CAUSWE Compiler, 2021

# Compiler Project #1 Documentation



Class#01 Team#59 안재형(20171248)

Class#02 Team#56 구건모(20181856)

## **Definition of Tokens**

```
Signed integer
  (SIGNED INTEGER) - for non-empty sequence of digit, starting from a non-zero digit
                                                                or minus sign symbol. And single zero digit e.g 0, 11, -123
Single character
  (CHARACTER) - for a single digit. English letter, block or any symbol
                                                    starting from and terminating with 'e.g 'o', 'a'
 Boolean String
  <BOOLSTR> - for true and false
 Literal String
<LITERAL STRING> - for any combination of digits, letters. blank.
                                                       starting from and terminating with " e.g "Hello"
 An idetifier of variables and functions
  (ID) - for a non-empty sequence of English letters. digits, and underscore
                         starting from an English letter or a underscore e.g. i , i , _ Sunc_,
 Keywords for special statements
   (KEYWORD) - for if else while class return keword
 Arithmetic operators
   <ARITHMETICOP> - for `+'`-', *' /' operator
Assignment operator
   (OP) - for assign operator =
Comparison operator
   (COMPARISON) - for compare operator '>'>=', '< ', '<=', '==', '!='
 A terminating symbol of statements,
 A pair of symbols for defining area/scope of variables and functions,
 for indicating a function/statement, for using an array,
for seperating input arguments in functions.
  <!-- <!-- Compare the control of the control o
```

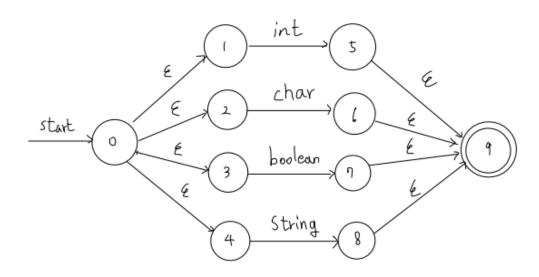
# **Regular Expression**

```
Regular expression
 (VARTYPE) = int | char | boolean | String
 DIGIT = 0/1/2/ ... 1819 Non-zero = 1/2/3/ ... 19
 LETTER = albi...lylzlAIBI...lYlZ
 OTHER: any other symbol ex)'$', ' ( ) W'
 (SIGNEDINTEGER) = (C-18) Non-zero (Non-zero 10)*)10
<(CHARACTER) = `DIGITILETTERIOTHER'</pre>
 (BOOLSTR) = true | false
 <LITERAL STRING> = "(DIGITILETTER)""
 <ID>= (LETTER | -) (LETTER | DIGIT | -)*
 <KEYWORD> = if | else | while | class | return
 <ARITHMETICOP> = +1-1*1/
\langle OP \rangle = =
(< (> |<) (= | E)) | ((! |=) = )
```

# **NFAs**

## <VARTYPE>

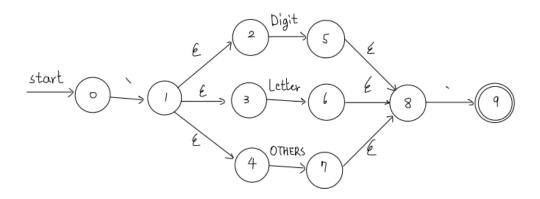
(VARTYPE) = int | char| boolean | String



#### • <SIGNED INTEGER>

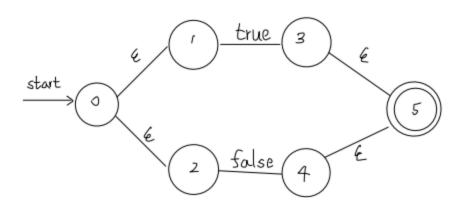
Non-zero =  $||2|3|4|\cdots |8|9$  $\langle SIGNED INTEGER \rangle = ((-|E) Non-zero (Non-zero |0)*)|0$  E Glart Gl

#### <CHARACTER>

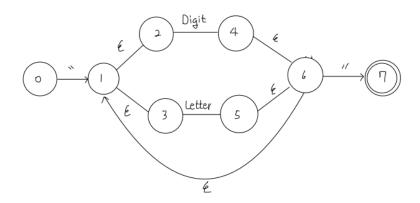


### <BOOLSTR>

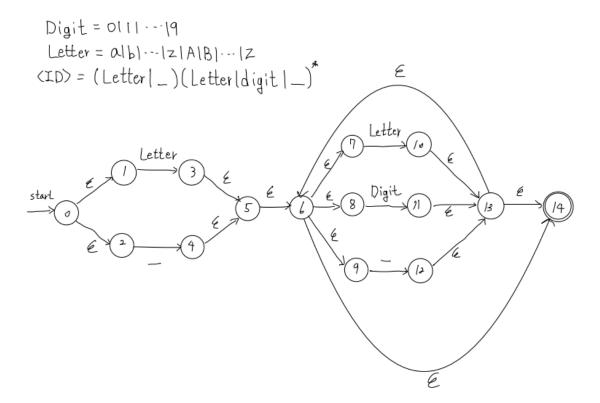
(BOOLSTR) = true | false



## • <LITERAL STRING>



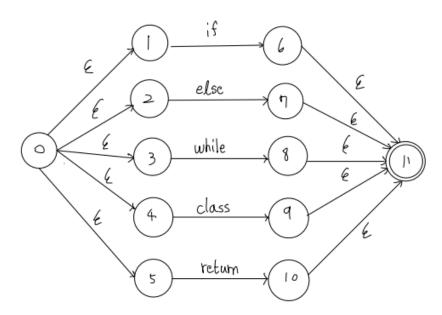
## • <ID>



## <KEYWORD>

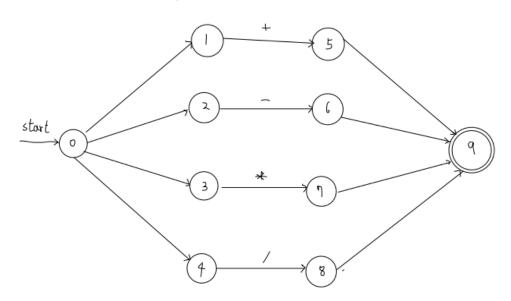
<KEYWORD> NFA

(KEYWORD) = if lelse | while | class | return



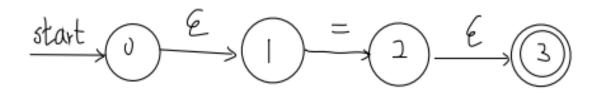
## <ARITHMETIC OP>

<ARITHMETICOP> = + | - | \* | /



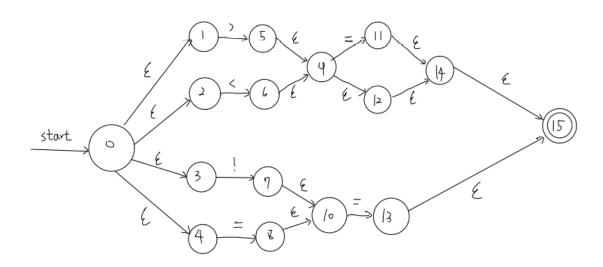
# • <OP>

$$\langle OP \rangle = =$$



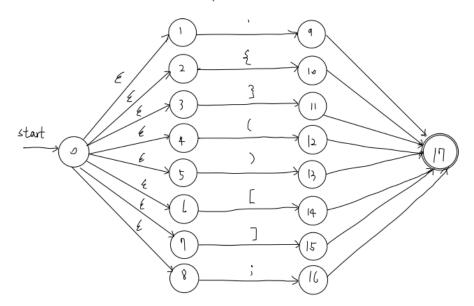
## <COMPARISON>

 $\langle COMPARISON \rangle = ((>|<)(=|E))|((! |=)=)$ 



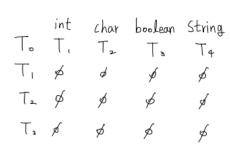
# <SYMBOL>

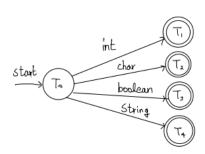
 $\langle SYMBOL \rangle = , | \{ | \} | ( | ) | [ | ] | ;$ 



## **DFAs and Transition table**

#### <VARTYPE>





#### <SIGNED INTEGER>

$$T_{0} = \ell - closure(0) = \{0.1, 2, 3, 5, 6\}$$

$$T_{1} = \ell - closure(\delta(T_{0}, -)) = \ell - closure(4) = \{4, 6\}$$

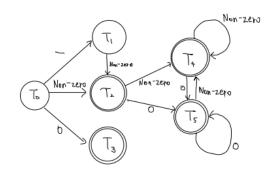
$$T_{2} = \ell - closure(\delta(T_{0}, Non-zero)) = \ell - closure(1) = \{1, 8, 9, 10, 13, 15\}$$

$$T_{3} = \ell - closure(\delta(T_{0}, 0)) = \ell - closure(14) = \{14, 15\}$$

$$\ell - closure(\delta(T_{1}, Non-zero)) = T_{2}$$

$$T_{4} = \ell - closure(\delta(T_{2}, Non-zero)) = \ell - closure(11) = \{8, 9, 10, 11, 13, 15\}$$

$$T_{4} = \ell - closure(\delta(T_{2}, Non-zero)) = \ell - closure(12) = \{8, 9, 10, 12, 13, 15\}$$



#### • <CHARACTER>

$$T_{0} = \text{\&-closure}(0) = \text{\&0}$$

$$T_{1} = \text{\&-closure}(8(T_{0}, ')) = \text{\&-closure}(1) = \text{\&1}, 2, 3, 4$$

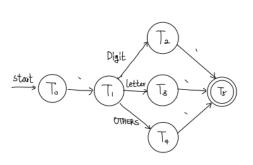
$$T_{2} = \text{\&-closure}(8(T_{1}, \text{Digit})) = \text{\&-closure}(5) = \text{\&5}, 8$$

$$T_{3} = \text{\&-closure}(8(T_{1}, \text{Letter})) = \text{\&-closure}(6) = \text{\&6}, 8$$

$$T_{4} = \text{\&-closure}(8(T_{1}, \text{OTHERS})) = \text{\&-closure}(9) = \text{\&1}, 8$$

$$T_{5} = \text{\&-closure}(8(T_{2}, ')) = \text{\&-closure}(8) = \text{\&8}, 9$$

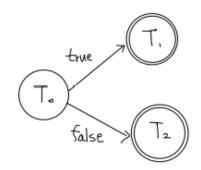
	\	Digit	Letter	OTHERS
T.	T,	ø	ø	$\phi$
$T_{i}$	ø	72	$T_3$	T4
T.	Ts	Ø	⋫	ø
$T_3$	$T_{5}$	Á	ø	ø
$T_{\bullet}$	12	ø	ø	ø
$T_{\mathfrak{s}}$	Þ	$\phi$	ø	ø



#### <BOOLSTR>

$$T_0 = \mathcal{E} - closure(0) = \{0, 1, 2\}$$
  
 $T_1 = \mathcal{E} - closure(\mathcal{B}(T_0, true)) = \mathcal{E} - closure(3) = \{3, 5\}$   
 $T_2 = \mathcal{E} - closure(\mathcal{B}(T_0, talse)) = \mathcal{E} - closure(4) = \{4, 5\}$ 

true false
To To To To
To 
$$abla$$
To  $abla$ 



#### <LITERAL STRING>

$$T_{0} = \mathcal{E} - closure(0) = \mathcal{E}0$$

$$T_{1} = \mathcal{E} - closure(\mathcal{E}(T_{0}, ")) = \mathcal{E} - closure(1) = \mathcal{E}1, 2, 3\mathcal{F}$$

$$T_{2} = \mathcal{E} - closure(\mathcal{E}(T_{1}, Digit)) = \mathcal{E} - closure(4) = \mathcal{E}1, 2, 3, 4, 6\mathcal{F}$$

$$T_{3} = \mathcal{E} - closure(\mathcal{E}(T_{1}, Letter)) = \mathcal{E} - closure(5) = \mathcal{E}1, 2, 3, 5, 6\mathcal{F}$$

$$\mathcal{E} - closure(\mathcal{E}(T_{2}, Digit)) = T_{2} \quad \mathcal{E} - closure(\mathcal{E}(T_{2}, Letter)) = T_{3}$$

$$\mathcal{E} - closure(\mathcal{E}(T_{2}, Digit)) = T_{1} \quad \mathcal{E} - closure(\mathcal{E}(T_{3}, Letter)) = T_{3}$$

$$\mathcal{E} - closure(\mathcal{E}(T_{2}, ")) = \mathcal{E} - closure(\eta) = \mathcal{E}1\mathcal{F}$$

$$\mathcal{E} - closure(\mathcal{E}(T_{2}, ")) = \mathcal{E} - closure(\eta) = \mathcal{E}1\mathcal{F}$$

$$\mathcal{E} - closure(\mathcal{E}(T_{2}, ")) = \mathcal{E} - closure(\eta) = \mathcal{E}1\mathcal{F}$$

$$\mathcal{E} - closure(\mathcal{E}(T_{2}, ")) = \mathcal{E} - closure(\eta) = \mathcal{E}1\mathcal{F}$$

$$\mathcal{E} - closure(\mathcal{E}(T_{2}, ")) = \mathcal{E} - closure(\eta) = \mathcal{E}1\mathcal{F}$$

$$\mathcal{E} - closure(\mathcal{E}(T_{2}, ")) = \mathcal{E} - closure(\eta) = \mathcal{E}1\mathcal{F}$$

$$\mathcal{E} - closure(\mathcal{E}(T_{2}, ")) = \mathcal{E} - closure(\eta) = \mathcal{E}1\mathcal{F}$$

$$\mathcal{E} - closure(\mathcal{E}(T_{2}, ")) = \mathcal{E} - closure(\eta) = \mathcal{E}1\mathcal{F}$$

$$\mathcal{E} - closure(\mathcal{E}(T_{2}, ")) = \mathcal{E} - closure(\mathcal{E}(T_{3}, Letter)) = \mathcal{E} - closure(\mathcal{E}(T_{3},$$

#### <ID>

(ID) NFA to DFA

To = 
$$E - closure(0) = 20.1.2$$
}

Ti =  $E - closure(86To, letter)) = E - closure(3) = 23.5.6.7.8, 9.14$ }

Ti =  $E - closure(86To, -1) = E - closure(4) = 24.5, 6.7.8, 9.14$ }

Ti =  $E - closure(66To, -1) = E - closure(10) = 26.7.8, 9.10.13.14$ }

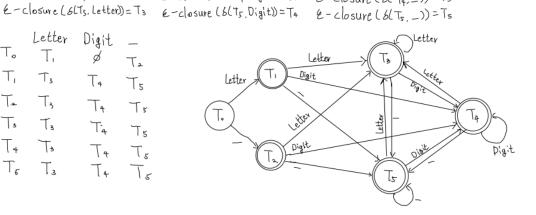
Ti =  $E - closure(6(To, letter)) = E - closure(10) = 26.7.8, 9.10.13.14$ }

Ti =  $E - closure(6(To, letter)) = E - closure(11) = 26.7.8, 9.11.13.14$ }

Ti =  $E - closure(6(To, letter)) = E - closure(12) = 26.7.8, 9.12.13.14$ }

Eight =  $E - closure(6(To, letter)) = E - closure(6(Eo, logit)) = E - closure(6(Eo, letter)) = E -$ 

Letter Digit T. Т. Τ,  $T_3$ T5 Ta Tz Ts T3 T4

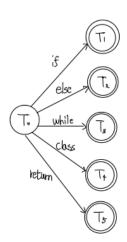


&-closure (6(74,-1)=Ts

#### <KEYWORD>

$$T_0 = \mathcal{E} - closure(0) = 20.1.2.3, 4.5$$
 $T_1 = \mathcal{E} - closure(6(T_0, 1f)) = \mathcal{E} - closure(6) = 26.11$ 
 $T_2 = \mathcal{E} - closure(6(T_0, else)) = \mathcal{E} - closure(9) = 29.11$ 
 $T_3 = \mathcal{E} - closure(8(T_0, while)) = \mathcal{E} - closure(8) = 28.11$ 
 $T_4 = \mathcal{E} - closure(6(T_0, class)) = \mathcal{E} - closure(9) = 29.11$ 
 $T_5 = \mathcal{E} - closure(8(T_0, return)) = \mathcal{E} - closure(10) = 20.11$ 

if else while class return 
$$T_s$$
  $T_s$   $T_s$ 



#### <ARITHMETIC OP>

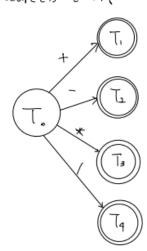
$$T_0 = \mathcal{E} - closure(0) = \{0, 1, 2, 3, 4\}$$

$$T_1 = \mathcal{E} - closure(\delta(T_0, +)) = \mathcal{E} - closure(1) = \{5, 9\}$$

$$T_2 = \mathcal{E} - closure(\delta(T_0, -)) = \mathcal{E} - closure(2) = \{6, 9\}$$

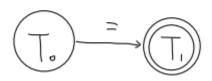
$$T_3 = \mathcal{E} - closure(\delta(T_0, *)) = \mathcal{E} - closure(3) = \{9, 9\}$$

$$T_4 = \mathcal{E} - closure(\delta(T_0, 1)) = \mathcal{E} - closure(4) = \{8, 9\}$$



## • <OP>

$$T_0 = \mathcal{E} - closure(0) = \{0,1\}$$
  
 $T_1 = \mathcal{E} - closure(8(T_0, =)) = \mathcal{E} - closure(2) = \{2,3\}$ 

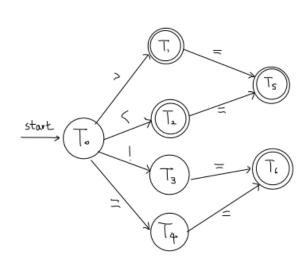


#### <COMPARISON>

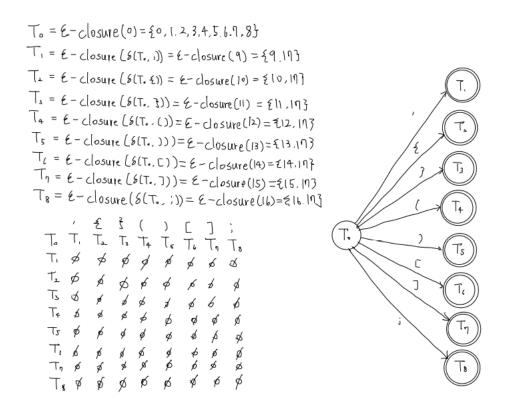
(COMPARISON) NFA to DFA

$$T_3 = \varepsilon - closure(S(T_0,!)) = \varepsilon - closure(\eta) = \xi \eta, 107$$

$$T_s = \mathcal{E} - closure(\delta(T_1, =)) = \mathcal{E} - closure(11) = \mathcal{E}11.14.153$$



#### <SYMBOL>



# **How This Lexical Analyzer Work?**

## Requirements

- Python >= 3.8.2
- Ubuntu or MacOS

#### **Architecture**

There's a DFA\_Graph class. This class contains a token's DFA Informations and current\_state. If it gets a symbol as an input, current\_state changes and returns its value. Of course, If it got a wrong symbol, current\_state changes to None and returns it until it resets.

There are syntaces as DFA\_Graph class in this code.

For the main function,

- 1. it reads a file by a single symbol.
- 2. Then pass it into every token DFA\_Graph's input.
- 3. Repeat it until every DFA\_Graph returns None.
- 4. Print token's information that returned a number until last time. If there were several tokens, choose the first one.
- 5. From the last symbol, do 2 and repeat until a file ends.

And If parsing ends up with an error, it reports where the error occurred. (Error Message is not written at output file.)

## **Running Table**

This table shows what happens for each symbol input.

		i	n	t		t	е	S	t	
DFA_ Graph	INT .current_state	1	1	2(f)	None →None	None →None	None	None	None	None →None
	ID .current_state	2(f)	2(f)	2(f)	None →None	None →1(f)	1(f)	1(f)	1(f)	None →None
	WS .current_state	None	None	None	None →1(f)	None →None	None	None	None	None →1(f)
		Every states becomes None† print INT(priority)								
						↑ print WS				
										↑ print ID

# Sample Result

```
input (sample2.java)

int func(int a) { return 0; }

INT int
ID func
LPAREN (
INT int
ID a
RPAREN )
LBRACE {
RETURN return
SIGNED INTEGER 0
TERMINATE;
RBRACE }
```

```
input (sample5.java)
                                    output (sample5.java_output.txt)
                                      IF if
 if(true) {
                                      LPAREN (
     "Hello World"
                                      BOOLEAN STRING true
                                      RPAREN )
 }else{
                                      LBRACE {
                                      STRING "Hello World"
     "It's Java Time!"
                                      RBRACE }
                                      ELSE else
                                      LBRACE {
                                      STRING "It's Java Time!"
                                      RBRACE }
```

```
input (sample_final.java)

output (sample_final.java_output.txt)

INT int CHAR char BOOLEAN boolean STRING String SIGNED INTEGER 0 SIGNED INTEGER 0 SIGNED INTEGER 2345 SINGLE CHARACTER 'a' SINGLE CHARACTER '1'
```

```
"Hello Java World"
                                                   SINGLE CHARACTER
                                                   SINGLE CHARACTER
                                                                          '&'
                                                   TERMINATE
                                                   BOOLEAN STRING true
a+b
                                                   BOOLEAN STRING false
                                                   STRING "Hello Java World"
b-c
                                                   CLASS class
1-5
                                                   ID
                                                          Foo
                                                   ID
                                                   OP_ARITHMATIC +
                                                   ID
                                                          b
int foo = -3;
                                                   ID
                                                          b
int bar 123 = 4;
                                                   OP_ARITHMATIC -
                                                   ID
                                                         С
if(true) {
                                                   SIGNED INTEGER 1
                                                   OP ARITHMATIC
                                                   SIGNED INTEGER 5
                                                   SIGNED INTEGER -17
                                                   INT
                                                          int
                                                   ID
                                                          foo
                                                   OP ASSIGNMENT =
                                                   SIGNED INTEGER -3
                                                   TERMINATE
int func(int a) {return a}
                                                   INT
                                                          int
                                                   ID
                                                           bar 123
                                                   OP ASSIGNMENT =
                                                   SIGNED INTEGER 4
                                                   TERMINATE
                                                   IF
                                                          if
                                                   LPAREN (
                                                   BOOLEAN STRING true
                                                   RPAREN )
                                                   LBRACE {
3 > 5
                                                   STRING "Hello World"
3 >= 5
                                                   TERMINATE
                                                   RBRACE }
                                                   ELSE
                                                          else
                                                  LBRACE {
STRING "It's Java Time!"
                                                   TERMINATE
                                                   RBRACE }
3 == 5
                                                   INT
                                                          int
                                                   ID
                                                          func
                                                   LPAREN (
                                                   INT
b = 3 / 5
                                                          int
                                                   ID
                                                   RPAREN )
                                                  LBRACE {
while (i < 5) {
                                                  RETURN return
                                                   ID
                                                          а
                                                   RBRACE }
                                                   INT
                                                          int
                                                   LBRACKET
                                                   RBRACKET
                                                   ID
                                                   OP ASSIGNMENT =
                                                   LBRACE {
                                                   SIGNED INTEGER 1
                                                   SEPARATE
                                                   SIGNED INTEGER 2
                                                   SEPARATE
                                                   SIGNED INTEGER 3
                                                   RBRACE }
                                                   TERMINATE
                                                   CLASS class
```

```
compiler
LBRACE {
STRING String
ID
       Teacher
TERMINATE
CHAR
       char
ID
       Grade
TERMINATE
RBRACE }
SIGNED INTEGER 3
OP COMPARISON >
SIGNED INTEGER 5
SIGNED INTEGER 3
OP COMPARISON >=
SIGNED INTEGER 5
SIGNED INTEGER 3
OP COMPARISON <
SIGNED INTEGER 5
SIGNED INTEGER 3
OP_COMPARISON <=
SIGNED INTEGER 5
SIGNED INTEGER 3
OP_COMPARISON !=
SIGNED INTEGER 5
SIGNED INTEGER 3
OP COMPARISON ==
SIGNED INTEGER 5
INT
       int
ID
       b
OP ASSIGNMENT =
SIGNED INTEGER 3
OP_ARITHMATIC *
SIGNED INTEGER 5
       b
OP ASSIGNMENT =
SIGNED INTEGER 3
OP_ARITHMATIC /
SIGNED INTEGER 5
INT
       int
ID
OP ASSIGNMENT =
SIGNED INTEGER 4
TERMINATE
WHILE while
LPAREN (
ID
OP COMPARISON <
SIGNED INTEGER 5
RPAREN )
LBRACE {
OP_ASSIGNMENT =
SIGNED INTEGER 10
TERMINATE
RBRACE }
STRING "ddddsdfasdfasdf"
       ghfhhdh
Lexical Analyzer Error: Cannot Parse at line 39
```