



중앙대학교
CHUNG-ANG UNIVERSITY

Student Number : 20200123
Student Name : Hwang seo-jin
Team : Team 29 (1 인)

1. My SLR parsing table

[illegible]

2. ALL about how my syntax analyzer works

```
SLR_grammar_dic = {'CODE': ['CODE'], 'CODE': ['FDECL CODE', 'FDECL', 'VDECL'], 'VDECL': ['vtype identifier semi']
SLR_grammar = {1: 'CODE -> VDECL CODE', 2: 'CODE -> FDECL CODE', 3: 'CODE -> FDECL', 4: 'CODE -> VDECL', 5: 'VDECL -> vtype ide

terminals = ['identifier', 'lparen', 'comma', 'add_sub', 'return', 'while', 'rparen', 'rbrace', 'mult_div', 'integer', 'lbrace'
nonterminals = ['CODE', 'FDECL', 'STMT', 'RHS', 'FACTOR', 'VDECL', 'ARG', 'RETURN', 'COND', 'EXPR', 'TERM', 'MOREARGS', 'BLOCK'

start_symbol = "CODE"

parsing_table = {0: {'vtype': 's4', 'identifier': '', 'semi': '', 'lparen': '', 'rparen': '', 'lbrace': '', 'comma': '', 'rbrace':
1: {'vtype': '', 'identifier': '', 'semi': '', 'lparen': '', 'rparen': '', 'lbrace': '', 'comma': '', 'rbrace':
2: {'vtype': 's4', 'identifier': '', 'semi': '', 'lparen': '', 'rparen': '', 'lbrace': '', 'comma': '', 'rbrace':
3: {'vtype': 's4', 'identifier': '', 'semi': '', 'lparen': '', 'rparen': '', 'lbrace': '', 'comma': '', 'rbrace':
4: {'vtype': '', 'identifier': 's7', 'semi': '', 'lparen': '', 'rparen': '', 'lbrace': '', 'comma': '', 'rbrace':
5: {'vtype': '', 'identifier': '', 'semi': '', 'lparen': '', 'rparen': '', 'lbrace': '', 'comma': '', 'rbrace':
6: {'vtype': '', 'identifier': '', 'semi': '', 'lparen': '', 'rparen': '', 'lbrace': '', 'comma': '', 'rbrace':
7: {'vtype': '', 'identifier': '', 'semi': 's8', 'lparen': 's9', 'rparen': '', 'lbrace': '', 'comma': '', 'rbrace':
8: {'vtype': 'r5', 'identifier': 'r5', 'semi': '', 'lparen': '', 'rparen': '', 'lbrace': '', 'comma': '', 'rbrace':
9: {'vtype': 's11', 'identifier': '', 'semi': '', 'lparen': '', 'rparen': 's12', 'lbrace': '', 'comma': '', 'rbrace':
10: {'vtype': 'r6', 'identifier': '', 'semi': '', 'lparen': '', 'rparen': '', 'lbrace': '', 'comma': '', 'rbrace':
11: {'vtype': '', 'identifier': 's13', 'semi': '', 'lparen': '', 'rparen': '', 'lbrace': '', 'comma': '', 'rbrace':
12: {'vtype': '', 'identifier': '', 'semi': '', 'lparen': '', 'rparen': 's14', 'lbrace': '', 'comma': '', 'rbrace':
13: {'vtype': '', 'identifier': '', 'semi': '', 'lparen': '', 'rparen': 's17', 'lbrace': 's16', 'rbrace':
14: {'vtype': 's27', 'identifier': 's23', 'semi': '', 'lparen': '', 'rparen': '', 'lbrace': '', 'comma': '', 'rbrace':
15: {'vtype': 'r7', 'identifier': 's23', 'semi': '', 'lparen': '', 'rparen': '', 'lbrace': '', 'comma': '', 'rbrace':
16: {'vtype': 's28', 'identifier': 's23', 'semi': '', 'lparen': '', 'rparen': '', 'lbrace': '', 'comma': '', 'rbrace':
17: {'vtype': '', 'identifier': 's23', 'semi': '', 'lparen': 's29', 'rparen': 's29', 'lbrace': 's29', 'comma': '', 'rbrace':
18: {'vtype': 'r8', 'identifier': 's23', 'semi': '', 'lparen': 's29', 'rparen': 's29', 'lbrace': 's29', 'comma': '', 'rbrace':
19: {'vtype': 's27', 'identifier': 's23', 'semi': '', 'lparen': 's29', 'rparen': 's29', 'lbrace': 's29', 'comma': '', 'rbrace':
20: {'vtype': 'r13', 'identifier': 's23', 'semi': '', 'lparen': 's29', 'rparen': 's29', 'lbrace': 's29', 'comma': '', 'rbrace':
21: {'vtype': 'r15', 'identifier': 's23', 'semi': '', 'lparen': 's29', 'rparen': 's29', 'lbrace': 's29', 'comma': '', 'rbrace':
22: {'vtype': 'r15', 'identifier': 's23', 'semi': '', 'lparen': 's29', 'rparen': 's29', 'lbrace': 's29', 'comma': '', 'rbrace':
23: {'vtype': 'r15', 'identifier': 's23', 'semi': '', 'lparen': 's29', 'rparen': 's29', 'lbrace': 's29', 'comma': '', 'rbrace':
24: {'vtype': 'r15', 'identifier': 's23', 'semi': '', 'lparen': 's29', 'rparen': 's29', 'lbrace': 's29', 'comma': '', 'rbrace':
25: {'vtype': 'r15', 'identifier': 's23', 'semi': '', 'lparen': 's29', 'rparen': 's29', 'lbrace': 's29', 'comma': '', 'rbrace':
26: {'vtype': 'r15', 'identifier': 's23', 'semi': '', 'lparen': 's29', 'rparen': 's29', 'lbrace': 's29', 'comma': '', 'rbrace':
27: {'vtype': 'r15', 'identifier': 's23', 'semi': '', 'lparen': 's29', 'rparen': 's29', 'lbrace': 's29', 'comma': '', 'rbrace':
```

Defined SLR_grammar_dic and SLR_grammar from CFG, terminals & nonterminals, start_symbol.
Most importantly, implemented parsing_table based on my SLR parsing table at before page of this documentation.

```
# input (using lexical_analyzer output!)
lexical_file = open('lexical_test.out','r')
# input lexemes for pointing out where error has occurred
lexeme_file = open('lexical_lexeme.out', 'r')

# output (syntax analyzer output : whether the grammar is accepted or not)
syntax_file = open('syntax_test.out','w')

# lexemes for pointing out where error has occurred
lexemes = lexeme_file.read()
# file input uses as a string.
lexical_output = lexical_file.read().lower()

token_stream = (lexical_output + " $").split()
lexeme_stream = (lexemes).split()

# for counting steps
step_num = 0
# stack needed for parsing
syntax_stack = ['0']

pointer = 0
token = token_stream[pointer]
```

Open files. Because we are using `lexical_analyzer.py`'s output for input, we open `'lexical_test.out'`, which only includes tokens.

Also for pointing out where the error has occurred, I opened `'lexical_lexeme.out'` which contains only lexemes in the same order with `'lexical_test.out'`

And I **defined and initialized** some values needed later on.

```

# Method to find First set.
def FIRST_SET(nonterminal):
    # if it is terminal, just return!
    if nonterminal in terminals:
        return {nonterminal}

    first = set({})

    first_list = [nonterminal]

    for next_symbol in SLR_grammar_dic[nonterminal]:
        for (i, next) in enumerate(next_symbol.split()):
            length = len(next_symbol)

            if next not in first_list:
                # check next's FIRST SET
                symbol_first = FIRST_SET(next)

                for temp in symbol_first:
                    if temp in terminals:
                        first.add(temp)

                if '^' not in symbol_first:
                    break
            else:

```

```

# Method to find FOLLOW set.
def FOLLOW_SET(nonterminal):
    follow = set({})
    follow_list = [nonterminal]

    # starting symbol add $
    if nonterminal == start_symbol:
        follow.add('$')

    for (before, next_symbols) in SLR_grammar_dic.items():
        for next in next_symbols:
            next = next.split()

            if nonterminal in next[:-1]:
                index_search = next.index(nonterminal) + 1

                first = FIRST_SET(next[index_search])
                follow = follow | (first - set('^'))

            if '^' in first:
                if before not in follow_list:
                    follow = follow | FOLLOW_SET(before)

    elif nonterminal in next[-1]:
        if before not in follow_list:

```

Defined **FIRST_SET** function, the method to find the **FIRST SET** of the symbol and **FOLLOW_SET** function, the method to find **FOLLOW SET**.

```

# SLR Parser to check whether the Grammar is accepted or not.
while 1:
    try:
        step_num += 1
        state = int(syntax_stack[-1])

        # If there is no Symbols in parsing table, print ERROR & break
        if token not in parsing_table[state].keys():
            error_message = "\nERROR occurred at : " + str(step_num) + "\n"
            print("\nREJECTED! THERE IS NO MATCHING SYMBOL NAMED" + token + "\n")
            print(error_message)
            break

        # If all grammar accepted, write and print accept message & break
        elif parsing_table[state][token] == "acc":
            accept_message = "\nACCEPTED! GRAMMATICALLY CORRECT\n"
            syntax_file.write(accept_message)
            print(accept_message)
            break

        # GOTO
        elif parsing_table[state][token][0] == "r":
            next_state = int(parsing_table[state][token][1:])
            grammar = SLR_grammar[next_state].split()

            if grammar[-1] != '^':
                after_index = grammar.index('->') + 1
                boundary_len = 2 * len(grammar[after_index:])

            syntax_stack = syntax_stack[:-boundary_len]

```

THE MAIN KEY PART! Implemented SLR Parser part to check whether the Grammar is accepted or not!

- If there is no corresponding symbol in parsing table, we print ERROR that there is no matched symbol.
- If all token's grammar are accepted, we write and print the accept message.

```

seojin@hwangseojin-ui-MacBookPro lexical_analyzer % python3 syntax_analyzer.py
ACCEPTED! GRAMMATICALLY CORRECT

```

- Apart from this, handled GOTO, and ACTION
- If IndexError occurs, it means grammatical error! (no corresponding symbol in particular state in parsing table) So we report ERROR and explain why and where (At ?th lexeme, lexeme name) the Error occurred.

```

REJECTED!
GRAMMATICAL ERROR found at : 12th lexeme(token) '('

```

3. Test input files and outputs which I used

- 1) First

○ Input : Test.c

```
int operationfunc(int a, int b)
{
    d = 0;

    while (b<=1){
        a = 1;
    }

    return 2;
}
```

○ Output :

ACCEPTED! GRAMMATICALLY CORRECT

- 2) Second

○ Input : Test2.c

```
int operationfunc1(int a, int b)
{
    print("Hello World");

    char c = (char)a - b;
    int d = -2134;
    d = 0;
}
```

○ Output : rejected because in CFG that we are using, there is no grammar that defines '(' comes after the identifier !

REJECTED!
GRAMMATICAL ERROR found at : 12th lexeme(token) '('

- 3) Third

○ Input : Test3.c

```
int operationfunc2(int a, int b)
{
    char c = (char)a - b;
    int d = -2134;
    d = 0;

    while (b<=1){
        a = 1;
    }

    return 2;
}
```

○ Output : rejected because in CFG that we are using, there is no grammar that defines 'vtype identifier assign' !

REJECTED!
GRAMMATICAL ERROR found at : 13th lexeme(token) '='

Just In case, I attach the whole capture of the
SLR Parser site(where I made parsing table)
next page!


```

SLR grammar ( ' ' is  $\epsilon$ ):
(0) CODE  $\rightarrow$  VDECL
(1) CODE  $\rightarrow$  VDECL
(2) VDECL  $\rightarrow$  vtype identifier
(3) semi
(4) VDECL  $\rightarrow$  vtype identifier
(5) lparen ARG
(6) ARG  $\rightarrow$  vtype identifier
(7) MOREARGS
(8) ARG  $\rightarrow$  rparen lbrace BLOCK
(9) MOREARGS  $\rightarrow$  comma vtype
(10) identifier MOREARGS
(11) MOREARGS  $\rightarrow$  rparen lbrace
    BLOCK
(12) BLOCK  $\rightarrow$  STMT BLOCK
(13) BLOCK  $\rightarrow$  RETURN rbrace
(14) BLOCK  $\rightarrow$  rbrace
(15) BLOCK  $\rightarrow$  rbrace else lbrace

```

[illegible][illegible]

kens):
 number of steps:

Trace			Tree
k	Input	Action	
	id + id + id §		