

Parser

Phase 2

Team members:

- Ahmed Mohamed Ali Hassan Aboemera (#7)
- Omar Mohamed Aly Abd Elsalam Barakat (#42)
- Omar Youssry Mahmoud Attia (#44)
- Karim Waguih El Azzouni (#49)

Contents:

- 1- Project Description
 - 2- Project Design
 - Parsing
 - **eliminating left recursion and left factoring the grammar**
 - Calculation of FIRST and FOLLOW
 - Building Predictive table
 - Receiving tokens from lexical analyzer
 - Simulating the code
 - 3- About the grammar
 - 4- Program Flow
 - 5- Code Snippets
 - 6- Algorithms and Data structures
-

1- Project Description:

Your task in this phase of the assignment is to design and implement an LL(1) parser generator tool.

The parser generator expects an LL (1) grammar as input. It should compute First and Follow sets and uses them to construct a predictive parsing table for the grammar.

The table is to be used to drive a predictive top-down parser. If the input grammar is not LL (1), an appropriate error message should be produced.

The generated parser is required to produce some representation of the leftmost derivation for a correct input. If an error is encountered, a panic-mode error recovery routine is to be called to print an error message and to resume parsing.

The parser generator is required to be tested using the given context free grammar of a small subset of Java. Of course, you have to modify the grammar to allow predictive parsing.

Combine the lexical analyzer generated in phase 1 and parser such that the lexical analyzer is to be called by the parser to find the next token. Use the simple program given in phase 1 to test the combined lexical analyzer and parser.

Bonus Task

Automatically eliminating grammar left recursion and performing left factoring before generating the parser will be considered a bonus work.

2- Project Design:

- **Parsing:**

At the very beginning , method **modify_grammar_file** is invoked to concatenate the production rules defined at multiple lines. The return of this method is a **vector<string>** each entry in this vector defines a production rule.

Then an instance of **initiator** class is made to initiate the data structures used afterwards in the project.

The data structures are the following:

- **vector<string> terminals** :: contains all terminals of the grammar.
- **vector<string> non_terminals** :: contains the names of all non_terminals in the grammar.
- **vector<vector<vector<string>>> non_terminals_defs** :: each entry in this vector defines the **production rule** of the corresponding non_terminal in non_terminals vector.

The vector<vector<string>> defines the production rule of a non_terminal as a SOP "Sum of products"

- **Eliminating left recursion and left factoring the grammar:**

First Elimination of Left recursion is performed according to the given algorithm in the lecture.

- Arrange non-terminals in some order: $A_1 \dots A_n$
- **for** i **from** 1 **to** n **do** {
 - **for** j **from** 1 **to** $i-1$ **do** {
 - replace each production

$$A_i \rightarrow A_j \gamma$$
 by

$$A_i \rightarrow \alpha_1 \gamma \mid \dots \mid \alpha_k \gamma$$
 where $A_j \rightarrow \alpha_1 \mid \dots \mid \alpha_k$
- eliminate immediate left-recursions among A_i productions

Afterwards Left Elimination is performed according to the rules in the lecture

- For each non-terminal A with two or more alternatives (production rules) with a common non-empty prefix, let say

$$A \rightarrow \alpha\beta_1 \mid \dots \mid \alpha\beta_n \mid \gamma_1 \mid \dots \mid \gamma_m$$

convert it into

$$A \rightarrow \alpha A' \mid \gamma_1 \mid \dots \mid \gamma_m$$

$$A' \rightarrow \beta_1 \mid \dots \mid \beta_n$$

- **Calculation of FIRST and FOLLOW:**

First and follow are calculated by functions provided in class Utilities.

The first is computed by the function “get_first(string non_terminal)” which returns all terminals in first of the passed non_terminal.

The follows are computed by the function “compute_follow()” which returns all terminals in first of the passed non_terminal. This function uses the two functions “compute_follow_rule1()” which applies rule 1 to the computed follows and “compute_follow_rule2()” which applies rule 2 to the computed follows.

The computed firsts and follows are stored in maps to be easily retrieved later.

- **Building Predictive Table:**

Predictive table is built using the FIRST and FOLLOW functions for each non-terminal calculated earlier, by combining algorithm 4.31 in the textbook and the panic mode error recovery method in LL(1) parser.

For the production $A \rightarrow \alpha$ is required to be input at the input symbol a , and a production $A \rightarrow \beta$ already exists, an exception is thrown to indicate that the grammar is not LL(1) and the while parsing operation is aborted.

- **Receiving tokens from the lexical analyzer:**

(Disclaimer: Since our lexical analyzer from phase 1 was buggy, we decided that the input to the parser would be a text file of the tokens that should be received by a correct lexical analyzer)

Input buffer of the parser is filled with tokens from the lexical analyzer by the function "getNextToken()".

- **Simulating the code:**

Code is simulated using algorithm 4.34 in the textbook, which is a nonrecursive predictive parsing technique that uses the predictive table and an explicit stack to parse the input tokens. The output is written into output.txt which is a text file that contains a step-by-step left-most derivation of the source code inputted to the lexical analyzer.

4- About the grammar:

Original Grammar

```
# METHOD_BODY = STATEMENT_LIST
# STATEMENT_LIST = STATEMENT | STATEMENT_LIST STATEMENT
# STATEMENT = DECLARATION
| IF
| WHILE
| ASSIGNMENT
# DECLARATION = PRIMITIVE_TYPE 'id' ';'
# PRIMITIVE_TYPE = 'int' | 'float'
# IF = 'if' '(' EXPRESSION ')' '{' STATEMENT '}' 'else' '{' STATEMENT '}'
# WHILE = 'while' '(' EXPRESSION ')' '{' STATEMENT '}'
# ASSIGNMENT = 'id' 'assign' EXPRESSION ';'
# EXPRESSION = SIMPLE_EXPRESSION
| SIMPLE_EXPRESSION 'relop' SIMPLE_EXPRESSION
# SIMPLE_EXPRESSION = TERM | SIGN TERM | SIMPLE_EXPRESSION 'addop' TERM
# TERM = FACTOR | TERM 'mulop' FACTOR
```

Grammar after eliminating left recursion

```
# METHOD_BODY = STATEMENT_LIST
# STATEMENT_LIST = STATEMENT STATEMENT_LIST`
# STATEMENT_LIST` = STATEMENT STATEMENT_LIST` | \L
# STATEMENT = DECLARATION | IF | WHILE | ASSIGNMENT
# DECLARATION = PRIMITIVE_TYPE 'id' ';'
# PRIMITIVE_TYPE = 'int' | 'float'
# IF = 'if' '(' EXPRESSION ')' '{' STATEMENT '}' 'else' '{' STATEMENT '}'
# WHILE = 'while' '(' EXPRESSION ')' '{' STATEMENT '}'
# ASSIGNMENT = 'id' 'assign' EXPRESSION ';'
# EXPRESSION = SIMPLE_EXPRESSION | SIMPLE_EXPRESSION 'relop' SIMPLE_EXPRESSION
# SIMPLE_EXPRESSION = TERM SIMPLE_EXPRESSION` | SIGN TERM SIMPLE_EXPRESSION`
# SIMPLE_EXPRESSION` = 'addop' TERM SIMPLE_EXPRESSION` | \L
# TERM = FACTOR TERM`
# TERM` = 'mulop' FACTOR TERM` | \L
# FACTOR = 'id' | 'num' | '(' EXPRESSION ')'
# SIGN = '+' | '-'
```

Grammar after left factoring

```
# METHOD_BODY = STATEMENT_LIST
# STATEMENT_LIST = STATEMENT STATEMENT_LIST`
# STATEMENT_LIST` = STATEMENT STATEMENT_LIST` | \L
# STATEMENT = DECLARATION | IF | WHILE | ASSIGNMENT
# DECLARATION = PRIMITIVE_TYPE 'id' ';'
# PRIMITIVE_TYPE = 'int' | 'float'
# IF = 'if' '(' EXPRESSION ')' '{' STATEMENT '}' 'else' '{' STATEMENT '}'
# WHILE = 'while' '(' EXPRESSION ')' '{' STATEMENT '}'
# ASSIGNMENT = 'id' 'assign' EXPRESSION ';'
# EXPRESSION = SIMPLE_EXPRESSION SIMPLE_EXPRESSION*
# SIMPLE_EXPRESSION* = \L | 'relop' SIMPLE_EXPRESSION
# SIMPLE_EXPRESSION = TERM SIMPLE_EXPRESSION` | SIGN TERM SIMPLE_EXPRESSION`
# SIMPLE_EXPRESSION` = 'addop' TERM SIMPLE_EXPRESSION` | \L
# TERM = FACTOR TERM`
# TERM` = 'mulop' FACTOR TERM` | \L
# FACTOR = 'id' | 'num' | '(' EXPRESSION ')'
# SIGN = '+' | '-'
```

FIRST

```
FIRST(METHOD_BODY):          {'int', 'float', 'if', 'while', 'id'}
FIRST(STATEMENT_LIST):         {'int', 'float', 'if', 'while', 'id'}
FIRST(STATEMENT_LIST`):        {'int', 'float', 'if', 'while', 'id', '\L'}
FIRST(STATEMENT):               {'int', 'float', 'if', 'while', 'id'}
FIRST(DECLARATION):            {'int', 'float'}
FIRST(IF):                      {'if'}
FIRST(WHILE):                   {'while'}
FIRST(ASSIGNMENT):              {'id'}
FIRST(PRIMITIVE_TYPE):         {'int', 'float'}
FIRST(EXPRESSION):              {'id', 'num', '(', '+', '-'}
FIRST(SIMPLE_EXPRESSION*):      {'\L', 'relop'}
FIRST(SIMPLE_EXPRESSION):       {'id', 'num', '(', '+', '-'}
FIRST(SIMPLE_EXPRESSION`):     {'addop', '\L'}
FIRST(TERM):                    {'id', 'num', '('}
FIRST(TERM`):                   {'mulop', '\L'}
FIRST(SIGN):                     {'+', '-'}
FIRST(FACTOR):                  {'id', 'num', '('}
```

FOLLOW

```
FOLLOW(METHOD_BODY)          : {'$'}
FOLLOW(STATEMENT_LIST)         : {'$'}
FOLLOW(STATEMENT_LIST`)       : {'$'}
FOLLOW(STATEMENT)               : {'int', 'float', 'if', 'while', 'id', '$', ''}
FOLLOW(DECLARATION)             : {'int', 'float', 'if', 'while', 'id', '$', ''}
FOLLOW(IF)                      : {'int', 'float', 'if', 'while', 'id', '$', ''}
FOLLOW(WHILE)                   : {'int', 'float', 'if', 'while', 'id', '$', ''}
FOLLOW(ASSIGNMENT)              : {'int', 'float', 'if', 'while', 'id', '$', ''}
FOLLOW(PRIMITIVE_TYPE)         : {'id'}
FOLLOW(EXPRESSION)              : {')', ';'}
FOLLOW(SIMPLE_EXPRESSION)       : {')', ';', 'relop'}
FOLLOW(SIMPLE_EXPRESSION*)      : {')', ';'}
FOLLOW(SIMPLE_EXPRESSION`)     : {')', ';', 'relop'}
FOLLOW(TERM)                    : {'addop', ')', ';', 'relop'}
FOLLOW(TERM`)                   : {'addop', ')', ';', 'relop'}
FOLLOW(SIGN)                     : {'id', 'num', '('}
FOLLOW(FACTOR)                  : {'mulop', 'addop', ')', ';', 'relop'}
```

Sample code:

```
int x;
x = 5;
if (x > 2)
{
    x = 0;
}
```

Tokenized code:

```
int
id
;
id
assign
num
;
if
(
id
relop
num
)
{
id
assign
num
;
}
```


Predictive Parsing Table (after enabling panic mode)

ASSIGNMENT:

\$:	synch
float:	synch
id:	'id' 'assign' EXPRESSION ';'
if:	synch
int:	synch
while:	synch
};	synch

DECLARATION:

\$:	synch
float:	PRIMITIVE_TYPE 'id' ';'
id:	synch
if:	synch
int:	PRIMITIVE_TYPE 'id' ';'
while:	synch
};	synch

EXPRESSION:

(:	SIMPLE_EXPRESSION EXPRESSION*
):	synch
+:	SIMPLE_EXPRESSION EXPRESSION*
-:	SIMPLE_EXPRESSION EXPRESSION*
::	synch
id:	SIMPLE_EXPRESSION EXPRESSION*
num:	SIMPLE_EXPRESSION EXPRESSION*

EXPRESSION*:

):	\L
::	\L
relop:	'relop' SIMPLE_EXPRESSION

FACTOR:

(:	'(' EXPRESSION ')'
):	synch
::	synch
addop:	synch
id:	'id'
mulop:	synch
num:	'num'
relop:	synch

IF:

\$:	synch
float:	synch
id:	synch

```

if:      'if' '(' 'EXPRESSION ')' '{' 'STATEMENT '}' 'else' '{' 'STATEMENT '}'
int:     synch
while:   synch
}:       synch
METHOD_BODY:
$:       synch
float:   STATEMENT_LIST
id:      STATEMENT_LIST
if:      STATEMENT_LIST
int:     STATEMENT_LIST
while:   STATEMENT_LIST
PRIMITIVE_TYPE:
float:   'float'
id:      synch
int:     'int'
SIGN:
(:       synch
+:       '+'
-:       '-'
id:      synch
num:     synch
SIMPLE_EXPRESSION:
(:       TERM SIMPLE_EXPRESSION'
):       synch
+:       SIGN TERM SIMPLE_EXPRESSION'
-:       SIGN TERM SIMPLE_EXPRESSION'
;:       synch
id:      TERM SIMPLE_EXPRESSION'
num:     TERM SIMPLE_EXPRESSION'
relop:   synch
SIMPLE_EXPRESSION':
):       \L
;:       \L
addop:   'addop' TERM SIMPLE_EXPRESSION'
relop:   \L
STATEMENT:
$:       synch
float:   DECLARATION
id:      ASSIGNMENT
if:      IF
int:     DECLARATION
while:   WHILE

```

```

    ):
        synch
STATEMENT_LIST:
    $:
        synch
    float:
        STATEMENT STATEMENT_LIST'
    id:
        STATEMENT STATEMENT_LIST'
    if:
        STATEMENT STATEMENT_LIST'
    int:
        STATEMENT STATEMENT_LIST'
    while:
        STATEMENT STATEMENT_LIST'
STATEMENT_LIST':
    $:
        \L
    float:
        STATEMENT STATEMENT_LIST'
    id:
        STATEMENT STATEMENT_LIST'
    if:
        STATEMENT STATEMENT_LIST'
    int:
        STATEMENT STATEMENT_LIST'
    while:
        STATEMENT STATEMENT_LIST'
TERM:
    (:
        FACTOR TERM'
    ):
        synch
    ;;
        synch
    addop:
        synch
    id:
        FACTOR TERM'
    num:
        FACTOR TERM'
    relop:
        synch
TERM':
    ):
        \L
    ;;
        \L
    addop:
        \L
    mulop:
        'mulop' FACTOR TERM'
    relop:
        \L
WHILE:
    $:
        synch
    float:
        synch
    id:
        synch
    if:
        synch
    int:
        synch
    while:
        'while' '(' EXPRESSION ')' '{' STATEMENT '}'
    ):
        synch

```

5- Project flow:

The input to the program is the grammar file, not necessarily left factored or with left recursion eliminated.

The first step is to put the grammar into LL(1) format by eliminating the left recursion and left factoring. Then, the first and follow of each non-terminal is calculated following the given algorithms in the lecture and in the reference.

Using the first and follow, we can now construct the parsing table.

After that, the tokenized java code can be parsed using a stack and the parsing table, producing a left derivation of the given code. Parser uses the Panic Mode recovery technique when errors occur to skip unmatched tokens.

6- Code Snippets:

Eliminator.h

```
1 /*
2  * Eliminator.h
3  *
4  * Created on: 13 Apr 2015
5  * Author: ahmedaboemera
6  */
7
8 #ifndef ELIMINATOR_H_
9 #define ELIMINATOR_H_
10
11 #include "Initiator.h"
12 using namespace std;
13
14 class Initiator;
15
16 class Eliminator{
17 public:
18     Initiator* initiator;
19     Eliminator(Initiator* i);
20     void eliminate_LF();
21     void eliminate_LR();
22     void print_grammar();
23 private:
24     bool find (vector<string>* taken, string s);
25     vector<vector<int>> > get_left_factors(int index);
26     void eliminate_LF(int index);
27     bool probe_for_LF(int index);
28     int find_index(string s);
29     vector<vector<string>> > replace(int i);
30     void eliminate_from_updated(vector<vector<string>> > replaced, int index);
31 };
32
33
34
35 #endif /* ELIMINATOR_H_ */
```

Initiator.h

```
1 /*
2  * Parser.h
3  * Created on: 12 Apr 2015
4  * Author: ahmedaboemera
5  */
6
7 #ifndef INITIATOR_H_
8 #define INITIATOR_H_
9
10 #include <vector>
11 #include <string>
12 #include <iostream>
13 #include <cstdlib>
14 #include <cstring>
15 #include <map>
16 #include "Utilities.h"
17
18 using namespace std;
19
20 class Utilities;
21
22 class Initiator {
23 public:
24     Initiator();
25     string starting;
26     vector<string>* terminals;
27     vector<string>* non_terminals;
28     vector<vector<vector<string>> > >* non_terminal_defs;
29     void get_terminals_and_nonterminals(vector<string>* lines);
30     bool is_terminal(string s);
31     bool is_non_terminal(string s);
32     string get_starting();
33     // map<string, map<string, vector<string>> > > built_predictive_table();
34     vector<vector<string>> > get_def(string s);
35     string trim(string x);
36 private:
37     // Utilities util;
38     void fill_map(vector<string>* lines);
39     void finalize(vector<vector<string>> > non_t_defs);
40     vector<string> split(string s, int start, char regex);
41     int get_eq(string s);
42     vector<string>* get_terminals(string s);
43     string get_nonterminal(string s);
44 };
45 #endif /* INITIATOR_H_ */
```

Parser.h

```
1 /*
2  * Parser.h
3  *
4  * Created on: 17 Apr 2015
5  * Author: ahmedaboemera
6  */
7
8 #ifndef SRC_PARSER_H_
9 #define SRC_PARSER_H_
10
11 #include "Utilities.h"
12 #include "Initiator.h"
13
14 class Parser{
15 public:
16     Parser(Initiator* i, Utilities* u);
17     void built_predictive_table();
18     vector<string> get_entry(string non_terminal, string terminal);
19     void print_predictive_table();
20     string remove_uni_quotes(string s);
21 private:
22     map<string, map<string, vector<string> > > predictive_table;
23     Initiator* i;
24     Utilities* u;
25 };
26
27
28 #endif /* SRC_PARSER_H_ */
```

Utilities.h

```
1
2 #ifndef UTILITIES_H_
3 #define UTILITIES_H_
4
5 #include <algorithm>
6 #include <vector>
7 #include <map>
8 #include <set>
9
10 #include "Initiator.h"
11
12 using namespace std;
13
14 class Initiator;
15
16 class Utilities{
17 public:
18     Utilities(Initiator* i);
19     map<string, vector<string> > compute_first();
20     void compute_follow();
21     vector<set<string> > get_first(string non_terminal);
22     set<string> get_follow(string non_terminal);
23     void print_set(set<string> firsts);
24     void print_first();
25     void print_follow();
26
27 private:
28     Initiator *initiator;
29     const string EPSILON = "\\L";
30     map<string, vector<set<string> > > non_terminals_first;
31     map<string, set<string> > non_terminals_follow;
32     set<string> compute_follow_rule1(string non_terminal);
33     void compute_follow_rule2(string non_terminal);
34     vector<string>::iterator vector_find(vector<string>* vec, string str);
35     set<string> union_sets(vector<set<string> > vec);
36 };
37
38
39 #endif UTILITIES_H_
```

7- Algorithms and data structures:

- C++ STL's structures; mainly:
 - map
 - vector
 - set
 - stack (for parsing)
- C++ STL's algorithms; mainly:
 - sort
 - insert (to union sets)

[illegible]