Théorie des langages : THL

CM 10

Uli Fahrenberg

EPITA Rennes

S5 2023

Aperçu

Programme du cours

1. Langages rationnels, automates finis
3. Langages algébriques, grammaires hors-contexte, automates à pile
   - TP 1 : `flex`
5. Parsage LL
   - TP 2 : LL
7. Parsage LR
   - TP 3 : `bison`
8. Conclusion
   - TP 4 : `flex` & `bison`

Re : parsage ascendant : the basics

function $\textsc{BULRP}(\alpha)$
    if $\alpha = S$ **then**
        return True
    for $i \leftarrow 1$ **to** $|\alpha|$ **do**
        for $j \leftarrow i$ **to** $|\alpha|$ **do** ▷ décalage / SHIFT
            for $A \in N$ **do**
                if $A \to \alpha_i \dots \alpha_j$ **then** ▷ réduction / REDUCE
                    if $\textsc{BULRP}(\alpha_1 \dots \alpha_{i-1} A \alpha_{j+1} \dots \alpha_n)$ **then**
                        return True
    return False

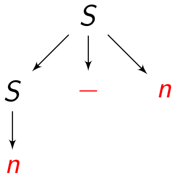
Re : parsage LR(0)

$$
\begin{aligned}
Z &\to S\$ && (0)\\
S &\to S-n && (1)\\
&\mid n && (2)
\end{aligned}
$$

état	action	n	$-$	$\$$	S
0	décaler	1			2
1	réduire 2				
2	décaler		4	3	
3	accepter				
4	décaler	5			
5	réduire 1				

parser $n - n\$$:

entrée	pile	action	
$n - n\$$	$\perp 0$	décaler	
$-n\$$	$\perp 01$	réduire 2	$S \to n$
$-n\$$	$\perp 02$	décaler	
$n\$$	$\perp 024$	décaler	
$\$$	$\perp 0245$	réduire 1	$S \to S-n$
$\$$	$\perp 02$	décaler	
	$\perp 023$	✓	



Re : parsage SLR(1)

1. calculer la table LR(0)
2. si conflits : conditionner l'action par le FOLLOW

Exemple :

$$
\begin{aligned}
Z &\to S\$ && (0)\\
S &\to n - S && (1)\\
&\mid n && (2)
\end{aligned}
$$

état	action	n	$-$	$\$$	S
0	décaler	2			1
1	décaler			4	
2	réd. 2, déc.		3		
3	décaler	2			5
4	accepter				
5	réduire 1				

$\Longrightarrow$

état	n	$-$	$\$$	S
0	d.2			d.1
1			d.4	
2		d.3	r.2	
3	d.2			d.5
4	— accepter —			
5			r.1	

Re : parsage LR(1)

- conditionner l'action par le contexte : les symboles qui peuvent suivre

Exemple :

$$
\begin{array}{rll}
Z \to & S\$ & (0) \\
S \to & L{=}E & (1) \\
\mid & E & (2) \\
L \to & x & (3) \\
\mid & *E & (4) \\
E \to & L & (5)
\end{array}
$$

état	productions pointées élargies
0	$Z \to \bullet S\$ \; [\varepsilon]$
	$S \to \bullet L{=}E \; [\$]$, $S \to \bullet E \; [\$]$
	$L \to \bullet x \; [=]$, $L \to \bullet {*}E \; [=]$
	$E \to \bullet L \; [\$]$
	$L \to \bullet x \; [\$]$, $L \to \bullet {*}E \; [\$]$
1	$Z \to S \bullet \$ \; [\varepsilon]$
2	$S \to L \bullet {=}E \; [\$]$, $E \to L \bullet \; [\$\checkmark]$

Exemple

$$
\begin{aligned}
Z &\to S\$ && (0)\\
S &\to L{=}E && (1)\\
&\mid E && (2)\\
L &\to x && (3)\\
&\mid {*}E && (4)\\
E &\to L && (5)
\end{aligned}
$$

état	x	$*$	$=$	$\$$	S	L	E
0	d.4	d.5			d.1	d.2	d.3
1				d.6			
2			d.7	r.5			
3				r.2			
4			r.3	r.3			
5	d.4	d.5				d.9	d.8
6			— accepter —				
7	d.12	d.13				d.11	d.10
8			r.4				
9			r.5				
10				r.1			
11				r.5			
12				r.3			
13	d.12	d.13				d.11	d.14
14				r.4			

Parsage LALR(1) et GLR

Exemple, bis

$$
\begin{aligned}
Z &\to S\$ && (0)\\
S &\to L{=}E && (1)\\
&\mid E && (2)\\
L &\to x && (3)\\
&\mid *E && (4)\\
E &\to L && (5)
\end{aligned}
$$

état	productions pointées élargies
0	$Z \to \bullet S\$\ [\varepsilon]$, $S \to \bullet L{=}E\ [\$]$, $S \to \bullet E\ [\$]$, $L \to \bullet x\ [=]$
	$L \to \bullet *E\ [=]$, $E \to \bullet L\ [\$]$, $L \to \bullet x\ [\$]$, $L \to \bullet *E\ [\$]$
1	$Z \to S\bullet\$\ [\varepsilon]$
2	$S \to L\bullet{=}E\ [\$]$, $E \to L\bullet\ [\$\checkmark]$
3	$S \to E\bullet\ [\$\checkmark]$
4	$L \to x\bullet\ [=\checkmark]$, $L \to x\bullet\ [\$\checkmark]$
5	$L \to *\bullet E\ [=]$, $L \to *\bullet E\ [\$]$, $E \to \bullet L\ [=]$, $L \to \bullet x\ [=]$
	$L \to \bullet *E\ [=]$, $E \to \bullet L\ [\$]$, $L \to \bullet x\ [\$]$, $L \to \bullet *E\ [\$]$
6	$Z \to S\$\bullet\ [\varepsilon\checkmark]$
7	$S \to L{=}\bullet E\ [\$]$, $E \to \bullet L\ [\$]$, $L \to \bullet x\ [\$]$, $L \to \bullet *E\ [\$]$
8	$L \to *E\bullet\ [=\checkmark]$, $L \to *E\bullet\ [\$\checkmark]$
9	$E \to L\bullet\ [=\checkmark]$, $E \to L\bullet\ [\$\checkmark]$
10	$S \to L{=}E\bullet\ [\$\checkmark]$
11	$E \to L\bullet\ [\$\checkmark]$
12	$L \to x\bullet\ [\$\checkmark]$
13	$L \to *\bullet E\ [\$]$, $E \to \bullet L\ [\$]$, $L \to \bullet x\ [\$]$, $L \to \bullet *E\ [\$]$
14	$L \to *E\bullet\ [\$\checkmark]$

Exemple, bis

$$
\begin{aligned}
Z &\to S\$ && (0)\\
S &\to L{=}E && (1)\\
&\mid E && (2)\\
L &\to x && (3)\\
&\mid {*}E && (4)\\
E &\to L && (5)
\end{aligned}
$$

état	productions pointées élargies
0	$Z \to \bullet S\$\ [\varepsilon]$, $S \to \bullet L{=}E\ [\$]$, $S \to \bullet E\ [\$]$, $L \to \bullet x\ [=]$ $L \to \bullet {*}E\ [=]$, $E \to \bullet L\ [\$]$, $L \to \bullet x\ [\$]$, $L \to \bullet {*}E\ [\$]$
1	$Z \to S \bullet \$\ [\varepsilon]$
2	$S \to L \bullet {=}E\ [\$]$, $E \to L \bullet\ [\$\checkmark]$
3	$S \to E \bullet\ [\$\checkmark]$
4	$L \to x \bullet\ [=\checkmark]$, $L \to x \bullet\ [\$\checkmark]$
5	$L \to {*}\bullet E\ [=]$, $L \to {*}\bullet E\ [\$]$, $E \to \bullet L\ [=]$, $L \to \bullet x\ [=]$ $L \to \bullet {*}E\ [=]$, $E \to \bullet L\ [\$]$, $L \to \bullet x\ [\$]$, $L \to \bullet {*}E\ [\$]$
6	$Z \to S\$ \bullet\ [\varepsilon\checkmark]$
7	$S \to L{=}\bullet E\ [\$]$, $E \to \bullet L\ [\$]$, $L \to \bullet x\ [\$]$, $L \to \bullet {*}E\ [\$]$
8	$L \to {*}E \bullet\ [=\checkmark]$, $L \to {*}E \bullet\ [\$\checkmark]$
9	$E \to L \bullet\ [=\checkmark]$, $E \to L \bullet\ [\$\checkmark]$
10	$S \to L{=}E \bullet\ [\$\checkmark]$
11	$E \to L \bullet\ [\$\checkmark]$
12	$L \to x \bullet\ [\$\checkmark]$
13	$L \to {*}\bullet E\ [\$]$, $E \to \bullet L\ [\$]$, $L \to \bullet x\ [\$]$, $L \to \bullet {*}E\ [\$]$
14	$L \to {*}E \bullet\ [\$\checkmark]$

Parsage LALR(1)

Définition

Deux productions pointées élargies $A \to \alpha \bullet \beta$ $[a]$ et $A \to \alpha' \bullet \beta'$ $[b]$ sont équivalent LALR(1) si $\alpha = \alpha'$ et $\beta = \beta'$.

- les items sont identiques, mais les contextes peuvent être différents

Définition

L'automate LALR(1) d'une grammaire hors-contexte G est le quotient de l'automate LR(1) de G sous équivalence LALR(1).

Exemple, ter

$$
\begin{array}{rcll}
Z & \to & S\$ & (0) \\
S & \to & L = E & (1) \\
 & | & E & (2) \\
L & \to & x & (3) \\
 & | & *E & (4) \\
E & \to & L & (5)
\end{array}
$$

état	x	$*$	$=$	$\$$	S	L	E
0	d.4	d.5			d.1	d.2	d.3
1				d.6			
2			d.7	r.5			
3				r.2			
4			r.3	r.3			
5	d.4	d.5				d.9	d.8
6			— accepter —				
7	d.12	d.13				d.11	d.10
8			r.4				
9			r.5				
10				r.1			
11				r.5			
12				r.3			
13	d.12	d.13				d.11	d.14
14				r.4			

Exemple, ter

$$
\begin{aligned}
Z &\to S\$ && (0)\\
S &\to L{=}E && (1)\\
&\mid E && (2)\\
L &\to x && (3)\\
&\mid *E && (4)\\
E &\to L && (5)
\end{aligned}
$$

état	x	$*$	$=$	$\$$	S	L	E
0	d.4	d.5			d.1	d.2	d.3
1				d.6			
2			d.7	r.5			
3				r.2			
4			r.3	r.3			
5	d.4	d.5				d.9	d.8
6			— accepter —				
7	d.12	d.13				d.11	d.10
8			r.4	r.4			
9			r.5	r.5			
10				r.1			

Résolution de conflits

Exemple :

$$\begin{aligned} Z &\to E\$ && (0) \\ E &\to E+E && (1) \\ &\mid E*E && (2) \\ &\mid n && (3) \end{aligned}$$

état	+	*	n	$	E
0			d.2		g.1
1	d.4	d.5		d.3	
2	r.3	r.3		r.3	
3	— accepter —				
4			d.2		g.6
5			d.2		g.7
6	d.4	d.5			
	r.1	r.1		r.1	
7	d.4	d.5			
	r.2	r.2		r.2	

- une grammaire ambiguë
- donc pas LR(k) pour n'importe quel k

Résolution de conflits

Exemple :

$$
\begin{aligned}
Z &\to E\$ && (0)\\
E &\to E+E && (1)\\
&\mid E*E && (2)\\
&\mid n && (3)
\end{aligned}
$$

état	+	*	n	$	E
0			d.2		g.1
1	d.4	d.5		d.3	
2	r.3	r.3		r.3	
3	— accepter —				
4			d.2		g.6
5			d.2		g.7
6	d.4	d.5			
	r.1	r.1		r.1	
7	d.4	d.5			
	r.2	r.2		r.2	

- une grammaire ambiguë
- donc pas LR(k) pour n'importe quel k
- associativité : d.4 $\Rightarrow$ $n+(n+n)$; r.1 $\Rightarrow$ $(n+n)+n$
- priorité : d.5 $\Rightarrow$ $n*(n+n)$; r.1 $\Rightarrow$ $(n*n)+n$

Résolution de conflits

Exemple :

$$
\begin{array}{lll}
Z \to E\$ & & (0) \\
E \to E+E & & (1) \\
\quad\;\; \mid E*E & & (2) \\
\quad\;\; \mid n & & (3)
\end{array}
$$

état	$+$	$*$	n	$\$$	E
0			d.2		g.1
1	d.4	d.5		d.3	
2	r.3	r.3		r.3	
3	— accepter —				
4			d.2		g.6
5			d.2		g.7
6	d.4	d.5			
	r.1	r.1		r.1	
7	d.4	d.5			
	r.2	r.2		r.2	

- une grammaire ambiguë
- donc pas LR(k) pour n'importe quel k
- associativité : d.4 $\Rightarrow n+(n+n)$; r.1 $\Rightarrow (n+n)+n$
- priorité : d.5 $\Rightarrow n*(n+n)$; r.1 $\Rightarrow (n*n)+n$
- solution : règles de priorité
- ici : r.1 $>$ d.4, r.2 $>$ d.5, r.2 $>$ d.4, d.5 $>$ r.1 $\Leftarrow$!

Parsage LR généralisé

- *embrace non-determinism!*
- parsage GLR : en cas de conflit, suivre tous les chemins en parallel
- « parsage parallel », « parsage Tomita »
- implémenter l'automate (non-déterministe) de parsage sans déterminisation
- états : productions pointées, pas de clôture
- algorithme en temps exponentiel, pas linéaire
- optimisation : partager préfixes et suffixes de piles

Résumé du cours

Hiérarchie de Chomsky

type	langages	grammaires	automates
4	finis	à choix finis	finis acycliques
	$\subsetneq$	$\Downarrow$	$\Downarrow$
3	réguliers	régulières	finis
	$\subsetneq$	$\Downarrow$	$\Downarrow$
2	algébriques	hors-contexte	à pile
	$\subsetneq$	$\Downarrow$	
1	contextuels	contextuelles	linéairement bornés
	$\subsetneq$	$\Downarrow$	$\Downarrow$
0	récursivement énumerables	syntagmatiques	de Turing

Hiérarchie de Chomsky

type	langages	grammaires	automates
4	finis	à choix finis	finis acycliques
	$\subsetneq$	$\Downarrow$	$\Downarrow$
3	réguliers	régulières	finis
	$\subsetneq$	$\Downarrow$	$\Downarrow$
2	algébriques	hors-contexte	à pile
	$\subsetneq$	$\Downarrow$	
1	contextuels	contextuelles	linéairement bornés
	$\subsetneq$	$\Downarrow$	$\Downarrow$
0	récursivement énumerables	syntagmatiques	de Turing

Zoom sur type 3

syntaxe		sémantique
aut. finis dét. complets		langages reconnaissables
$\subsetneq$		$=$
aut. finis déterministes		langages reconnaissables
$\subsetneq$	$L(\cdot)$	$=$
automates finis	$\longrightarrow$	langages reconnaissables
$\subsetneq$		$=$
aut. finis à trans. spontanées		langages reconnaissables
		$=$
expressions rationnelles		langages rationnelles
		$=$
grammaires régulières		langages réguliers

Zoom sur type 2

syntaxe		sémantique
grammaires hc forme Greibach		langages algébriques
$\subsetneq$		$=$
grammaires hors-contexte		langages algébriques
$\supsetneq$		$=$
grammaires hc forme Chomsky	$\xrightarrow{L(\cdot)}$	langages algébriques
		$=$
automates à pile		langages algébriques
$\supsetneq$		$=$
automates à pile sans trans. spont.		langages algébriques
$\supsetneq$		$\supsetneq$
automates à pile déterministes		langages algébriques déterministes

Zoom sur LR

syntaxe		sémantique
grammaires hors-contexte		langages algébriques
$\supsetneq$		$\supsetneq$
grammaires hc non-ambiguës		lang. alg. non-ambigués
$\supsetneq$		$\supsetneq$
grammaires hc déterministes		lang. alg. déterministes
$\supsetneq$		$=$
grammaires LR(k)	$\xrightarrow{L(\cdot)}$	lang. alg. déterministes
$\supsetneq$		$=$
$\cdots$		$\cdots$
$\supsetneq$		$=$
grammaires LR(1)		lang. alg. déterministes
$\supsetneq$		$\supsetneq$
grammaires LALR(1)		langages LALR(1)
$\supsetneq$		$\supsetneq$
grammaires SLR(1)		langages SLR(1)
$\supsetneq$		$\supsetneq$
grammaires LR(0)		langages LR(0)

Zoom sur LL

syntaxe		sémantique
grammaires hors-contexte		langages algébriques
$\supsetneq$		$\supsetneq$
grammaires hc non-ambiguës		lang. alg. non-ambigués
$\supsetneq$		$\supsetneq$
grammaires hc déterministes		lang. alg. déterministes
$\supsetneq$		$\supsetneq$
grammaires LL(k)	$\xrightarrow{L(\cdot)}$	langages LL(k)
$\supsetneq$		$\supsetneq$
$\cdots$		$\cdots$
$\supsetneq$		$\supsetneq$
grammaires LL(2)		langages LL(2)
$\supsetneq$		$\supsetneq$
grammaires LL(1)		langages LL(1)


That's all Folks!