***Compilers***

***Principle ,Techniques, and Tools***

***Anbar University – Computer College***

**LR Parsers**

This section presents an efficient Bottom-Up syntax analysis technique that can be used to parse a large class of context-free grammars. The technique is called LR(k) parsing, the "L" is for left to right scanning of input, the "R" for constructing a rightmost derivation in reverse, and "k" for the number of input symbols of lookahead that are used in making parsing decisions-when "k" is omitted , k is assumed to be 1). LR parsing is attractive for a variety of reasons:-

1. LR parsers can be constructed to recognize virtually all programming language constructs for which context-free grammars can be written.
2. The LR parsing method is the most general *nonbacktracking* shift-reduce parsing method known, yet itcan be implemented as efficiently as other shift-reduce methods.
3. The class of grammars that can be parsed using LR methods is a proper superset of the class of grammars that can be parsed with predictive parsers.

The schematic form of an LR parser is shown in following figure .It consists of **an Input**,**an Output**, **a Stack**,**a Driver** **program**,and **a Parsing table** that has two parts (*action*and*goto*) .

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Input | | a1 | …… | ai | ……. | an | $ |
|  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |



|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | LR | | | |  |  |
| Sm |  |  |  |  |  | Output |
|  |  | Parsing program | | | | | |  |
|  |  |  |  |
| Xm |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Sm-1 |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| Xm-1 |  |  |  |  |  |  |  |  |  |  |
| ……… |  |  | *action* | |  |  | *goto* | |  |  |
| S0 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |



Stack

**Model of an LR parser**

***With My Best Wishes*** *35* ***Esam***

***Compilers***

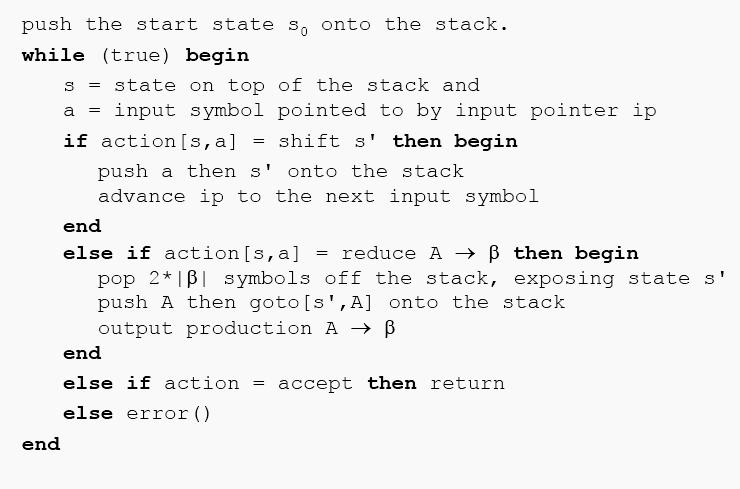
***Principle ,Techniques, and Tools***

***Anbar University – Computer College***

There are three techniques for LR parser depending on the construct of LR parsing table for a grammar :

1. **Simple LR parser (SLR for short):**Is the easiest toimplement but the least powerful of the three.It may be fail to produce a parsing table for certain grammars on which the other methods succeed.
2. **Canonical LR parser:** It is most powerful, and mostexpensive.
3. **Lookahead LR parser (LALR for short):**It isintermediate in power and cost between other two. The LALR method will work on most programming-language grammars and ,with some effort ,can be implemented efficiently.

**The LR parsing Algorithm :-** The LR program is the same forall LR parsers, only the parsing table changes from one parser to another.



***With My Best Wishes*** *36* ***Esam***

***Compilers***

***Principle ,Techniques, and Tools***

***Anbar University – Computer College***

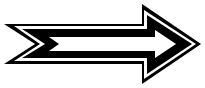
**Implementation of SLR parser:-**

The SLR-parser Extremely tedious to build by hand, so need a generator. The following steps represents the main stages, which are used to build system that is used for implementing the SLR-parser:

1. **Input stage :** Inthis state the grammar has been readingand the symbols of grammar (*terminals* and *nonterminals*) could be specified and each production of grammar must be on one straight line. Finally, the productions has been numbered.

For example, consider the grammar

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| E | E+T | |  | T | 1 E | E+T |
|  | 2 E | T |
|  |
| T | T\*F |  |  | F | 3 T | T\*F |
|  |  |
|  |  | 4 T | F |
| F | ( E ) | |  | id |
|  | 5 F | ( E ) |
|  |
|  |
|  |  |  |  |  | 6 F | id |



1. **Compute First & Follow stage :** Through this state First
   * Follow could be detected for each *nonterminal*.
2. **Construct DFA stage:**By using a deterministic finiteautomaton ( DFA )the SLR-parser know when to *shift* and when to *reduce.* the edges of DFA are labeled by symbols of grammar( terminals & nonterminals).In this state, where the input begins with S'(root),that means that it begins with any possible right-hand side of an S-production we indicate

|  |  |  |
| --- | --- | --- |
| that by |  |  |
| S' | **.**S$ |
|  |
|  | S | **.** |
|  | **.** |  |
|  | **.** |  |



Call this **state1** or **state0**,a productions combined with the **dot**(**.**) that indicates a position of parser.Firstly ,for eachproduction in state1 we exam the symbol that occur after dot, there are three cases :

1. If the symbol is ***null*** (the dot has been occurred in the end of right side of production),then there are no new state .
2. If the symbol is **"$"** sign, then there are no new state.

***With My Best Wishes*** *37* ***Esam***

***Compilers***

***Principle ,Techniques, and Tools***

***Anbar University – Computer College***

1. If the symbol is a ***terminal*** or ***nonterminal***, then there are new state, this state start with current production after the dot has been proceeded one step forward. If the symbol has been occurred after the dot(in new position)is *nonterminal* such as ***A***, then we add all possible right hand side of ***A*** to a new state, and so on.

You must know that any new state must built firstly in a buffer, and we compare it with a previous states in **DFA**, if there are no similarity situation then the new state is added to **DFA** and give it a new number equal to number of states in DFA plus one. Finally ,we repeat this steps on all new states until the **DFA** completed.

**Example** : consider grammar

**E  T+E**

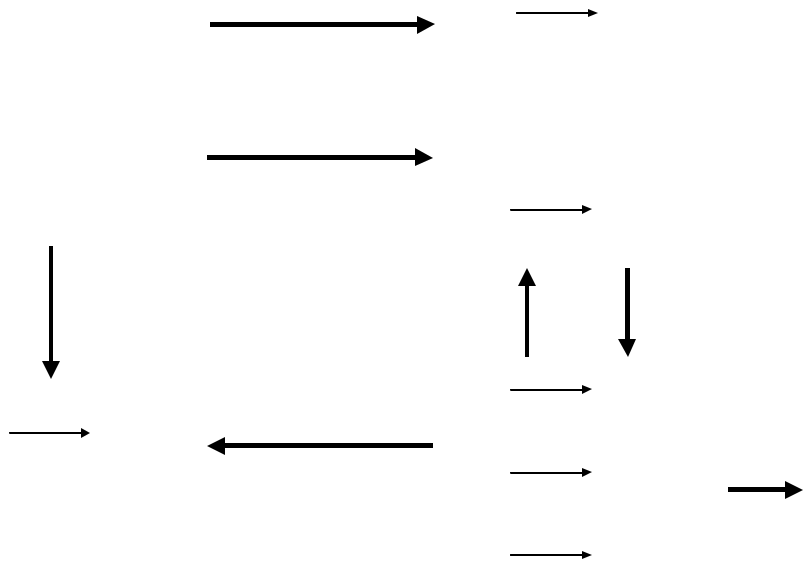
**E  T**

**T  x**

Initially, it will have an empty stack, and the input will be a complete S-sentence followed by $;that is the right-hand side of the S' rule will be on the input. we indicate this as S' . S$ where the dot indicates the current position of the parser. So:



|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | **0 S'** | | **E$** |  |  |  |  |
|  |  |  |  |  | **1 E** | | **T+E** |  |  |  |  |
|  |  |  |  |  | **2 E** | | **T** |  |  |  |  |
|  |  | 0 |  |  | **3 T** | | **x** | 1 |  |  |  |
|  |  | E |  |  |  |  |  |  |  |
|  |  |  |  | **S'** | **E .** |  |  |  |  |
| **S'** | **. E$** |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| **E** | **.** |  | T |  |  |  |  | 2 |  |  |  |
| **T+E** |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| **E** | **. T** |  |  |  |  | **E** | **T .** |  |  |  |  |
|  |  |  |  |  |  | **+E** |  |  |  |  |  |
| x |  |  |  |  |  | T | + | 3 |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  | 4 |  |  |  | **E** | **T+** |  |  |  | 5 |
|  |  |  |  |  |  |  |  |
| **T** | **x .** |  |  | x |  | **.E** |  |  |  | **E** | **T+ E** |
|  |  |  |  |  |  | **E** | **.** |  | E |
|  |  |  |  |  |  |  | **.** |  |
|  |  |  |  |  |  | **T+E** | |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | **E** | **.T** |  |  |  |  |
|  | **Deterministic finite automaton (DFA)** | | | | | | | | |  | ***Esam*** |
| ***With My Best Wishes*** | | | |  |  |  | *38* |  |  |  |



***Compilers***

***Principle ,Techniques, and Tools***

***Anbar University – Computer College***

* 1. **Construct SLR table stage:**The SLR-table is a datastructure consist of many rows equal to the number of the states in DFA,also many columns equal to the number of grammar symbols plus "$" sign .As know ,data structure presents fast in information treatment and information retrieve .In this stage SLR-table is constructed .this table had seen as two subtables:

1. **The Action table:** consist of many rows equal to numberof states in DFA,and many columns equal to number of terminals plus "$" sign(the end of input).
2. **The Goto table:** consist of many rows equal to number ofstates in DFA,and many columns equal to number of nonterminals.

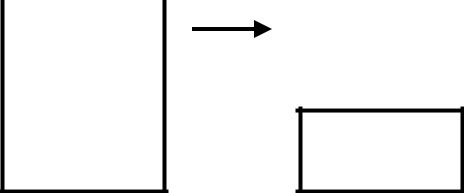
the elements(entries) in the SLR-table are labeled with four kinds of actions:

|  |  |  |
| --- | --- | --- |
|  | Sn | shift into state n |
|  | gn | goto state n |
|  | rk | reduce by production k |
|  | a | accept |

* error (denoted by blank entry in the table)

For the construction of this table and the contribution the actions on the tables cells must pass to each state in DFA individually :

|  |  |  |
| --- | --- | --- |
|  | t |  |
|  | State |
| State | | n+1 |
| n-1 | |
|  |



* 1. State

n

* Shift action & Goto action could be specified according to the edge which has been moved from the current state(n) to the new state.

If the edge was terminal symbol (t) then Cell[n-1,t]= sn

If the edge was nonterminal symbol (N) then Cell[n-1,N]= gn



***With My Best Wishes*** *39* ***Esam***

***Compilers***

***Principle ,Techniques, and Tools***

***Anbar University – Computer College***

 If there are production in current state has the form

**A** **ß.** (the dot in the end of right hand side, **ß** is any

string ),then the action is reduce

Cell[n-1,f]=rk { **f** in Follow(A), **k** is the no. of

production}

 If there are production in current state has the

**A** **ß.**$ {the dot occurred before $ sign, **ß** is any string

the action is accept

Cell[n-1,$]=a

form

},then

* Finally, any empty cell in row n-1 means error action. Repeat the above steps for each states in DFA.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **state** | **x** | **+** | **$** | **E** | **T** |
| **0** | S4 |  |  | g1 | g2 |
| **1** |  |  | Accept |  |  |
| **2** |  | S3 | r2 |  |  |
| **3** | S4 |  |  | g5 | g2 |
| **4** |  | r3 | r3 |  |  |
| **5** |  |  | r1 |  |  |

**5.Implement LR Algorithm :** Suppose input string is **x+x**.After insert input string the **LR-program** is executed, as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| Stack | Input | Action |  |
| 0 | x+x $ | shift |  |
| 0S4 | +x$ | Reduce by **T** | **x** |
| 0T2 | +x$ | shift |  |
| 0T2S3 | x$ | Shift |  |
| 0T2S3S4 | $ | Reduce by **T** | **x** |
| 0T2S3T2 | $ | Reduce by **E** | **T** |
| 0T2S3E5 | $ | Reduce by **E** | **T+E** |
| 0E1 | $ | Accept |  |



***With My Best Wishes*** *40* ***Esam***