REPORT

2021-1 Compiler Term-Project

Implement Syntax Analyzer

과목명 | 컴파일러 01분반

담당교수 | 김효수 교수님

학과 | 소프트웨어학부

학년 | 3학년

학번/이름 | 20194538 이나혁

| 20193418 이하윤

제출일 | 2021년 6월 5일

Contents

- I. Find CFG's Ambiguity
- II. SLR PARSING TABLE
- III. Implementation
 - 1. Definition of Rules
 - 2. Definition of SLR Table
 - 3. SLR Stack
 - 4. Shift
 - 5. Reduce & GOTO
 - 6. Accept
 - 7. Others
- IV. Experiment
 - 1. Accept
 - 2. Reject (Error)

I. Find CFG's Ambiguity

1. 주어진 CFG:

```
CFG G:
        01: CODE \rightarrow VDECL CODE | FDECL CODE | CDECL CODE | \epsilon
        02: VDECL → vtype id semi | vtype ASSIGN semi
        03: ASSIGN → id assign RHS
        04: RHS → EXPR | literal | character | boolstr
        05: EXPR → EXPR addsub EXPR | EXPR multdiv EXPR
        06: EXPR → Iparen EXPR rparen | id | num
        07: FDECL → vtype id Iparen ARG rparen Ibrace BLOCK RETURN rbrace
        08: ARG → vtype id MOREARGS | \epsilon
        09: MOREARGS \rightarrow comma vtype id MOREARGS | \epsilon
        10: BLOCK → STMT BLOCK | €
        11: STMT → VDECL | ASSIGN semi
        12: STMT → if Iparen COND rparen Ibrace BLOCK rbrace ELSE
        13: STMT → while Iparen COND rparen Ibrace BLOCK rbrace
        14: COND → COND comp COND | boolstr
        15: ELSE → else Ibrace BLOCK rbrace | \epsilon
        16: RETURN → return RHS semi
        17: CDECL → class id Ibrace ODECL rbrace
        18: ODECL \rightarrow VDECL ODECL | FDECL ODECL | \epsilon
```

2. 모호함을 제거한 CFG:

```
CFG G:

O0: S → CODE

O1: CODE → VDECL CODE | FDECL CODE | CDECL CODE| €

O2: VDECL → vtype id semi | vtype ASSIGN semi

O3: ASSIGN → id assign RHS

O4: RHS → EXPR | literal | character | boolstr

O5: EXPR → TERM addsub EXPR | TERM

O6: TERM → FACTOR multdiv TERM | FACTOR

O7: FACTOR → | paren EXPR rparen | id | num

O8: FDECL → vtype id | paren ARG rparen | brace BLOCK RETURN rbrace

O9: ARG → vtype id MOREARGS | €

10: MOREARGS → comma vtype id MOREARGS | €

11: BLOCK → STMT BLOCK | €
```

12: STMT → VDECL | ASSIGN semi

13: STMT → if Iparen COND rparen Ibrace BLOCK rbrace ELSE

14: STMT → while Iparen COND rparen Ibrace BLOCK rbrace

15: COND → **FACTOR** comp **FACTOR** | boolstr

16: ELSE → else Ibrace BLOCK rbrace | ϵ

17: RETURN → return RHS semi

18: CDECL → class id Ibrace ODECL rbrace

19: ODECL \rightarrow VDECL ODECL | FDECL ODECL | ϵ

- 00: CFG 의 Start 를 S -> CODE 로 수정하였다.
- 05~07: 연산자의 우선순위가 모호해져 우선순위를 가지도록 수정하였다.
- 15: 비교연산자의 우선순위가 모호해져 우선순위를 가지도록 수정하였다.

II. SLR PARSING TABLE

우리는 주어진 웹 사이트를 통해 모호하지 않은 CFG 에 대한 SLR Parsing Table 을 구성했다.

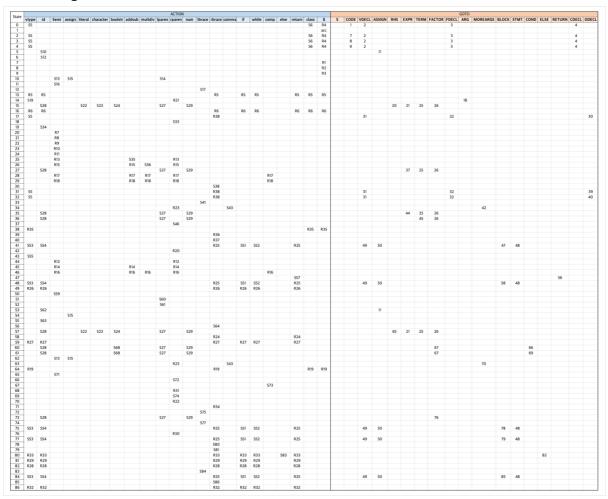
(URL: http://jsmachines.sourceforge.net/machines/slr.html)

- SLR Grammar:

ODECL -> epsilon

S -> CODE CODE -> VDECL CODE CODE -> FDECL CODE CODE -> CDECL CODE CODE -> epsilon VDECL -> vtype id semi VDECL -> vtype ASSIGN semi ASSIGN -> id assign RHS RHS -> EXPR RHS -> literal RHS -> character RHS -> boolstr EXPR -> TERM addsub EXPR EXPR -> TERM TERM -> FACTOR multdiv TERM TERM -> FACTOR FACTOR -> Iparen EXPR rparen FACTOR -> id FACTOR -> num FDECL -> vtype id Iparen ARG rparen Ibrace BLOCK RETURN rbrace ARG -> vtype id MOREARGS ARG -> epsilon MOREARGS -> comma vtype id MOREARGS MOREARGS -> epsilon BLOCK -> STMT BLOCK BLOCK -> epsilon STMT -> VDECL STMT -> ASSIGN semi STMT -> if Iparen COND rparen Ibrace BLOCK rbrace ELSE STMT -> while Iparen COND rparen Ibrace BLOCK rbrace COND -> FACTOR comp FACTOR COND -> boolstr ELSE -> else Ibrace BLOCK rbrace ELSE -> epsilon RETURN -> return RHS semi CDECL -> class id Ibrace ODECL rbrace ODECL -> VDECL ODECL ODECL -> FDECL ODECL

- SLR Parsing Table



III. Implementation

0. Developing Environmemt

Name	OS	Language	IDE
이나혁	Mac OS 11 Big Sur (Intel Processor)	Python3.6	PyCharm
이하윤	Mac OS 11 Big Sur (M1 Silicon)	Python3.8	PyCharm

1. Definition of Rules

```
RULE = [
    ['S', 'CODE'],
    ['CODE', 'FDECL CODE'],
    ['CODE', 'CDECL CODE'],
    ['RHS', 'EXPR'],
    ['RHS', 'literal'],
    ['RHS', 'character'],
    ['RHS', 'boolstr'],
    ['EXPR', 'TERM addsub EXPR'],
    ['EXPR', 'TERM'],
    ['TERM', 'FACTOR'],
    ['FACTOR', 'id'],
    ['FDECL', 'vtype id Iparen ARG rparen Ibrace BLOCK RETURN rbrace'],
    ['MOREARGS', 'comma vtype id MOREARGS'],
    ['BLOCK', 'STMT BLOCK'],
    ['BLOCK', 'epsilon'],
    ['STMT', 'if Iparen COND rparen Ibrace BLOCK rbrace ELSE'],
```

```
['COND', 'FACTOR comp FACTOR'],
['COND', 'boolstr'],
['ELSE', 'else Ibrace BLOCK rbrace'],
["ELSE', 'epsilon"],
['RETURN', 'return RHS semi'],
['CDECL', 'class id Ibrace ODECL rbrace'],
['ODECL', 'VDECL ODECL'],
['ODECL', 'FDECL ODECL'],
['ODECL', 'epsilon']
```

위와 같이 정의한 Rule을 리스트 자료형을 사용해 정의했다.

2. Definition of SLR Parsing Table

```
{'vtype': 'S5', 'class': 'S6', '$': 'R4', 'CODE': 1, 'VDECL': 2, 'FDECL': 3, 'CDECL': 4, }, {\$': 'ACC', },
```

위와 같이 딕셔너리형 리스트를 사용해 SLR Parsing Table 을 정의했다.

3. SLR Stack

```
def parser():
    if len(terminal_list) == 1:
        return True, "

SLR_stack = [0]
    splitter = 0
    error_line = 1
    check = 0
```

```
while True:
    next_input_symbol = terminal_list[splitter]
    check += 1
    current_state = SLR_stack[-1]
```

parser()함수를 정의했으며, SLR Stack, splitter 를 선언했다. 초기 stack 에는 0 이 들어있으며, splitter 는 0 에서 시작한다. check 와 error_line 은 reject 시 에러 메시지 출력에 사용될 변수이다. current_state 는 항상 SLR stack 의 마지막 원소와 같다.

4. Shift

파서는 현재 state 에서 다음 입력 symbol 을 읽어 Shift, Reduce, GOTO, Accept 중에 한 가지 명령을 수행한다.

```
# 판단할 수 없는 symbol 이 있으면 error

if next_input_symbol not in SLR_TABLE[current_state].keys():

error = "[REJECTED] Error in line " + str(error_line) + ", " + error_checker[error_line-1]

print(error)

return False, error
```

SLR Table 에 input 으로 정의되지 않은 symbol 이 있으면 에러이므로, 가장 먼저 SLR table 에 해당 symbol 이 존재하는지 판단한다.

그 후, SLR Table 의 이동 상태가 S로 시작하면 Shift 를 수행하도록 한다.

```
# shift
if SLR_TABLE[current_state][next_input_symbol][0] == 'S':
    splitter += 1
    error_line += 1
    SLR_stack.append(int(SLR_TABLE[current_state][next_input_symbol][1:]))
```

Shift 는 symbol 을 읽고 지나가는 것이므로 splitter 에 1을 더하고, SLR Stack 에 이동 상태의 숫자(State)를 삽입한다.

5. Reduce & GOTO

1. Reduce

SLR Table 의 이동 상태가 R로 시작하면 Reduce 를 수행하도록 한다.

```
# Reduce
elif SLR_TABLE[current_state][next_input_symbol][0] == 'R':
    table_num = SLR_TABLE[current_state][next_input_symbol][1:]

non_termi = RULE[int(table_num)][0] # ex) CODE → epsilon 에서 VDECL 의미
termi = RULE[int(table_num)][1] # ex) CODE → epsilon 에서 epsilon 의미
termi_length = len(str(termi).split("")) # 롤의 어절 개수 ex) VDECL → vtype id semi : 3

for i in range(termi_length):
    if termi != 'epsilon':
        SLR_stack.pop()
        terminal_list.pop(splitter-i-1)

if non_termi not in SLR_TABLE[SLR_stack[-1]].keys():
    error = "[REJECTED] Error in line " + str(error_line) + ", " + error_checker[error_line - 1]
```

```
print(error)
```

Reduce 는 정의한 Rule 에 따라 terminal 혹은 non-terminal 을 non-terminal 로 줄이는 과정이다. 만약 SLR Table 에서 읽은 Reduce 의 number 에 해당하는 Rule 이 epsilon 을 non-terminal 로 바꾸는 것이 아니라면 rule 의 어절의 개수만큼 SLR stack 에서 pop 을 실행한다. (ex. Rule 이 VDECL -> vtype id semi 라면 SLR Stack 에서 pop 을 3 번 실행한다.)

만약 정의한 rule 에서 읽은 non-terminal 이 SLR Table 에 없다면 에러이므로 에러 메시지를 출력하도록 한다.

```
# GOTO
SLR_stack.append(SLR_TABLE[SLR_stack[-1]][non_termi])

if termi != 'epsilon':
    splitter = splitter - termi_length + 1

else:
    splitter += 1

terminal_list.insert(splitter-1, non_termi)
```

Reduce 를 실행하게 되면 non-terminal 을 terminal_list 에 삽입하게 된다. 이 non-terminal 을 SLR Table 에서 읽은 결과를 SLR Stack 에 삽입한다. (GOTO 과정)

Rule 에 따라 terminal_list 의 terminal 과 non-terminal 을 줄이는 개수가 달라지므로, 그에 맞게 spliter 위치를 조절한다.

그 후 Reduce 한 Rule의 non-terminal을 terminal list에 삽입한다.

6. Accept

```
if current_state == 1 and next_input_symbol == '$':
    print("ACCEPTED!!!")
    return True
```

현재 State 가 1이고 next_input_symbol 이 '\$'이면 Accept 되는 경우이므로 command 창에 Accept 메시지를 출력한다.

7. Others

1. lexical analyzer

```
### 결과를 새로운 파일에 출력 및 저장하는 함수

def save_result():
  global token_value
  global token_key

# Output File 에 대한 이름은 '<input_file_name>_output.txt'로 저장됨
  save = open(file_path + '_output.txt', 'w')
  for i, j in zip(token_value, token_key):
    # White Space 를 제외하고 저장
    if j != 'WHITE_SPACE':
```

```
#save.write("<" + j + ", " + i + ">\n")
save.write(j + "\n")
# print("<"+j+",", i+">")
print("Successfully token list saved")
save.close
```

Project 1 에서 설계했던 lexical analyzer 를 syntax analyzer 에 사용하기 위해 lexical analyzer 의 출력 형식을 token 의 분석 결과만 출력하도록 수정하였다.

2. 전체 process

Input file 은 Command 에서 Input File 명을 받아와 읽게 된다.

```
# Command 내 인자 개수 확인

if len(sys.argy)!= 2:
    print("Insufficient arguments")
    sys.exit()

# input File 열기
file_path = sys.argv[1]
f = open(file_path, 'r')

END_MARK = '$'
terminal_list = []

lines = f.readlines()
for line in lines:
    terminal = line.split()[0]
    terminal_list.append(terminal)

terminal_list.append(END_MARK)
error_checker = copy.copy(terminal_list)

parser()
```

파일을 읽은 후, 위에서 정의한 parser()함수를 실행한다. parser() 함수는 input terminal list 를 shift, reduce 작업을 반복해 최종적으로 accept 와 reject 을 판단하므로 True 를 반환하거나 False 를 반환할 때까지 shift 와 reduce 를 반복한다.

IV. Experiment

1. Aceept

- Case 1:

Input.java	Input.java_output.txt
	class
	id
	Ibrace
class Person {	vtype
int a = 1;	id
}	assign
	num
	semi
	rbrace

Result:

```
(base) hayunlee@Hayunui-MacBookAir syntax analyzer % python lexical_analyzer.py input.java
Successfully token list saved
(base) hayunlee@Hayunui-MacBookAir syntax analyzer % python syntax_analyzer.py input.java_output.txt
ACCEPTED!!!
```

- Case 2:

Input2.java	Input2.java_output.txt
	vtype
	id
	assign
	num
float a = 0;	semi
	class
class Person {	id
int a = 1;	Ibrace
}	vtype
	id
	assign
	num
	semi
	rbrace

Result:

```
(base) hayunlee@Hayunui-MacBookAir syntax analyzer % python lexical_analyzer.py input2.java
Successfully token list saved
(base) hayunlee@Hayunui-MacBookAir syntax analyzer % python syntax_analyzer.py input2.java_output.txt
ACCEPTED!!!
```

- Case 3:

Input3.java	Input3.java_output.txt
	vtype
	id
	assign
	num
	semi
	vtype
	id
int a = 3;	assign
int b = 5;	num
int c = a+b;	semi
	vtype
	id
	assign
	id
	addsub
	id
	semi

Result:

(base) hayunlee@Hayunui-MacBookAir syntax analyzer % python lexical_analyzer.py input3.java
Successfully token list saved
(base) hayunlee@Hayunui-MacBookAir syntax analyzer % python syntax_analyzer.py input3.java_output.txt
ACCEPTED!!!

2. Reject

- Case 1:

Input4.java	Input4.java_output.txt
	class
	id
	Ibrace
class Person {	vtype
int a = 1;	id
	assign
	num
	semi

Result:

```
(base) hayunlee@Hayunui-MacBookAir syntax analyzer % python lexical_analyzer.py input4.java
Successfully token list saved
(base) hayunlee@Hayunui-MacBookAir syntax analyzer % python syntax_analyzer.py input4.java_output.txt
[REJECTED] Error in line 9, $
```

- Case 2:

Input5.java	Input5.java_output.txt
	vtype
	id
	assign
	num
float a = 0;	semi
	class
class Person{	id
int = 1;	Ibrace
}	vtype
	assign
	num
	semi
	rbrace

Result:

```
(base) hayunlee@Hayunui-MacBookAir syntax analyzer % python lexical_analyzer.py input5.java
Successfully token list saved
(base) hayunlee@Hayunui-MacBookAir syntax analyzer % python syntax_analyzer.py input5.java_output.txt
[REJECTED] Error in line 10, assign
```

- Case 3:

Input6.java	Input6.java_output.txt
	vtype
	id
	assign
	num
	semi
	vtype
	id
int a = 3;	assign
int b = 5;	num
a+b = int c;	semi
	id
	addsub
	id
	assign
	vtype
	id
	semi

Result:

(base) hayunlee@Hayunui-MacBookAir syntax analyzer % python lexical_analyzer.py input6.java Successfully token list saved (base) hayunlee@Hayunui-MacBookAir syntax analyzer % python syntax_analyzer.py input6.java_output.txt [REJECTED] Error in line 11, id