

PROVE IN ITINERE

Corso Automi, Linguaggi e Traduttori (Laurea e Laurea on-line) 2007-08

Docente: Alberto Pettorossi

Si raccomanda la precisione e la chiarezza dell'esposizione.

Prova 1. Parsing per un linguaggio di tipo 2 esteso.

Scrivere e documentare un programma in Java o in C++ che data una qualsiasi grammatica G_2 di tipo 2 estesa (o context-free estesa) e una parola w , faccia il parsing di w secondo la grammatica G_2 . Si ammette che nella grammatica data ci possano essere anche ε -productions. In particolare, si implementino gli algoritmi: (i) "from above", (ii) "from below", (iii) avoiding unit-production, (iv) avoiding ε -productions, e (v) avoiding left-recursion.

La grammatica viene data in input come una stringa che ne codifica le produzioni. Tale stringa è generata dalla seguente grammatica con assioma G (ma non tutte le stringhe generate da G codificano le produzioni di una grammatica di tipo 2 estesa):

$$\begin{aligned} G &\rightarrow H. \\ H &\rightarrow P \mid P ; H \quad (P \text{ denota tutte le produzioni per lo stesso nonterminale } L) \\ P &\rightarrow L \rightarrow R_s \quad (R_s \text{ denota una sequenza di right-hand-side di produzioni}) \\ L &\rightarrow \mathbf{A} \mid \mathbf{B} \mid \dots \mid \mathbf{Z} \\ R_s &\rightarrow R \mid R \mid R_s \quad (R \text{ denota il right-hand-side di una produzione}) \\ R &\rightarrow \mathbf{e} \mid R_{ne} \quad (R_{ne} \text{ denota il right-hand-side non-empty di una produzione}) \\ R_{ne} &\rightarrow U \mid U R_{ne} \\ U &\rightarrow \mathbf{a} \mid \mathbf{b} \mid \mathbf{c} \mid \mathbf{d} \mid \mathbf{f} \mid \dots \mid \mathbf{z} \mid \mathbf{A} \mid \mathbf{B} \mid \dots \mid \mathbf{Z} \end{aligned}$$

Il simbolo nonterminale a sinistra della prima produzione è l'assioma della grammatica. I simboli nonterminali sono caratteri maiuscoli da \mathbf{A} a \mathbf{Z} . I simboli terminali sono caratteri minuscoli da \mathbf{a} a \mathbf{z} , escluso il carattere \mathbf{e} che è riservato per la codifica della stringa vuota ε (cioè la empty string fatta da zero caratteri). Per esempio, la grammatica di tipo 2 estesa le cui produzioni sono:

$$S \rightarrow aAb \mid SA \quad A \rightarrow aA \mid \varepsilon \mid S$$

e il cui assioma è S , è data come la stringa: $\mathbf{S} \rightarrow \mathbf{aAb} \mid \mathbf{SA}; \mathbf{A} \rightarrow \mathbf{aA} \mid \mathbf{e} \mid \mathbf{S}$.

Prova 2. Membership per un linguaggio di tipo 1.

Scrivere e documentare un programma in Java o in C++ che, data una grammatica G_1 di tipo 1 e una parola w , verifichi se w appartiene al linguaggio $L(G_1)$ generato da G_1 .

Si ricorda che una grammatica è di tipo 1 se e solo se le sue produzioni sono del tipo:

$$\text{stringa1} \rightarrow \text{stringa2}$$

ove: (i) stringa1 e stringa2 sono stringhe di terminali e non-terminali, e (ii) la lunghezza della stringa stringa2 è maggiore o uguale a quella della stringa stringa1.

Si ricorda anche che per la ricerca di un nodo in un albero di possibili alternative, si può usare lo schema di backtracking. La grammatica viene data in input come una stringa che ne codifica le produzioni. Tale stringa è generata dalla seguente grammatica con assioma G (ma non tutte le stringhe generate da G codificano le produzioni di una grammatica di tipo 1):

$$\begin{aligned} G &\rightarrow H. \\ H &\rightarrow P \mid P ; H \quad (P \text{ denota tutte le produzioni per lo stesso left-hand-side } L) \\ P &\rightarrow L \rightarrow R_s \quad (L \text{ denota il left-hand-side di una produzione}) \\ L &\rightarrow U_s \\ R_s &\rightarrow R \mid R \mid R_s \quad (R \text{ denota il right-hand-side di una produzione}) \\ R &\rightarrow U_s \\ U_s &\rightarrow U \mid U U_s \\ U &\rightarrow \mathbf{a} \mid \mathbf{b} \mid \dots \mid \mathbf{z} \mid \mathbf{A} \mid \mathbf{B} \mid \dots \mid \mathbf{Z} \end{aligned}$$

Il simbolo nonterminale a sinistra della prima produzione è l'assioma della grammatica. I simboli nonterminali sono caratteri maiuscoli da \mathbf{A} a \mathbf{Z} . I simboli terminali sono caratteri minuscoli da \mathbf{a} a \mathbf{z} .

Per esempio, la grammatica di tipo 1 le cui produzioni sono:

$$S \rightarrow ABc \mid a \quad AB \rightarrow acB \quad Bc \rightarrow ac \quad ac \rightarrow cBc \mid bc \mid cb$$

e il cui assioma è S , è data come la stringa $\mathbf{S} \rightarrow \mathbf{ABc} \mid \mathbf{a}; \mathbf{AB} \rightarrow \mathbf{acB}; \mathbf{Bc} \rightarrow \mathbf{ac}; \mathbf{ac} \rightarrow \mathbf{cBc} \mid \mathbf{bc} \mid \mathbf{cb}$.