

Esercizi: algoritmo di Earley

Lunedì 5 Novembre

Esercizio

- Siano date la grammatica G e la stringa $aaaa$. Si esegua il riconoscimento della stringa utilizzando l'algoritmo di Earley.

$G: \quad S \rightarrow aaS \mid Saaa \mid a \mid \epsilon$

Soluzione

G: $S \rightarrow aaS \mid Saaa \mid a \mid \epsilon$

$S' \rightarrow S\$$

$x = aaaa\$$

Esercizio

- Siano date la grammatica G e la stringa $aabb$. Si esegua il riconoscimento della stringa utilizzando l'algoritmo di Earley.

G :

$S \rightarrow aAbB \mid C$

$A \rightarrow aA \mid a$

$B \rightarrow bB \mid b$

$C \rightarrow aCb \mid ab$

Soluzione

$G:$

$$\begin{aligned} S &\rightarrow aAbB \mid C \\ A &\rightarrow aA \mid a \\ B &\rightarrow bB \mid b \\ C &\rightarrow aCb \mid ab \\ S' &\rightarrow S\$ \end{aligned}$$

$x = aabb\$$

Un compito importante di un parser

Gestione degli errori sintattici

Gestione degli errori in un parser

- Un parser deve essere in grado di *scoprire*, *diagnosticare* e correggere gli errori in maniera efficiente, per *riprendere* l'analisi e scoprire nuovi errori.
- Alcuni parser (LL e LR) hanno la proprietà “viable prefix”: sono in grado di rilevare un errore non appena si presenta perché sono in grado di riconoscere i prefissi validi del linguaggio

Strategie di riparazione

- “panic mode”: scoperto l’errore il parser riprende l’analisi in corrispondenza di alcuni token sincronizzanti predefiniti (es.: delimitatori begin end) *scartando* alcuni caratteri. Svantaggi: può essere *scartato* molto input.
- “phrase level”: correzioni locali ottenute inserendo, modificando, cancellando alcuni terminali per poter riprendere l’analisi (es.: ‘;’ -> ‘;’)
Svantaggi: difficoltà quando la distanza dall’errore è notevole.
- “error productions”: uso di produzioni che estendono la grammatica per generare gli errori più comuni. Metodo efficiente per la diagnostica.
- “global correction”: si cerca di “calcolare” la migliore correzione possibile alla derivazione errata (minimo costo di interventi per inserzioni/cancellazioni). Metodo globale poco usato in pratica, ma tecnica usata per ottimizzare la strategia “phrase level”.

USO di EOF in Flex

<<EOF>>

- La regola speciale <<EOF>> indica azioni che sono eseguite quando si incontra un end-of-file e yywrap() restituisce un valore non nullo, ovvero indica che non ci sono altri file da processare. Le azioni devono terminare preferibilmente facendo una delle seguenti cose:
 - Assegnare yyin a un nuovo file in input.
 - Eseguire un'azione di return
 - Eseguire yyterminate().
- Nel caso di input con file multipli, si deve gestire la funzione yywrap.
- La regola <<EOF>> non dovrebbe essere usata con altri pattern. Dovrebbe solo essere qualificata con una lista di start conditions. Se non è qualificata da nessuna start condition allora si applica a tutte le start condà azioni <<EOF>>.

Esercizio con Flex

- Scrivere in Flex un tokenizzatore per il linguaggio C che riconosca i seguenti lessemi:

```
main  
int  
void  
return  
{  
}  
(  
)  
;  
=  
+  
*
```

identificatori (che cominciano con lettera), con token ID.

Costanti intere con token INTCONST

Esercizio con Flex

- Il programma deve restituire l'elenco dei token, i relativi attributi (ovvero nel caso degli identificatori, il nome dell'identificatore, nel caso delle costanti, il numero) e il numero di riga in cui il token compare.
- I commenti devono essere riconosciuti e devono essere contate le righe che contengono commenti, sia quelli racchiusi tra `/*` e `*/` che il commento preceduto da `//`.

Esercizio con Flex

- Per esempio, se l'input è
int main(void)

```
{ int x;  
x = 19;  
x = x * x;  
/* commento  
su due righe*/  
return 0; }
```

- Verrà prodotto il seguente output

TOKEN	LESSEMA	LINENO
tok_int		1
tok_main		1
tok_lparen		1
tok_void		1
tok_rparen		1
tok_lbrace		2
tok_int		2
ID	x	2

Esercizio con Flex

tok_semicolon		2
ID	x	3
tok_equal		3
INTCONST	19	3
tok_semicolon		3
ID	x	4
tok_equal		4
ID	x	4
tok_mult		4
ID	x	4
tok_semicolon		4
tok_rbrace		4
tok_return		7
INTCONST	0	7
tok_semicolon		7
tok_rbrace		7

2 righe contengono commenti.