Name: Saloni Vishwakarma
Roll No.C1-13

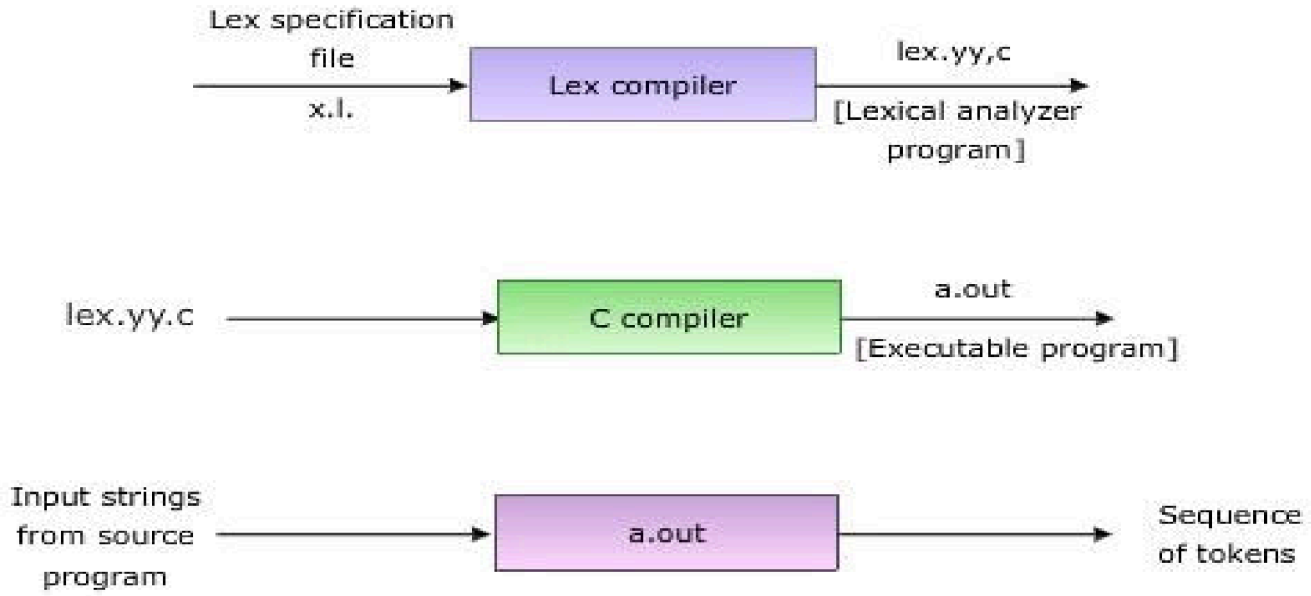
Practical No. 1

Theory

LEX: Lex is a program generator designed for lexical processing of character input streams. It accepts a high level, problem-oriented specification for character string matching, and produces a program in a general purpose language which recognizes regular expressions. The regular expressions are specified by the user in the source specifications given to Lex. The Lex written code recognizes these expressions in an input stream and partitions the input stream into strings matching the expressions. At the boundaries between strings program sections provided by the user are executed. The Lex source file associates the regular expressions and the program fragments. As each expression appears in the input to the program written by Lex, the corresponding fragment is executed.

Lex is not a complete language, but rather a generator representing a new language feature which can be added to different programming languages, called ``host languages." Just as general purpose languages can produce code to run on different com puter hardware, Lex can write code in different host languages.

Lex turns the user's expressions and actions (called source in this pic) into the host general-purpose language; the generated program is named yylex. The yylex program will recognize expressions in a stream (called input in this pic) and perform the specified actions for each expression as it is detected.



Generation of Lexical Analyzer using Lex

Diagram of LEX

Format for Lex file

The general format of Lex source is:

{definitions}
%%
{rules}
%%
{user subroutines}

where the definitions and the user subroutines are often omitted. The second %% is optional, but the first is required to mark the beginning of the rules. The absolute minimum Lex program is thus %% (no definitions, no rules) which translates into a program which copies the input to the output unchanged.

Regular Expression

A regular expression (or RE) specifies a set of strings that matches it; the functions in this module let you check if a particular string matches a given regular expression (or if a given regular expression matches a particular string, which comes down to the same thing).

Regular expressions can be concatenated to form new regular expressions; if A and B are both regular expressions, then AB is also a regular expression. In general, if a string p matches A and another string q matches B, the string pqwill match AB. This holds unless A or B contain low precedence operations; boundary conditions between A and B; or have numbered group references. Thus, complex expressions can easily be constructed from simpler primitive expressions. Regular expressions can contain both special and ordinary characters. Most ordinary characters, like "A", "a", or "0", are the simplest regular expressions; they simply match themselves. You can concatenate ordinary characters, so last matches the string 'last'. (In the rest of this section, we'll write RE's in this special style, usually without quotes, and strings to be matched 'in single quotes'.)

Some characters, like "|" or "(", are special. Special characters either stand for classes of ordinary characters or affect how the regular expressions around them are interpreted.

Lex Library Routines

Lex library routines are those functions which have a detailed knowledge of the lex functionalities and which can be called to implement various tasks in a lex program.

The following table gives a list of some of the lex routines.

Lex Routine	Description
Main()	Invokes the lexical analyzer by calling the yylex subroutine.
yywrap()	Returns the value 1 when the end of input occurs.
yymore()	Appends the next matched string to the current value of the yytext array rather than replacing the contents of the yytext array.

yyless(int n)	Retains n initial characters in the yytext array and returns the remaining

	characters to the input stream.
yyreject	Allows the lexical analyzer to match multiple rules for the same
	input string. (The yyreject subroutine is called when the special action REJECT is used.)
yylex()	The default main () contains the call of yylex()

Answer the Questions:

- Use of yywrap:
  Ans: yywrap is a function that allows you to control what happens when the end of input is reached.

- Use of yylex function:
  Ans:The yylex() function is the core function in Lex (or Flex) lexical analyzers. It's responsible for scanning the input stream and identifying tokens based on the rules defined in the lexer specification file.

- What does lex.yy.c. do?
  Ans:The primary purpose of lex.yy.c is to perform lexical analysis of input text. It contains the code that recognizes patterns defined in the Lex file and generates tokens accordingly.

Practical No.1 E1
Aim:Design a lexical analyzer to identify the tokens such as keywords, identifiers, operators, constants (Int & float), special symbols and strings for C language using LEX. Use File for the input.

Program:

```
%{
#include <stdio.h>

int identifiers = 0, operators = 0, keywords = 0, ints = 0, floats = 0, chars = 0, strings = 0;

%}

digit   [0-9]
letter  [A-Za-z]

%%

begin|if|else|for|while|int|return|char|float|case|switch|static|goto|struct|continue|break|void
 {keywords++; printf("Keyword: %s\n", yytext);}

{letter}({letter}|{digit})*      {identifiers++; printf("Identifier: %s\n", yytext);}
```

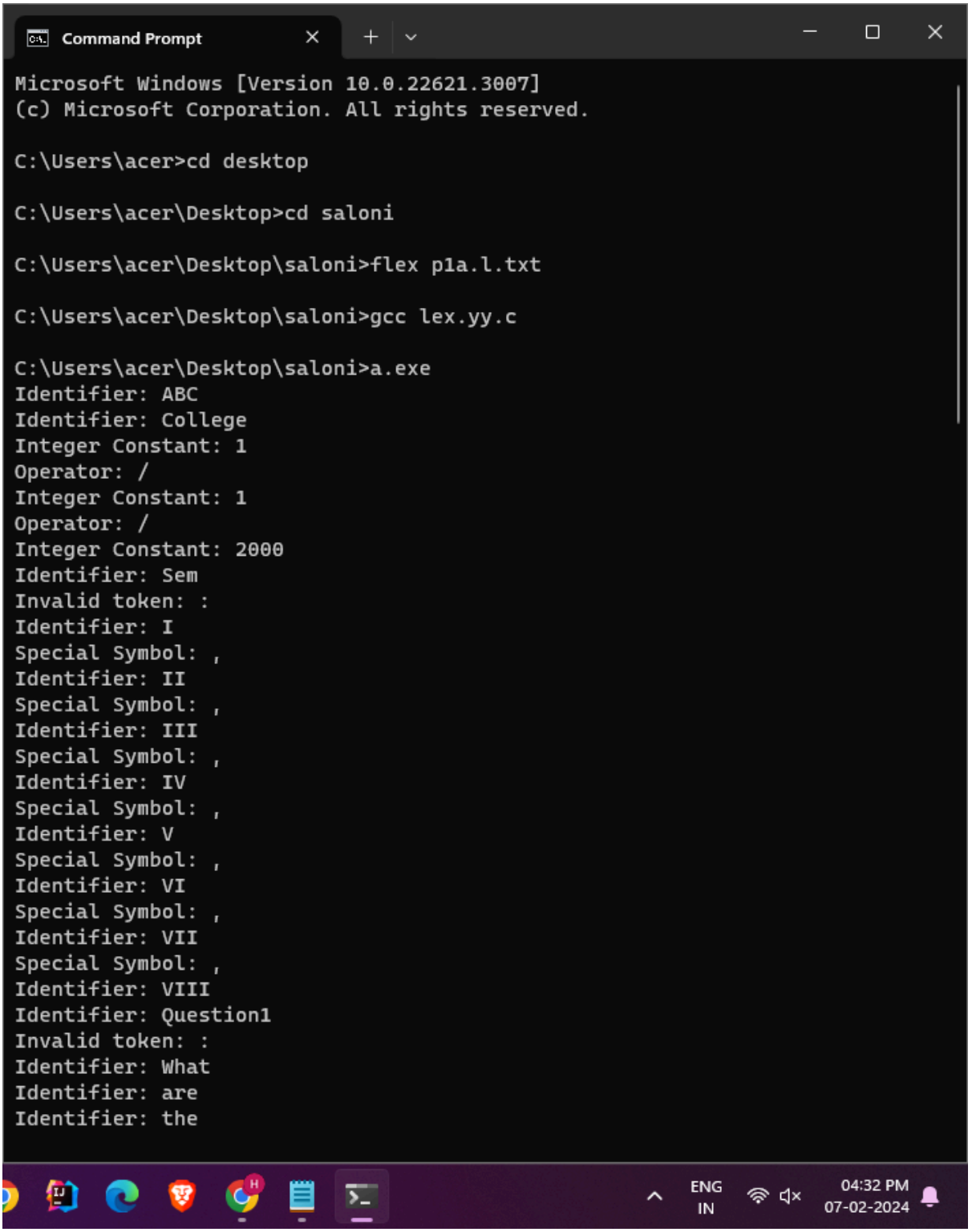
```
{digit}+    {ints++; printf("Integer Constant: %s\n", yytext);}

{digit}+"."{digit}*   {floats++; printf("Float Constant: %s\n", yytext);}

\'(\\.|[^\\'])\'    {chars++; printf("Character Constant: %s\n", yytext);}

\"(\\.|[^\\"])*\"    {strings++; printf("String: %s\n", yytext);}

[+\-*/]     {operators++; printf("Operator: %s\n", yytext);}

[(){}\[\];,#]    {printf("Special Symbol: %s\n", yytext);}

[=><]|(\+\+|\-\-|\+\=|\*\=|\/\=|\=\>) {operators++; printf("Operator: %s\n", yytext);}

\.  {operators++; printf("Operator: %s\n", yytext);}

[ \t\n]+    /* Ignore whitespace and newline */

.   {printf("Invalid token: %s\n", yytext);}

%%

int main() {
   yyin = fopen("input.txt", "r");
   yylex();

   printf("%d Identifiers, %d Operators, %d Keywords\n", identifiers, operators, keywords);
printf("%d Integer Constants, %d Float Constants, %d Character Constants, %d Strings\n", ints,
floats, chars, strings);

   return 0;
}

int yywrap()
{
   return (1);
}
```

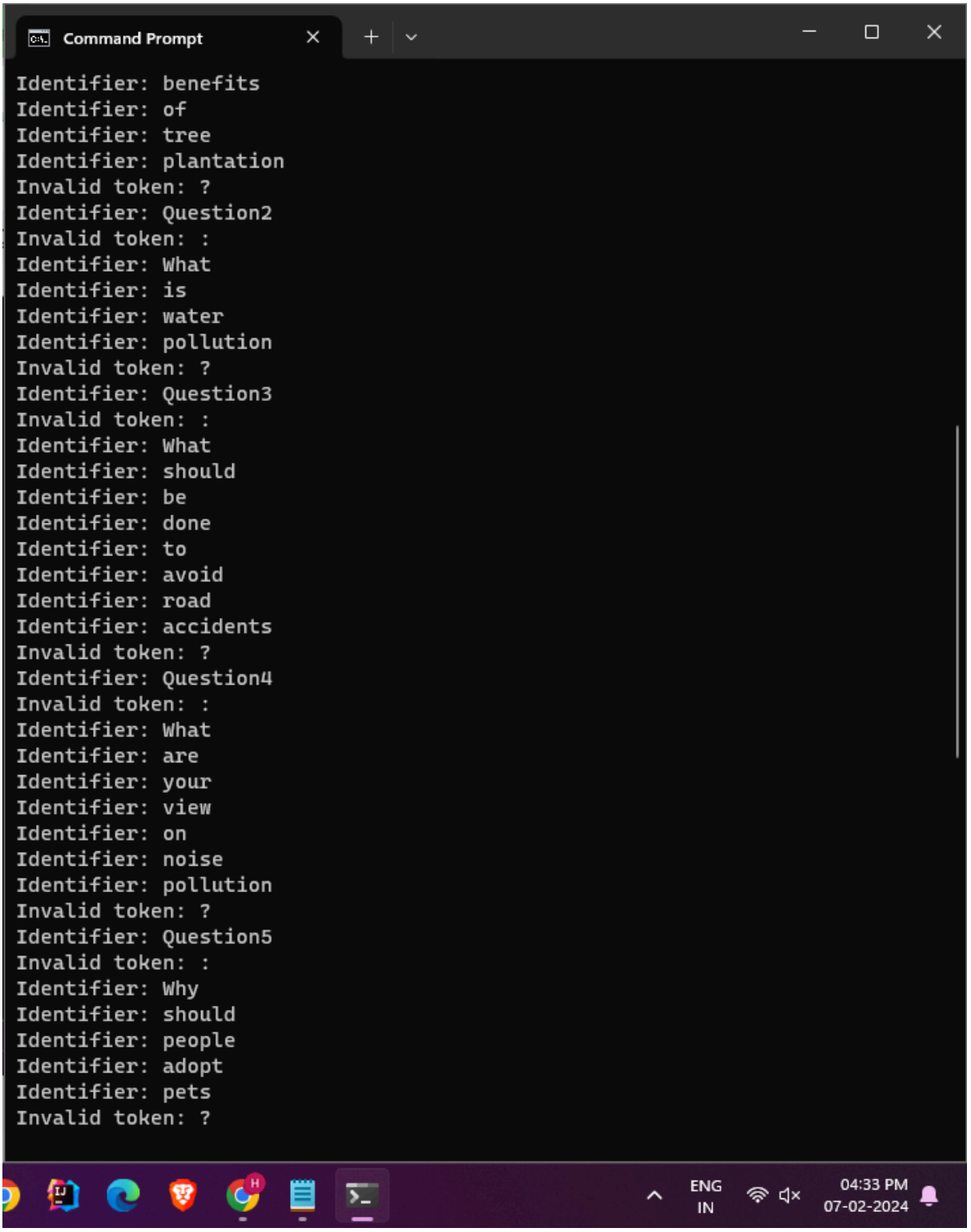
Input file:

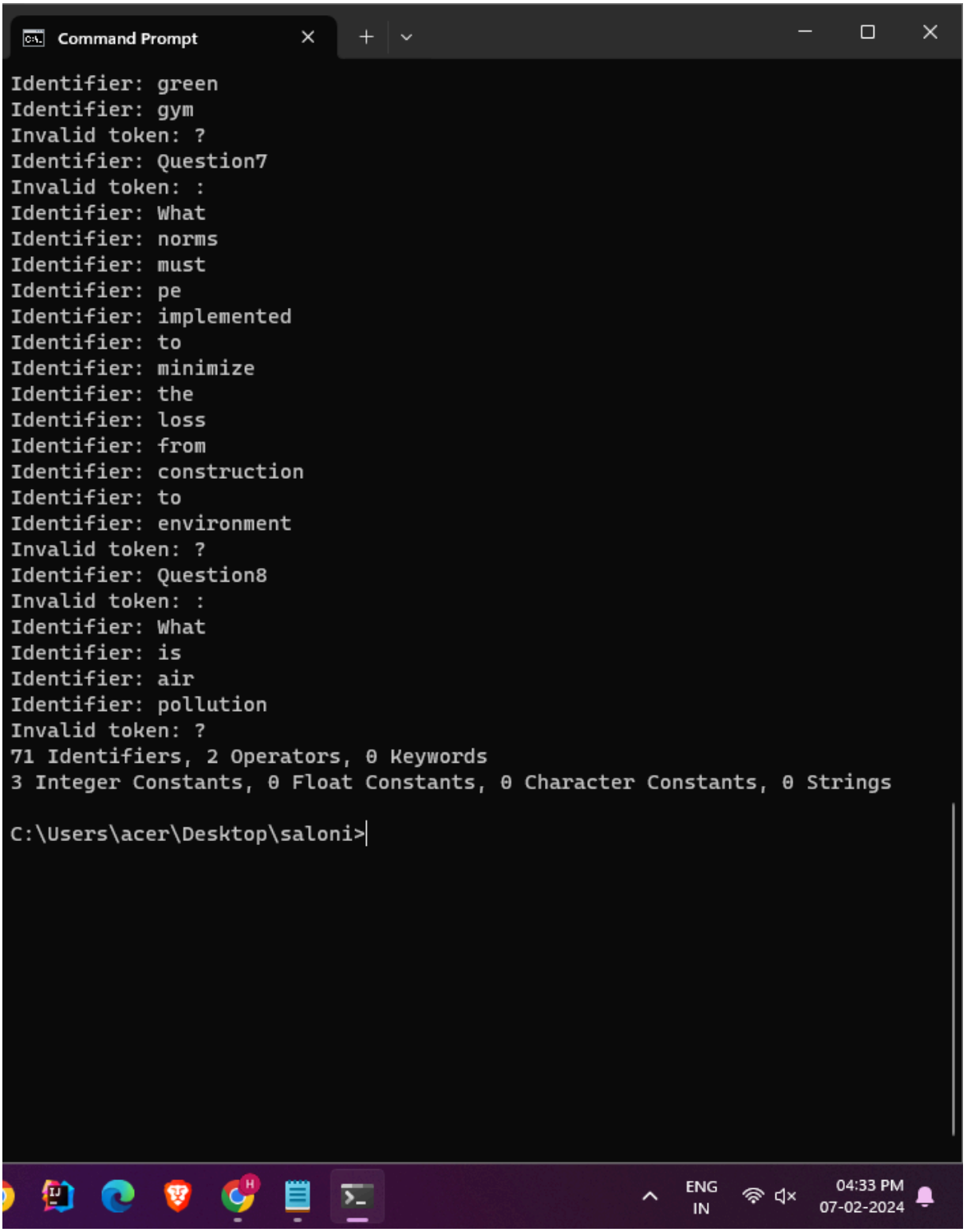
ABC College

1/1/2000 Sem: I, II, III, IV, V, VI, VII, VIII Question1
: What are the benefits of tree plantation?
Question2 : What is water pollution?
Question3 : What should be done to avoid road accidents?
Question4 : What are your view on noise pollution?
Question5 : Why should people adopt pets?
Question6 : What is green gym?
Question7 : What norms must pe implemented to minimize the loss from construction to environment?
Question8 : What is air pollution?

Output:


Command Prompt
Microsoft Windows [Version 10.0.22621.3007]
(c) Microsoft Corporation. All rights reserved.
C:\Users\acer>cd desktop
C:\Users\acer\Desktop>cd saloni
C:\Users\acer\Desktop\saloni>flex p1a.l.txt
C:\Users\acer\Desktop\saloni>gcc lex.yy.c
C:\Users\acer\Desktop\saloni>a.exe
Identifier: ABC
Identifier: College
Integer Constant: 1
Operator: /
Integer Constant: 1
Operator: /
Integer Constant: 2000
Identifier: Sem
Invalid token: :
Identifier: I
Special Symbol: ,
Identifier: II
Special Symbol: ,
Identifier: III
Special Symbol: ,
Identifier: IV
Special Symbol: ,
Identifier: V
Special Symbol: ,
Identifier: VI
Special Symbol: ,
Identifier: VII
Special Symbol: ,
Identifier: VIII
Identifier: Question1
Invalid token: :
Identifier: What
Identifier: are
Identifier: the
ENG
IN
04:32 PM
07-02-2024


Command Prompt
Identifier: benefits
Identifier: of
Identifier: tree
Identifier: plantation
Invalid token: ?
Identifier: Question2
Invalid token: :
Identifier: What
Identifier: is
Identifier: water
Identifier: pollution
Invalid token: ?
Identifier: Question3
Invalid token: :
Identifier: What
Identifier: should
Identifier: be
Identifier: done
Identifier: to
Identifier: avoid
Identifier: road
Identifier: accidents
Invalid token: ?
Identifier: Question4
Invalid token: :
Identifier: What
Identifier: are
Identifier: your
Identifier: view
Identifier: on
Identifier: noise
Identifier: pollution
Invalid token: ?
Identifier: Question5
Invalid token: :
Identifier: Why
Identifier: should
Identifier: people
Identifier: adopt
Identifier: pets
Invalid token: ?
ENG
IN
04:33 PM
07-02-2024

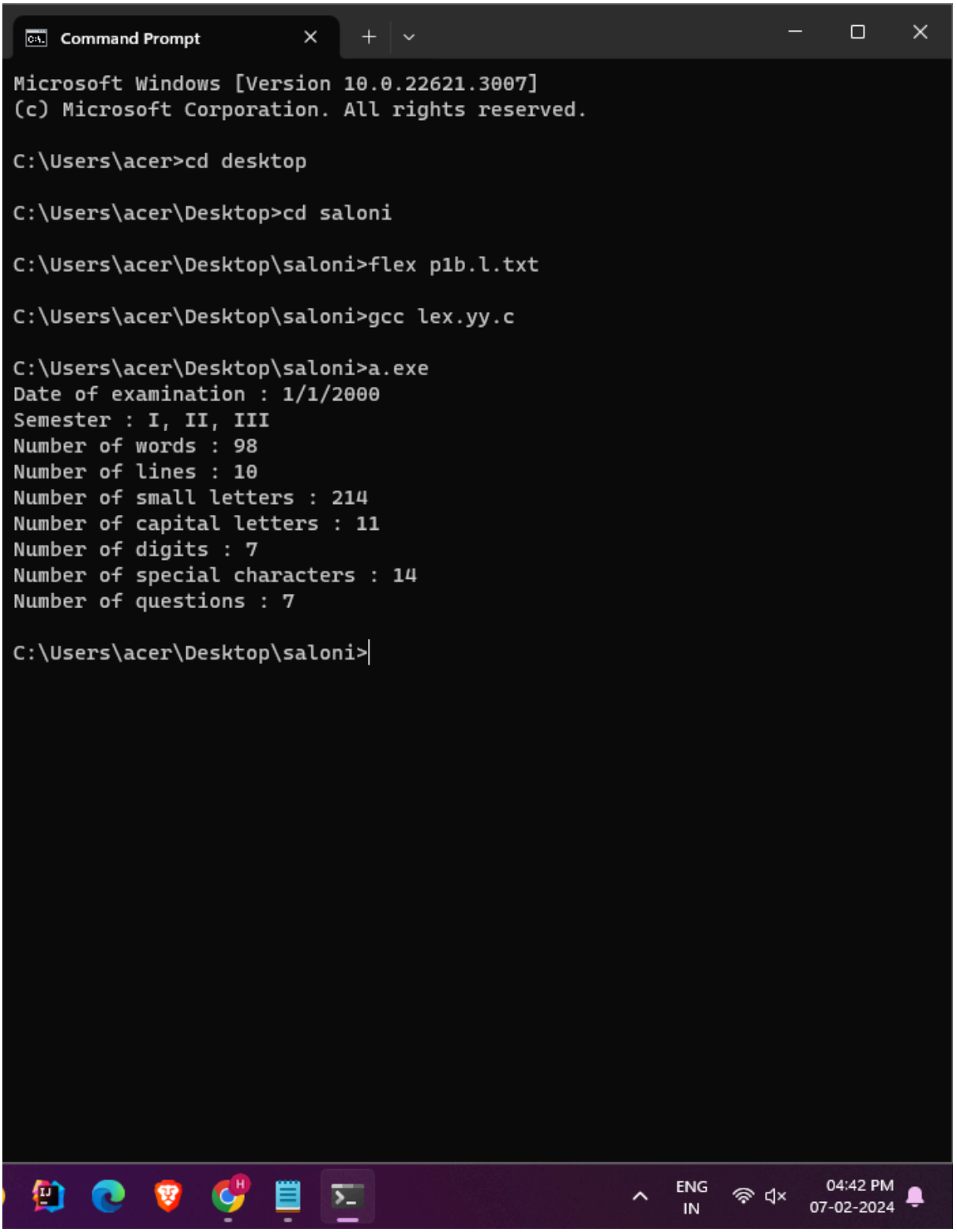

Command Prompt
Identifier: green
Identifier: gym
Invalid token: ?
Identifier: Question7
Invalid token: :
Identifier: What
Identifier: norms
Identifier: must
Identifier: pe
Identifier: implemented
Identifier: to
Identifier: minimize
Identifier: the
Identifier: loss
Identifier: from
Identifier: construction
Identifier: to
Identifier: environment
Invalid token: ?
Identifier: Question8
Invalid token: :
Identifier: What
Identifier: is
Identifier: air
Identifier: pollution
Invalid token: ?
71 Identifiers, 2 Operators, 0 Keywords
3 Integer Constants, 0 Float Constants, 0 Character Constants, 0 Strings
C:\Users\acer\Desktop\saloni>
ENG
IN
04:33 PM
07-02-2024

Practical No. E2

Aim:E2: Write a Lex program to find the parameters given below. Consider as input a question paper of an examination and find:
Date of examination, semester, number of questions, numbers of words, lines, small letters, capital letters, digits, and special characters.

Program:
Output:

```
%{
#include<stdio.h>
#include<string.h>
int words=0, lines=0, small=0, capital=0, digits=0, special=0, questions=0;
char date[10], sem[10];
%}

%%

[A-Z] { capital++; }
[a-z] { small++; }
[0-9] { digits++; }
[ \t ' '] { words++; }
\n { lines++; words++; }
"Question" { questions++; }
[^A-Za-z0-9 \t\n] { special++; }
[0-9]+"/"[0-9]+"/"[0-9]+ { strncpy(date, yytext, sizeof(date)); }
"Sem: ".* { strncpy(sem, yytext+5, sizeof(sem)); }

%%

int main() {
   yyin= fopen("input.txt","r");
   yylex();
   printf("Date of examination : %s\n", date);
   printf("Semester : %s\n", sem);
   printf("Number of words : %d\n", words);
   printf("Number of lines : %d\n", lines);
   printf("Number of small letters : %d\n", small);
   printf("Number of capital letters : %d\n", capital);
   printf("Number of digits : %d\n", digits);
   printf("Number of special characters : %d\n", special);
   printf("Number of questions : %d\n", questions);
   return 0;
}

int yywrap() {
   return 1;
}
```


Command Prompt
Microsoft Windows [Version 10.0.22621.3007]
(c) Microsoft Corporation. All rights reserved.
C:\Users\acer>cd desktop
C:\Users\acer\Desktop>cd saloni
C:\Users\acer\Desktop\saloni>flex p1b.l.txt
C:\Users\acer\Desktop\saloni>gcc lex.yy.c
C:\Users\acer\Desktop\saloni>a.exe
Date of examination : 1/1/2000
Semester : I, II, III
Number of words : 98
Number of lines : 10
Number of small letters : 214
Number of capital letters : 11
Number of digits : 7
Number of special characters : 14
Number of questions : 7
C:\Users\acer\Desktop\saloni>
ENG
IN
04:42 PM
07-02-2024