow} w_j$
3. Inoltre $w_j = X_j \in T$ oppure $X_j \overset{*}{\Rightarrow} w_j$ in meno di n passi.

Osserviamo che:

Osserviamo che:

1. $A \to X_1 X_2 \cdots X_k$ è una produzione in G (per definizione di derivazione diretta)

Osserviamo che:

1. $A \rightarrow X_1 X_2 \cdots X_k$ è una produzione in G (per definizione di derivazione diretta)
2. Per la **proprietà della fattorizzazione**, esiste una fattorizzazione $w = w_1 w_2 \cdots w_k$ di w, con $w_j \in T^*$, $1 \leq j \leq k$ tale che $X_j \stackrel{*}{\Rightarrow} w_j$

Osserviamo che:

1. $A \rightarrow X_1 X_2 \cdots X_k$ è una produzione in G (per definizione di derivazione diretta)
2. Per la **proprietà della fattorizzazione**, esiste una fattorizzazione $w = w_1 w_2 \cdots w_k$ di w, con $w_j \in T^*$, $1 \leq j \leq k$ tale che $X_j \stackrel{*}{\Rightarrow} w_j$
3. Inoltre $w_j = X_j \in T$ oppure $X_j \stackrel{*}{\Rightarrow} w_j$ in meno di n passi.

Osserviamo che:

1. $A \to X_1 X_2 \cdots X_k$ è una produzione in G (per definizione di derivazione diretta)
2. Per la **proprietà della fattorizzazione**, esiste una fattorizzazione $w = w_1 w_2 \cdots w_k$ di w, con $w_j \in T^*$, $1 \leq j \leq k$ tale che $X_j \overset{*}{\Rightarrow} w_j$
3. Inoltre $w_j = X_j \in T$ oppure $X_j \overset{*}{\Rightarrow} w_j$ in meno di n passi.

Se $w_j = X_j \in T$, denotiamo con $\mathcal{T}_j$ l'albero ridotto al solo nodo X_j. Altrimenti, se $X_j \overset{*}{\Rightarrow} w_j$ in meno di n passi, con $w_j \neq X_j$, esiste un albero sintattico $\mathcal{T}_j$ di radice X_j e prodotto w_j, per ipotesi induttiva.

Osserviamo che:

1. $A \to X_1 X_2 \cdots X_k$ è una produzione in G (per definizione di derivazione diretta)
2. Per la **proprietà della fattorizzazione**, esiste una fattorizzazione $w = w_1 w_2 \cdots w_k$ di w, con $w_j \in T^*$, $1 \leq j \leq k$ tale che $X_j \stackrel{*}{\Rightarrow} w_j$
3. Inoltre $w_j = X_j \in T$ oppure $X_j \stackrel{*}{\Rightarrow} w_j$ in meno di n passi.

Se $w_j = X_j \in T$, denotiamo con $\mathcal{T}_j$ l'albero ridotto al solo nodo X_j. Altrimenti, se $X_j \stackrel{*}{\Rightarrow} w_j$ in meno di n passi, con $w_j \neq X_j$, esiste un albero sintattico $\mathcal{T}_j$ di radice X_j e prodotto w_j, per ipotesi induttiva.

Consideriamo l'albero $\mathcal{T}$ costruito mediante A e gli alberi $\mathcal{T}_j$, $1 \leq j \leq k$, cioè l'albero di radice A, con figli da sinistra, etichettati $X_1, X_2, \ldots, X_k$, e dove ogni X_j è la radice del sottoalbero $\mathcal{T}_j$.

Osserviamo che:

1. $A \to X_1 X_2 \cdots X_k$ è una produzione in G (per definizione di derivazione diretta)
2. Per la **proprietà della fattorizzazione**, esiste una fattorizzazione $w = w_1 w_2 \cdots w_k$ di w, con $w_j \in T^*$, $1 \leq j \leq k$ tale che $X_j \stackrel{*}{\Rightarrow} w_j$
3. Inoltre $w_j = X_j \in T$ oppure $X_j \stackrel{*}{\Rightarrow} w_j$ in meno di n passi.

Se $w_j = X_j \in T$, denotiamo con $\mathcal{T}_j$ l'albero ridotto al solo nodo X_j. Altrimenti, se $X_j \stackrel{*}{\Rightarrow} w_j$ in meno di n passi, con $w_j \neq X_j$, esiste un albero sintattico $\mathcal{T}_j$ di radice X_j e prodotto w_j, per ipotesi induttiva.

Consideriamo l'albero $\mathcal{T}$ costruito mediante A e gli alberi $\mathcal{T}_j$, $1 \leq j \leq k$, cioè l'albero di radice A, con figli da sinistra, etichettati $X_1, X_2, \ldots, X_k$, e dove ogni X_j è la radice del sottoalbero $\mathcal{T}_j$.

$\mathcal{T}$ è un albero sintattico con radice etichettata A e con prodotto w.

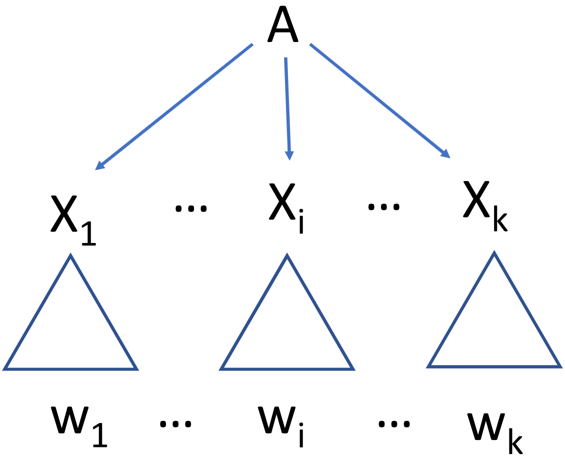


Figura: Passo Ricorsivo.

Definizione

Una grammatica context-free $G = (V, T, P, S)$ *è ambigua se esiste almeno una stringa* w *in* T^* *per la quale possiamo trovare due alberi sintattici distinti, ciascuno con radice etichettata* S *e con prodotto* w.

Definizione

Una grammatica context-free $G = (V, T, P, S)$ *è ambigua se esiste almeno una stringa* w *in* T^* *per la quale possiamo trovare due alberi sintattici distinti, ciascuno con radice etichettata* S *e con prodotto* w*.*

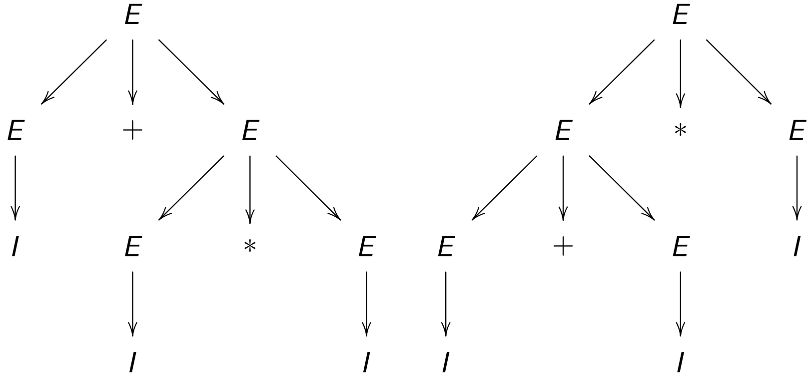
- **Nota**. $G = (V, T, P, S)$ è non ambigua se per ogni w in T^* possiamo trovare al più un albero sintattico con radice etichettata S e con prodotto w.

Definizione

Una grammatica context-free $G = (V, T, P, S)$ è ambigua se esiste almeno una stringa w in T^* per la quale possiamo trovare due alberi sintattici distinti, ciascuno con radice etichettata S e con prodotto w.

- **Nota**. $G = (V, T, P, S)$ è non ambigua se per ogni w in T^* possiamo trovare al più un albero sintattico con radice etichettata S e con prodotto w.

  La figura seguente mostra due alberi sintattici distinti della grammatica $G_E = (\{E, I\}, T, P, E)$ per espressioni aritmetiche che hanno come prodotto $I + I * I$.



Possiamo facilmente ricavare dai due alberi precedenti due alberi sintattici distinti di G_E con prodotto una stringa di terminali, ad esempio $a + a * a$, e quindi G_E è ambigua.

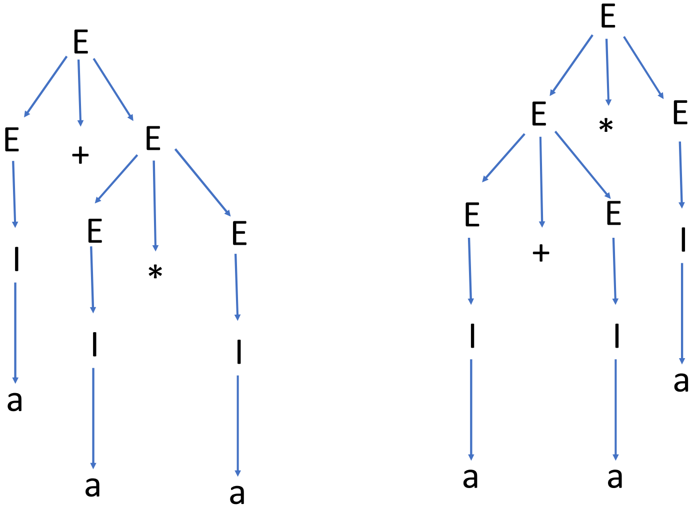


Figura: Due alberi sintattici con lo stesso prodotto.

- **Nota**. $G = (V, T, P, S)$ è non ambigua se per ogni w in T^* possiamo trovare al più un albero sintattico con radice etichettata S e con prodotto w.

- **Nota**. $G = (V, T, P, S)$ è non ambigua se per ogni w in T^* possiamo trovare al più un albero sintattico con radice etichettata S e con prodotto w.

  Gli alberi sintattici permettono di associare una struttura a un programma. Se questa struttura non è unica questo diventa un problema.

Siccome gli alberi sintattici sono legati alle derivazioni si potrebbe sospettare che l'esistenza di più alberi sintattici di una grammatica con lo stesso prodotto dipenda dal numero di possibili derivazioni di una stringa.

Siccome gli alberi sintattici sono legati alle derivazioni si potrebbe sospettare che l'esistenza di più alberi sintattici di una grammatica con lo stesso prodotto dipenda dal numero di possibili derivazioni di una stringa.

- **Nota**. L'ambiguità di G non dipende dal fatto che una stringa ammette due derivazioni distinte poiché esse possono dar luogo allo stesso albero sintattico.

Siccome gli alberi sintattici sono legati alle derivazioni si potrebbe sospettare che l'esistenza di più alberi sintattici di una grammatica con lo stesso prodotto dipenda dal numero di possibili derivazioni di una stringa.

- **Nota**. L'ambiguità di G non dipende dal fatto che una stringa ammette due derivazioni distinte poiché esse possono dar luogo allo stesso albero sintattico.

- Esempio. In G_E le due seguenti derivazioni danno luogo allo stesso albero sintattico.

$$
\begin{array}{lll}
E & \Rightarrow & E+E \Rightarrow I+E \Rightarrow a+E \Rightarrow a+I \Rightarrow a+b \\
E & \Rightarrow & E+E \Rightarrow E+I \Rightarrow I+I \Rightarrow I+b \Rightarrow a+b
\end{array}
$$

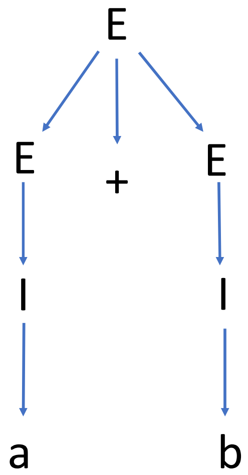


Figura: Lo stesso albero sintattico.

Teorema

Per ogni CFG $G = (V, T, P, S)$ e per ogni $w \in T^*$, la stringa w ha due alberi sintattici distinti, con radice etichettata S e con prodotto w, se e solo se w ha due distinte derivazioni sinistre da S.

Teorema

Per ogni CFG $G = (V, T, P, S)$ e per ogni $w \in T^*$, la stringa w ha due alberi sintattici distinti, con radice etichettata S e con prodotto w, se e solo se w ha due distinte derivazioni sinistre da S.

- **Nota**. Quindi una grammatica context-free $G = (V, T, P, S)$ è ambigua se esiste almeno una stringa w in T^* che ha due distinte derivazioni sinistre da S.

Eliminare le ambiguità da una grammatica?

Definizione

Due grammatiche context-free G, G' *sono* *equivalenti* *se* $L(G) = L(G')$

Eliminare le ambiguità da una grammatica?

Definizione

Due grammatiche context-free G, G' sono equivalenti se $L(G) = L(G')$

- **Nota**. Data una grammatica context-free ambigua G, a volte è possibile trasformarla in una grammatica context-free G', dove G' è equivalente a G e G' è non ambigua.

Eliminare le ambiguità da una grammatica?

Definizione

Due grammatiche context-free G, G' *sono* *equivalenti* *se* $L(G) = L(G')$

- **Nota**. Data una grammatica context-free ambigua G, a volte è possibile trasformarla in una grammatica context-free G', dove G' è equivalente a G e G' è non ambigua.

  Nota. È possibile ad esempio trasformare G_E in una grammatica context-free G'_E, dove G'_E è equivalente a G_E e G'_E è non ambigua.

Eliminare le ambiguità da una grammatica?

Una grammatica non ambigua che genera lo stesso linguaggio di G_E è $G'_E = (V', \Sigma, P', E)$, dove $V' = \{E, F, T, I\}$, $\Sigma = \{a, b, 0, 1, (,), +, *\}$ e P' consiste delle produzioni

$$
\begin{array}{rcl}
E & \rightarrow & T \mid E + T \\
T & \rightarrow & F \mid T * F \\
F & \rightarrow & I \mid (E) \\
I & \rightarrow & a \mid b \mid Ia \mid Ib \mid I0 \mid I1
\end{array}
$$

Una grammatica non ambigua

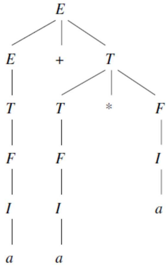


Figura: L'unico albero sintattico di $a + a * a$ in G'_E

Eliminare le ambiguità da una grammatica?

Sarebbe utile poter disporre di un algoritmo per trasformare una CFG ambigua in una CFG equivalente e non ambigua.

Purtroppo non esiste neppure un algoritmo per stabilire se una CFG è ambigua.

Eliminare le ambiguità da una grammatica?

Sarebbe utile poter disporre di un algoritmo per trasformare una CFG ambigua in una CFG equivalente e non ambigua. Purtroppo non esiste neppure un algoritmo per stabilire se una CFG è ambigua.

- Esiste un algoritmo per decidere se una CFG è ambigua?

Eliminare le ambiguità da una grammatica?

Sarebbe utile poter disporre di un algoritmo per trasformare una CFG ambigua in una CFG equivalente e non ambigua. Purtroppo non esiste neppure un algoritmo per stabilire se una CFG è ambigua.

- Esiste un algoritmo per decidere se una CFG è ambigua?

  No

  $$L = \{\langle G \rangle \mid G \text{ è una CFG e } G \text{ è ambigua}\}$$

  è indecidibile.

Eliminare le ambiguità da una grammatica?

- È sempre possibile trasformare una CFG ambigua in una CFG equivalente e non ambigua?

Eliminare le ambiguità da una grammatica?

- È sempre possibile trasformare una CFG ambigua in una CFG equivalente e non ambigua?

  No.

Eliminare le ambiguità da una grammatica?

- È sempre possibile trasformare una CFG ambigua in una CFG equivalente e non ambigua?

  No.

  Esistono CFG G ambigue e tali che ogni grammatica equivalente a G è ambigua.

Eliminare le ambiguità da una grammatica?

- È sempre possibile trasformare una CFG ambigua in una CFG equivalente e non ambigua?

  No.

  Esistono CFG G ambigue e tali che ogni grammatica equivalente a G è ambigua.

  In tal caso $L(G)$ si dice inerentemente ambiguo.

Definizione

Un linguaggio context-free L è inerentemente ambiguo se ogni grammatica che lo genera è ambigua.

Definizione

Un linguaggio context-free L è inerentemente ambiguo se ogni grammatica che lo genera è ambigua.

- Un CFL L non è inerentemente ambiguo se esiste una CFG non ambigua G tale che $L = L(G)$.

Definizione

Un linguaggio context-free L è inerentemente ambiguo se ogni grammatica che lo genera è ambigua.

- Un CFL L non è inerentemente ambiguo se esiste una CFG non ambigua G tale che $L = L(G)$.
- Il linguaggio $L(G_E)$ delle espressioni aritmetiche non è inerentemente ambiguo.

Definizione

Un linguaggio context-free L è inerentemente ambiguo se ogni grammatica che lo genera è ambigua.

- Un CFL L non è inerentemente ambiguo se esiste una CFG non ambigua G tale che $L = L(G)$.
- Il linguaggio $L(G_E)$ delle espressioni aritmetiche non è inerentemente ambiguo.
- Vedremo che i linguaggi regolari sono linguaggi context-free non inerentemente ambigui.

Definizione

Un linguaggio context-free L è inerentemente ambiguo se ogni grammatica che lo genera è ambigua.

- Un CFL L non è inerentemente ambiguo se esiste una CFG non ambigua G tale che $L = L(G)$.
- Il linguaggio $L(G_E)$ delle espressioni aritmetiche non è inerentemente ambiguo.
- Vedremo che i linguaggi regolari sono linguaggi context-free non inerentemente ambigui.
- Un linguaggio inerentemente ambiguo:

  $$L = \{a^n b^n c^m d^m \mid n \geq 1, m \geq 1\} \cup \{a^n b^m c^m d^n \mid n \geq 1, m \geq 1\}$$

$$L = \{a^n b^n c^m d^m \mid n \geq 1, m \geq 1\} \cup \{a^n b^m c^m d^n \mid n \geq 1, m \geq 1\}$$

è inerentemente ambiguo ed è generato dalla seguente grammatica context-free

$$S \rightarrow AB \mid C$$

$$A \rightarrow aAb \mid ab \quad B \rightarrow cBd \mid cd \quad C \rightarrow aCd \mid aDd \quad D \rightarrow bDc \mid bc$$

Esercizio: provare a generare la stringa *aabbccdd* con le due derivazioni canoniche sinistre che iniziano con le due produzioni per S.

Un linguaggio inerentemente ambiguo

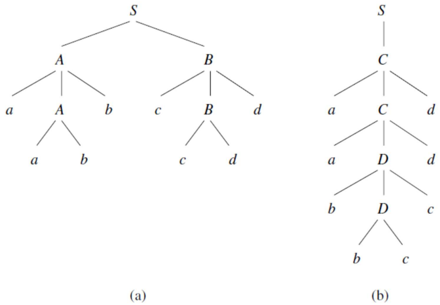


Figura: Due alberi sintattici per *aabbccdd*