

Lecture 9

Syntax Analysis

Awanish Pandey

Department of Computer Science and Engineering Indian Institute of Technology Roorkee

February 7, 2025



Calculation of follow set



- Calculation of follow set
- Creation of parse table



- Calculation of follow set
- Creation of parse table
- Error recovery technique.



- Calculation of follow set.
- Creation of parse table
- Error recovery technique.
- Bottom up parsing



- Calculation of follow set.
- Creation of parse table
- Error recovery technique.
- Bottom up parsing
- Shift-Reduce parser





• Assume grammar is $E \rightarrow E + E|E*E|id$



- Assume grammar is $E \rightarrow E + E|E*E|id$
- Parse id*id+id



- Assume grammar is $E \rightarrow E + E|E*E|id$
- Parse id*id+id

String	Action



- Assume grammar is $E \rightarrow E + E|E*E|id$
- Parse id*id+id

String	Action
.id*id+id	shift



- Assume grammar is $E \rightarrow E + E|E*E|id$
- Parse id*id+id

String	Action
.id*id+id	shift
id.*id+id	reduce $E o id$



- Assume grammar is $E \rightarrow E + E|E*E|id$
- Parse id*id+id

String	Action
.id*id+id	shift
id.*id+id	reduce $E o id$
E.*id+id	shift



- Assume grammar is $E \rightarrow E + E|E*E|id$
- Parse id*id+id

String	Action
.id*id+id	shift
id.*id+id	reduce $ extbf{ extit{E}} ightarrow extit{id}$
E.*id+id	shift
$E^*.id+id$	shift



- Assume grammar is $E \rightarrow E + E|E*E|id$
- Parse id*id+id

String	Action
.id*id+id	shift
id.*id+id	reduce $E o id$
E.*id+id	shift
$E^*.id+id$	shift
E*id.+id	reduce $ extbf{ extit{E}} ightarrow extit{id}$



- Assume grammar is $E \rightarrow E + E|E*E|id$
- Parse id*id+id

String	Action
.id*id+id	shift
id.*id+id	reduce $E o id$
E.*id+id	shift
$E^*.id+id$	shift
E*id.+id	reduce $E o id$
E*E.+id	reduce $E o E * E$



- Assume grammar is $E \rightarrow E + E|E*E|id$
- Parse id*id+id

String	Action
.id*id+id	shift
id.*id+id	reduce $E o id$
E.*id+id	shift
$E^*.id+id$	shift
E*id.+id	reduce $E o id$
E*E.+id	reduce $E \to E * E$
E.+id	shift



- Assume grammar is $E \rightarrow E + E|E*E|id$
- Parse id*id+id

Action
shift
$reduce\; E \to \mathit{id}$
shift
shift
reduce $E o id$
$reduce\; E \to E * E$
shift
shift



- Assume grammar is $E \rightarrow E + E|E * E|id$
- Parse id*id+id

Action
shift
reduce $E o id$
shift
shift
reduce $E o id$
reduce $E \rightarrow E * E$
shift
shift
reduce $ extbf{ extit{E}} ightarrow extit{id}$



- Assume grammar is $E \rightarrow E + E|E * E|id$
- Parse id*id+id

String	Action
.id*id+id	shift
id.*id+id	reduce $E o id$
E.*id+id	shift
$E^*.id+id$	shift
E*id.+id	reduce $E o id$
E*E.+id	reduce $E \rightarrow E * E$
E.+id	shift
E+.id	shift
E+id.	reduce $E o id$
E+E.	reduce $E \rightarrow E + E$



- Assume grammar is $E \rightarrow E + E|E*E|id$
- Parse id*id+id

Ctring	Action
String	Action
.id*id+id	shift
id.*id+id	reduce $E o id$
E.*id+id	shift
$E^*.id+id$	shift
E*id.+id	reduce $E o id$
E*E.+id	reduce $E \rightarrow E * E$
E.+id	shift
E+.id	shift
E+id.	reduce $E o id$
E+E.	reduce $E \rightarrow E + E$
E.	ACCEPT





• Symbols on the left of "." are kept on a stack



- Symbols on the left of "." are kept on a stack
 - ► Top of the stack is at "."



- Symbols on the left of "." are kept on a stack
 - ► Top of the stack is at "."
 - ► Shift pushes a terminal on the stack



- Symbols on the left of "." are kept on a stack
 - ► Top of the stack is at "."
 - ► Shift pushes a terminal on the stack
 - ▶ Reduce pops symbols (rhs of production) and pushes a non terminal (lhs of production) onto the stack



- Symbols on the left of "." are kept on a stack
 - ► Top of the stack is at "."
 - Shift pushes a terminal on the stack
 - ▶ Reduce pops symbols (rhs of production) and pushes a non terminal (lhs of production) onto the stack
- The most important issue: when to shift and when to reduce



- Symbols on the left of "." are kept on a stack
 - ► Top of the stack is at "."
 - Shift pushes a terminal on the stack
 - Reduce pops symbols (rhs of production) and pushes a non terminal (lhs of production) onto the stack
- The most important issue: when to shift and when to reduce
- Reduce action should be taken only if the result can be reduced to the start symbol



• A more powerful parsing technique



- A more powerful parsing technique
- LR grammars more expensive than LL



- A more powerful parsing technique
- LR grammars more expensive than LL
- Can handle left recursive grammars



- A more powerful parsing technique
- LR grammars more expensive than LL
- Can handle left recursive grammars
- Can handle virtually all the programming languages



- A more powerful parsing technique
- LR grammars more expensive than LL
- Can handle left recursive grammars
- Can handle virtually all the programming languages
- Natural expression of programming language syntax



- A more powerful parsing technique
- LR grammars more expensive than LL
- Can handle left recursive grammars
- Can handle virtually all the programming languages
- Natural expression of programming language syntax
- Automatic generation of parsers (Yacc, Bison etc.)



- A more powerful parsing technique
- LR grammars more expensive than LL
- Can handle left recursive grammars
- Can handle virtually all the programming languages
- Natural expression of programming language syntax
- Automatic generation of parsers (Yacc, Bison etc.)
- Detects errors as soon as possible



- A more powerful parsing technique
- LR grammars more expensive than LL
- Can handle left recursive grammars
- Can handle virtually all the programming languages
- Natural expression of programming language syntax
- Automatic generation of parsers (Yacc, Bison etc.)
- Detects errors as soon as possible
- Allows better error recovery





How do we know which action to take



- How do we know which action to take
 - whether to shift or reduce



- How do we know which action to take
 - whether to shift or reduce
 - which production to use for reduction?



- How do we know which action to take
 - whether to shift or reduce
 - which production to use for reduction?
- Sometimes parser can reduce but it should not: $X \to \epsilon$ can always be reduced!



- How do we know which action to take
 - whether to shift or reduce
 - which production to use for reduction?
- Sometimes parser can reduce but it should not: $X \to \epsilon$ can always be reduced!
- ullet Given stack δ and input symbol 'a', should the parser



- How do we know which action to take
 - whether to shift or reduce
 - which production to use for reduction?
- Sometimes parser can reduce but it should not: $X \to \epsilon$ can always be reduced!
- \bullet Given stack δ and input symbol 'a', should the parser
 - ▶ Shift a onto stack (making it δa)



- How do we know which action to take
 - whether to shift or reduce
 - which production to use for reduction?
- Sometimes parser can reduce but it should not: $X \to \epsilon$ can always be reduced!
- ullet Given stack δ and input symbol 'a', should the parser
 - ▶ Shift a onto stack (making it δa)
 - ▶ Reduce by some production $A \rightarrow \beta$ assuming that stack has form $a\beta$ (making it aA)



- How do we know which action to take
 - whether to shift or reduce
 - which production to use for reduction?
- Sometimes parser can reduce but it should not: $X \to \epsilon$ can always be reduced!
- ullet Given stack δ and input symbol 'a', should the parser
 - ▶ Shift a onto stack (making it δa)
 - ▶ Reduce by some production $A \rightarrow \beta$ assuming that stack has form $a\beta$ (making it aA)
 - ▶ How to keep track of length of β ?





• A string that matches right hand side of a production and whose replacement gives a step in the reverse right most derivation



- A string that matches right hand side of a production and whose replacement gives a step in the reverse right most derivation
- If $S \to^{rm*} aAw \to^{rm} a\beta w$ then β (corresponding to production $A \to \beta$) in the position following 'a' is a handle of $a\beta w$. The string w consists of only terminal symbols



- A string that matches right hand side of a production and whose replacement gives a step in the reverse right most derivation
- If $S \to^{rm*} aAw \to^{rm} a\beta w$ then β (corresponding to production $A \to \beta$) in the position following 'a' is a handle of $a\beta w$. The string w consists of only terminal symbols
- We only want to reduce handle and not any rhs



- A string that matches right hand side of a production and whose replacement gives a step in the reverse right most derivation
- If $S \to^{rm*} aAw \to^{rm} a\beta w$ then β (corresponding to production $A \to \beta$) in the position following 'a' is a handle of $a\beta w$. The string w consists of only terminal symbols
- We only want to reduce handle and not any rhs
- Handle pruning: If β is a handle and $A \to \beta$ is a production then replace β by A



- A string that matches right hand side of a production and whose replacement gives a step in the reverse right most derivation
- If $S \to^{rm*} aAw \to^{rm} a\beta w$ then β (corresponding to production $A \to \beta$) in the position following 'a' is a handle of $a\beta w$. The string w consists of only terminal symbols
- We only want to reduce handle and not any rhs
- Handle pruning: If β is a handle and $A \to \beta$ is a production then replace β by A
- A right most derivation in reverse can be obtained by handle pruning



- A string that matches right hand side of a production and whose replacement gives a step in the reverse right most derivation
- If $S \to^{rm*} aAw \to^{rm} a\beta w$ then β (corresponding to production $A \to \beta$) in the position following 'a' is a handle of $a\beta w$. The string w consists of only terminal symbols
- We only want to reduce handle and not any rhs
- Handle pruning: If β is a handle and $A \to \beta$ is a production then replace β by A
- A right most derivation in reverse can be obtained by handle pruning
- Handles always appear at the top of the stack and never inside it. This makes stack a suitable data structure



- A string that matches right hand side of a production and whose replacement gives a step in the reverse right most derivation
- If $S \to^{rm*} aAw \to^{rm} a\beta w$ then β (corresponding to production $A \to \beta$) in the position following 'a' is a handle of $a\beta w$. The string w consists of only terminal symbols
- We only want to reduce handle and not any rhs
- Handle pruning: If β is a handle and $A \to \beta$ is a production then replace β by A
- A right most derivation in reverse can be obtained by handle pruning
- Handles always appear at the top of the stack and never inside it. This makes stack a suitable data structure
- Bottom up parsing is based on recognizing handles





• The general shift-reduce technique is:



- The general shift-reduce technique is:
 - ▶ if there is no handle on the stack then shift



- The general shift-reduce technique is:
 - ▶ if there is no handle on the stack then shift
 - ▶ If there is a handle then reduce



- The general shift-reduce technique is:
 - ▶ if there is no handle on the stack then shift
 - ▶ If there is a handle then reduce
- However, what happens when there is a choice



- The general shift-reduce technique is:
 - ▶ if there is no handle on the stack then shift
 - ▶ If there is a handle then reduce
- However, what happens when there is a choice
 - What action to take in case both shift and reduce are valid. Shift-Reduce conflict



- The general shift-reduce technique is:
 - if there is no handle on the stack then shift
 - ▶ If there is a handle then reduce
- However, what happens when there is a choice
 - What action to take in case both shift and reduce are valid. Shift-Reduce conflict
 - Which rule to use for reduction if reduction is possible by more than one rule?
 Reduce-Reduce conflict



- The general shift-reduce technique is:
 - if there is no handle on the stack then shift
 - ▶ If there is a handle then reduce
- However, what happens when there is a choice
 - ▶ What action to take in case both shift and reduce are valid. Shift-Reduce conflict
 - Which rule to use for reduction if reduction is possible by more than one rule? Reduce-Reduce conflict
- Conflicts come either because of ambiguous grammars or parsing method is not powerful enough





• Consider the grammar $E \rightarrow E + E|E * E|id$



- Consider the grammar $E \rightarrow E + E|E*E|id$
- input id + id * id



ullet Consider the grammar E o E + E|E*E|id

• input id + id * id

input	action		
*id	reduce $E \rightarrow E + E$		
*id	shift		
id	shift		
	reduce E→id		
	reduce E→E*E		
	*id *id		



• Consider the grammar $E \rightarrow E + E|E * E|id$

• input id + id * id

stack	input	action
E+E	*id	reduce $E \rightarrow E + E$
Е	*id	shift
E*	id	shift
E*id		reduce E→id
E*E		reduce E→E*E
Е		

stack	input	action
E+E	*id	shift
E+E*	id	shift
E+E*id		reduce E→id
E+E*E		reduce $E \rightarrow E*E$
E+E		reduce $E \rightarrow E + E$
Е		





ullet Consider the grammar M o R + R|R + c|R R o c



- ullet Consider the grammar M o R + R|R + c|R R o c
- input c + c



ullet Consider the grammar M o R + R|R + c|R R o c

• input c + c

input	action
c+c	shift
+c	reduce $R{ ightarrow}C$
+c	shift
С	shift
	reduce R→c
	reduce $M \rightarrow R + R$
	c+c +c +c



• Consider the grammar $M \to R + R|R + c|R$ $R \to c$

• input c + c

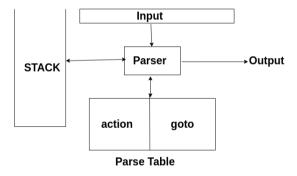
input	action
c+c	shift
+c	reduce $R{ ightarrow}C$
+c	shift
С	shift
	reduce R $ ightarrow$ c
	reduce $M \rightarrow R + R$
	c+c +c +c

stack	input	action
	c+c	shift
С	+c	reduce $R{ ightarrow}C$
R	+c	shift
R+	С	shift
R+c		reduce $M \rightarrow R + c$
М		

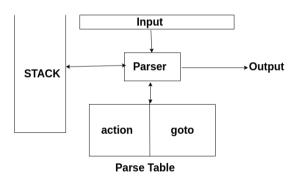


LR Parsing



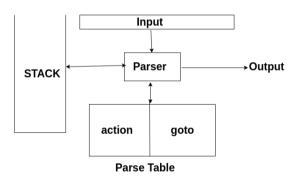






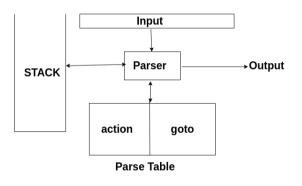
• Input contains the input string.





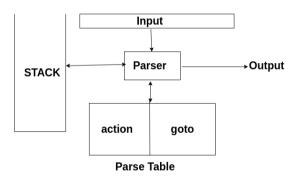
- Input contains the input string.
- Stack contains a string of the form $S_0X_1S_1X_2\cdots X_nS_n$ where each X_i is a grammar symbol and each S_i is a state.





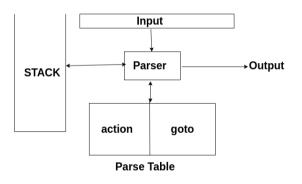
- Input contains the input string.
- Stack contains a string of the form $S_0X_1S_1X_2\cdots X_nS_n$ where each X_i is a grammar symbol and each S_i is a state.
- Tables contain **action** and **goto** parts.





- Input contains the input string.
- Stack contains a string of the form $S_0X_1S_1X_2\cdots X_nS_n$ where each X_i is a grammar symbol and each S_i is a state.
- Tables contain action and goto parts.
- action table is indexed by state and terminal symbols.





- Input contains the input string.
- Stack contains a string of the form $S_0X_1S_1X_2\cdots X_nS_n$ where each X_i is a grammar symbol and each S_i is a state.
- Tables contain action and goto parts.
- action table is indexed by state and terminal symbols.
- goto table is indexed by state and non terminal symbols.





ullet Assume S_i is top of stack and a_j is current input symbol



- Assume S_i is top of stack and a_i is current input symbol
- Action $[S_i, a_i]$ can have four values



- Assume S_i is top of stack and a_j is current input symbol
- Action $[S_i, a_i]$ can have four values
 - **1** shift a_j to the stack and goto state S_k



- Assume S_i is top of stack and a_i is current input symbol
- Action $[S_i, a_i]$ can have four values
 - \bullet shift a_j to the stack and goto state S_k
 - reduce by a rule



- Assume S_i is top of stack and a_i is current input symbol
- Action $[S_i, a_i]$ can have four values
 - lacktriangledown shift a_j to the stack and goto state S_k
 - reduce by a rule
 - accept



- Assume S_i is top of stack and a_i is current input symbol
- Action $[S_i, a_i]$ can have four values
 - **1** shift a_j to the stack and goto state S_k
 - reduce by a rule
 - accept
 - error





$$Stack: S_0X_1S_1X_2\cdots X_mS_m$$
 Input: $a_ia_{i+1}\cdots a_n$ \$

• If $action[S_m, a_i] = shift S$



Stack: $S_0X_1S_1X_2\cdots X_mS_m$ Input: $a_ia_{i+1}\cdots a_n$ \$

• If $action[S_m,a_i]=shift\ S$ Then the configuration becomes

Stack: $S_0X_1S_1\cdots X_mS_ma_iS$ Input: $a_{i+1}\cdots a_n$ \$



 $Stack: S_0X_1S_1X_2\cdots X_mS_m$ Input: $a_ia_{i+1}\cdots a_n$ \$

• If $action[S_m, a_i] = shift S$ Then the configuration becomes $Stack: S_0X_1S_1\cdots X_mS_ma_iS$ Input: $a_{i+1}\cdots a_n$ \$

• If $action[S_m, a_i] = reduce A \rightarrow \beta$



$$Stack: S_0X_1S_1X_2\cdots X_mS_m$$
 Input: $a_ia_{i+1}\cdots a_n$ \$

- If $action[S_m, a_i] = shift S$ Then the configuration becomes
 - Stack: $S_0X_1S_1\cdots X_mS_ma_iS$ Input: $a_{i+1}\cdots a_n$ \$
- If $action[S_m, a_i] = \text{reduce } A \rightarrow \beta$ Then the configuration becomes $Stack : S_0X_1S_1 \cdots S_{m-r}X_{m-r}S$ Input : $a_ia_{i+1} \cdots a_n$ \$



- If $action[S_m, a_i] = shift\ S$ Then the configuration becomes $Stack: S_0X_1S_1\cdots X_mS_ma_iS \quad Input: a_{i+1}\cdots a_n$ \$
- If $action[S_m, a_i] = \text{reduce } A \rightarrow \beta$ Then the configuration becomes $Stack: S_0X_1S_1 \cdots S_{m-r}X_{m-r}S$ Input: $a_ia_{i+1} \cdots a_n$ \$ Where $r = |\beta|$ and $S = goto[S_{m-r}, A]$



$$Stack: S_0X_1S_1X_2\cdots X_mS_m \quad Input: a_ia_{i+1}\cdots a_n$$
\$

- If $action[S_m, a_i] = shift\ S$ Then the configuration becomes $Stack: S_0X_1S_1\cdots X_mS_ma_iS \quad Input: a_{i+1}\cdots a_n$
- If $action[S_m, a_i] = \text{reduce } A \rightarrow \beta$ Then the configuration becomes $Stack: S_0X_1S_1 \cdots S_{m-r}X_{m-r}S$ Input: $a_ia_{i+1} \cdots a_n$ \$ Where $r = |\beta| and$ $S = goto[S_{m-r}, A]$
- If $action[S_m, a_i] = accept$



$$Stack: S_0X_1S_1X_2\cdots X_mS_m$$
 Input: $a_ia_{i+1}\cdots a_n$ \$

- If $action[S_m, a_i] = shift\ S$ Then the configuration becomes $Stack: S_0X_1S_1\cdots X_mS_ma_iS \quad Input: a_{i+1}\cdots a_n$ \$
- If $action[S_m, a_i] = \text{reduce } A \rightarrow \beta$ Then the configuration becomes $Stack: S_0X_1S_1 \cdots S_{m-r}X_{m-r}S$ Input: $a_ia_{i+1} \cdots a_n$ \$ Where $r = |\beta|$ and $S = goto[S_{m-r}, A]$
- If $action[S_m, a_i] = accept$ Then parsing is completed. *HALT*



- If $action[S_m, a_i] = shift\ S$ Then the configuration becomes $Stack: S_0X_1S_1\cdots X_mS_ma_iS \quad Input: a_{i+1}\cdots a_n$
- If $action[S_m, a_i] = \text{reduce } A \rightarrow \beta$ Then the configuration becomes $Stack: S_0X_1S_1 \cdots S_{m-r}X_{m-r}S$ Input: $a_ia_{i+1} \cdots a_n$ \$ Where $r = |\beta|$ and $S = goto[S_{m-r}, A]$
- If $action[S_m, a_i] = accept$ Then parsing is completed. HALT
- If $action[S_m, a_i] = error$



$$Stack: S_0X_1S_1X_2\cdots X_mS_m$$
 Input: $a_ia_{i+1}\cdots a_n$ \$

- If $action[S_m, a_i] = shift\ S$ Then the configuration becomes $Stack: S_0X_1S_1\cdots X_mS_ma_iS \quad Input: a_{i+1}\cdots a_n$
- If $action[S_m, a_i] = \text{reduce } A \rightarrow \beta$ Then the configuration becomes $Stack: S_0X_1S_1 \cdots S_{m-r}X_{m-r}S$ Input: $a_ia_{i+1} \cdots a_n$ \$ Where $r = |\beta|$ and $S = goto[S_{m-r}, A]$
- If $action[S_m, a_i] = accept$ Then parsing is completed. HALT
- If $action[S_m, a_i] = error$ Then invoke error recovery routine



LR parsing Algorithm

Algorithm $LR_parsing_Algorithm$

```
1: Initial State Stack : So Input : w$
 2. for TRUE do
      if action[S, a] = shift S' then
        push(a); push(S'); ip++
      else if action[S,a] = reduce A \rightarrow \beta then
        pop (2 * |\beta|) symbols:
6:
        push(A); push (goto[S",A]) {S" is the state after popping symbols}
      else if action[S,a] = accept then
8:
9:
        exit
10:
      else
11:
        error()
      end if
12:
13: end for
```



Example

• Consider the grammar



Example

• Consider the grammar $E \to E + T|T$ $T \to T * F|F$ $F \to (E)|id$

state	id	+	*	()	\$	Е	Т	F
0	s5			s4			1	2	3
1		s6				acc			
2		r2	s7		r2	r2			
3		r4	r4		r4	r4			
4	s5			s4			8	2	3
5		r6	r6		r6	r6			
6	s5			s4				9	3
7	s5			s4					10
8		s6			s11				
9		r1	s7		r1	r1			
10		r3	r3		r3	r3			
11		r5	r5		r5	r5			



Parse id + id * id

stack	input	action					
0	id + id * id\$	s5					
0id5	+ id*id\$	reduce $F{ ightarrow}id$					
0F3	+id*id\$	$reduce\;T {\to}\;F$					
0T2	+id*id\$	reduce $E{ ightarrow}T$					
0E1	+id*id\$	shift 6					
0E1+6	id*id \$	shift 5					
0E1+6id5	*id \$	reduce F→id					
0E1+6F3	*id \$	reduce $T{ ightarrow} F$					
0E1+6T9	*id \$	shift 7					
0E1+6T9*7	id \$	shift 5					
0E1+6T9*7id5	\$	$reduce\;F\toid$					
0E1+6T9*7F10	\$	reduce $T \rightarrow T*F$					
0E1+6T9	\$	reduce $E \rightarrow E + T$					
0E1	\$	acc					

