COMP3438 Assignment 2

JAHJA Darwin, 16094501d
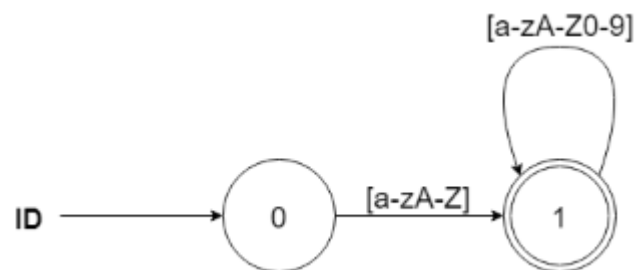
---

1. Regex

```
KEYWORD    -> var|begin|end
COMMA      -> ,
SEMICOLON  -> ;
ASSIGN     -> =
PERIOD     -> \.
NUM        -> [0-9]*(\.([0-9]+))?
PLUS       -> +
MINUS      -> -
MUL        -> *
DIV        -> /
LBRACE     -> (
RBRACE     -> )
ID         -> [a-zA-Z][a-zA-Z0-9]*
```

2. Finite Automata (FA)

The following constructed minimized-FAs are used to identify each token type:

KEYWORD & ID

A keyword is identifies by the above FA as an ID. Then, after an ID is obtained, the program will check a keyword table to see whether it is an ID or keyword.

In this way, the number of states can be reduced.

NUM — [0-9] loop on state 0, state 0 → state 1 on `.`, state 1 → state 2 on [0-9], [0-9] loop on state 2. States 0 and 2 are accepting.

COMMA — state 0 → state 1 on `,`. State 1 accepting.

SEMICOLON — state 0 → state 1 on `;`. State 1 accepting.

ASSIGN — state 0 → state 1 on `=`. State 1 accepting.

PERIOD — state 0 → state 1 on `.`. State 1 accepting.

PLUS — state 0 → state 1 on `+`. State 1 accepting.

MINUS — state 0 → state 1 on `-`. State 1 accepting.

MUL — state 0 → state 1 on `*`. State 1 accepting.

DIV — state 0 → state 1 on `/`. State 1 accepting.

LBRACE — state 0 → state 1 on `(`. State 1 accepting.

RBRACE — state 0 → state 1 on `)`. State 1 accepting.