BAsimpl manual

Contents

Introduction………………………………………………………..…………..…………..…………..…………..…………..……………………1

[Introduction 2](#_Toc449724393)

[Executing the code 2](#_Toc449724394)

[Source Code 6](#_Toc449724395)

[Lexical analysis 6](#_Toc449724396)

[Parsing 8](#_Toc449724397)

[Grammar 8](#_Toc449724398)

[Symbol table 12](#_Toc449724399)

[Interpreter 12](#_Toc449724400)

[Runtime/Virtual Machine 12](#_Toc449724401)

[Interpreter 15](#_Toc449724402)

[Sample Programs 16](#_Toc449724403)

[Factorial\_Iterative 16](#_Toc449724404)

[Factorial\_Recursive 16](#_Toc449724405)

[Hemachandra\_Fibonacci 16](#_Toc449724406)

[Sum\_Of\_N 17](#_Toc449724407)

[Max\_Min\_Avg\_Sum 17](#_Toc449724408)

# Introduction

BaSimPL Language has been influenced by TinyADA, Python and C languages, learning from all three and imbibing their best and most useful features. The concept of stacks implemented in our language was influenced by that in python, the separation of definition and declaration from TinyADA and the bacis programming construct from C. We have implemented it using Python language. The Compiler and Interpreter have been organized into two separate steps, as they are supposed to be, and the sub-parts of a Compiler – the Parser and the Lexer, have also been divided into two distinct steps to provide a neat flow. The grammar we built has been listed later on in the manual.

Further in-depth details of the BaSimPL language can be found in this manual, or on our Wiki Page:

<https://github.com/harsha-kadekar/BaSimPL/wiki>

The YouTube link to our presentation video is:

<https://www.youtube.com/watch?v=HJjdUxsjX6w&feature=youtu.be>

Github repository link –

<https://github.com/harsha-kadekar/BaSimPL>

# Executing the code

To run our code, you need to execute **BaSimPL\_Execute.py .** It can be passed multiple arguments

Usuage: **python BaSimPL\_Execute.py [inputfile=filename.smpl] [debug=1] [outputfile=filename.bspl]**

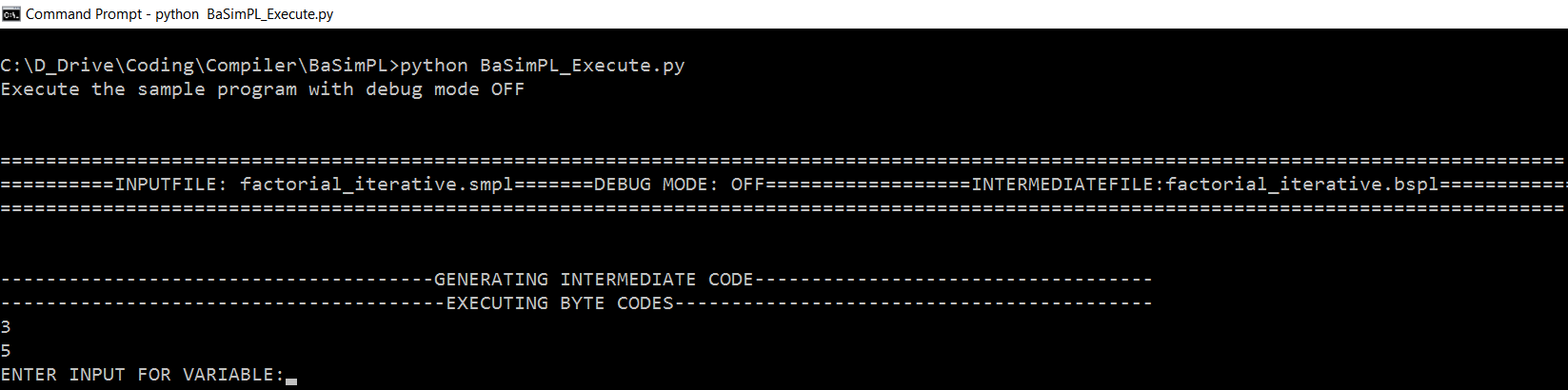
All our source program are having file type as .smpl and all our intermediate byte code files have type .bspl

These files can be opened in the notepad.

All the arguments present inside square brackets are optional. So you can just give

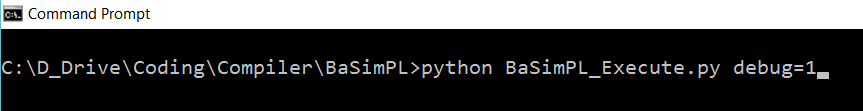
**python BaSimPL\_Execute.py**

In this case it will just run the test cases that is, it will execute multiple sample programs without debug mode.

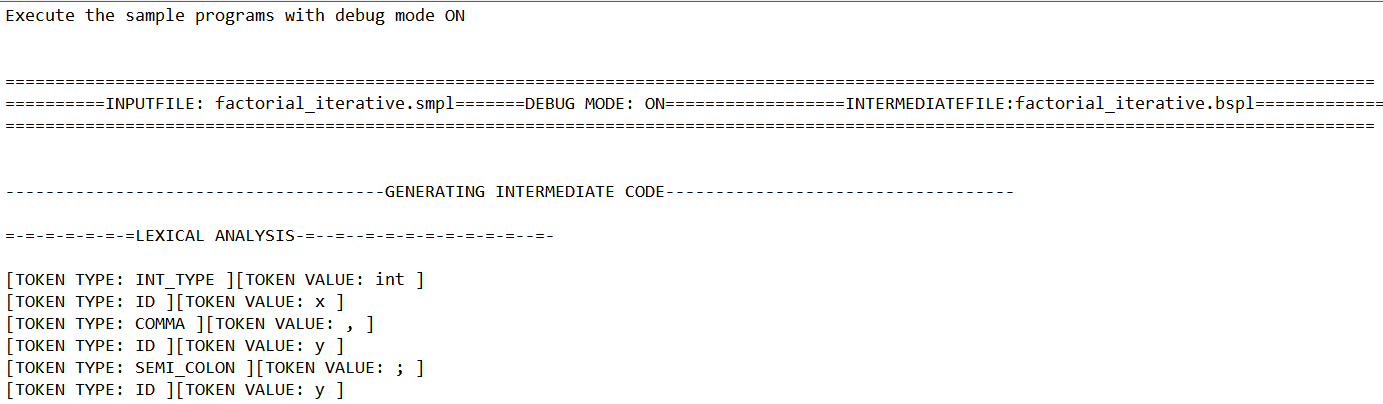


**Python BaSimPL\_Execute.py debug=1**

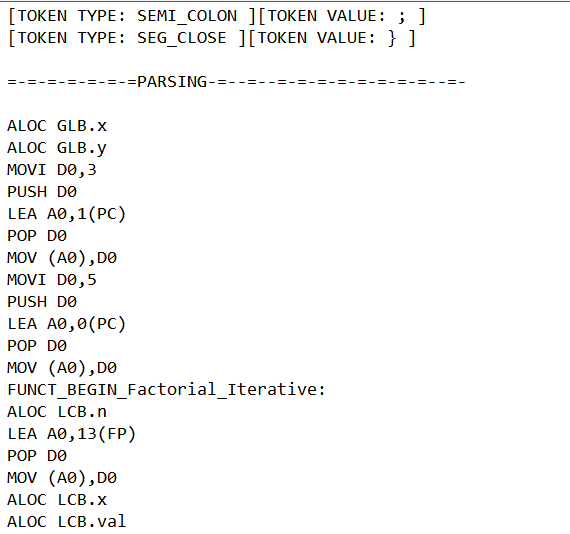
In this case, all the tests will be performed with debug mode ON. So you will get lot of information like intermediate code generated, tokens generated, runtime informations.

****

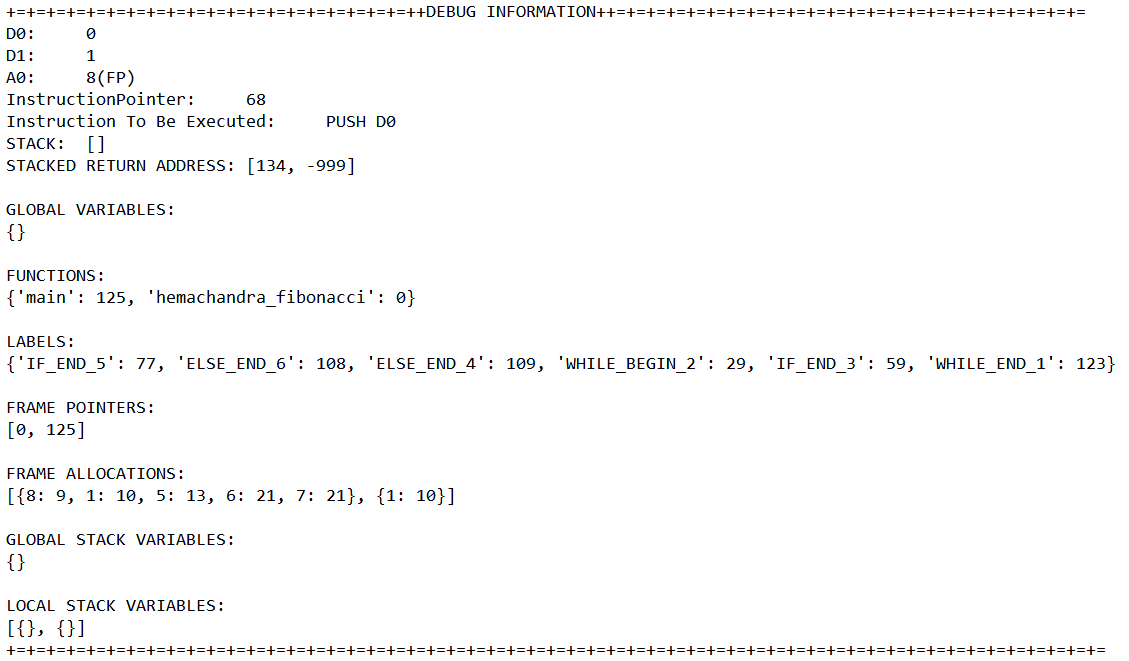
Tokens generated –



Intermediate Code generated



Runtime information –

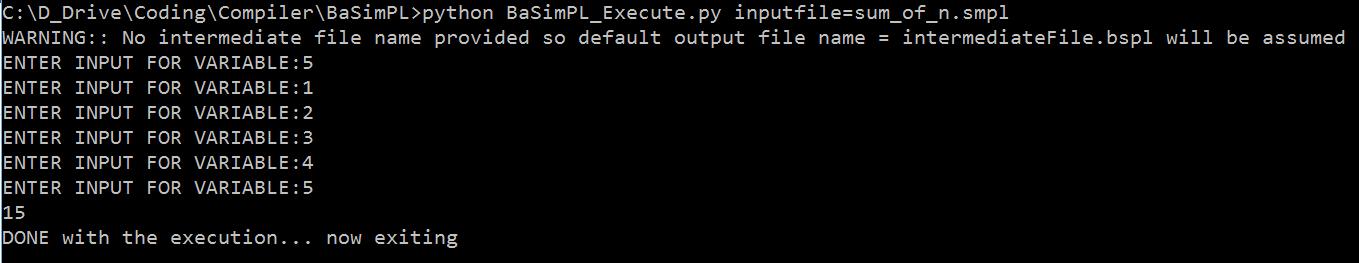


**python BaSimPL\_Execute.py inputfile=sum\_of\_n.smpl**

**python BaSimPL\_Execute.py inputfile=sum\_of\_n.smpl debug=1**

**python BaSimPL\_Execute.py inputfile=sum\_of\_n.smpl debug=1 outputfile=sum\_of\_n.bspl**

In this we are executing a simple program. Suppose you have written a program in a file sum\_of\_n.smpl and you want to execute it then use this command. You can enable debug command by adding debug=1 to the above argument. As outputfile is not provided, it will generate byte code in a file called intermediateFile.bspl. If you want to provide the output file name then use outputfile=example.bspl argument to it.

****

****

****

# Source Code

Following are the important source files.

1. BaSimPL\_Lexxer.py

This has classes and functions related to lexical analysis. It uses regular expressions mainly to identifying tokens

1. BaSimPL\_Parser.py
2. This has classes and functions for syntax and sematic analysis. In this tokens are parsed to generate the intermediate code.
3. BaSimPL\_Execute.py
4. This is the entry point of the program. Based on the argument it will call the interpreter. It also has test functions to test different components.
5. BaSimPL\_Compiler.py

This is our compiler. It uses the objects of lexxer and parser. This finally generates the intermediate byte code file.

1. BaSimPL\_DummyRuntime.py

This is our virtual machine or runtime. It has the structure of our machine. It will execute instructions each line by line.

1. BaSimPL\_Interpreter.py

This utilizes the compiler and runtime to first generate the byte code and finally to execute the generated bytecode.

1. Entry.py

This represents each entry in the symbol table.

1. SymbolTable.py

This is our symbol table which has entries based on the program. It is used by the parser while generating the intermediate code.

# Lexical analysis

This is the list of tokens and its regular expressions –

INT = 'INT'

IDENTIFIER = 'ID'

ADD\_OPERATOR = 'ADD'

SUB\_OPERATOR = 'SUB'

MUL\_OPERATOR = 'MUL'

DIV\_OPERATOR = 'DIV'

EQUALS\_OPERATOR = 'EQUALS'

ASSIGNMENT\_OPERATOR = 'ASSIGN'

GREATERTHAN\_OPERATOR = 'GREATER'

LESSERTHAN\_OPERATOR = 'LESSER'

GREATEREQUAL\_OPERATOR = 'GREATEREQUAL'

LESSEREQUAL\_OPERATOR = 'LESSEREQUAL'

NOTEQUAL\_OPERATOR = 'NOTEQUAL'

OPEN\_BRACE = 'OPEN\_BRACE'

CLOSE\_BRACE = 'CLOSE\_BRACE'

LOG\_AND\_OPERATOR = 'AND'

LOG\_OR\_OPERATOR = 'OR'

LOG\_NOT\_OPERATOR = 'NOT'

IF = 'IF'

ELSE = 'ELSE'

WHILE = 'WHILE'

SEG\_OPEN = 'SEG\_OPEN'

SEG\_CLOSE = 'SEG\_CLOSE'

SEMICOLON = 'SEMI\_COLON'

COMMA = 'COMMA'

FUNCTION\_DEFINITION = 'FUNCT\_DEF'

VOID = 'VOID'

INT\_TYPE = 'INT\_TYPE'

EOF = 'EOF'

RETURN = 'FUN\_RETURN'

STACK\_TYPE = 'STACK\_TYPE'

BOOLEAN = 'BOOLEAN'

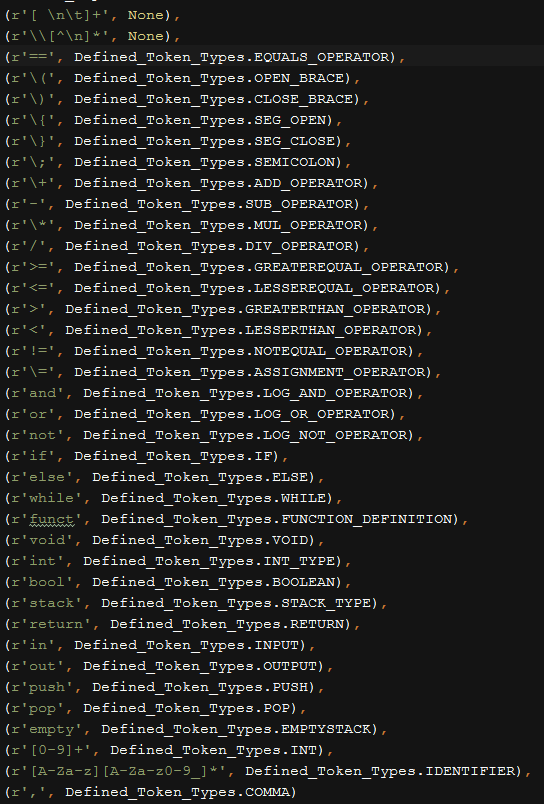
EMPTYSTACK = 'STACK\_EMPTY'

POP = 'STACK\_POP'

PUSH = 'STACK\_PUSH'

INPUT = 'CONSOLE\_INPUT'

OUTPUT = 'CONSOLE\_OUTPUT'



# Parsing

For parsing we are using recursive decent parsing in a top down way.

## Grammar

This is the grammar of our programing language.

program -> topLevelDeclarations\*

topLevelDeclaration -> functionDecl | GlobalDataDecl | globalAssignment | globalPushStatement | globalPopStatement | globalStackEmptyStatement

GlobalDecl -> TYPE\_DECL declList SEMICOLON

DataDecl -> TYPE\_DECL declList SEMI\_COLON

functionDecl -> FUNCT\_DEF funcReturnType ID OPEN\_BRACE PARAMETER\_LIST CLOSE\_BRACE funBlock

funReturnType -> VOID | INT\_TYPE

declList -> ID { COMA ID }

funBlock -> SEG\_OPEN declRegion SEQSTATMENTS SEG\_CLOSE

declRegion -> DataDecl\*

globalAssignment -> ID ASSIGN\_OPERATOR INT SEMI\_COLON

PARAMETER\_LIST -> TYPE\_DECL ID { COMA TYPE\_DECL ID }

TYPE\_DECL -> INT\_TYPE | STACK\_TYPE

SEQSTATEMENTS -> STATEMENT {STATEMENT}

STATEMENT -> simpleSTATEMENT | compoundSTATEMENT

simpleSTATEMENT -> assignmentSTATEMENT | procedureSTATEMENT | returnStatement | pushStatement | popStatement | stackemptyStatement

returnStatement -> FUN\_RETURN ReturnValue SEMI\_COLON

ReturnValue -> ID | INT | NULL

compoundSTATEMENT -> ifSTATEMENT | whileSTATEMENT

assignmentSTATEMENT -> ID ASSIGNT\_OPERATOR expression SEMI\_COLON

ifSTATEMENT -> IF OPEN\_BRACE expression CLOSE\_BRACE SEG\_OPEN SEQSTATEMENTS SEG\_CLOSE {ELSE SEG\_OPEN SEQSTATEMENTS SEG\_CLOSE}

whileSTATEMENT -> WHILE OPEN\_BRACE expression CLOSE\_BRACE SEG\_OPEN SEQSTATEMENTS SEG\_CLOSE

Expression -> bterm [ OR\_OPR bterm]\*

bterm -> notfactor [AND\_OPR notfactor]\*

notfactor -> [NOT] bfactor

bfactor -> INT | ID | relation

relation -> simpleExpression { relationalOperator simpleExpression}

simpleExpression -> term { ADDSUB\_OPERATOR term}

term -> factor { MULDIV\_OPERATOR factor}

factor -> OPEN\_BRACE simpleExpression CLOSE\_BRACE | ID | INT | procedureSTATEMENT

ADDSUB\_OPEATOR -> ADD\_OPERATOR | SUB\_OPERATOR

MULDIV\_OPERATOR -> MUL\_OPERATOR | DIV\_OPERATOR

relationalOperator -> EQUALS | GREATER | LESSER | GREATEREQUAL | LESSEREQUAL | NOTEQUAL

INT -> [0-9]{[0-9]}

ID -> [a-zA-Z]{[a-zA-Z0-9\_]}

OPEN\_BRACE -> "("

CLOSE\_BRACE -> ")"

SEG\_OPEN -> "{"

SEG\_CLOSE -> "}"

IF -> "if"

ELSE -> "else"

WHILE -> "while"

ADD\_OPERATOR -> "+"

SUB\_OPERATOR -> "-"

MUL\_OPERATOR -> "\*"

DIV\_OPERATOR -> "/"

ASSIGNT\_OPERATOR -> "=="

EQUALS -> "="

SEMI\_COLON -> ";"

FUNCT\_DEF -> "funct"

VOID -> "void"

INT\_TYPE -> "int"

COMMA -> ","

FUN\_RETURN -> "return"

pushStatement -> ID PUSH EXPRESSION SEMICOLON

popStatement -> ID POP ID SEMICOLON

emptyStatement -> ID EMPTY ID SEMICOLON

PUSH -> "push"

POP -> "pop"

EMPTY -> "empty"

globalPushStatement -> ID PUSH EXPRESSION SEMICOLON

globalPopStatement -> ID POP ID SEMICOLON

globalEmptyStatement -> ID EMPTY ID SEMICOLON

## Symbol table

Every symbol table has list of entries. Each entry has name, value, type, location, return address, parameter list. Only for functions return address and parameter list will be used. For integer variables type is of INT\_VAR where as for stack variable we have type as STACK\_VAR.

Apart from entries list, symbol table is linked to previous parent symbol table. Example if I am entering a function, then a new symbol table will be created and that symbol table is linked to global symbol table. By this we are handling multilevel scope blocks.

# Interpreter

## Runtime/Virtual Machine

Registers:

D0 - General Purpose Register

D1 - General Purpose Register

A0 - Register that holds address

Instructions

MOVI - Used to load an intermediate value to a register. Example if we want to load integer constant 5 to register D0 then we use MOVI D0, 5 . After executing this instruction D0 will have value 5.

PUSH - Used to put value into the stack. Suppose we want to put the value in register D0 to stack then we use PUSH D0.

POP - Used to take a value out of the stack. POP D0, here whatever value was present at the top of the stack will be put in register D0.

LEA - This will load the address to register A0. Suppose we are trying to assign some value to a variable, then that value will be stored in the address allocated to the variable. LEA will help to load the address location of that variable. LEA A0, n(PC)

MOV - This instruction will help to move the value from one register location another another register. Example MOV D0, D1 - This will move the value in D1 to D0. Similarly if we want to move the value in a memory location that is variable to a register also we use MOV. Suppose we want to move the value of variable n to register D0 then, First we need to get the address of the variable n using LEA that is LEA A0, n(PC). Now we use MOV D0, (A0) to move the value to register D0.

CGT - This instruction will be used to check for greater than condition between two registers. CGT D0, D1 - Here if D0 > D1 then value 1 will be stored in register D0, else 0 will be stored in the register D0.

CNE - This instruction will be used to check for not equal to condition between two registers. CNE D0, D1 - Here if D0 is not equal to D1 then value 1 will be stored in register D0, else 0 will be stored in the register D0

CEQ - This instruction will be used to check for equal to condition between two registers. CEQ D0, D1 - Here if D0 is equal to D1 then value 1 will be stored in register D0, else 0 will be stored in the register D0

CGE - This instruction will be used to check for greater than or equal to condition between two registers. CGE D0, D1 - Here if D0 is greater than or equal to D1 then value 1 will be stored in register D0, else 0 will be stored in the register D0.

CLE - This instruction will be used to check for lesser than or equal to condition between two registers. CLE D0,D1 - Here if D0 is lesser than or equal to D1 then value 1 will be stored in register D0, else 0 will be stored in the register D0.

CLT - This instruction will be used to check for lesser than condition between two registers. CLT D0, D1 - Here if D0 is lesser than D1 then value 1 will be stored in register D0, else 0 will be stored in the register D0.

ADD - This instruction will add the values of two registers. ADD D0,D1 - Here we will be adding two registers values D0 and D1 and result will be stored back to D0.

SUB - This instruction will subtract values of two registers. SUB D0,D1 - Here we will be subtraction D0 from D1 that is D0-D1 and result will be stored back to D0.

MUL - This instruction will multiply values of two registers. MUL D0,D1 - Here we will be multiplying D0 and D1 and product will be stored back to D0.

DIV - This instruction will divide values of two registers. DIV D0,D1 - Here D0 will be divided by D1 and quotient will be stored back to D0.

CALL - this instruction is used for calling/executing a function. CALL func1 - here we are calling function func1

BEQ label - This instruction is used for branching. Here if 0 is stored in register D0, then branching will take place and next instruction will be after label. If 1 or any value other than 0 is stored then no branching will occur as a result, next instruction will be immediately after BEQ instruction.

BNE label - This instruction is used for branching. It is opposite of BEQ. If 0 is stored in register D0 then branching will not take placed so next instruction is the instruction which is immediate to the BNE. If 1 or any value other than 0 is stored in D0, then branch will take place and next instruction is the one after label.

JMP label - This instruction is used for branching without condition. When used next instruction will be the instruction following label and not the instruction which is next to JMP instruction.

OR - this will do a logical or of two values. That is D0 || D1. If D0 0 and D1 is 100 then after OR D0, D1 instruction D0 will have 1 as any value other than 0 is considered as logical 1.

AND - this will do a logical and of two values. That is D0 && D1. If D0 0 and D1 is 100 then after AND D0, D1 instruction D0 will have 0. Similarly if D0 100 and D1 is 1 then after operation D1 will have 1 as any value other than 0 is considered as 1.

RET - this will return from function. Before returning it will clear up activation records.

ALOCSTACK - This will allocate the stack either in global namespace or in local namespace.

PUSHST - This will push the element into a stack variable. PUSHST (A0), D0 where D0 has the values to be pushed into the stack. A0 has the address of the stack.

POPST - this will pop the element from the stack variable. POPST (A0), D0 where A0 will have the address of the stack variable. and popped value will reside in D0

EMPTYST - this will check if a stack variable is empty. EMPTYST (A0), D0 where A0 will have the address of the stack variable. and if stack is empty then 1 will be stored in D0 else 0 will be stored in D0.

## Interpreter

It is a two phase interpreter. In the first phase it will identify the labels/functions and assigns the addresses to those labels. In the second phase it will execute the actual bytecode. It has a machine stack where we can push or pop numbers during operation. For all the global integer variables, allocation will be done in globalVariables dictionary. Similarly for all the global stack variables, allocation will done in a globalStack dictionary. There is frames stack where for each function call new activation record will be created and pushed into the stack. In that record local integer variables will be allocated. Similarly for each functions we are creating a localStackVariables record and pushed into a separate stack. For every function call a function pointer/environment pointer pointing to the start of the function will be pushed into a framepointer stack. Return address of the caller function will be pushed into a separate ReturnAddress stack. On encountering ret opcode, we are popping all these things and returning back to the normal caller function. We also have an instruction pointer pointing to the current instruction to be executed.

There is a debug flag which will give you detailed information of these things while executing the program.

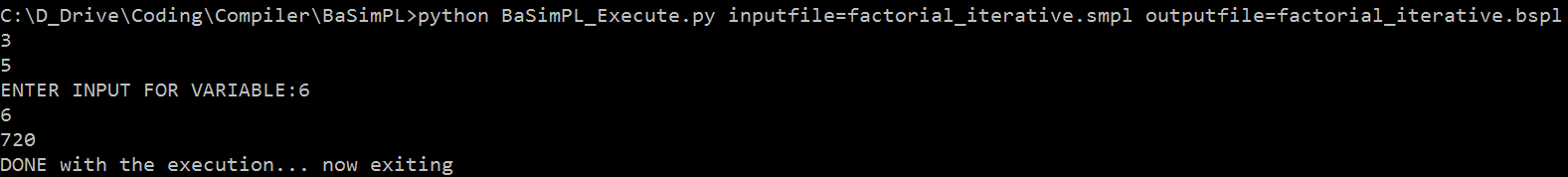
# Sample Programs

## Factorial\_Iterative

This program will calculate the factorial of the given number. Program is present in the file factorial\_iterative.smpl . This can be opened in the notepad. Corresponding bytecode can be found in the file factorial\_iterative.bspl . To execute this program use this command in the command prompt:

**python BaSimPL\_Execute.py inputfile=factorial\_iterative.smpl outputfile=factorial\_iterative.bspl**

It will first print 3 followed by 5. Both are global variables. Then it will ask input from the user, value of n whose factorial needs to be found out. Given input will be printed again. It will then print factorial of the value.



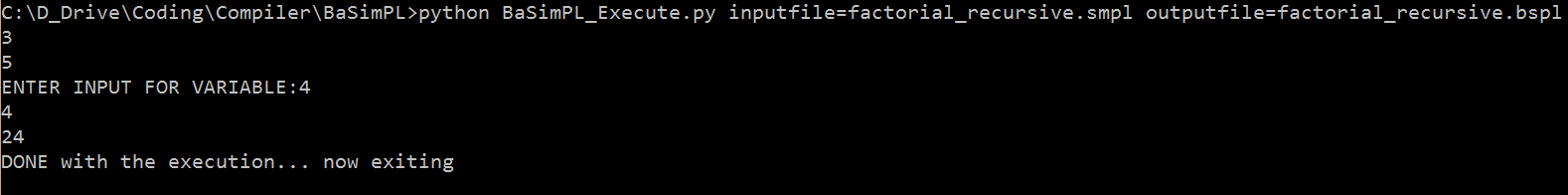
In this program we have demonstrated – global integer variable, local integer variable, Boolean expression, while loop, iterative function execution, parameterized functions, input and output to console.

## Factorial\_Recursive

This program will calculate the factorial of the given number in a recursive way. Program is present in the file factorial\_recursive.smpl . This can be opened in the notepad. Corresponding bytecode can be found in the file factorial\_recursive.bspl . To execute this program use this command in the command prompt:

**python BaSimPL\_Execute.py inputfile=factorial\_recursive.smpl outputfile=factorial\_recursive.bspl**

It will first print 3 followed by 5. Both are global variables. Then it will ask input from the user, value of n whose factorial needs to be found out. Given input will be printed again. It will then print factorial of the value.



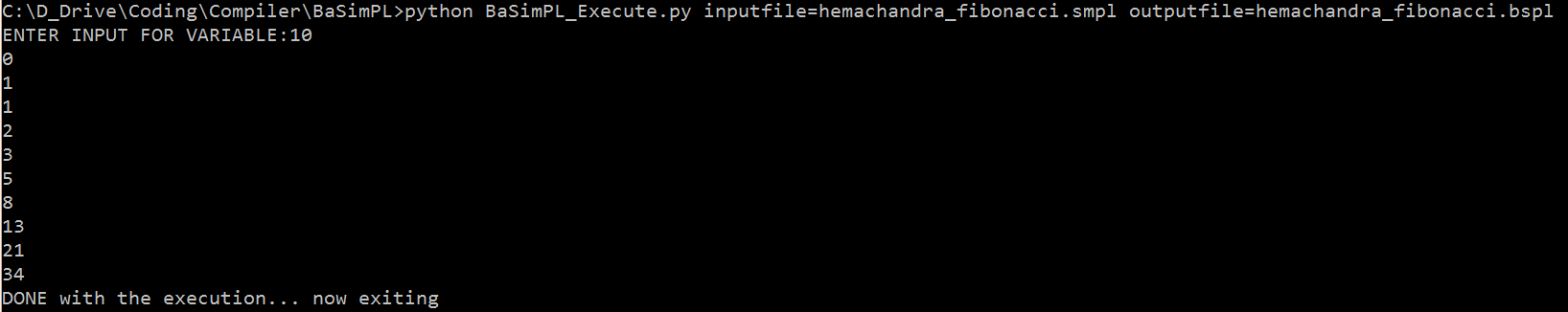
In this program we have demonstrated – global integer variable, local integer variable, Boolean expression, if else constructs, recursive function execution, input and output to console.

## Hemachandra\_Fibonacci

This program will calculate the Fibonacci numbers. Program is present in hemachandra\_fibonacci.smpl. This can be opened in the notepad. Corresponding bytecode can be found in the file hemachandra\_fibonacci.bspl. To execute this program use this command in the command prompt:

**python BaSimPL\_Execute.py inputfile=hemachandra\_fibonacci.smpl outputfile=hemachandra\_fibonacci.bspl**

It will first ask for an integer variable which determines how many Fibonacci numbers needs to be generated. Based on the given input will generate the Fibonacci numbers.



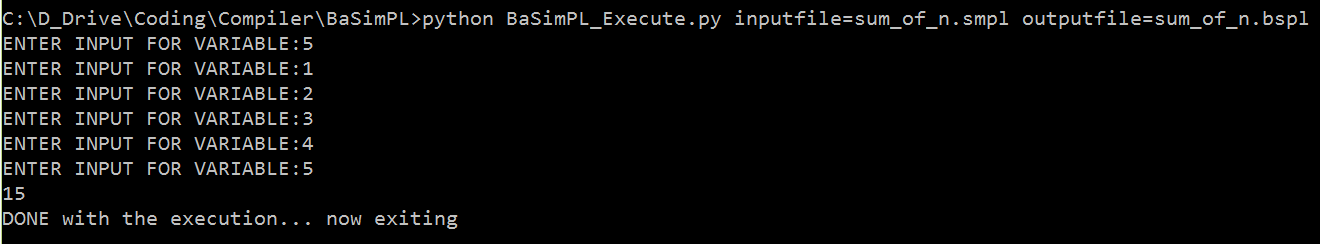
In this we have demonstrated function calls, local variables, if else constructs, in and out, assignment statements, arithmetic expression evaluation.

## Sum\_Of\_N

This program will calculate the sum of n numbers. It uses stack variable. Program can be found in file sum\_of\_n.smpl . Corresponding bytecode of the program can be found in file sum\_of\_n.bspl . To execute this program use this command

**python BaSimPL\_Execute.py inputfile=sum\_of\_n.smpl outputfile=sum\_of\_n.bspl**

It will first ask an input. This will be your n. Based on the given input it will ask that many input variables. Finally it will give you the sum of those n numbers.



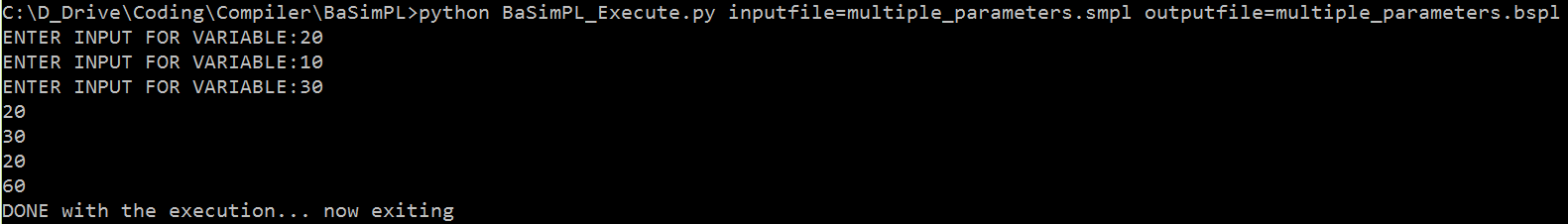
In this we have demonstrated stack variable, local variables, function calls, stack operations, while operations. Relational operators.

## Max\_Min\_Avg\_Sum

This program finds the maximum, minimum, average and sum of the given 3 numbers. Program can be found in the file multiple\_parameters.smpl . Corresponding bytecode can be found in file multiple\_parameters.bspl . To execute the program we can use following command –

**python BaSimPL\_Execute.py inputfile=multiple\_parameters.smpl outputfile=multiple\_parameters.bspl**

It will first ask three numbers. Then it will show average of 3 numbers, maximum of 3 numbers, minimum of 3 numbers and finally sum of 3 numbers.



In this we have demonstrated function calls. 3 level function calls Main -> Average -> Summation. Multiple parameters to functions: 3 parameters. Logical operations, if else constructs. Arithmetic operations.