Sean Cox
Cory Caprioli
Caitlin San Pedro

CS323 Documentation

About 2 pages

1. Problem Statement

To write a lexical analyzer that reads an input file and tokenize the contents, printing out the token and lexeme in an output file using finite state machines

2. How to use your program

(how i did it on tuffix)

1. download zip folder on to desktop
2. unzip folder
3. open folder in editor
4. open terminal and run commands
5. “cd Desktop”
6. “cd CPSC323-project1-main/”
7. “clang++ -std=c++11 main.cpp lexer.cpp -o main”
8. execute with ./main
9. enter the input file and the output will be printed in output.txt
10. add your own input files

3. Design of your program

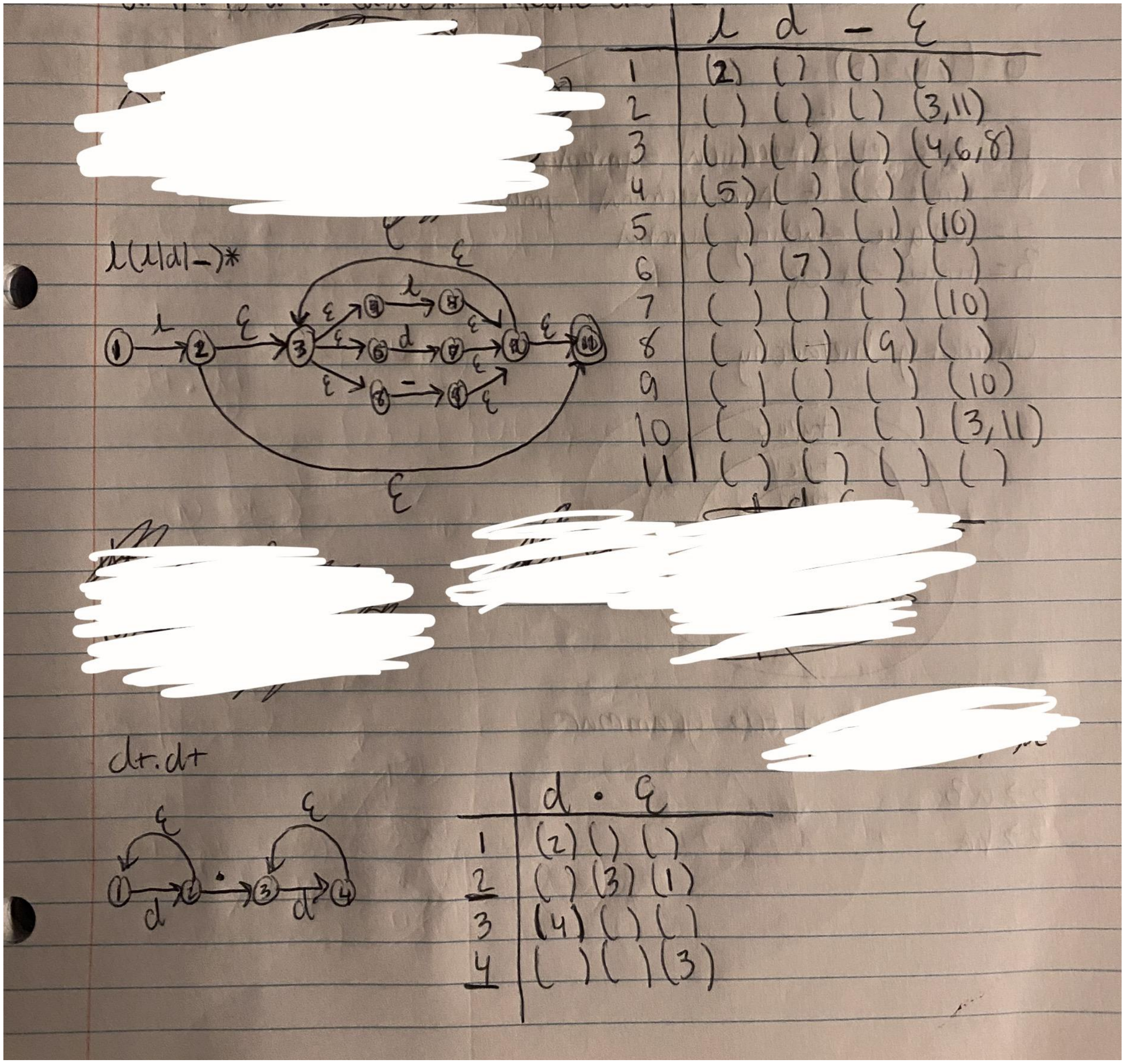
Data Structures: vector, 2d array

REs:
id = l(l | d | _)*
real = d+.d+
int = d+

2 DFSMs:

one for integers and real numbers
one for identifiers

NFSM (Thompson Construction Method): next page



[redacted]

l(l|d|_)*

1
l
2
ε
3
ε
4
l
5
ε
ε
6
d
7
ε
ε
8
_
9
ε
10
ε
11
ε
ε

	l	d	_	ε
1	(2)	( )	( )	( )
2	( )	( )	( )	(3,11)
3	( )	( )	( )	(4,6,8)
4	(5)	( )	( )	( )
5	( )	( )	( )	(10)
6	( )	(7)	( )	( )
7	( )	( )	( )	(10)
8	( )	( )	(9)	( )
9	( )	( )	( )	(10)
10	( )	( )	( )	(3,11)
11	( )	( )	( )	( )

[redacted]

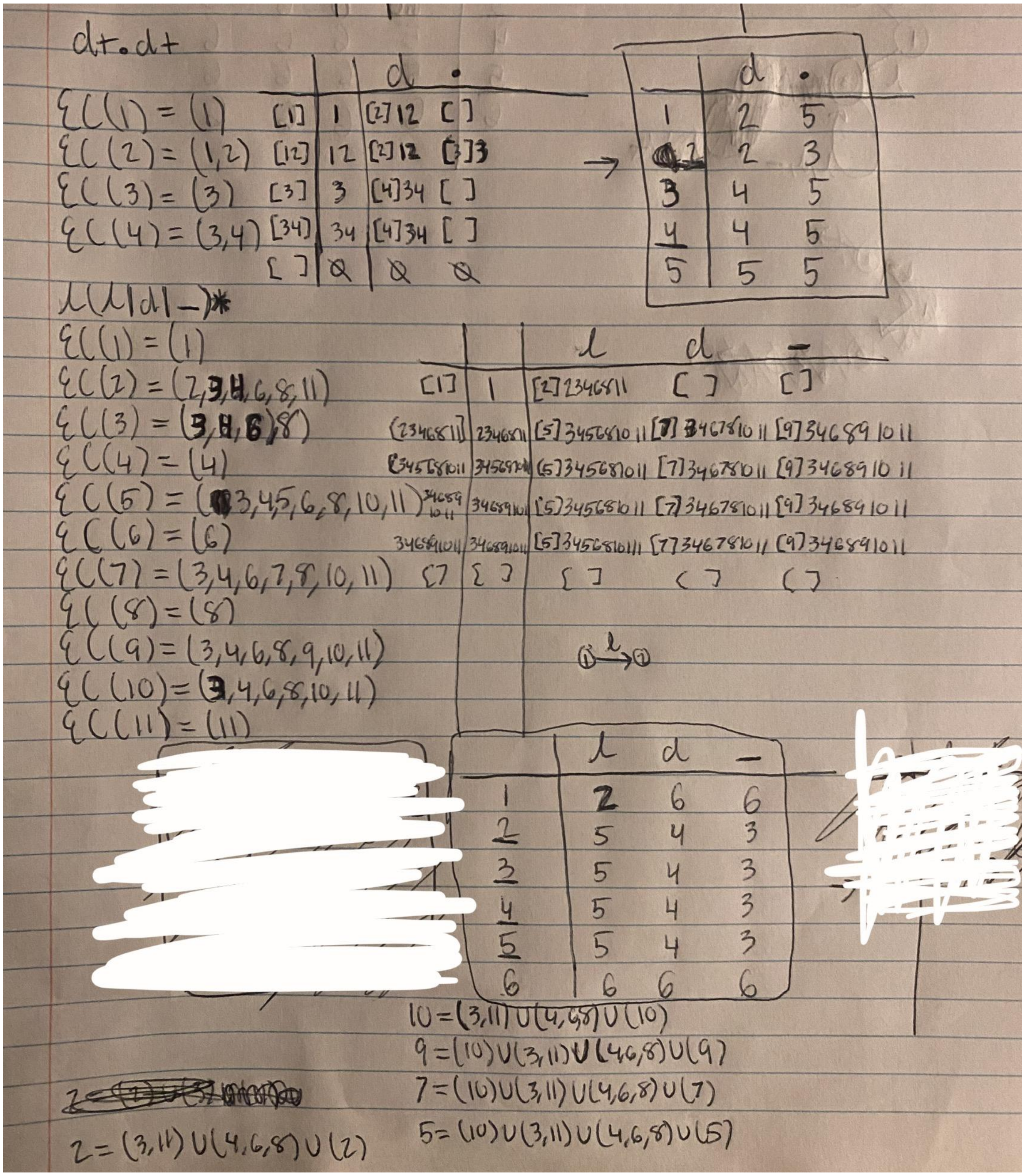
[redacted]

[redacted]

d+.d+

ε
ε
1
d
2
.
3
d
4

	d	.	ε
1	(2)	( )	( )
2	( )	(3)	(1)
3	(4)	( )	( )
4	( )	( )	(3)



dt∘dt

		d	•
ε(L(1) = (1) [1]	1	[2] 12	[ ]
ε(L(2) = (1,2) [12]	12	[2] 12	[3] 3
ε(L(3) = (3) [3]	3	[4] 34	[ ]
ε(L(4) = (3,4) [34]	34	[4] 34	[ ]
[ ]	∅	∅	∅

→

	d	•
1	2	5
2	2	3
3	4	5
4	4	5
5	5	5

l(l|d|_)*

ε(L(1) = (1)

ε(L(2) = (2,3,4,6,8,11)

ε(L(3) = (3,4,6,8)

ε(L(4) = (4)

ε(L(5) = (3,4,5,6,8,10,11)

ε(L(6) = (6)

ε(L(7) = (3,4,6,7,8,10,11)

ε(L(8) = (8)

ε(L(9) = (3,4,6,8,9,10,11)

ε(L(10) = (3,4,6,8,10,11)

ε(L(11) = (11)

		l	d	_
[1]	1	[2] 2346811	[ ]	[ ]
[2346811]	2346811	[5] 345681011	[7] 346781011	[9] 346891011
[345681011	345681011	[5] 345681011	[7] 346781011	[9] 346891011
346891011	346891011	[5] 345681011	[7] 346781011	[9] 346891011
346891011	346891011	[5] 3456810111	[7] 346781011	[9] 346891011
[7	[ ]	[ ]	( ]	( ]

$1 \xrightarrow{l} 2$

[redacted]

	l	d	_
1	2	6	6
2	5	4	3
3	5	4	3
4	5	4	3
5	5	4	3
6	6	6	6

[redacted]

10 = (3,11) ∪ (4,6,8) ∪ (10)

9 = (10) ∪ (3,11) ∪ (4,6,8) ∪ (9)

7 = (10) ∪ (3,11) ∪ (4,6,8) ∪ (7)

5 = (10) ∪ (3,11) ∪ (4,6,8) ∪ (5)

~~2 = (7) (3)~~ [illegible]

2 = (3,11) ∪ (4,6,8) ∪ (2)

4. Any Limitation

1. some keywords, operators, or separators may be missing from the lists, we included as many as we could think of

2. cannot read the "::" and '\n' operators (eg. std::cout << "hello\n")

3. “++” and “--” operators need a space between former identifier or identifier won’t be read (eg. i++ needs to be i ++ to read the i)

5. **Any shortcomings**
   none