PSG COLLEGE OF TECHNOLOGY

DEPARTMENT OF APPLIED MATHEMATICS AND COMPUTATIONAL SCIENCES

MSC SS – VI SEMESTER

18XW93 PRINCIPLES OF COMPILER DESIGN LAB

PROBLEM SHEET -I

LEXICAL ANALYSIS

The first stage in compilation process is lexical analysis. This is the process through which a stream of text is converted into a sequence of tokens. Additionally, it does some preprocessing like

a. insert line numbers
b. remove comment lines
c. compress multiple spaces/tabs/new lines into a single space/tab/new line.

Implement the following using C/C++/JAVA/Python. Try all operations for C/C++ language.

1. Write a program to scrap the psgtech website and faculty publication details of all the departments and create a csv file containing faculty name, google scholar url and research interests.
2. Read a HTML file, remove the tags and convert into text file
3. Write a program to read a text file and count the number of alphabets, digits, space, special characters, lines and words in the text file. Also find the size of file.

4. Write a program to read a C program (with simple macros using # define statements) from a file and expand the source program by replacing macro calls with the corresponding macro definitions in the source file. Print the expanded source program and store it in a separate file.
5. Write a program to read the source program which is stored in a file and insert line numbers at the beginning of every line in a file. Store the source program with line numbers in a separate file. Print the contents of output file.
6. Write a program to read a source program (C++) from a file and remove single or multi line comments.
7. Write a program to read a source program from a file and compress multiple spaces/tabs/lines into a single space/tab/line.
8. Write a program to create a file of keywords of C language. Then, read a file of keywords and store them in an efficient data structure. The data structures under consideration are :
    a. Array
    b. Linked list
    c. Binary search tree
    d. Hash table
9. Write a program to read a C / C++ program file and identify the following different tokens of C/C++ programming language.
    a. Keywords (Use the data structure created from Qn. 4)
    b. Identifiers
    c. Integer Constant
    d. Floating point constant

e. String constant
f. Labels
g. Operators (Arithmetic, relational, logical)
h. Octal
i. Backslash character constant

10. Integrate the above modules to construct a lexical analyser that generates a sequence of tokens and generates errors if
    a. Any strange characters that are not in the character set of a language. However, these characters within a quoted string are allowed.
    b. Invalid numbers such as floating point constant 123.67.89 exist.
    c. Long quoted string that exceeds a single line.

11. Given a non-deterministic finite automaton of an identifier, design an equivalent deterministic Finite Automata which recognizes the identifier.

Problem Sheet – II (LEX)

1. Write a lex program to read a text file and count the number of alphabets, digits, space, special characters, lines and words in the text file.
2. Write a program to read the source program which is stored in a file and insert line numbers at the beginning of every line in a file. Store the source program with line numbers in a separate file. Print the contents of output file.

3. Write a program to read a source program (C++) from a file and remove single or multi line comments.
4. Write a program to read a source program from a file and compress multiple spaces/tabs/lines into a single space/tab/line.
5. Write a program to read a C / C++ program file and identify the following different tokens of C/C++ programming language.
    a. Keywords (Use the list of keywords as patterns)
    b. Identifiers
    c. Integer Constant
    d. Floating point constant
    e. String constant
    f. Labels
    g. Operators (Arithmetic, relational, logical)
    h. Octal
    i. Backslash character constant
6. Write a program to read file containing a list of valid and invalid URLs and create an output file for storing valid urls .
7. Write a lex program to read a file containing ip and mac addresses and store valid (ip and mac) addresses in a file.

Problem Sheet – III (YACC)

1. Write a YACC program to check the syntax of the following language constructs:
    a) Arithmetic expression

b) Relational expression
c) Boolean expression
d) If statement
e) For loop
f) While statement
g) Switch statement
h) SQL Queries (SELECT, INSERT, UPDATE)
i) Declaration statement
j) Assignment statement
k) Class declaration and definitions