NORTH WESTERN UNIVERSITY

Report

Course Code: CSE-4104
Course Title: Compiler Design Sessional

Developed by:

Tonoy Das

Student Id: 20201087010

Riaj Hossen

Student Id: 20201071010

Department Of CSE

North Western University

Khulna, Bangladesh

Table of Contents

1.Introduction

Lexical analysis in compiler design is a software where token analysis of C programming code is done using Java language.

It is the first phase of a compiler or interpreter that converts the source code into a more manageable form for subsequent analysis. A lexical token is a sequence of characters that can be treated as a unit in the grammar of the programming languages. Example of tokens:

Keywords; Examples-for, while, if etc.

Identifier; Examples-Variable name, function name, etc.
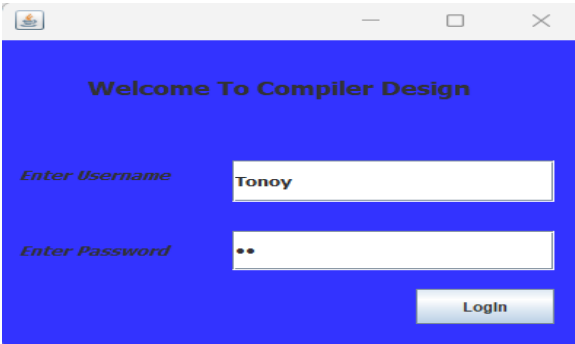
Operators; Examples '+', '++', '-' etc.

Separators; Examples ',' ';' etc.

2.Objectives

The purpose of this lab project was to design and implement a Lexical Analysis is the first phase of the compiler also known as a scanner. It converts the High level input program into a sequence of Tokens. The objective of this project is to create a program that can analyze the source code of a programming language and break it down into individual tokens. The program should accurately recognize and categorize tokens such as keywords, identifiers, literals, operators, and punctuation symbols and get the line number.

3.Description:

Log In: This the login page of Lexical analysis in compiler design. After pressing username Tonoy and password 87 the page login.

By Pressing Login Button users are able to go Lexical Analysis page.

User Input: A code of C programming should be pasted here.

Tokenization: By Pressing Analyze Button users are able to go different Tokenization.
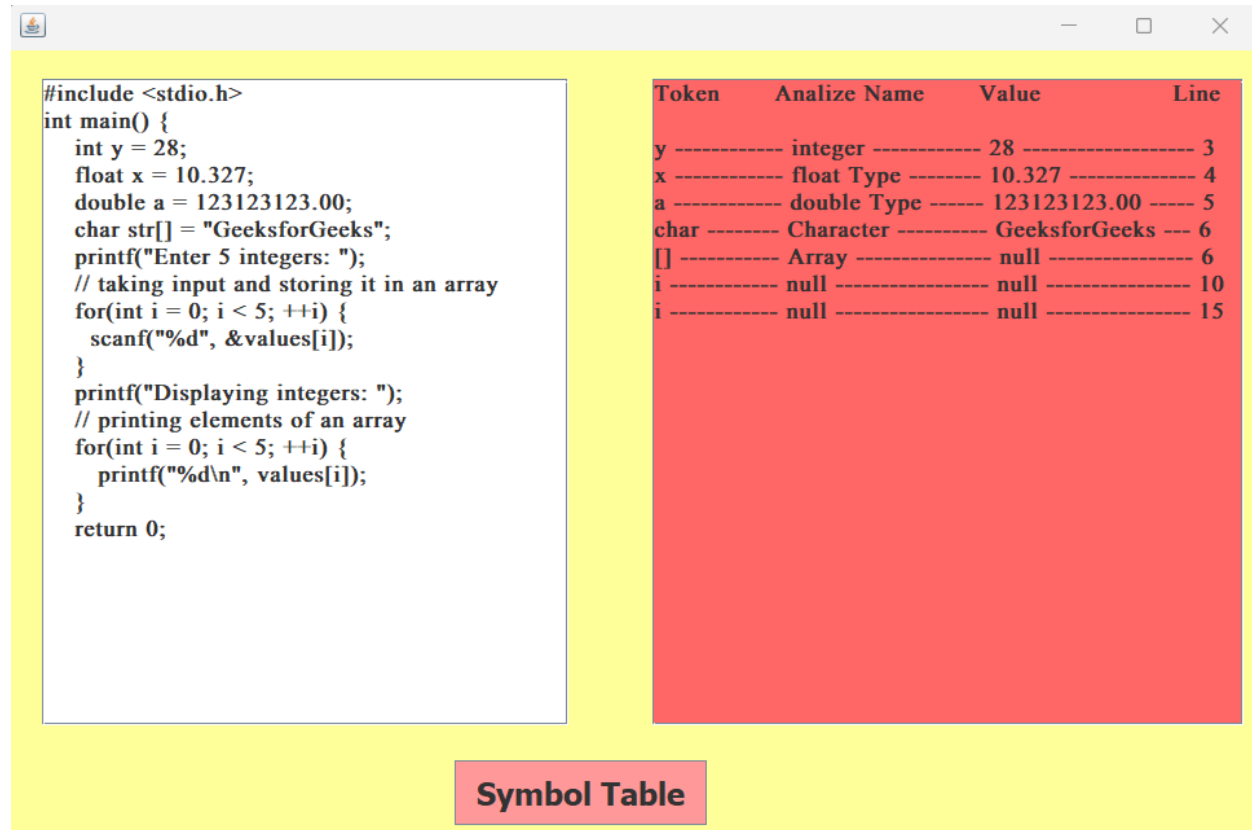

By Pressing Symbol Button users are able to go Symbol Analysis page.

The sequence of characters matched by a pattern to form the corresponding token or a sequence of input characters that comprises a single token is called a lexeme. A C program consists of various tokens and a token is either a keyword, an identifier, a constant, a string literal, or a symbol. **Token** is the smallest entity of the code, it is either a keyword, identifier, constant, string literal, symbol. Lexeme is a sequence of characters in the source code that are matched by given

predefined language rules for every lexeme to be specified as a valid token. Pattern specifies a set of rules that a scanner follows to create a token.

Symbol Table: By Pressing Button users are able to go different Variable Analysis.

4. Dependencies:

NetBeans IDE 8.2: NetBeans IDE lets you quickly and easily develop Java desktop. For built this project we use Java language and Java Built in Packages. After Install all packages the project will run successfully.

C Language: We perform token analysis of C Language from the user.

05 June, 2023

North Western University

Khulna, Bangladesh.