Helwan University

Faculty of computers and artificial intelligence

Computer Science department

Spring 2020-2021

Compilers

Compiler #1

Compiler #1

Scanner

Description:

Compiler #1 is a **case sensitive** imperative Computer programming language Like C-Language.

Lexical Analysis:

Compiler#1 Scanner is a lexical analysis stage used to locate and identify language's lexemes. It must recognize the following Lexemes and returns Return Token according to the following table:

Lexeme	Meaning	Return Token
Yesif-Otherwise	conditional statements	Condition
Omw	Integer type	Integer
SIMww	Signed Integer type	SInteger
Chji	Character Type	Character
Seriestl	Group of characters	String
IMwf	Float type	Float
SIMwf	Signed Float type	SFloat
NOReturn	Does not return a value	Void
RepeatWhen/Reiterate	repeatedly execute code as long as condition is true	Loop
GetBack	Return a value from a function	Return
OutLoop	Break immediately from a loop	Break
Loli	grouped list of variables placed under one name	Struct
(+, -, *, /,)	Used to add, subtract, multiply and divide respectively	Arithmetic Operation
(&&, ||, ~)	Used to and, or and not repectively	Logic operators
(==, <, >, !=, <=, >=)	Used to describe relations	relational operators
=	Used to describe Assignment operation	Assignment operator
->	Used in loli to access loli elements	Access Operator
{,},[,],(,)	Used to group statements ,array index or loop condition respectively	Braces
[0-9] and any combination	Used to describe numbers	Constant

“,’	Used in defining strings and single character respectively	Quotation Mark
Include	Used to include one file in another	Inclusion
/@	Used to Comment some portion of code (Multiple Lines)	Comment
@/	Used to a matcher to Comment left side (Multiple Lines)	Comment
/^	Used to Comment some portion of code (Single Line)	Comment
$	Used as Token Delimiter	Token Delimiter
.	Used as Line Delimiter	Line Delimiter
Start	Program Starting Statement	Start
Last	Used as Line Delimiter	End

Com#1_1: Tokens Description

The Scanner also recognizes identifiers which are not listed in table Com#1_1. An identifier is a sequence of letters and digits, starting with a letter. The underscore ‘_’ counts as a letter. For each identifier, Compiler #1 Scanner returns the token IDENTIFIER. Compiler#1 language allows many identifiers to be identified by one type separated by comma (,)

Comments in Compiler#1:

Compiler #1 includes two types of comments single line comments are prefixed by /^ and multiple line comment are written between /@ and @/.Your scanner must ignore all comments and white.

Include file command:

In order to facilitate the inclusion of multiple files, your Compiler#1 scanner is also responsible for directly handling the include file command. When encountering the include directive placing at the first column of a given line, the scanner must open the file indicated by the file name in the directive and start processing its contents. Once the included file has been processed the scanner must return to processing the original file. An included file may also include another file and so forth. If the file names does not exist in the local directory you should simply ignore the include command and proceed with the tokens in the current file.

Compiler#1 Input:

A Source Code Written on Editor or included from external files.

Compiler#1 Output format:

You must build a dictionary to save lexemes that are defined in Compiler#1 language. Dictionary structure is according to the following:

Line NO	Lexeme	Return Token	Word NO in Line	matchability

Total NO of Errors:

Note: Matchability must be either Matched or Not Matched

Compiler#1 Language Delimiters (words and lines):

The lexemes are delimited by Dollar Sign ($) and lines are delimited by Dot (.).

Sample Input and output:

Input:

1-/^This is main function

2- Imw$decrease(){

3- **Imw**$3num=5.

4- **Reiterate** (counter<num){

5-reg3=reg3-1.} }

Scanner Output:

Line NO	Lexeme	Return Token	Lexeme NO in Line	matchability
1	/^	Comment	1	Matched
2	Imw	Integer type	1	Matched
2	$	Token Delimiter	2	Matched
2	Decrease	Identifier	3	Matched

----------Etc.

Total NO of errors:1 if found

Best Wishes Prof.Amal Aboutabl , Eng.Ahmed Badawy