

Milestone #1: Benchmarks

Overview

The purpose of this assignment is mostly organizational. The tasks are quite simple, but are intended to start the whole project rolling.

This document describes the language that we will be working with throughout the quarter: Extraordinarily Vexing Insipid Language. This language is similar in many respects to C, but limited in features.

Over the course of the term, you will implement an optimizing compiler for this language. Your first task is to familiarize yourself with the language.

Extraordinarily Vexing Insipid Language

The following grammar partially describes the language's syntax. In the EBNF below, non-terminals are typeset in **bold font** and terminals are typeset in **typewriter font**.

```

program → types declarations functions
types → {type_declaration}*
type_declaration → struct id { nested_decl } ;
nested_decl → decl ; {decl ;}*
decl → type id
type → int | bool | struct id
declarations → {declaration}*
declaration → type id_list ;
id_list → id { , id }*
functions → {function}*
function → fun id parameters return_type { declarations statement_list }
parameters → ( {decl { , decl }* }opt )
return_type → type | void
statement → block | assignment | print | read | conditional | loop | delete | ret | invocation
block → { statement_list }
statement_list → {statement}*
assignment → lvalue = expression ;
print → print expression {endl}opt ;
read → read lvalue ;
conditional → if ( expression ) block {else block}opt
loop → while ( expression ) block
delete → delete expression ;
ret → return {expression}opt ;
invocation → id arguments ;
lvalue → id { . id }*
expression → boolterm { {&& | ||} boolterm }*
boolterm → simple { {== | < | > | != | <= | >= } simple }opt
simple → term { {+ | -} term }*
term → unary { { * | / } unary }*
unary → { ! | - } * selector
selector → factor { . id }*
factor → ( expression ) | id { arguments }opt | number | true | false | new id | null

```

arguments \rightarrow $(\{ \text{expression } \{, \text{expression} \}^* \}_{opt})$

The following rules complete the syntactic definition.

- A valid program is followed by an end-of-file indicator; extra text is not legal.
- The terminal (token) “id” represents a nonempty sequence (beginning with a letter) of letters and digits other than one of the keywords. Similarly, the terminal (token) “number” represents a nonempty sequence of digits.
- As is the case in most languages, tokens are formed by taking the longest possible sequences of constituent characters. For example, the input “abcd” represents a single identifier, not several identifiers. Whitespace (i.e., one or more blanks, tabs, or newlines) may precede or follow any token. E.g., “x=10” and “x = 10” are equivalent. Note that whitespace delimits tokens; e.g., “abc” is one token whereas “a bc” is two.
- A comment begins with “#” and consists of all characters up to a newline.
- Local declarations and parameters may hide global declarations (and functions), but local declarations cannot hide parameters.
- Structure names are in a separate namespace from variables and functions.

Semantics

The semantics for the language are given informally.

- Program execution begins in the function named `main` that takes no arguments and that returns an `int`. Every valid program must have such a function.
- The scope of each structure type is from the point of definition to the end of the file (this means that a structure type can only include elements of the primitive types and the structure types defined before it, though it should be able to include a member of its own type).
- The scope of each function is from the point of definition to the end of the file (though recursion must be supported, this restriction precludes mutual recursion).
- `if` and `while` have semantics equivalent to those of Java. They both require boolean guards.
- Assignment requires that the left-hand side and right-hand side have compatible types (equal in all cases except for `null`).
- A declaration with a structure type declares a reference to a structure (the structure itself must be dynamically allocated).
- `null` may be assigned to any variable of structure type.
- The `.` operator is used for field access (as in C and Java).
- All arguments are passed by value. For a structure reference, the reference itself is passed by value.
- `print` requires an integer argument and outputs it to standard out.
- `read` reads an integer value from standard in and stores it in the provided argument.
- `new` dynamically allocates a new structure, but does not initialize it, and evaluates to a reference to the newly allocated structure.
- `delete` deallocates the referenced structure.
- Arithmetic and relational operators require integer operands.

- Equality operators require operands of integer or structure type. The operands must have matching type. Structure references are compared by address (i.e., the references themselves are compared).
- Boolean operators require boolean operands.
- Boolean operators are *not* short-circuit.
- Each function with a non-void return type must return a valid value along all paths. Each function with a void return type must not return a value.

Part 1: Partner

Include with your submission a file called `cohort` that includes your name and, if you have one, the name of your partner for the term project. Only a single submission must be made per team.

Part 2: Benchmark

Write and submit *one* benchmarking program (one per team). These programs will be used by all students in the course as input to their compilers. The results will be gathered and compared at the end of the course. The “winning” team (or teams) will get a prize.

So, more details.

1. Take input from the user (provide a sample input file), but only enough to get the benchmark running. For example, the benchmark might read a few integers in order to vary its behavior, but time spent reading input should not dominate the runtime.
2. Use functions.
3. Consider using structures.
4. No fewer than 15 lines of actual code, unless you can justify the limited size.
5. A test case should run for at least a few second, but no longer than a minute.
6. Name your benchmarks. Benchmark1 is boring. Bob is better. But try to avoid name clashes with other teams.
7. Be creative.

Note: Read the grammar carefully. Your benchmark should be syntactically valid and should run without bugs.

If you are working with a partner, you and your partner need only submit one program.

Part 3: C Translation

Submit a semantically equivalent C translation of your benchmark.

Part 4: Input and Output

Submit an input file for one test case and a corresponding output file against which subsequent runs can be compared.

Notes

- This should be a simple assignment. Don’t forget to do it.
- Grading will be divided as follows.

Part	Percentage
1	10
2	50
3	10
4	30

- A message giving exact details on how to submit your files will be posted to the website.