Analyseur descendant non-récursif

Analyseur non-récursif LL(1)

Préambule

Vous avez découvert dans ce module deux types d'analyseurs syntaxiques descendants :

- les analyseurs récursifs
- les analyseurs non-récursifs

Un analyseur non-récursif est un automate à pile dont le but est de reconnaître les mots du langage engendré par la grammaire à partir de laquelle on a construit l'analyseur. Une analyse syntaxique non-récurive LL(1) se décompose en deux parties :

(i) **construction de l'analyseur** : à partir d'une grammaire algébrique on construit un automate à pile déterministe qui reconnaît le langage engendré par la grammaire

(ii) **reconnaissance d'un mot** : pour un mot donné, on simule le fonctionnement de l'automate à pile précédemment construit pour savoir si le mot est reconnu

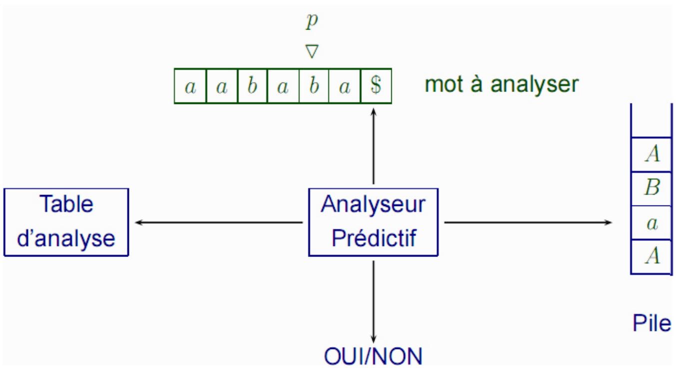
Construction de l'analyseur

Construire l'automate à pile consiste à déterminer la table de transition de l'automate dans la mesure ou son fonctionnement est toujours le même (l'algorithme de simulation de l'automate est le même quelque soit la grammaire). Cette table est construite à partir des symboles directeurs des règles de la grammaire. La détermination de ces derniers ne peut cependant se faire qu'à condition que la grammaire vérifie un certain nombre de propriétés : elle ne doit pas être ambiguë, elle doit être réduite, non récursive-gauche et factorisée. En résumé, pour construire un analyseur non-récursif d'une grammaire, il faut :

- vérifier que la grammaire n'est pas ambiguë
- réduire la grammaire si nécessaire
- dérecursiver la grammaire si nécessaire
- factoriser la grammaire si nécessaire
- déterminer les non-terminaux produisant le vide
- calculer les premiers et les suivants des non-terminaux
- en déduire les symboles directeurs des règles
- construire la table d'analyse

Reconnaissance d'un mot

Le reconnaissance d'un mot consiste simplement à simuler le fonctionnement de l'automate à pile constuit à partir de la table d'analyse déterminée précédemment.



- si le sommet de la pile est un non-terminal E alors
  - si $p \triangleright a$ et que $table[E, a] = erreur$, alors le mot est refusé
  - sinon, dépiler E et empiler $table[E, a]$

- si le sommet de la pile est un terminal a alors
  – si $p \triangleright a$ alors dépiler a et avancer p
  – si $p \triangleright b$ avec $b \neq a$ alors le mot est refusé
- Si la pile est vide alors
  – si $p \triangleright \$$ alors le mot est accepté
  – sinon le mot est refusé

Exemple 1

Soit $G = \left(N = \{E, T, F\}, T = \{\$, +, *, (,), i\}, \left\{ \begin{array}{c} E \rightarrow E + T \mid T \\ T \rightarrow T * F \mid F \\ F \rightarrow (E) \mid i \end{array} \right\}, E \right)$ une grammaire algébrique dont on souhaite construite un analyseur syntaxique descendant LL(1) non récursif.

(*i*) on réduit, dérécursive et factorise la grammaire. On obtient la grammaire équivalente suivante :

$$G' = \left(N' = \{E, T, F, A, B\}, T' = \{\$, +, *, (,), i\}, \left\{ \begin{array}{l} E \rightarrow TA \\ T \rightarrow FB \\ F \rightarrow (E) \mid i \\ A \rightarrow +TA \mid \epsilon \\ B \rightarrow *FB \mid \epsilon \end{array} \right\}, E \right)$$

(*ii*) on calcule les ensembles $Vide$, $Premier$, $Suivant$:

N'	$Vide$	$Premier$	$Suivant$
E		$(, i$	$\$,)$
T		$(, i$	$\$, +,)$
F		$(, i$	$\$, *, +,)$
A	$\times$	$+$	$\$,)$
B	$\times$	$*$	$\$, +,)$

(*iii*) on en déduit la table d'analyse :

	$\$$	$+$	$*$	$($	$)$	i
E	*erreur*	*erreur*	*erreur*	$E \rightarrow TA$	*erreur*	$E \rightarrow TA$
T	*erreur*	*erreur*	*erreur*	$T \rightarrow FB$	*erreur*	$T \rightarrow FB$
F	*erreur*	*erreur*	*erreur*	$F \rightarrow (E)$	*erreur*	$F \rightarrow i$
A	$A \rightarrow \epsilon$	$A \rightarrow +TA$	*erreur*	*erreur*	$A \rightarrow \epsilon$	*erreur*
B	$B \rightarrow \epsilon$	$B \rightarrow \epsilon$	$B \rightarrow *FB$	*erreur*	$B \rightarrow \epsilon$	*erreur*

Notre analyseur est enfin prêt (sous réserve que la procédure de simulation de l'automate à pile soit implantée). On cherche donc à savoir si pour un mot m donné, $m \in \mathcal{L}(\mathcal{G})$.

∇			
i	$*$	i	$\$$

Pile : E

Commentaire : i est symbole directeur de la règle $E \rightarrow TA$
On dépile donc E et on empile TA

∇			
i	$*$	i	$\$$

Pile : T / A

Commentaire : i est symbole directeur de la règle $T \rightarrow FB$
On dépile donc T et on empile FB

∇			
i	$*$	i	$\$$

Pile :

F
B
A

Commentaire : i est symbole directeur de la règle $F \to i$
On dépile donc F et on empile i

∇			
i	$*$	i	$\$$

Pile :

i
B
A

Commentaire : Le sommet de la pile désigne le même caractère que le pointeur.
On poursuit donc la lecture (*i.e.* on décale le pointeur et on dépile i).

	∇		
i	$*$	i	$\$$

Pile :

B
A

Commentaire : $*$ est symbole directeur de la règle $B \to *FB$
On dépile donc B et on empile $*FB$

	∇		
i	$*$	i	$\$$

Pile :

$*$
F
B
A

Commentaire : Le sommet de la pile désigne le même caractère que le pointeur.
On poursuit donc la lecture (*i.e.* on décale le pointeur et on dépile $*$).

		∇	
i	$*$	i	$\$$

Pile :

F
B
A

Commentaire : i est symbole directeur de la règle $F \to i$
On dépile donc F et on empile i

		∇	
i	$*$	i	$\$$

Pile :

i
B
A

Commentaire : Le sommet de la pile désigne le même caractère que le pointeur.
On poursuit donc la lecture (*i.e.* on décale le pointeur et on dépile i).

			∇
i	$*$	i	$\$$

Pile :

B
A

Commentaire : $ est symbole directeur de la règle $B \to \epsilon$
On dépile donc B et on empile ϵ (autrement dit, on n'empile rien !)

			∇
i	$*$	i	$\$$

Pile :

A

Commentaire : $ est symbole directeur de la règle $A \to \epsilon$
On dépile donc A et on empile ϵ (autrement dit, on n'empile rien !)

			∇
i	$*$	i	$\$$

Pile :

Commentaire : La pile est vide et le pointeur désigne le marqueur « fin de mot »
$\Rightarrow$ le mot **est** engendré par la grammaire.

Exemple 2

Soit $G = \left(N = \{X, A\}, T = \{\$, (, a,), b\}, \left\{ \begin{array}{c} X \to (XaX) \mid A \mid (X) \\ A \to bX \mid b \end{array} \right\}, X \right)$ une grammaire algébrique dont on souhaite construite un analyseur syntaxique descendant LL(1) non récursif.

(i) on réduit, dérécursive et factorise la grammaire. On obtient la grammaire équivalente suivante : $G' = \left(N' = \{X, A, B, C\}, T' = \{\$, (, a,), b\}, \left\{ \begin{array}{l} X \to (XB \mid A \\ A \to bC \\ B \to aX) \mid) \\ C \to X \mid \epsilon \end{array} \right\}, X \right)$

(ii) on calcule les ensembles $Vide$, $Premier$, $Suivant$:

N'	$Vide$	$Premier$	$Suivant$
X		$b, ($	$\$, a,)$
A		b	$\$, a,)$
B		$a,)$	$\$, a,)$
C	$\times$	$b, ($	$\$, a,)$

(iii) on en déduit la table d'analyse :

	$\$$	(	a	)	b
X	*erreur*	$X \to (XB$	*erreur*	*erreur*	$X \to A$
A	*erreur*	*erreur*	*erreur*	*erreur*	$A \to bC$
B	*erreur*	*erreur*	$B \to aX)$	$B \to)$	*erreur*
C	$C \to \epsilon$	$C \to X$	$C \to \epsilon$	$C \to \epsilon$	$C \to X$

Notre analyseur est enfin prêt (sous réserve que la procédure de simulation de l'automate à pile soit implantée). On cherche donc à savoir si pour un mot m donné, $m \in \mathcal{L}(\mathcal{G})$.

$$\begin{array}{cccccc} \nabla & & & & & \\ \hline \multicolumn{1}{|c|}{(} & \multicolumn{1}{c|}{b} & \multicolumn{1}{c|}{a} & \multicolumn{1}{c|}{b} & \multicolumn{1}{c|}{)} & \multicolumn{1}{c|}{\$} \\ \hline \end{array} \quad \textit{Pile} : \begin{array}{|c|} \\ \hline X \\ \hline \end{array}$$

Commentaire : (est symbole directeur de la règle $X \to (XB$
On dépile donc X et on empile $(XB$

$$\begin{array}{cccccc} \nabla & & & & & \\ \hline \multicolumn{1}{|c|}{(} & \multicolumn{1}{c|}{b} & \multicolumn{1}{c|}{a} & \multicolumn{1}{c|}{b} & \multicolumn{1}{c|}{)} & \multicolumn{1}{c|}{\$} \\ \hline \end{array} \quad \textit{Pile} : \begin{array}{|c|} \\ \hline (\\ \hline X \\ \hline B \\ \hline \end{array}$$

Commentaire : Le sommet de la pile désigne le même caractère que le pointeur.
On poursuit donc la lecture (*i.e.* on décale le pointeur et on dépile ().

$$\begin{array}{cccccc} & \nabla & & & & \\ \hline \multicolumn{1}{|c|}{(} & \multicolumn{1}{c|}{b} & \multicolumn{1}{c|}{a} & \multicolumn{1}{c|}{b} & \multicolumn{1}{c|}{)} & \multicolumn{1}{c|}{\$} \\ \hline \end{array} \quad \textit{Pile} : \begin{array}{|c|} \\ \hline X \\ \hline B \\ \hline \end{array}$$

Commentaire : b est symbole directeur de la règle $X \to A$
On dépile donc X et on empile A

$$\begin{array}{cccccc} & \nabla & & & & \\ \hline \multicolumn{1}{|c|}{(} & \multicolumn{1}{c|}{b} & \multicolumn{1}{c|}{a} & \multicolumn{1}{c|}{b} & \multicolumn{1}{c|}{)} & \multicolumn{1}{c|}{\$} \\ \hline \end{array} \quad \textit{Pile} : \begin{array}{|c|} \\ \hline A \\ \hline B \\ \hline \end{array}$$

Commentaire : b est symbole directeur de la règle $A \to bC$
On dépile donc A et on empile bC

$$\begin{array}{cccccc} & \nabla & & & & \\ \hline \multicolumn{1}{|c|}{(} & \multicolumn{1}{c|}{b} & \multicolumn{1}{c|}{a} & \multicolumn{1}{c|}{b} & \multicolumn{1}{c|}{)} & \multicolumn{1}{c|}{\$} \\ \hline \end{array} \quad \textit{Pile} : \begin{array}{|c|} \\ b \\ \hline C \\ \hline B \\ \hline \end{array}$$

Commentaire : Le sommet de la pile désigne le même caractère que le pointeur.
On poursuit donc la lecture (*i.e.* on décale le pointeur et on dépile b).

		∇			
(	b	a	b	)	\$

Pile :

C
B

Commentaire : a est symbole directeur de la règle $C \to \epsilon$
On dépile donc C et on empile ϵ (autrement dit, on n'empile rien !)

		∇			
(	b	a	b	)	\$

Pile :

B

Commentaire : a est symbole directeur de la règle $B \to aX$)
On dépile donc B et on empile aX)

		∇			
(	b	a	b	)	\$

Pile :

a
X
)

Commentaire : Le sommet de la pile désigne le même caractère que le pointeur.
On poursuit donc la lecture (*i.e.* on décale le pointeur et on dépile a).

			∇		
(	b	a	b	)	\$

Pile :

X
)

Commentaire : b est symbole directeur de la règle $X \to A$
On dépile donc X et on empile A

			∇		
(	b	a	b	)	\$

Pile :

A
)

Commentaire : b est symbole directeur de la règle $A \to bC$
On dépile donc A et on empile bC

			∇		
(	b	a	b	)	\$

Pile :

b
C
)

Commentaire : Le sommet de la pile désigne le même caractère que le pointeur.
On poursuit donc la lecture (*i.e.* on décale le pointeur et on dépile b).

				∇	
(	b	a	b	)	\$

Pile :

C
)

Commentaire :) est symbole directeur de la règle $C \to \epsilon$
On dépile donc C et on empile ϵ (autrement dit, on n'empile rien !)

				∇	
(	b	a	b	)	\$

Pile :

)

Commentaire : Le sommet de la pile désigne le même caractère que le pointeur.
On poursuit donc la lecture (*i.e.* on décale le pointeur et on dépile)).

					∇
(	b	a	b	)	\$

Pile :

Commentaire : La pile est vide et le pointeur désigne le marqueur « fin de mot »
$\Rightarrow$ le mot **est** engendré par la grammaire.