* Assignment No. 7 * (GroupB)

* Title:- Lexical analysis to Count number of words, lines and characters.

* Problem statement:- Write a program using lex specifications to implement lexical analysis phase of compiler to Count no. of words, lines and characters in given input file.

* Learning outcomes:-

① Able to Count no. of words, lines and characters in text file.

* SW Package and Hardware apparatus:-

64 bit open source Fedora 20.
Eclipse IDE.
Lex and Yacc.

* Learning objectives:-

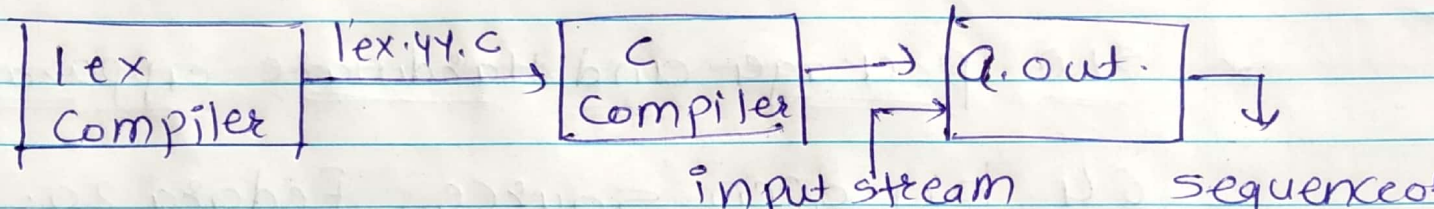
① Analyze source code.
② Identify tokens required in lexical analysis phase.
③ Count the lines, words and characters.

Theory:-

Lex:-

lex is a program that generates lexical analyzer.

* It is used with YACC parser generator.
* The lexical analyzer is a program that transforms an input stream, into a sequence of tokens.
* It reads the input stream and produces the source code as output through implementing the lexical analyzer in the C program.



Functions of lex:-

① Firstly lexical analyzer creates a program lex.l in the language of lex. Then lex compiler runs the lex.l program and produces a C program lex.yy.c.
② Finally C compiler runs the lex.yy.c program and produces an object program a.out.

③ a.out is lexical analyzer that transform an input stream into sequence of tokens.

Lex file format:-

* A lex program is seprated into 3 sections by %% delimetery.
* The format of lex source is as follows.

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

* Program structure:-

* Definition section:-

The definition section contains the declaration of variables, regular definitions, manifest constants.

* In the definition section, text is enclosed in %{ %} brackets.
* Anything written in this brackets is copied successfully to file yy.c

* Rules section:-

* The rules section contains a series of rules in the form: pattern action and pattern must

be unintended and action begin on the same line in { } brackets.
* The rule section is enclosed in "%% ... %%".

* User Code Section:-

* This section contains C statements and Conditional functions.
* We can also compile these functions seperately and load this with lexical analyze.

* Advantages of lexical Analysis:-

① Lexical analyzer method is used by programm like compilers which can use the parsed data from a programmer's code to create a compiled binary executable code.
② It is used by web browsers to format and display a web browsers with the help of parsed data from js, html and css.
③ A seperate lexical analyzer helps you to construct a specliazed and potentially more efficient processor for the task.

* Disadvantages:-

① You need to spend significant significant time reading the source program and partitioning it in the form of tokens.

② More effort is needed to develop and debug the lexer and its token descriptions.

* Steps to run the program:-

① To run the program, it should be first saved with the extention .l or .lex Run the below commands on terminal in order to run the program file.

Commands:-

```
lex filename.l
gcc lex.yy.c
./a.out
```

Test Case:-

This is my 1st lex program.

The output will

2 lines
9 words
30 small characters
3 Capital letters
1 digit
9 special character.
In total 43 characters.

Conclusion:- Thus, implemented lex program successfully to calculate number of words, characters and lines in the given text file.