Lexical Analysis

Assignment-1

Due : 9th February, 2021

Design a lexical analyzer in C++ to accept the languages mentioned below. The analyzer should remove whitespaces from the input code, ignore the text within comments (anything enclosed within /* */) and recognize the tokens. The description of the patterns are the following.

- Identifier: A string starting with an underscore or a letter and followed by any number of underscores, letters and digits.
- Keywords: if, then, else, int, return
- Num : Numerical constants

A file containing input code will be fed to the lexical analyzer. The lexical analyzer will print the keywords or num if it finds keywords or numerical constants, respectively in the input file. If it finds any identifier, it enters that into symbol table. The symbol table must contain two fields which are used to store (a) the lexeme itself and (b) the type (int or float).

Suppose, the following file is given as input to the lexical analyzer.

```
 /* The function is used to calculate the sum of two numbers */
int           add(int a, int b)
{
   return a+b;
}
int main()
{
   int x = add(5, 7);
}
```

The following output should be printed on the console.

```
int                              Keyword
add                              Identifier
a                                Identifier
b                                Identifier
```

return	Keyword
main	Identifier
x	Identifier
5	Num
7	Num

The lexical analyzer should also show the symbol table as output. It is given below.

a	int
b	int
x	int