CIIC 4030/ICOM 4036 Programming Languages
Assignment #1

Using PLY, design a Lexical Analyzer that recognizes the following valid tokens:

- **ID**: a string consisting of a letter (in the range a-z or A-Z) followed by zero or more letters and digits (in the range 0-9), but not equal to any of the keywords def, var, Int, if, else.
- **DEF**: the string (keyword) def
- **VAR**: the string (keyword) var
- **INT**: the string (keyword) Int
- **IF**: the string (keyword) if
- **ELSE**: the string (keyword) else
- **NUM**: a string consisting of a single digit (in the range 0-9) or two or more digits the first of which is not 0
- **LPAREN**: the string (
- **RPAREN**: the string)
- **LBRACE**: the string {
- **RBRACE**: the string }
- **BECOMES**: the string =
- **EQ**: the string ==
- **NE**: the string !=
- **LT**: the string <
- **GT**: the string >
- **LE**: the string <=
- **GE**: the string >=
- **PLUS**: the string +
- **MINUS**: the string -
- **STAR**: the string *
- **SLASH**: the string /
- **PCT**: the string %
- **COMMA**: the string ,
- **SEMI**: the string ;
- **COLON**: the string :
- **ARROW**: the string =>
- **COMMENT**: the string // followed by any characters other than the newline character (ascii 10, \n)
- **WHITESPACE**: one of the following characters: tab (ascii 9, \t), newline (ascii 10, \n), carriage return (ascii 13, \r), space (ascii 32)

A few additional instructions for this assignment

- Tokens that contain letters are case-sensitive; for example, Int is an INT token, while int is not.
- WHITESPACE and COMMENT tokens are irrelevant to the syntactic specification of the programming language in this assignment. Therefore, they should be removed (ignored) from the sequence of tokens.
- Your lexical analyzer must print each token and the row where the token appears in the test program (provided).
- You must submit via Moodle only your python code.