The uC Language Specification

This document was generated on 2016-12-04 at 21:52.

Contents

Lexical Structure		2
Whitespace		2
Comments		2
Keywords		3
Identifiers		3
Literals		4
Integer Literals		4
Floating-Point Literals		4
String Literals		4
Boolean Literals		5
Null Literal		5
Operators		5
Delimiters		5
Deminiers		_
Program Structure		5
Type Declarations		6
Function Declarations		6
Types		7
Primitive Types		7
User-Defined Types		7
Arrays		7
Type Compatibility		8
Functions		8
Built-in Functions		8
User-Defined Functions		9
Statements		9
Blocks		9
Conditionals		
Loops		
Return Statements		
Expression Statements		U
Expressions	1	በ
Simple Expressions		
Function Calls		
Allocation Expressions		
Field Access		
Array Indexing		
Unary Operations		
Binary Operations		
Arithmetic Operations		
Logical Operations		
Comparisons		
Equality Tests		4

Assignment	14
Array Operations	14
Associativity and Precedence	15
Memory Management	15
Default Initialization	15

Lexical Structure

Excepting comments and string literals, programs must be written in the subset of ASCII consisting of the following 91 characters:

- the 62 alphanumeric characters: 0-9, a-z, A-Z
- the 23 symbols: + * / % | & ! < > = () [] { } , . ; " \ _
- the 6 whitespace characters: space, horizontal tab, carriage return, new line, vertical tab, form feed

A source program may consist of whitespace, comments, and tokens.

```
ProgramText:
    TextElements

TextElements:
    TextElement
    TextElements TextElement

TextElement:
    Whitespace
    Comment
    Token
```

Tokens consist of keywords, identifiers, literals, operators, and delimiters.

```
Token:
Keyword
Identifier
Literal
Operator
Delimiter
```

Whitespace

Tokens may be separated by any of the 6 whitespace characters.

```
Whitespace:
    space
    horizontal tab
    carriage return
    vertical tab
    form feed
    NewLine

NewLine: the new-line character
```

Comments

There are two kinds of comments, a delimited comment and an end-of-line comment.

```
Comment:
    DelimitedComment
    EndOfLineComment

DelimitedComment:
```

```
/ * CommentChars * /
    / * * /
CommentChars:
    * CommentCharsNoSlash
    CommentCharNoStar CommentChars
    CommentCharNoStar
CommentCharNoStar: any ASCII character except *
CommentCharsNoSlash:
    CommentCharNoSlash CommentChars
    CommentCharNoSlash
CommentCharNoSlash: any ASCII character except /
EndOfLineComment:
    / / CharactersNoNewLine NewLine
    / / NewLine
CharactersNoNewLine:
    CharacterNoNewLine
    CharactersNoNewLine CharacterNoNewLine
CharacterNoNewLine: any ASCII character except the new line
```

Keywords

The following character sequences are reserved keywords and may not be used as identifiers.

```
Keyword: one of
    if else while struct break continue return new
```

Identifiers

Identifiers consist of a sequence of characters begininning with a lower-case letter (a-z) or upper-case letter (A-z). The remaining characters may consist of underscores, lower-case letters, upper-case letters, and digits (0-9).

```
Identifier: except Keyword, BooleanLiteral, and NullLiteral
   IdentifierStartCharacter
   IdentifierStartCharacter IdentifierCharacters
IdentifierStartCharacter:
   LowerCaseLetter
   UpperCaseLetter
IdentifierCharacters:
   IdentifierCharacter
   IdentifierCharacters IdentifierCharacter
IdentifierCharacter:
   IdentifierStartCharacter
   Digit
LowerCaseLetter: one of
   abcdefghijklmnopqrstuvwxyz
UpperCaseLetter: one of
   ABCDEFGHIJKLMNOPQRSTUVWXYZ
```

```
Digit: one of 0 1 2 3 4 5 6 7 8 9
```

Literals

Literals include integer, floating-point, string, boolean literals, and the null literal.

```
Literal:
    IntegerLiteral
    FloatingLiteral
    StringLiteral
    BooleanLiteral
    NullLiteral
```

Integer Literals

Integer literals must be expressed in decimal format. They consist of a sequence of one or more decimal digits, followed by an optional lower- or upper-case L. Integer literals with an 1 or L suffix are of type long, while those without are of type int.

```
IntegerLiteral:
    Digits IntegerSuffix
    Digits

Digits:
    Digit
    Digits Digit

IntegerSuffix: one of
    1 L
```

Floating-Point Literals

A floating-point literal consists of a digit sequence consisting of a *significand* and optional *exponent*. The significand is a sequence of digits containing a single period (.). If an exponent is provided, the period (.) may be elided from the significand. The exponent consists of the character e followed by an optional sign and a non-empty digit sequence. Floating-point literals are of type float.

```
FloatingLiteral:
    SignificandWithPeriod
    SignificandWithPeriod Exponent
    Digits Exponent

SignificandWithPeriod:
    Digits .
    Digits
    Digits
    Digits

Exponent:
    e Sign Digits
    e Digits

Sign: one of
    + -
```

String Literals

A string literal consists of a sequence of characters surrounded by double quotes ("). Any ASCII character except for the new-line character may appear in a string literal. The backslash (\) and double-quote (") characters may only appear as part of an escape sequence. A string literal has type string.

```
StringLiteral:
    "StringCharacters "
    ""

StringCharacter:
    StringCharacter
    StringCharacter
    StringCharacter
StringCharacter:
    UnescapedCharacter
    EscapedCharacter

UnescapedCharacter:
    any ASCII character except \, ", or new line

EscapedCharacter: any of the escape sequences below
```

The following escape sequences represent special characters:

Escape Sequence	Character
\"	double quote
\\	backslash
\a	audible bell
\b	backspace
\n	new line
\t	horizontal tab
\f	form feed
\r	carriage return

Boolean Literals

The character sequences true and false represent boolean literals, which have type bool.

```
BooleanLiteral: one of true false
```

Null Literal

The character sequence null represents a null literal, which has indeterminate reference type.

```
NullLiteral: null
```

Operators

The following 18 characters and character sequences represent operators.

```
Operator: one of + - * / % || \&\& ! <> <=>==!==++-- <<>>
```

Delimiters

The following characters are delimiters.

```
Delimiter: one of
     ( ) [ ] { } , . ;
```

Program Structure

A uC program consists of a single source file, conventionally with extension .uc. A program consists of a sequence of type and function declarations.

```
Program:
    Declarations

Declarations:
    Declarations Declaration
    Empty

Declaration:
    FunctionDeclaration
    TypeDeclaration

Empty: empty

A program must have exactly one function with the following declaration:

void main(string[] args)(...) {
    ...
}
```

The main () function is the entry point of a uC program.

Type Declarations

A type declaration consists of the struct keyword, followed by the name of the type, followed by a comma-separated list of field declarations, terminated by a semicolon.

```
TypeDeclaration:
    struct Identifier ( FieldDeclarationsOpt );
FieldDeclarationsOpt:
    VariablesOpt

VariablesOpt:
    Variables
    Empty

Variables:
    Variable
    Variable
    Type Identifier
```

As an example, the following declares a type containing an integer and a floating-point field:

```
struct foo(int x, float y);
```

It is illegal for multiple fields to be given the same name. It is illegal for a type declaration to declare a type with the same name as an existing built-in or user-defined type.

A type can be used anywhere in the source program, including prior to its declaration.

Function Declarations

A function declaration consists of a return type, followed by the name of the function, a parameter list, a list of local-variable declarations, and a body consisting of a block of statements.

```
FunctionDeclaration:
    Type Identifier ( ParametersOpt ) ( VariablesOpt ) Block
ParametersOpt:
    VariablesOpt
```

As an example, the following declares a function that takes an integer argument, uses a floating-point local variable, and returns a string:

```
string bar(int i)(float f) {
   ...
}
```

It is illegal for a parameter or local-variable name to be used for another parameter or local variable, or for a local variable to have the same name as parameter. It is illegal for a function declaration to declare a function with the same name as an existing built-in or user-defined function. It is *not* illegal for a function to have the same name as a built-in or user-defined type.

A function can be used anywhere in the source program, including prior to its declaration.

Types

The types in a uC program consist of primitive types, user-defined types, and array types.

```
Type:
Identifier
ArrayType
```

Primitive Types

The following primitive types are available to uC programs:

- int: a 32-bit integer, representing values in the range $[-2^{31}, 2^{31} 1]$.
- long: a 64-bit integer, representing values in the range $[-2^{63}, 2^{63} 1]$.
- float: an IEEE754-compliant, double-precision floating-point number.
- boolean: a truth value, either true or false.
- string: a sequence of zero or more characters.
- void: denotes the return type of a function that does not return a value. It is illegal to use void in any other context.
- null: the type of the null literal. This type is syntactically prevented from being used anywhere other than the null literal.

Primitive types have *value semantics*, meaning that they are stored directly the variables, parameters, fields, and array elements. It is illegal to use the new keyword with a primitive type.

User-Defined Types

User-defined types are those introduced by Type Declarations. Such types have *reference semantics*, meaning that a variable, parameter, field, or array element of user-defined type is either an indirect reference (i.e. a pointer) to an object of the given type or contains the special null reference.

Arrays

An array is a homogeneous, sequential collection of values. An array type is denoted by an element type followed by an empty pair of square brackets.

```
ArrayType:
    Type [ ]
```

The element type may be any primitive, user-defined type, or array type, excluding the void and null types. Arrays have reference semantics, so a variable, parameter, field, or array element of array type holds either an indirect reference to an array object or the special null reference.

An array has a length field, which contains the number of elements in the array.

Type Compatibility

The following implicit conversions are allowed:

- conversion of a value of type int to long or float type
- conversion of a value of type long to type float
- conversion of the null literal to any reference type (user-defined and array types)

Except for the conversions above, the type of a value must exactly match the type expected by the context om which the value is used. Specifically:

- the type of a return value must match or be implicitly covertible to the function's return type
- the type of an argument passed to a function call must match or be implicitly convertible to the type of the corresponding function parameter
- the type of the value on the right-hand side of an assignment must match or be implicitly convertible to the type of the object on the left-hand side

Functions

The functions in a uC program consist of built-in functions, which are available to all uC uC programs, and user-defined functions.

Built-in Functions

The following functions convert between numeric types:

- long int_to_long(int)
- float int_to_float(float)
- int long_to_int(long)
- float long_to_float(long)
- int float_to_int(float)
- long float_to_long(float)

The following functions convert between strings and other primitive types:

- string int_to_string(int)
- string long_to_string(long)
- string float_to_string(float)
- string boolean_to_string(boolean)
- int string_to_int(string)
- long string_to_long(string)
- float string_to_float(string)
- boolean string_to_boolean(string)

It is illegal to pass a string that does not consist of a valid representation of the input type to the functions that convert to string.

The following functions are defined on strings:

- int length(string): returns the length of the input string.
- string substr(string, int, int): produces a string that contains a subsequence of the input string. The second argument is the start position, which must be in the range [0,length(string)-1]. The third argument is the length of the subsequence, which must be non-negative. The length is truncated if the input string contains insufficient characters.

- int ordinal(string): returns the ASCII value of the given single-character string. If the given string does not consist of exactly one character, -1 is returned.
- string character(int): returns a string containing the character with the given ASCII value. If the given argument is not in the range [1,127], an empty string is returned.

The following numerical functions are defined:

- float pow(float, float): computes the first argument raised to the power of the second.
- float sqrt (float): computes the square root of the argument. The argument must be non-negative.
- float ceil (float): computes a floating-point representation of the ceiling of the argument.
- float floor (float): computes a floating-point representation of the floor of the argument.

The following printing functions are defined:

- void print (string): prints the given string to standard output, without a trailing newline.
- void println(string): prints the given string to standard output, with a trailing newline.

User-Defined Functions

User-defined functions are introduced by Function Declarations. A user-defined function consists of a return type, a name, parameters, local-variable declarations, and a block of statements constituting the body. Calling a user-defined function binds the values of the argument expressions to the parameter names in a fresh environment and executes the body in the subsequent environment. Functions are lexically scoped and do not have access to the caller's environment.

Within the body of a function, it is illegal to use a variable before it has been assigned a value. It is illegal for control to reach the end of a function whose return type is not void. A function that returns void may only have bare return statements without an expression. A function that returns any other type must have return statements with an expression whose type matches or is implicitly convertible to the return type.

Statements

Statements in uC include conditionals, loops, control statements (break, continue, and return), and expression statements.

```
Statement:
    IfStatement
    WhileStatement
    BreakStatement
    ContinueStatement
    ReturnStatement
    ExpressionStatement
```

Blocks

A block is not itself a statement, but it consists of a sequence of zero or more statements, enclosed by curly braces. A block executes each of its statements in order. It terminates normally when control reaches the end of its last statement. A control statement within a block may cause *abrupt termination*, where execution of the block is immediately terminated.

```
Block:
    { StatementsOpt }

StatementsOpt:
    Statements
    Empty

Statements:
    Statement
    Statemen
```

Conditionals

A conditional consists of the keyword if, a parenthesized expression, a block, and optionally, an else keyword followed by a another block.

```
IfStatement:
   if (Expression ) Block
   if (Expression ) Block else Block
```

The expression constitutes the test of the condtional, and its type must be boolean. If the expression evaluates to true, the block immediately following the test is executed. Otherwise, the second block is executed, if present.

Loops

A loop consists of the while keyword, a parenthesized expression, and a block.

```
WhileStatement: while (Expression ) Block
```

The expression constitutes the test of the loop, and its type must be boolean. If the test expression evaluates to true, the block is executed, after which control normally returns to the beginning of the loop. If the test expression evaluates to false, execution proceeds past the loop.

If execution reaches a return statement within the loop, the loop is abruptly terminated.

Execution of a loop can also be controlled with a break or continue statement, which can only appear within the body of a loop.

```
BreakStatement:
    break;

ContinueStatement:
    continue;
```

A break statement causes execution of the loop that immediately contains it to abruptly terminate. A continue statement causes execution of the loop that immediately contains it to abruptly return to the beginning of the loop.

Return Statements

A return statement consists of the return keyword, followed by an optional expression, followed by a semicolon.

```
ReturnStatement:
    return Expression ;
    return ;
```

If execution reaches a return statement, the current function is abruptly terminated. If the return type of the current function is void, then a return statement must not be given an expression. If the return type is any other type, then a return statement must be given an expression whose type matches or is implicitly convertible to the return type.

Expression Statements

An expression statement consists of an expression followed by a semicolon.

An expression statement evaluates the given expression and discards its value.

Expressions

An expression is evaluated to produce a value. All expressions in a uC program have a type, and the type must be valid for the context in which the expression is used.

```
Expression:

ParenthesizedExpression
Literal
NameExpression
CallExpression
NewExpression
NewArrayExpression
FieldAccessExpression
ArrayIndexExpression
UnaryPrefixOperation
BinaryOperation
```

Any expression can be parenthesized, which can be used to override the default associativity and precedence of a sequence of operations.

```
ParenthesizedExpression:
    ( Expression )
```

Simple Expressions

A literal is an expression that evaluates to the value represented by the literal.

```
Literal:
    IntegerLiteral
    FloatingLiteral
    StringLiteral
    BooleanLiteral
    NullLiteral
```

The type of an integer literal is long if it ends with the l or L suffix, otherwise it is int. The type of a floating-point literal is float, a string literal is string, a boolean literal is boolean, and the null literal is null.

An identifier is an expression, and it must name a variable or parameter in the local scope. It is an l-value and can be used in an l-value context. When used in a context that expects a value, it evaluates to the value to which the variable or parameter is bound.

```
NameExpression: Identifier
```

The type of an identifier expression is the type of the variable or parameter it names.

Function Calls

A function call consists of an identifier followed by a parenthesized list of arguments, separated by commas.

```
CallExpression:
    Identifier ( ArgumentsOpt )

ArgumentsOpt:
    Arguments
    Empty

Arguments:
    Expression
    Arguments , Expression
```

The identifier must name a built-in or user-defined function. The number of arguments must match the number of function parameters, and each argument type must match or be implicitly convertible to the corresponding parameter type.

A function call evaluates the arguments in some arbitrary order, binds the argument values to the function parameters in a fresh environment, and executes the body of the function in the given environment.

The type of a function call is the return type of the invoked function.

Allocation Expressions

A simple allocation expression allocates space for an object of user-defined type and initializes the object. It consists of the keyword new, followed by an identifier, followed by a list of zero or more argument expressions enclosed by parentheses and separated by commas.

```
NewExpression:
   new Identifier ( ArgumentsOpt )
```

The identifier must name a user-defined type. The arguments may be empty, in which case the fields of the newly created object undergo Default Initialization. Otherwise, the number of arguments must match the number of fields in the given type, and the argument types must match or be implicitly convertible to the corresponding field types. The fields of the newly created object are then initialized with the argument values. The allocation expression evaluates to a reference to the new object, and the type of the expression is that named by the identifier.

An array allocation expression consists of the new keyword, followed by a type, followed by a list of zero or more argument expressions enclosed by curly braces and separated by commas.

```
NewArrayExpression:
   new Type { ArgumentsOpt }
```

An array allocation creates an array whose element type is the given type. The type of each argument must match or be implicitly convertible to the element type. The array is initialized with the argument values, and it has initial length equal to the number of arguments. The array allocation expression evaluates to a reference to the new array, and the type of the expression is an array of the given element type.

Field Access

A field-access expression consists of an expression denoting the receiver, followed by a period, followed by an identifier

```
FieldAccessExpression:
Expression . Identifier
```

The receiver expression preceding the period must be of user-defined or array type. It is a runtime error if the receiver expression evaluates to a null reference.

If the receiver is of array type, only the identifier length is valid, in which case the field-access expression has type int and evaluates to the length of the array produced by the receiver expression. The field-access expression is not an l-value in this case.

If the receiver is of user-defined type, the identifier must name a field defined by the type. The field-access expression has type associated with the given field. It evaluates to the l-value associated with the field in the object produced by the receiver expression. If used in a context that requires a value, then it evaluates to the value bound to that field.

Array Indexing

An array-indexing expression consists of an expression denoting the receiver, following by an open square bracket, an index expression, and a close square bracket.

The receiver expression preceding the square brackets must be of array type. It is a runtime error if the receiver evaluates to a null reference.

The index expression enclosed by the square brackets must be of type int. It is a runtime error if it evaluates to a value that is negative or greater than or equal to the length of the receiver array.

The array-indexing expression has type corresponding to the element type of the receiver, and it evaluates to the l-value associated with the array element at the given index. If used in a context that requires a value, then it evaluates to the value of the given element.

The order of evaluation of the receiver and index expressions is indeterminate.

Unary Operations

Unary prefix operations consist of sign operations, logical complement, increment, and decrement.

The sign operations, + and -, require the given expression to be of numeric type (int, long, or float), and the type of the sign expression is the same as its subexpression. The + operator has no effect on its operand, while the - operator negates it.

The complement operator! requires its operand expression to be of boolean type, and it results in a boolean. The complement expression evaluates to the complement of the value produced by the operand.

The increment and decrement operators, ++ and --, require the given operand to be an 1-value of numeric type. The increment and decrement and decrement, respectively, the object denoted by the operand. The result is the same 1-value denoted by the operand, and the type is the same as the operand type. The increment or decrement occurs before the resulting 1-value is produced, so that it contains the new value.

Binary Operations

Binary infix operations consist of arithmetic operations, logical operations, arithmetic comparisons, equality comparisons, assignment, and array push and pop operations.

```
BinaryOperation:
   ArithmeticOperation
   LogicalOperation
   Comparison
   EqualityTest
   Assignment
   ArrayOperation
```

The order of evaluation of the operands in a binary operation is indeterminate, except in logical operations.

Arithmetic Operations

Arithmetic operations include addition, subtraction, multiplication, division, and modulo.

```
ArithmeticOperation:

Expression + Expression
Expression - Expression
Expression * Expression
Expression / Expression
Expression % Expression
```

The operands to binary -, *, /, and % must be of numeric type. If the two operands have the same type, then the type of the overall expression also has that type. Otherwise, it is the case that the type of one operand is implicitly convertible to the type of the other, in which case the type of the overall expression is the latter.

The division operator /, when the result is of integer type (int or long), truncates the result. In other words, it rounds the result toward zero. The result value is undefined if the second operand evaluates to zero.

The operands of the modulo operator % must have integer type (int or long). The result is the remainder from dividing the first operand value by the second, preserving the sign of the first operand value. The result is undefined if the second operator evaluates to zero.

The operands to the + operator must be of primitive type, but cannot be of null or void type. If the operands are both of numeric type, the result type is as for the other arithmetic operators. If one operand is of boolean type, then the other operand must be of string type. If one or both operands is of string type, then the operator performs string concatenation, and the non-string operand, if there is one, is implicitly converted to a string using the appropriate built-in conversion function. The result of string concatenation has type string.

Logical Operations

Logical operations consist of disjunction $(|\cdot|)$ and conjunction (&&).

```
LogicalOperation:
    Expression || Expression
    Expression && Expression
```

The operands must be of type boolean, and the result is also of type boolean.

The logical operations are short circuiting. The left-hand operand is evaluated first, and if it is true in the case of disjunction or false in conjunction, the right-hand operand is not evaluated, and the result is the value of the left-operand. Otherwise, the right-hand operand is evaluated, and its value is the result of the operation.

Comparisons

Comparisons consist of the operators <, <=, >, and >=.

```
Comparison:
    Expression < Expression
    Expression <= Expression
    Expression > Expression
    Expression >= Expression
```

The operands must both be of numeric type or both be of string type. The result is of type boolean. Numeric values are compared numerically, while strings are compared lexicographically. Comparison operations are prohibited from being chained together, both syntactically and by type rules.

Equality Tests

Equality comparisons include the == and != operators.

```
EqualityTest:
    Expression == Expression
    Expression != Expression
```

The operands must either have the same type, or the type of one operand must be implicitly convertible to the type of the other. The result is of type boolean.

An equality comparison for a reference type compares the two objects by contents rather than by pointer equality. If one of the two operands evaluates to a null reference, the result is true if the other operand also evaluates to a null reference and false otherwise.

For user-defined types, objects are compared field by field, which are recursively compared. Two objects of a user-defined type with no fields always compare equal. Arrays are compared by length and then by the individual elements. Two empty arrays always compare equal.

Assignment

The assignment operation is as follows:

```
Assignment:
   Expression = Expression
```

The type of the right-hand expression must match or be implicitly convertible to the type of the left-hand side. The left-hand expression must produce an l-value. Assignment replaces the value of object produced by the left-hand operand with the value produced by the right-hand. The assignment operation as a whole has the same type as the left-hand operand, and it produces the same l-value as the left-hand operand. The assignment occurs before the resulting l-value is produced, so that it contains the new value.

Array Operations

Arrays support push (<<) and pop (>>) operations.

```
ArrayOperation:
    Expression << Expression
    Expression >> Expression
```

A push operation appends the value produced by the right-hand expression to the end of the array produced by the left-hand expression. The left-hand operand must be of array type, and the type of the right-hand operand must match or be implicitly convertible to the element type of the left-hand array type. The result of the push operation has the same type as the left-hand operand, and the value is the same array as that produced by the left-hand expression. It is a runtime error if the left-hand operand evaluates to null.

A pop operation removes the last item from the left-hand array and assigns it to the object produced by the right-hand operand, or discards it if the right-hand operand is the null literal. The left-hand expression must be of array

type. The right-hand operand must be either null or an l-value. In the latter case, the element type of the array must match or be implicitly convertible to the type of the right-hand expression. The result of the pop operation has the same type as the left-hand operand, and the value is the same array as that produced by the left-hand expression. It is a runtime error if the left-hand operand evaluates to null or an empty array.

Associativity and Precedence

Operators in uC have the following associativity and precedence, ordered from lowest precedence to highest:

Precedence Class	Operators	Associativity
1	<<,>>	left
2	=	right
3	11	left
4	& &	left
5	==, !=	non-associative
6	<, <=, >, >=	non-associative
7	infix +, -	left
8	*, /, %	left
9	prefix +, -, !, ++,	right
10	.,[]	left

Non-associative operators of a single precedence class cannot be chained together within a single expression.

Memory Management

Implementations of uC are required to perform automatic memory management, i.e. garbage collection, for objects of reference type.

Default Initialization

The initial value of a local variable within a user-defined function is undefined.

Objects of user-defined type constructed by the zero-argument default constructor are default initialized. In default initialization, an object's fields are set to default values. A field of primitive numeric type is set to zero, a field of boolean type is set to false, and a field of string type is set to an empty string. Fields of reference type, including array type, are set to null.