

GROUP - 8

→ There are two functions in the DFA:

↳ getFuncID(): On reaching this in DFA, after applying longest match, this can return two tokens:

- ① TK-MAIN: if input is -main
- ② TK-FUNID: if input is not -main

↳ getField(): On reaching this in DFA, after applying longest match, we do the following:

- Step1: Search for lexeme in lookup table.
- Step2: If lexeme is found, return the corresponding token, else goto step 3.
- Step3: Insert into lookup table; LEXEME: current match
- Step4: return TK-FIELDID

→ Algorithm used by Lexers: Modern compiler implementation in C, Reference Book 3.
Section 2.3: Finite Automata
Page 23: Recognising the longest match

→ Algorithm: We keep track of two variables:

- ① LAST-FINAL: The state number of most recent final state.
- ② INPUT-FINAL-POS: Position of input pointer at most recent final state.

along with

- ③ INPUT-POS: Position of last char read from input.
- ④ LEN: input chars read since last longest match token.

→ Do the following:

- ① make transition for next input char and increment INPUT-POS by 1.
- ② If the current state is final state, then LAST-FINAL is set to current final state
INPUT-FINAL-POS ← INPUT-POS

- ③ If transition for next character goes to dead state,

cur-token ← get token according to state or call the function, as described at beginning

start-of-token ← INPUT-POS-LEN

len-token ← INPUT-FINAL-POS-start-of-token

if (len-token == 0)

report lexical error

if (cur-token ≠ TK-WHITESPACE and cur-token ≠ TK-COMMENT)

report (cur-token, start-of-token, len-token) as a TOKEN

LEN ← 0

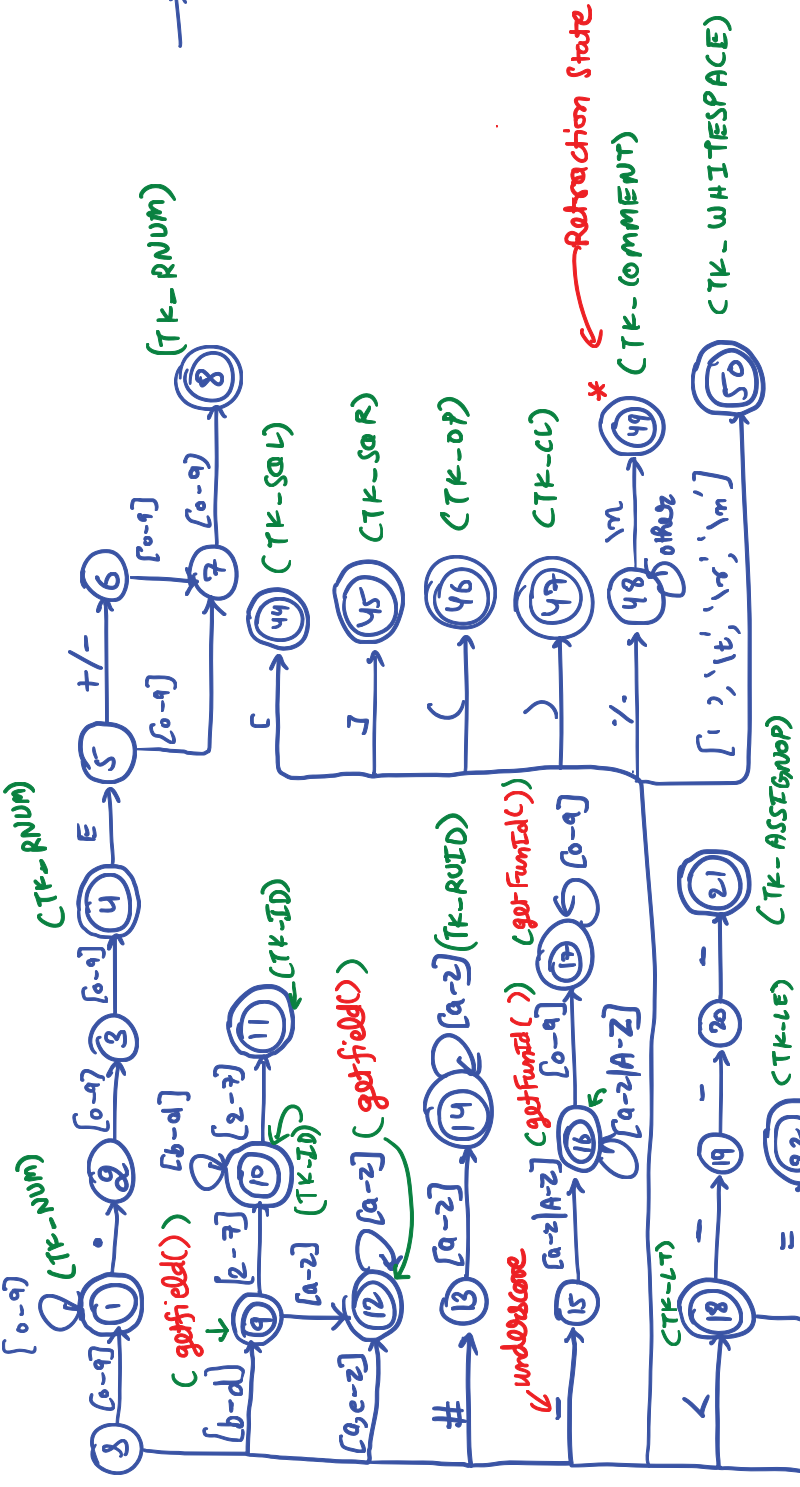
INPUT-POS ← INPUT-FINAL-POS

continue

LEXICAL

ERROR

whitespace and commands



Lookup Table	
LEXEME	TOKEN
with	TK-WITH
parameters	TK-PARAMETERS
end	TK-END
while	TK-WHILE
union	TK-UNION
endunion	TK-ENDUNION
defintype	TK-DEFINETYPE
as	TK-AS
type	TK-TYPE
-main	TK-MAIN
global	TK-GLOBAL
parameter	TK-PARAMETER
list	TK-LIST
input	TK-INPUT
output	TK-OUTPUT
int	TK-INT
real	TK-REAL
endwhile	TK-ENDWHILE
if	TK-IF
then	TK-THEN
endif	TK-ENDIF
read	TK-READ
write	TK-WRITE
return	TK-RETURN
call	TK-CALL
record	TK-RECORD
endrecord	TK-ENDRECORD
else	TK-ELSE

