CSE306L

Compiler Design Project

R programming language

Phase 1 Report

AP19110010393 - Akash Perla
AP19110010424 - Kreethi Mishra
AP19110010426 - Hanish Vaitla
AP19110010432 - Lalith Veerla
AP19110010442 - Harshini Chintalacheruvu
AP19110010448 - Lehar Sancheti
AP19110010449 - Sumana Bandarupalli

CSE C
Group 8

Contents

Data types available in the R language

R supports five data types- Numeric, Integer, Complex, Character(a.k.a. String), Logical(a.k.a. Boolean).

There are three number types in R:

1. Numeric - A numeric data type is the most common type in R, and contains any number with or without a decimal, like: 10.5, 55, 787.
2. Integers - They are numeric data without decimals. This is used when you are certain that you will never create a variable that should contain decimals. To create an integer variable, you must use the letter L after the integer value (500L, 4000000L)
3. Complex - A complex number is written with an "i" as the imaginary part: 3+4i, 15+i.
4. Character(a.k.a. String) - R supports character data types where you have all the alphabets and special characters. It stores character values or strings. Strings in R can contain alphabets, numbers, and symbols. The easiest way to denote that a value is of character type in R is to wrap the value inside single or double inverted commas. (“Hello World”)
5. Logical(a.k.a. Boolean) - R has logical data types that take either a value of TRUE or FALSE. A logical value is often created via a comparison between variables. A conditional expression that evaluates to true or false is also of boolean type.

To find the data type of an object you have to use the *class()* function.

Syntax of variable declaration & assumptions for R language

Ways of Declaring a Variable:

- In R, a variable always starts with a letter or with a period. A variable if started with a dot cannot be succeeded by a number.
- Variables cannot be created with keywords that are already predefined in R.
- A variable in R can be defined using just letters or an underscore with letters, dots along with letters. We can even define variables as a mixture of

digits, dots, underscore, and letters.

Assigning values to variables:

- In R, assigning values to a variable can be specified or achieved using the syntax of left angular brackets signifying the syntax of an arrow.

  Ex: a <- 12
  b <- "Hello"

- After assigning values to a variable, these values can be printed using the predefined cat() function and print() function.
- To integrate multiple statements assigned to a single variable and separated by double quotation marks and commas, we can use the function of cat().

CODE	OUTPUT
X <- "Hi, This is CD Phase 1 Report By Group 8" cat(X)	Hi, This is CD Phase 1 Report By Group 8

Decision-making statements in R language

Decision-making is about deciding the order of execution of statements based on certain conditions.
In decision making programmer needs to provide some condition that is evaluated by the program, along with it there also provided some statements which are executed if the condition is true and optionally other statements if the condition is evaluated to be false.

- if-else statement
  If-else provides us with an optional else block which gets executed if the condition for if block is false. If the condition provided to if block is true then the statement within the if block gets executed, else the statement within the else block gets executed.

- Syntax of if-else:

```
if(condition is true) {
        execute this statement
    }  else {
      execute this statement
    }
```

Iterative statements in R language

- for Loop:
  This type of loop is to be used when someone knows exactly how many times you want the code to repeat. The for loop accepts an iterator variable and a vector. It repeats the loop, giving the iterator each element from the vector in turn. In the simplest case, the vector contains integers:

  EX:
  ```
  x <- c(1,9,3,5,8,7,2)
  count <- 0
  for (val in x) {
  if(val %% 2 == 0)  count = count+1
  }
  print(count)

  OUTPUT: [1] 2
  ```

  In the above example, the loop iterates 7 times as the vector x has got 7 elements. In each iteration, the variable takes on the value of the corresponding element of x. Here we have used a counter to count the number of even numbers in x. We can see that x contains 2 (2 and 8) even numbers.

- while Loop:
  'While' loops are more like backward repeat loops. Instead of executing some code and then checking to see if the loop should end or not, this type of loop checks first and then (maybe) executes. Since the check happens at the very beginning, the contents of the loop may never be executed (unlike in a repeat loop).

EX:
```
i <- 20
while (i < 30) {
print(i)
i = i+1
}
```

In general, it is always possible to convert a ‘repeat’ loop to a ‘while’ loop or a ‘while’ loop to a ‘repeat’ loop, but usually the syntax is much cleaner one way or the other. If you know that the contents must execute at least once, use repeat; otherwise, use while.

CFG for constructs in R Language

Data Types

identifier ⟶ < . ><letter><identifier> | (<letter> | <digit>| _ | .) <identifier>) | λ
numeric ⟶ <integer> | <double>
integer ⟶ (+ | -) <digit> ‘L’
double ⟶ (+ | -) <digit> . <digit> | (+ | -) <digit> .
letter ⟶ a | b | c | d |.....| y | z | A | B | C | D |.....| Y | Z
digit ⟶ [0-9] <digit> | [0-9]

if statement

if_statement ⟶ if (<condition>) { <statements> }
condition ⟶ <expr> <op> <expr>
op ⟶ < | > | == | >= | <= | !=
statements ⟶ <identifier> <- <expr> <statements> | <identifier> <ops> <expr> <statements> |print(<identifier>) <statements> | λ
ops ⟶ += | -= | *= | *= | /=
expr ⟶ <identifier> | <numeric>

if else statement

if_else_statement → if (<condition>) { <statements> } else { <statements> }
condition → <expr> <op> <expr>
op → < | > | == | >= | <= | !=
statements → <identifier> <- <expr> <statements> | <identifier> <ops> <expr> <statements> |print(<identifier>) <statements> | λ
ops → += | -= | *= | *= | /=
expr → <identifier> | <numeric>

for loop

for_loop → for(<identifier> in <range>) { <statements> }
range → <integer> : <integer> | seq(<integer> , <integer>) | seq(<integer> , <integer> , <numeric>)
statements → <identifier> <- <expr> <statements> | <identifier> <ops> <expr> <statements> |print(<identifier>) <statements> | λ
ops → += | -= | *= | *= | /=

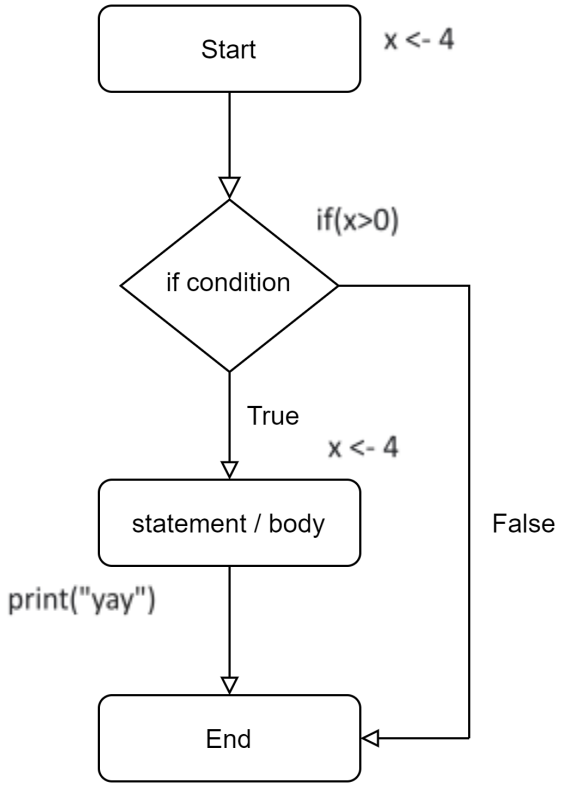
while loop

while_loop → while (<condition>) { <statements> }
condition → <expr> <op> <expr>
op → < | > | == | >= | <= | !=
statements → <identifier> <- <expr> <statements> | <identifier> <ops> <expr> <statements> |print(<identifier>) <statements> | λ
ops → += | -= | *= | *= | /=
expr → <identifier> | <numeric>

Examples with flow graphs:

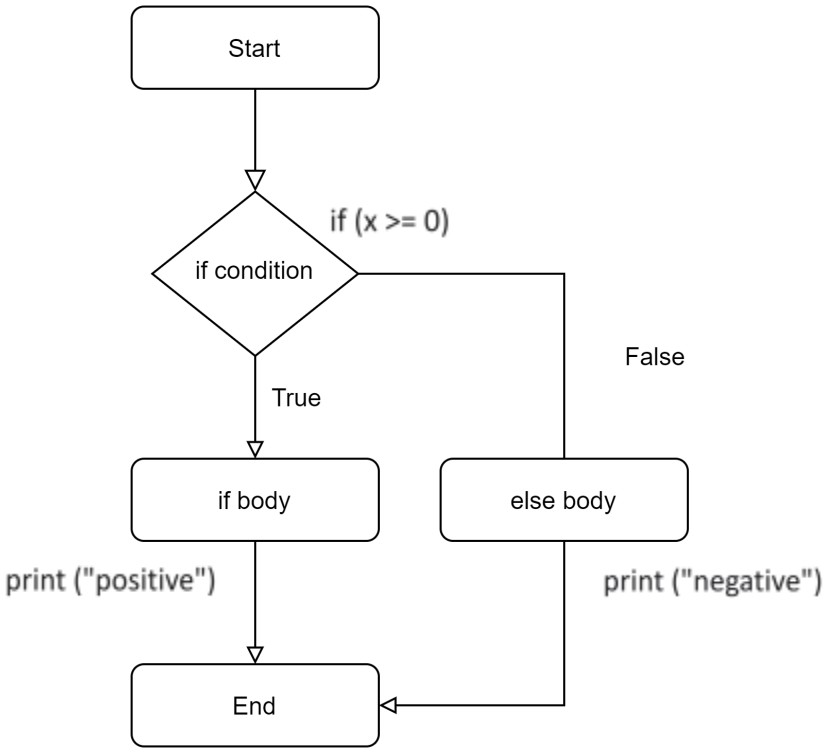
if statement

```
x <- 4
if(x>0) {
print("yay")
}
```



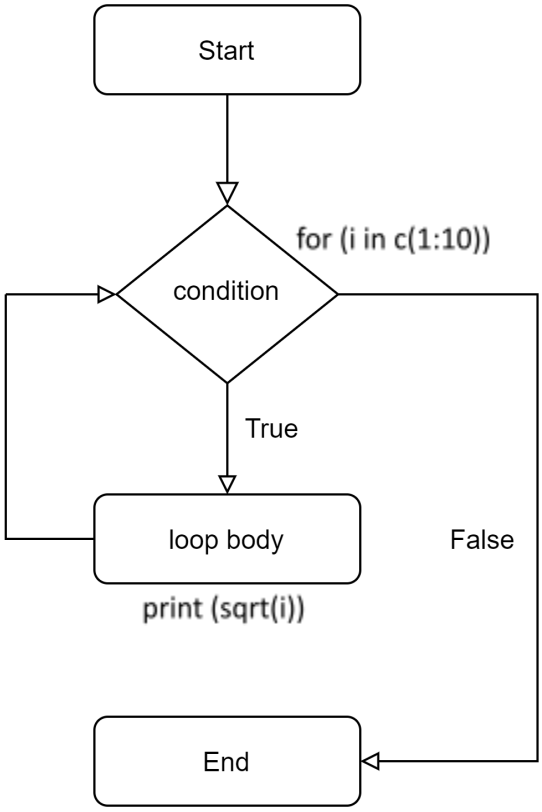
if-else statement

```
x <- 4
if (x >= 0) {
print ("positive")
} else {
print ("negative")
}
```



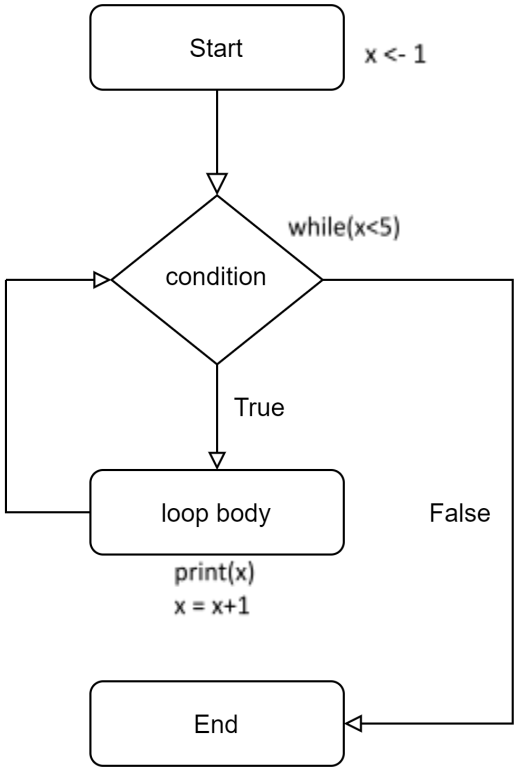
for Loop:

```
for (i in c(1:10)) {
print (sqrt(i))
}
```



while Loop:

```
x <- 1
while(x<5) {
print(x)
x = x+1
}
```



Design of Parser

Syntax Analysis or parsing is the second phase after the lexical analyzer. A syntax analyzer or parser takes the input from the lexical analyzer in the form of tokens. The parser analyses the source code against the production rules to detach the errors. It checks whether the given input is incorrect syntax. It dies it by building parse tree. The parse tree is defined by using predefined Grammer. If the given string can be used by predefined Grammer then the string is accepted as correct syntax.

Parse Tree
A parse tree is the graphical depiction of a derivation. The start symbol becomes the root node of the tree

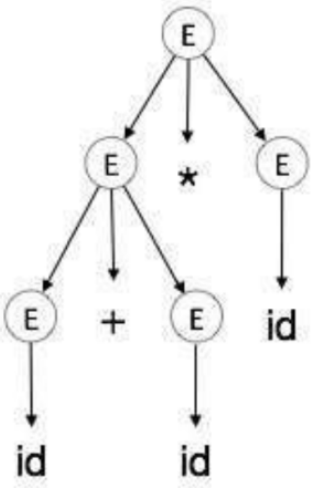
Consider the following example with given production rules

Input String:id+id*id
Production rules:

E->E+E
E->E*E
E->id

The leftmost Derivation is
E->E*E
E->E+E*E
E->id+E*E
E->id+id*E
E->id+id*id

The parse tree for the following is



In a Parse tree:

1)All leaf nodes are terminal

2)All interior nodes are non-terminals

3)In order terminal gives original output

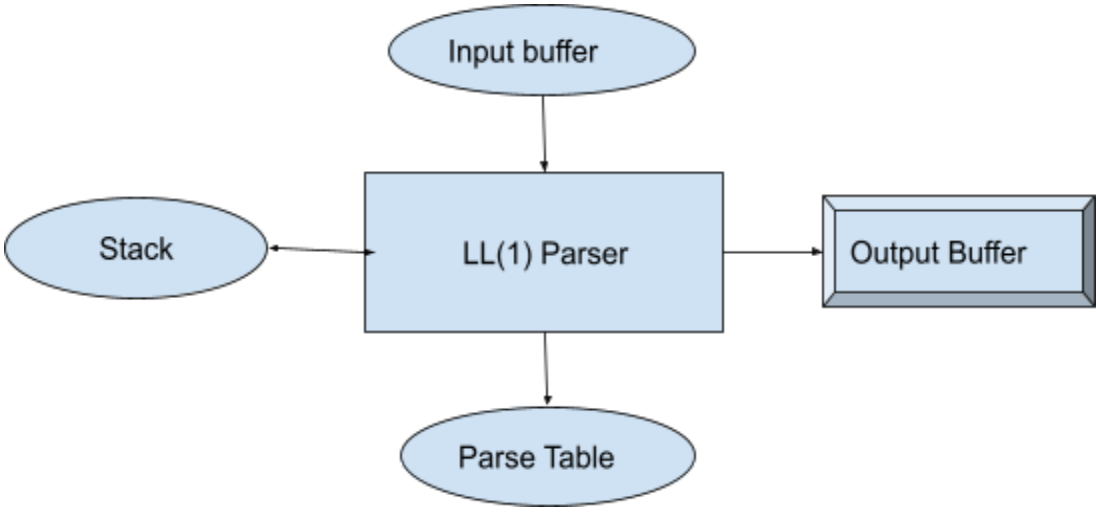
LL(1) Parsing

It is a non descent recursive parser used for formation of parse tree by using CFG.The charatersick of LL(1) Parser are as follows-:

1)It reads from Left to Right

2)It is top to down parser

3)It reads one token at a time from lexical analyser output



To Explain the working of above diagram let's take an example:let the steam of token me **a+b**

1)The tokens will be stored in input buffer 

2)From input each token is pushed in stack and popped out when operation is done
3)A parsing table if formed with the help of First and Follow method
4)All these create an output thus a parse tree

Construction of LL(1) Parser

There are five factor responsible for construction of LL(1) Parser
1)Elimination of Left Recursion
2)Elimination of Left Factoring
3)Find First and Follow
4)Construction of Parsing Table
5)To check whether String matches ie accepted by Parse tree

For example let the CFG of the language be-:

if_statement ⟶ if (condition) { statements }
condition ⟶ expr op expr
op ⟶ < | > | == | >= | <= | !=
statements ⟶ identifier <- exprstatements| expr ops expr statements |print(identifier) statements| λ
ops ⟶ + | - | * | *| /
expr ⟶ identifier | numeric
identifier ⟶ . letter identifier| letter | digit| _ | .) identifier| λ
numeric ⟶ integer | double

Step 1-First Check the Cfg shouldn't have any left recursion or factoring ,the above CFG doesn't have any

Step2-Find the first and follow of the CFG

First Function

First(if-statement)={if}
First(condition)=First(expr)={identifier,numeric}
First(op)={<,>,=,>,!}
First(statement)={identifier,print,$}
First(ops)={+,-,*,/}
First(expr)=First(identifier)={.,(,$}
First(identifier)={.,(,$}
First(numeric)={integer,double}

Follow Function

Follow(if-statement)={$}
Follow(condition)={(}
Follow(op)=First(expr)={identifier,numeric}
Follow(statements)={ } }
Follow of ops=First(expr)={identifier,numeric}
Follow(expr)=Follow(condition)UFollow(statements)={ (, } }
Follow(identifier)={(}UFollow(expr)U{<-}={ (, } , <- }
Follow(numeric)=Follow(expr)=Follow(condition)UFollow(statements)={ (, } }

Step3-> Create Parse table with help of First and Follow
Step4->Do the Stack implementation

Stack Implementation

String required->if(5>=3){
c<-a+b
print(c)
}

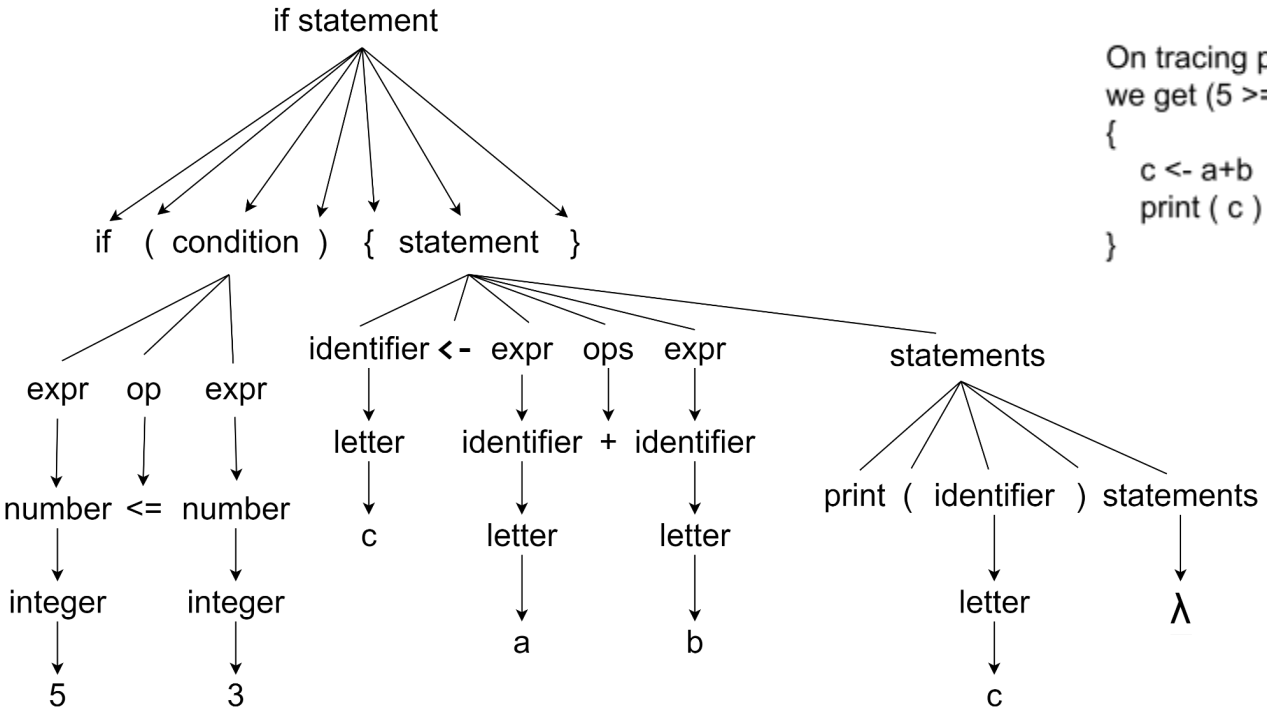
Stack	Implementation	Production rule
if_statements	if(5>=3){c=a+b print(c)}	if_statement → if (condition) { statements}
if (condition) {statements }	if(5>=3){c=a+b print(c)}	Pop if, Pop(
condition) {statements}	5>=3){c=a+b print(c)}	condition →expr op expr
expr op expr){statements }	5>=3){c=a+b print(c)}	expr->numeric
numeric op expr){statements }.	5>=3){c=a+b print(c)}	numeric->integer
integer op expr) {statements }.	5>=3){c=a+b print(c)}	integer->5
5 op expr) {statements }.	5>=3){c=a+b print(c)}	Pop 5
op expr) {statements }.	>=3){c=a+b print(c)}	Op- > >=
>= expr) {statements }.	>=3){c=a+b print(c)}	Pop >=

expr) {statements }	3){c=a+b print(c)}	expr->numeric
numeric){statements}	3){c=a+b print(c)}	numeric->integer
integer){statements}	3){c<-a+b print(c)}	integer->3
3){statements}	3){c<-a+b print(c)}	Pop 3
){statements}	){c<-a+b print(c)}	Pop(
{statements}	{c<-a+b print(c)}	Pop{
statements}	c<-a+b print(c)}	statements → identifier <- expr ops

		expr statements
identifier <- expr ops expr statements}	c<-a+b print(c)}	identifier->letter
letter <- expr ops expr statements}	c<-a+b print(c)}	letter->c
c <- expr ops expr statements}	c<-a+b print(c)}	Pop c
<- expr ops expr statements}	<-a+b print(c)}	Pop <-
expr ops expr statements}	a+b print(c)}	expr->identifier
identifier ops expr statements}	a+b print(c)}	identifier->letter
letter ops expr statements}	a+b print(c)}	letter->a
a ops expr statements}	a+b print(c)}	Pop a
ops expr statements}	+b print(c)}	ops->+
+expr statements}	+b print(c)}	Pop +
expr statements}	b print(c)}	expr->identifier
identifier statements}	b print(c)}	identifier->letter
letter statements}	b print(c)}	letter->b
b statements}	b print(c)}	Pop b
statements}	print(c)}	statements->print(<identifier>) statements
print(identifier) statements}	print(c)}	Pop print
(identifier) statements}	(c)}	Pop (
identifier) statements}	c)}	identifier->letter
Letter)statements}	c)}	letter->c
c)statements}	c)}	Pop c

)statements}	)}	Pop (
statements}	}	statements->$
$}	}	Pop }
$$		String Accepted

Step5->Create a Parse tree to find whether string is readable



The following list of features have been included in the parser

1)Syntax checking for Arithmetic, Logical and Relational Expression

2)If, if….else and nested if statements

3)Validation of Unary Operators

4)Validation for loops

For loop

While loop

5)Unbalanced Parentheses

Semantic Analysis

The semantics of a language provide meaning to its constructs, like tokens and syntax structure.
Semantics help to interpret their types, and their relations with each other.
Semantic analysis judges whether the syntax structure constructed in the source program derives any meaning or not

The semantics of a language provide meaning to its constructs, like tokens and syntax structure. Semantics help interpret symbols, their types, and their relations with each other. Semantic analysis judges whether the syntax structure constructed in the source program derives any meaning or not

CFG+semantics rule=Syntax Directed Derivation
For Example:

int a-> ”value”

The above declaration won’t show error in the lexical analyzer and syntax analyzer, as it is lexically and syntactically correct, but it should give an error as a string is assigned to int

Semantics Analysis is the task to check the declaration in which statement and structure and data type are supposed to be used.

It includes the following problems:
Type checking-Data types and the declaration and assignment
Scope-Resolution-Need to find the correct scope in which variable to use
Flow-Control->Control structure must be used properly

Symbol Table
It contains fields:

1) Name :identifier name
2) Token: token is constant or identifier

3) Type :type of identifier whether int,float,char
4) Scope: Scope could be global or function
5) Scope-id: Unique value is given to each block of code
6) Function-id :Unique for each functions

Syntax of the source language R

An R program is made up of three things: Variables, Comments, and Keywords.

- R does not have a command for declaring a variable. A variable is created the moment you first assign a value to it.
- In many other programming languages, we commonly use = as an assignment operator. In R, we can use both = and <-.
- However, <- is preferred in most cases because the = operator can be forbidden in some context in R.
- The operator <- can be used anywhere, whereas the operator = is only allowed at the top level
- While assigning a value to a variable, we use the <- sign.
  To output (or print) the variable value, just type the variable name.

  Ex:
  ```
  name <- "John"
  age <- 40
  name
  age
  ```

  Out:
  ```
  [1] "John"
  [1] 40
  ```

- To output text in R, use single or double quotes and to output numbers, just type the number.

  Ex:
  ```
  "Hello World!"
  55
  ```

  Out:
  ```
  [1] "Hello World!"
  [1] 55
  ```

- To output integers, we need to represent them by the letter L.

  Ex:2L,40L,1000L.

- Comments start with a #. When executing the R-code, R will ignore anything that starts with #.

  Ex:
  ```
  # This is a comment
  #More than a line
  "Hello World!"

  Out:  [1] "Hello World!"
  ```

- A keyword can't be used as a variable name, function name, etc.
  Ex: Keywords in R are if, else, while, for, function, break, TRUE, FALSE, etc.

- To call a function, use the function name followed by parenthesis, like my_function().

  Ex:
  ```
  my_function() <- function(){
  print("Yay!!")
  }
  my_function()

  Out:  [1] Yay!!
  ```

- Variables that are created outside of a function are known as global variables. Global variables can be used by everyone, both inside of functions and outside.

  Ex:
  ```
  txt <-"hot"
  my_function <-function(){
  paste("tea is",txt)
  }
  my_function()
  ```

Out: [1] tea is hot

Syntax of the target language C++

- C++ language defines several headers, which contain information that is either necessary or useful to your program. Header files add functionality to C++ programs.
- #include <iostream> is a header file library that lets us work with input and output objects
- The header using namespace std means that we can use names for objects and variables from the standard library.

Ex:
```
#include <iostream>
using namespace std;

int main()
{
    cout << "Hello World!";
    return 0;
}
```

- C++ ignores whitespaces.
- int main() is the main function where program execution begins. Any code inside curly brackets {} will be executed.
- 'cout' is an object used together with the insertion operator (<<) to output/print text.

  Ex:
```
cout << "Hello World!";
```
- Every C++ statement ends with a semicolon ';'

  Ex:
```
x = y;
y = y + 1;
```

```
add(x, y);
```

- Single-line comments begin with // and stop at the end of the line.
  Ex:

```
cout << "Hello World!"; //prints the statement
return 0;
```

- return 0; terminates main( ) function and makes it to return value 0.
- A block is a set of logically connected statements that are surrounded by opening and closing braces.

Ex:

```
{
        cout << "Hello World!";
        return 0;
}
```

- To create a variable, you must specify the type and assign it a value
  type variable = value;

Ex:

```
int myNum = 6;
double myFloatNum = 6.01;
char myLetter = 'A';
```

- The cout object is used together with the << operator to display variables.

Ex:

```
int myAge = 35;
cout << "I am " << myAge << " years old.";
```

- We will use ‘cin’ to get user input. ‘cin’ is a predefined variable that reads data from the keyboard with the extraction operator >>.

Ex: int x;
cin >> x;

- Keywords may not be used as constant or variable or any other identifier names.
  Ex: if, else, for, return, case, break, etc