Lexical Analysis

* Token Classes:

1) Keyword: Case-insensitive reserved words
   - load
   - filter
   - map
   - aggregate
   - where
   - for
   - in

2) Identifiers: Table names, variable names
   e.g. employees, salary

3) Literals:
   - String literals e.g. "employees.csv"
   - Number literals: e.g. 10, 1500.5

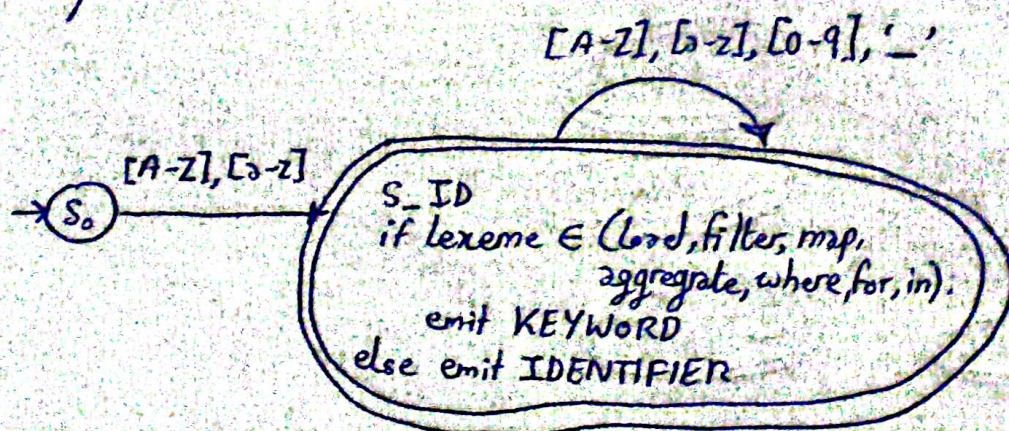4) Operators: +, -, *, /, =, !=, <, >, <=, >=, ==, /, AND, OR, ., comma(,)
   &&, ||

5) Grouping Symbols:
   LBRACE ({)
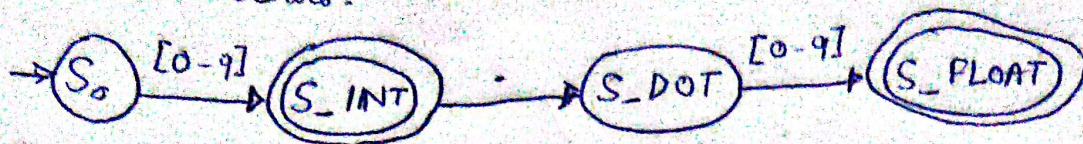   RBRACE (})
   LPAREN (( )
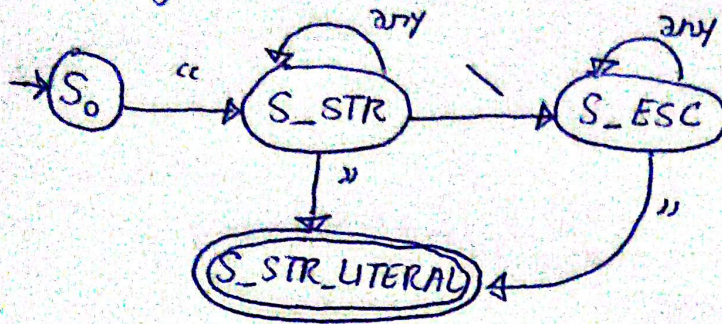   RPAREN ( ))

* DFAs
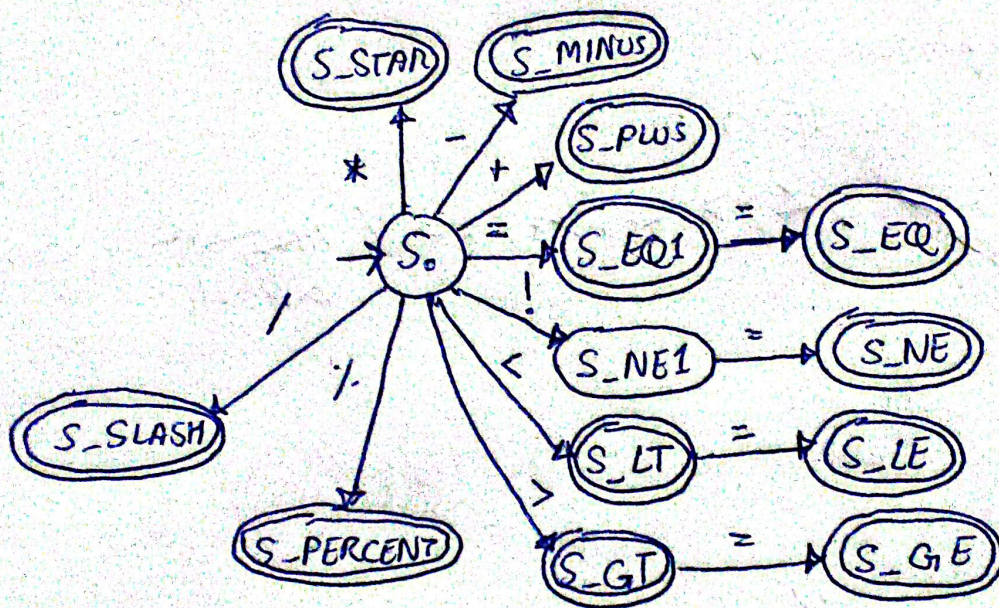   (For each token class and combined)

1) Keywords and Identifiers:

S_0 —[A-Z], [a-z]→

[A-Z], [a-z], [0-9], '_'

S_ID
if lexeme ∈ (load, filter, map, aggregate, where, for, in).
   emit KEYWORD
else emit IDENTIFIER

2) Number literals:

S_0 →[0-9]→ S_INT →.→ S_DOT →[0-9]→ S_PLOAT

3) String literals:

S_0 →"→ S_STR (any) →\→ S_ESC (any)

S_STR →"→ S_STR_LITERAL

S_ESC →"→ S_STR_LITERAL

-) Operators:

S_STAR   S_MINUS

S_PLUS

S_0:
- * → S_STAR
- - → S_MINUS
- + → S_PLUS
- = → S_EQ1 →=→ S_EQ
- ! → S_NE1 →=→ S_NE
- < → S_LT →=→ S_LE
- > → S_GT →=→ S_GE
- / → S_SLASH
- % → S_PERCENT

5: Grouping Symbols:

S_0:
- (→ S_LPAREN
-) → S_RPAREN
- { → S_LBRACE
- } → S_RBRACE