SRM UNIVERSITY AP
Andhra Pradesh

Department of Computer Science and Engineering
Compiler Design Lab (CS 306)

Week 5: Lexical analyser – Lab assignment

Instructions: This is optional and is given for self-practice. Implement following program and upload into your Github accounts under the folder **Week5-Lab-exercise**

Week 5 Lab Assignment: Consider the following mini Language, a simple procedural high-level language, only operating on integer data, with a syntax looking vaguely like a simple C crossed with Pascal. The syntax of the language is defined by the following BNF grammar:

<program> ::= <block>
<block> ::= { <variabledefinition> <slist> } | { <slist> }
<variabledefinition> ::= int<vardeflist>;
<vardeflist> ::= <vardec> | <vardec>, <vardeflist>
<vardec> ::= <identifier> | <identifier> [<constant>]
<slist> ::= <statement> | <statement>; <slist>
<statement> ::= <assignment> | <ifstatement> | <whilestatement> | <block> | <printstatement> | <empty>
<assignment> ::= <identifier> = <expression> | <identifier> [<expression>] = <expression>
<ifstatement> ::= <bexpression> then <slist> else <slist> endif | if <bexpression> then <slist> endif
<whilestatement> ::= while <bexpression> do <slist> enddo
<printstatement> ::= print (<expression>)
<expression> ::= <expression> <additionop> <term> | <term> | addingop> <term>
<bexpression> ::= <expression> <relop> <expression>
<relop> ::= < | <= | == | >= | > | !=
<addingop> ::= + | -
<term> ::= <term><mulitop> <factor> | <factor>
<multop> ::= * | /
<factor> ::= <constant> | <identifier> | <identifier> [<expression>] | (<expression>)
<constant> ::= <digit> | <digit> <constant>
<identifier> ::= <identifier> <letterordigit> | <letter>
<letterordigit> ::= <letter> | <digit>
<letter> ::= a|b|c|d|e|f|g|h|i|j|k|l|m|n|o|p|q|r|s|t|u|v|w|x|y|z
<digit> ::= 0|1|2|3|4|5|6|7|8|9
<empty> has the obvious meaning
Comments (zero or more characters enclosed between the standard C / Java style comment brackets /*...*/) can be inserted. The language has rudimentary support for 1-dimensional arrays. The declaration int a[3] declares an array of three elements, referenced as a[0], a[1] and a[2]. Note also that you should worry about the scoping of names.

A simple program written in this language is:

```
{ int a[3], t1, t2;
  t1 = 2; a[0] = 1; a[1] = 2; a[t1] = 3;
  t2 = -(a[2] + t1 * 6)/ a[2] -t1);
```

```
  if t2 > 5 then
    print(t2);
  else  {
    int t3;
    t3 = 99;
    t2 = -25;
    print(-t1 + t2 * t3); /* this is a comment on 2 lines */
  } endif
}
```

Design a Lexical analyser for the above language. The lexical analyser should ignore redundant spaces, tabs, and newlines. It should also ignore comments. Although the syntax specification states that identifiers can be arbitrarily long, you may restrict the length to some reasonable value.