Come descrivere un Linguaggio di Programmazione

Sommario: 2 marzo, 2015

- Sintassi, Semantica, Pragmatica e Implementazione
- Grammatiche e Sintassi
- Grammatiche Context Free: Derivazione, Alberi, Ambiguità
- Sintassi: Proprietà contestuali
- Compilatore: Le fasi di analisi sintattica, semantica e di generazione del codice
- Semantica: Formalismi
- Semantica operazionale strutturata: Stato, Transizione, Computazione.

Sintassi

- Un programma si presenta come una "lunga" sequenza di caratteri che rappresenta simboli e strutture del linguaggio con cui il programma è espresso.

  $\#\texttt{include}<\texttt{stdio.h}>\ \#\texttt{include}<\texttt{stdlib.h}>\ \texttt{voidXSwap}(\texttt{intA}[], \texttt{intB}[])\{\texttt{A}[0] = \ldots$

- **Lessico**. La definizione di come e quali sequenze di caratteri formano simboli (inclusi i separatori di simboli) affidata al Lessico.

- **Sintassi**. La definizione di come e quali sequenze di simboli formano costrutti e la struttura del programma affidata alla Sintassi.

Grammatiche

- **Grammatiche**. Lessico e Sintassi possono essere completamente definite mediante grammatiche:
  - Lineari, per il lessico, $\mathcal{G}^{\mathcal{R}}$.
  - Libere da contesto (Context Free), per la sintassi, $\mathcal{G}^{\mathcal{F}}$.
  - $\mathcal{G}^{\mathcal{R}} \subset \mathcal{G}^{\mathcal{F}}$.

- Sia A insieme finito di simboli;
  - $\mathbf{A}^\star$ insieme di tutte le sequenze finite di simboli su A.
  - $\mathrm{A}^\star$ è chiuso rispetto all'operatore binario, associativo, $\_.\_$, chiamato *giustapposizione*[1]: esempio $\mathrm{Pi}\mathbf{.}\mathrm{sa} = \mathrm{Pisa}$.
  - $\epsilon \in \mathtt{A}^\star$ seq. vuota: $\epsilon.\mathrm{q} = \mathrm{q}, \forall \mathrm{q} \in \mathtt{A}^\star$

Definition (Linguaggio (formale) su A)

Un linguaggio (formale) su A è un sottoinsieme di $\mathrm{A}^\star$

[1] la sua notazione è omessa quando non vi è ambiguità

Grammatiche Libere, $\mathcal{G}^{\mathcal{F}}$ /1

Definition (Grammatica Libera)

Una grammatica libera è una quadrupla (NT,T,R,S) tale che:

- **NT** insieme finito dei non terminali.
- **T** insieme finito dei terminali.
- **S** simbolo iniziale, S$\in$NT.
- **R** insieme finito delle produzioni,

$$\mathsf{R} \equiv \{\mathtt{V_i} \to \mathtt{w_i} \mid \mathtt{w_i} \in (\mathtt{T} \cup \mathtt{NT})^*, \mathtt{i} \in [1, \mathtt{k}]\}$$

Esempio

$\mathsf{G} = (\{\mathtt{E}\}, \{\mathtt{I}, +, *, (,)\}, \mathtt{E}, \mathtt{R})$
$\mathsf{R} = \{\mathtt{E} \to \mathtt{E} + \mathtt{E}, \mathtt{E} \to \mathtt{E} * \mathtt{E}, \mathtt{E} \to \mathtt{I}, \mathtt{E} \to (\mathtt{E})\}$

Notazione. $\mathtt{V} \to \alpha_1|...|\alpha_k$ invece di $\mathtt{V}_1 \to \alpha_1, ..., \mathtt{V}_k \to \alpha_k$ quando, $\mathtt{V} = \mathtt{V}_1 = ... = \mathtt{V}_k$.
Esempio

$$\mathsf{R} = \{\mathtt{E} \to \mathtt{E} + \mathtt{E} \mid \mathtt{E} * \mathtt{E} \mid \mathtt{I} \mid (\mathtt{E})\}$$

Grammatiche Libere, $\mathcal{G}^{\mathcal{F}}$ /2

Esempio

$\mathtt{G} = (\{\mathtt{E}\}, \{\mathtt{I}, +, *, (,)\}, \mathtt{E}, \mathtt{R})$

$\mathtt{R} = \{\mathtt{E} \longrightarrow \mathtt{E} + \mathtt{E}, \mathtt{E} \longrightarrow \mathtt{E} * \mathtt{E}, \mathtt{E} \longrightarrow \mathtt{I}, \mathtt{E} \longrightarrow (\mathtt{E})\}$

Qual'è il significato di G? [2]

Definition (La relazione $\Rightarrow$)

Ogni grammatica libera $\mathtt{G} \equiv (\mathtt{NT}, \mathtt{T}, \mathtt{R}, \mathtt{S})$ definisce una relazione binaria $\Rightarrow_{\mathtt{G}}$ (o, semplicemente $\Rightarrow$) su $(\mathtt{T} \cup \mathtt{NT})^*$ tale che:

$$\forall\, \alpha.\mathtt{V}.\rho,\ \alpha.\gamma.\rho \in (\mathtt{T} \cup \mathtt{NT})^*, \quad \alpha \mathtt{V} \rho \Rightarrow \alpha\gamma\rho \text{ sse } \ \mathtt{V} \longrightarrow \gamma \in \mathtt{R}$$

Definition (Linguaggio generato)

Sia $\mathtt{G} \equiv (\mathtt{NT}, \mathtt{T}, \mathtt{R}, \mathtt{S})$ libera. Il linguaggio definito da $\mathtt{G}$:

$$\mathcal{L}(\mathtt{G}) = \{\alpha \in \mathtt{T}^\star | \mathtt{S} \Rightarrow^\star \alpha\}$$

dove, $\Rightarrow^\star$ è la chiusura transitiva (e riflessiva) di $\Rightarrow_{\mathtt{G}}$

[2] "sse" sta per "se e solo se"

Grammatiche Libere, $\mathcal{G}^{\mathcal{F}}$ /3

Definition (Derivazione)

Sia $\mathtt{G} \equiv (\mathtt{NT}, \mathtt{T}, \mathtt{R}, \mathtt{S})$ libera. Siano $\mathtt{v}, \mathtt{w} \in (\mathtt{T} \cup \mathtt{NT})^*$. Sia $\mathtt{v} \Rightarrow^\star \mathtt{w}$. Una derivazione, in $\mathtt{G}$, da $\mathtt{v}$ a $\mathtt{w}$, è una sequenza $\mathtt{v} \Rightarrow \mathtt{w}_0 \Rightarrow \mathtt{w}_1 \Rightarrow \ldots \Rightarrow \mathtt{w}$

Esempio
La derivazione ci fornisce una prova di appartenenza. Si dimostri che $\mathtt{I} * \mathtt{I} + \mathtt{I} \in \mathcal{L}(\mathtt{G})$.

$\mathsf{G} = (\{\mathtt{E}\}, \{\mathtt{I}, +, *\}, \mathtt{E}, \mathtt{R})$
$\mathsf{R} = \{\mathtt{E} \longrightarrow \mathtt{E} + \mathtt{E}, \mathtt{E} \longrightarrow \mathtt{E} * \mathtt{E}, \mathtt{E} \longrightarrow \mathtt{I}, \mathtt{E} \longrightarrow (\mathtt{E})\}$

Procediamo applicando una $\Rightarrow$ alla volta a partire da $\mathtt{E}$

$$\begin{aligned} \mathtt{E} &\Rightarrow_2 \mathtt{E} * \mathtt{E} \\ &\Rightarrow_3 \mathtt{I} * \mathtt{E} \\ &\Rightarrow_1 \mathtt{I} * \mathtt{E} + \mathtt{E} \\ &\Rightarrow_3 \mathtt{I} * \mathtt{I} + \mathtt{E} \\ &\Rightarrow_3 \mathtt{I} * \mathtt{I} + \mathtt{I} \end{aligned}$$ [3]

Aggiungendo induzione possiamo dimostrare che:

$$\{\mathtt{I}.\alpha_1.\cdots.\alpha_\mathtt{n} \mid \alpha_\mathtt{i} \in \{+\mathtt{I}, *\mathtt{I}\}, \mathtt{n} \geq 0\} \subset \mathcal{L}(G).$$

[3] Un pedice può indicare la produzione usata. In questo caso, usiamo la posizione a partire da 1 per la produzione più a sinistra

Grammatiche Libere: La relazione $\Rightarrow$ su alberi /1

Esempio

$G = (\{\mathtt{E}\}, \{\mathtt{I}, +, *, (,)\}, \mathtt{E}, \mathtt{R})$

$R = \{\mathtt{E} \rightarrow \mathtt{E} + \mathtt{E}, \mathtt{E} \rightarrow \mathtt{E} * \mathtt{E}, \mathtt{E} \rightarrow \mathtt{I}, \mathtt{E} \rightarrow (\mathtt{E})\}$

- La $\Rightarrow$ su sequenze di simboli (stringhe) grammaticali ci dice proprio tutto?

$\mathtt{E} \Rightarrow_2 \mathtt{E} * \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{E} \Rightarrow_1 \mathtt{I} * \mathtt{E} + \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{I} + \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{I} + \mathtt{I}$

$\mathtt{E} \Rightarrow_2 \mathtt{E} * \mathtt{E} \Rightarrow_1 \mathtt{E} * \mathtt{E} + \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{E} + \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{I} + \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{I} + \mathtt{I}$

$\mathtt{E} \Rightarrow_1 \mathtt{E} + \mathtt{E} \Rightarrow_2 \mathtt{E} * \mathtt{E} + \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{E} + \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{I} + \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{I} + \mathtt{I}$

- In realtà NO.
  - 2 delle 3 derivazioni sopra ci dicono la stessa cosa, 1 dice una cosa.
  - Quali? In cosa, sono diverse?
  - Useremo la struttura degli alberi per rispondere.
- La sequenza di simboli grammaticali non è abbastanza espressiva

Grammatiche Libere: La relazione $\Rightarrow$ su alberi /2

Definition

- $[] \in \mathtt{Tree}^{\mathtt{A}}$
- $[\mathtt{a} - (\mathtt{t}_1, ..., \mathtt{t}_\mathtt{n})] \in \mathtt{Tree}^{\mathtt{A}}$, per $\mathtt{a} \in \mathtt{A}, \mathtt{t}_\mathtt{i} \in \mathtt{Tree}^{\mathtt{A}}, \mathtt{n} \geq 0$

Notazione. [a - ()] può essere scritto [a].

- Gli alberi possono essere utilizzati come relazioni sulla struttura di un termine
  - Ogni arco una coppia della relazione *sottotermine*
- Hanno una rappresentazione grafica.
  - mostriamo quella di [E - ([E],[+],[E])]

Grammatiche Libere: La relazione $\Rightarrow$ su alberi /3

Definition (La relazione $\Rightarrow^{\mathtt{Tree}}$)

Ogni grammatica libera $\mathtt{G} \equiv (\mathtt{NT}, \mathtt{T}, \mathtt{R}, \mathtt{S})$ definisce una relazione binaria $\Rightarrow^{\mathtt{Tree}}_{\mathtt{G}}$ (o, semplicemente $\Rightarrow^{\mathtt{Tree}}$, ovvero $\Rightarrow$) su $\mathtt{Tree}^{\mathtt{T}\cup\mathtt{NT}}$ tale che:

- $[\mathtt{V}] \Rightarrow [\mathtt{V} - ([\mathtt{u}_1], ..., [\mathtt{u}_n])]$
  sse $\mathtt{V} \longrightarrow \mathtt{u}_1. \dots .\mathtt{u}_n \in \mathtt{R}$ and $\mathtt{u}_i \in (\mathtt{T} \cup \mathtt{NT})$
- $[\mathtt{V} - (\mathtt{t}_1, ..., \mathtt{t}_j, ..., \mathtt{t}_n)] \Rightarrow [\mathtt{V} - (\mathtt{t}_1, ..., \mathtt{r}_j, ..., \mathtt{t}_n)]$
  sse $1 \le \mathtt{j} \le \mathtt{n} \wedge (\forall \mathtt{i})\mathtt{t}_\mathtt{i} \in \mathtt{Tree}^{\mathtt{T}\cup\mathtt{NT}} \wedge \mathtt{t}_\mathtt{j} \Rightarrow \mathtt{r}_\mathtt{j}$

Definition (Linguaggio generato su $\mathtt{Tree}^{\mathtt{T}\cup\mathtt{NT}}$ - Parse Tree set)

Sia $\mathtt{G} \equiv (\mathtt{NT}, \mathtt{T}, \mathtt{R}, \mathtt{S})$ libera. Il linguaggio definito da $\mathtt{G}$:

$$\mathcal{L}(\mathtt{G}) = \{\alpha \in \mathtt{Tree}^{\mathtt{T}\cup\mathtt{NT}} \mid [\mathtt{S}] \Rightarrow^\star \alpha\}$$

dove, $\Rightarrow^\star$ è la chiusura transitiva (e riflessiva) di $\Rightarrow$

Definition (Derivazione su $\mathtt{Tree}^{\mathtt{T}\cup\mathtt{NT}}$)

Sia $\mathtt{G} \equiv (\mathtt{NT}, \mathtt{T}, \mathtt{R}, \mathtt{S})$ libera. Siano $\mathtt{u}, \mathtt{q} \in \mathtt{Tree}^{\mathtt{T}\cup\mathtt{NT}}$. Sia $\mathtt{u} \Rightarrow^{\star} \mathtt{q}$.
Una derivazione, in $\mathtt{G}$, da $\mathtt{u}$ a $\mathtt{q}$, è una sequenza
$\mathtt{u} \Rightarrow \mathtt{q}_0 \Rightarrow \mathtt{q}_1 \Rightarrow \ldots \Rightarrow \mathtt{q}$

Esempio

$G = (\{\mathtt{E}\}, \{\mathtt{I}, +, *, (,)\}, \mathtt{E}, \mathtt{R})$
$R = \{\mathtt{E} \rightarrow \mathtt{E} + \mathtt{E}, \mathtt{E} \rightarrow \mathtt{E} * \mathtt{E}, \mathtt{E} \rightarrow \mathtt{I}, \mathtt{E} \rightarrow (\mathtt{E})\}$

- La $\Rightarrow$ su stringhe ci dice proprio tutto?

$\mathtt{E} \Rightarrow_2 \mathtt{E} * \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{E} \Rightarrow_1 \mathtt{I} * \mathtt{E} + \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{I} + \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{I} + \mathtt{I}$
$\mathtt{E} \Rightarrow_2 \mathtt{E} * \mathtt{E} \Rightarrow_1 \mathtt{E} * \mathtt{E} + \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{E} + \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{I} + \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{I} + \mathtt{I}$
$\mathtt{E} \Rightarrow_1 \mathtt{E} + \mathtt{E} \Rightarrow_2 \mathtt{E} * \mathtt{E} + \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{E} + \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{I} + \mathtt{E} \Rightarrow_3 \mathtt{I} * \mathtt{I} + \mathtt{I}$

- La $\Rightarrow$ su alberi?

$[\mathtt{E}] \Rightarrow_2 [\mathtt{E} - ([\mathtt{E}], [*], [\mathtt{E}])] \Rightarrow_3 [\mathtt{E} - ([\mathtt{E} - ([\mathtt{I}])], [*], [\mathtt{E}])]$
$\Rightarrow^{\star} [\mathtt{E} - ([\mathtt{E} - ([\mathtt{I}])], [*], [\mathtt{E} - ([\mathtt{E} - ([\mathtt{I}])], [+], [\mathtt{E} - ([\mathtt{I}])])])]$
$[\mathtt{E}] \Rightarrow^{\star} \ldots$
$\mathtt{E} \Rightarrow_1 [\mathtt{E} - ([\mathtt{E}], [+], [\mathtt{E}])] \Rightarrow_2 [\mathtt{E} - ([\mathtt{E} - ([\mathtt{E}], [*], [\mathtt{E}]], [+], [\mathtt{E}])]$
$\Rightarrow^{\star} [\mathtt{E} - ([\mathtt{E} - ([\mathtt{E} - ([\mathtt{I}])], [*], [\mathtt{E} - ([\mathtt{I}])]], [+], [\mathtt{E} - ([\mathtt{I}])])]$

Grammatiche Libere: Frontiera di un albero /5

Definition (Albero di derivazione sintattica o Parse Tree)

Sia $\mathtt{G} \equiv (\mathtt{NT}, \mathtt{T}, \mathtt{R}, \mathtt{S})$ libera. Sia $\mathtt{Q}_\mathtt{G} = \{\mathtt{q} \in \mathtt{Tree}^{\mathtt{T} \cup \mathtt{NT}} \mid [\mathtt{S}] \Rightarrow^\star \mathtt{q}\}$.
$\mathtt{Q}_\mathtt{G}$ è l'insieme di tutti e soli gli alberi di derivazione sintattica, detti anche, parse tree, di $\mathtt{G}$.

Definition (Frontiera di un albero /1)

La frontiera di un albero $\mathtt{q} \in \mathtt{Tree}^\mathtt{A}$ è una stringa di $\mathtt{A}^\star$, ovvero $\mathtt{fron}(\mathtt{q}) \in \mathtt{A}^\star$ per ogni q, così definita:

- $\mathtt{fron}([]) = \epsilon$
- $\mathtt{fron}([\mathrm{a}]) = \mathrm{a}$
- $\mathtt{fron}([\mathrm{a} - (\mathtt{t}_1, \ldots, \mathtt{t}_\mathtt{n})]) = \mathtt{fron}(\mathtt{t}_1).\cdots.\mathtt{fron}(\mathtt{t}_\mathtt{n})$

Definition (Frontiera di un albero)

La frontiera di un albero $q \in \mathtt{Tree}^{\mathtt{A}}$ è una stringa di $\mathtt{A}^\star$, ovvero $\mathtt{fron}(q) \in \mathtt{A}^\star$ per ogni q, così definita:

- $\mathtt{fron}([]) = \epsilon$
- $\mathtt{fron}([a]) = a$
- $\mathtt{fron}([a - (t_1, ..., t_n)]) = \mathtt{fron}(t_1).\cdots.\mathtt{fron}(t_n)$

Esempio

$$
\begin{aligned}
&\mathtt{fron}([\mathtt{E} - ([\mathtt{E} - ([\mathtt{I}])], [*], [\mathtt{E} - ([\mathtt{E} - ([\mathtt{I}])], [+], [\mathtt{E} - ([\mathtt{I}])])])])]) = \\
&\qquad = \mathtt{fron}([\mathtt{E} - ([\mathtt{I}])]).\mathtt{fron}([*]).\mathtt{fron}([\mathtt{E} - ([\mathtt{E} - ([\mathtt{I}])], [+], [\mathtt{E} - ([\mathtt{I}])])])]) \\
&\qquad = \mathtt{fron}([\mathtt{I}]).\mathtt{fron}([*]).\mathtt{fron}([\mathtt{E} - ([\mathtt{I}])]).\mathtt{fron}([+]).\mathtt{fron}([\mathtt{E} - ([\mathtt{I}])]) \\
&\qquad = \mathtt{fron}([\mathtt{I}]).\mathtt{fron}([*]).\mathtt{fron}([\mathtt{I}]).\mathtt{fron}([+]).\mathtt{fron}([\mathtt{I}]) \\
&\qquad = \mathtt{I}. * .\mathtt{I}. + .\mathtt{I} = \mathtt{I} * \mathtt{I} + \mathtt{I}
\end{aligned}
$$

Definition (Linguaggio generato utilizzando Parse Tree)

Sia $\mathtt{G} \equiv (\mathtt{NT}, \mathtt{T}, \mathtt{R}, \mathtt{S})$ libera. Il linguaggio definito da $\mathtt{G}$:

$$\mathcal{L}(\mathtt{G}) = \{\mathtt{fron}(\mathtt{q}) \in \mathtt{T}^\star \mid [\mathtt{S}] \Rightarrow^\star \mathtt{q}\}$$

dove, $\Rightarrow^\star$ è la chiusura transitiva (e riflessiva) di $\Rightarrow_{\mathtt{G}}^{\mathtt{Tree}^{\mathtt{T}\cup\mathtt{NT}}}$

Definition (Grammatica Ambigua)

$\mathtt{G} \equiv (\mathtt{NT}, \mathtt{T}, \mathtt{R}, \mathtt{S})$ è ambigua se $\exists \mathtt{q}_1 \neq \mathtt{q}_2 \in \mathtt{Tree}^{\mathtt{T}\cup\mathtt{NT}}$:

- $[\mathtt{S}] \Rightarrow^\star \mathtt{q}_1, [\mathtt{S}] \Rightarrow^\star \mathtt{q}_2$
- $\mathtt{fron}([\mathtt{q}_1]) = \mathtt{fron}([\mathtt{q}_2]) \in \mathtt{T}^*$

Grammatiche Libere: Ambiguità /7

- La sintassi di un LP deve essere espressa mediante una grammatica libera non ambigua
- Ogni termine di LP deve avere un'unica struttura sintattica che ne mostra i sottotermini componenti

Esempio. Grammatica Ambigua

$\mathsf{G} = (\{\mathtt{E}\}, \{\mathtt{I}, +, *\}, \mathtt{E}, \mathtt{R})$

$\mathsf{R} = \{\mathtt{E} \longrightarrow \mathtt{E} + \mathtt{E}, \mathtt{E} \longrightarrow \mathtt{E} * \mathtt{E}, \mathtt{E} \longrightarrow \mathtt{I}, \mathtt{E} \longrightarrow (\mathtt{E})\}$

Esempio. Grammatica non Ambigua

$\mathsf{G'} = (\{\mathtt{E}, \mathtt{F}, \mathtt{T}\}, \{\mathtt{I}, +, *\}, \mathtt{E}, \mathtt{R}')$

$$
\begin{aligned}
\mathsf{R'} = \{\mathtt{E} &\longrightarrow \mathtt{E} + \mathtt{F} \mid \mathtt{F} \\
\mathtt{F} &\longrightarrow \mathtt{F} * \mathtt{T} \mid \mathtt{T} \\
\mathtt{T} &\longrightarrow \mathtt{I} \mid (\mathtt{E})\}
\end{aligned}
$$

$\mathcal{L}(\mathtt{G}) = \mathcal{L}(\mathtt{G}')$