

COSC 461 – Fall 2020
csem - semantic routines for C programs
DUE: by midnight on 11/10

csem reads a C program (actually a subset of C) from its standard input and compiles it into LLVM intermediate representation (IR) on its standard output. It should support the following program constructs and statements in the C programming language:

program construct	example
<i>variable declaration</i>	<code>int foo;</code>
<i>array declaration</i>	<code>int foo[20];</code>
<i>function definition</i>	<code>double max(double x, double y);</code>
<i>function call</i>	<code>x = max(y, z);</code>
<i>label definition</i>	<code>LOOP1:</code>
<i>goto statement</i>	<code>goto LOOP1;</code>
<i>if statement</i>	<code>if (x < y) { x = x + 1; }</code>
<i>if-else statement</i>	<code>if (x < y) { x = x + 1; } else { y = y + 1; }</code>
<i>while statement</i>	<code>while (x < y) { x = x + 1; }</code>
<i>do-while statement</i>	<code>do { x = x + 1; } while (x < y);</code>
<i>for statement</i>	<code>for (x = 0; x < 10; x++) { y = y + 1; }</code>
<i>continue statement</i>	<code>continue;</code>
<i>break statement</i>	<code>break;</code>
<i>return statement</i>	<code>return x;</code>

You will also need to support arithmetic and conditional expressions as well as assignment statements with the following operators. The bitwise operators (`|`, `^`, `&`, `<<`, `>>`, `~`) and mod (`%`) only operate on integer types. Your implementation should assume that all other operators support integer and floating point types and make appropriate conversions between types when necessary.

operator	description
<code>==</code> <code>!=</code> <code><=</code> <code>>=</code> <code><</code> <code>></code> <code>=</code> <code> </code> <code>^</code> <code><<</code> <code>>></code> <code>+</code> <code>-</code> <code>*</code> <code>/</code> <code>%</code>	binary operators
<code>~</code> <code>-</code>	unary operators (inversion, negation)
<code>[]</code>	index into array

Your assignment is to write the semantic actions for the `csem` program to produce the appropriate LLVM IR. The following files, which should be downloaded from the course website (on Canvas), will comprise part of your program:

<code>makefile</code>	- <code>csem</code> makefile
<code>src/cgram.y</code>	- yacc grammar for subset of C
<code>src/main.cpp</code>	- main entry point for <code>csem</code>
<code>src/scan.c</code>	- lexical analyzer
<code>src/sem.cpp</code>	- semantic action routines
<code>src/semutil.c</code>	- utility routines for the semantic actions
<code>src/sym.c</code>	- symbol table management
<code>include/cc.h</code>	- defines several macros and globals for <code>csem</code>
<code>include/scan.h</code>	- defines prototypes for routines in <code>scan.c</code>
<code>include/sem.h</code>	- defines prototypes for routines in <code>sem.c</code>
<code>include/semutil.h</code>	- defines prototypes for routines in <code>semutil.c</code>
<code>include/sym.h</code>	- defines prototypes for routines in <code>sym.c</code>

The makefile will create an executable called `csem` in the current directory. For your assignment, you will fill in the definitions for the semantic action routines. This `sem.cpp` file contains stubs for each of the semantic action routines you will need to implement. While I have provided you access to the other `*.c`, `*.h`, and `*.cpp` files, you should not need to modify them. The yacc file `cgram.y` contains the entire grammar specification for the source language and also should not be modified. You can make additional functions and data structures in `sem.cpp` to abstract common operations and complete your implementation. When making your executable, refer to the makefile provided, which uses the other source files to produce the `csem` executable.

We have also included the executable file `ref_csem` that contains our working implementation. The executable was built on and will work on all of the hydra machines at UTK. The makefile also includes several targets to facilitate building and testing example input programs with different compilers. The following provides a brief explanation of the capabilities included in the makefile:

- To build the `csem` executable from source, simply type `make`:

```
> make
```

- To build an example input with `csem`, use `csem_build`:

```
> make csem_build INPUT=input4
```

This command will build the program `inputs/input4.c` with `csem`. The output files from the build will be written to the test folder. The executable file corresponding to the built input program will be in the top-level of your assignment directory and will be named `csem_exe`.

- If your implementation does not yet generate correct LLVM IR, you can use the `csem_ll` target to stop compilation after running `csem`:

```
> make csem_ll INPUT=input4
```

After running this command, the LLVM IR generated by `csem` will be in `test/csem.input4.ll`.

- To build an example input with `ref_csem`, use `ref_build`:

```
> make csem_build INPUT=input4
```

The output files from this build will also be written to the test folder. The built executable file will be named `ref_csem_exe`.

- To ensure your compiler is working correctly with a given input, use the `diff` target. This command will build the given input with both `ref_csem` as well as with `csem`, and will then run the resulting programs and compare their outputs with the Linux `diff` utility:

```
> make diff INPUT=input4
```

- To clean out the built program files in the test folder, use `test_clean`:

```
> make test_clean
```

- To clean out everything, including the built files associated with `csem`, use `clean`:

```
> make clean
```

You will also submit a project report, similar to the reports you submitted for previous projects. Specifically, you will submit a short (one to two page, single-spaced) document that describes:

1. (in your own words) the program you set out to write,
2. your approach (i.e. design and relevant implementation details) for writing this program,
3. how you debugged and tested your solution, and
4. any issues you had in completing the assignment.

You should upload a gzipped tar file (created with `tar cvzf ...`) with your source files and a pdf of your report to the Canvas course website before midnight on Tuesday, November 10th. Partial credits will be given for incomplete efforts. However, a program that does not compile or run will get 0 points. Point breakdown is below:

• backpatching	(10)
• loops, if statements	(10)
• conditional expressions	(10)
• type casting	(10)
• function call / return	(5)
• continue / break	(5)
• goto / labels	(5)
• logical expressions	(5)
• unary operators	(5)
• bitwise operators	(5)
• binary operators	(5)
• relational operators	(5)
• set expressions	(5)
• string code generation	(2)
• array indexing	(3)
• project report	(10)

Example 1 input (input2.c):

```
int main()
{
    int a, c;
    double b, d;

    a = 3;
    b = 4;
    c = 5;
    d = 6;

    print("%d %3.2f %d %3.2f\n", a, b, c, d);
    return 0;
}
```

Example 1 output:

```
; ModuleID = '<stdin>'
source_filename = "<stdin>"
```

```
@0 = private unnamed_addr constant [19 x i8] c"%d %3.2f %d %3.2f\0A\00"
```

```
declare i32 @print(i8*, ...)
```

```
define i32 @main() {
    %a = alloca i32
    %c = alloca i32
    %b = alloca double
    %d = alloca double
    store i32 3, i32* %a
    store double 4.000000e+00, double* %b
    store i32 5, i32* %c
    store double 6.000000e+00, double* %d
    %1 = load i32, i32* %a
    %2 = load double, double* %b
    %3 = load i32, i32* %c
    %4 = load double, double* %d
    %5 = call i32 @print(i8*, ...) @print(i8* getelementptr inbounds ([19 x i8], [19 x i8]* @0, i32
double %2, i32 %3, double %4)
    ret i32 0
}
```

Example 2 input (input5.c):

```
double m[6];

int scale(double x) {
    int i;

    if (x == 0)
        return 0;
    for (i = 0; i < 6; i += 1)
        m[i] *= x;
    return 1;
}

int main() {
    int i;
    double x;

    i = 0;
    while (i < 6) {
        m[i] = i;
        i = i + 1;
    }

    print("%5.1f %5.1f %5.1f\n", m[0], m[2], m[5]);

    x = 5;
    scale (x);

    print("%5.1f %5.1f %5.1f", m[0], m[2], m[5]);

    return 0;
}
```

Example 2 output:

```
; ModuleID = '<stdin>'
source_filename = "<stdin>"

@m = global [6 x double] zeroinitializer
@0 = private unnamed_addr constant [19 x i8] c"%5.1f %5.1f %5.1f\0A\00"
@1 = private unnamed_addr constant [18 x i8] c"%5.1f %5.1f %5.1f\00"

declare i32 @print(i8*, ...)

define internal i32 @scale(double %x) {
    %x1 = alloca double
    store double %x, double* %x1
    %i = alloca i32
    %1 = load double, double* %x1
    %2 = fcmp oeq double %1, 0.000000e+00
```

```

    br i1 %2, label %L0, label %L1

L0:                                     ; preds = %0
    ret i32 0

L1:                                     ; preds = %0
    store i32 0, i32* %i
    br label %L2

L2:                                     ; preds = %L1, %L3
    %3 = load i32, i32* %i
    %4 = icmp slt i32 %3, 6
    br i1 %4, label %L4, label %L5

L3:                                     ; preds = %L4
    %5 = load i32, i32* %i
    %6 = add i32 %5, 1
    store i32 %6, i32* %i
    br label %L2

L4:                                     ; preds = %L2
    %7 = load i32, i32* %i
    %8 = getelementptr [6 x double], [6 x double]* @m, i32 0, i32 %7
    %9 = load double, double* %x1
    %10 = load double, double* %8
    %11 = fmul double %10, %9
    store double %11, double* %8
    br label %L3

L5:                                     ; preds = %L2
    ret i32 1
}

define i32 @main() {
    %i = alloca i32
    %x = alloca double
    store i32 0, i32* %i
    br label %L6

L6:                                     ; preds = %0, %L7
    %1 = load i32, i32* %i
    %2 = icmp slt i32 %1, 6
    br i1 %2, label %L7, label %L8

L7:                                     ; preds = %L6
    %3 = load i32, i32* %i
    %4 = getelementptr [6 x double], [6 x double]* @m, i32 0, i32 %3
    %5 = load i32, i32* %i
    %6 = sitofp i32 %5 to double
    store double %6, double* %4
    %7 = load i32, i32* %i

```

```

%8 = add i32 %7, 1
store i32 %8, i32* %i
br label %L6

L8:                                     ; preds = %L6
%9 = load double, double* getelementptr inbounds ([6 x double], [6 x double]* @m, i32 0, i
%10 = load double, double* getelementptr inbounds ([6 x double], [6 x double]* @m, i32 0,
%11 = load double, double* getelementptr inbounds ([6 x double], [6 x double]* @m, i32 0,
%12 = call i32 @i8*, ...) @print(i8* getelementptr inbounds ([19 x i8], [19 x i8]* @0, i32
%9, double %10, double %11)
store double 5.000000e+00, double* %x
%13 = load double, double* %x
%14 = call i32 @scale(double %13)
%15 = load double, double* getelementptr inbounds ([6 x double], [6 x double]* @m, i32 0,
%16 = load double, double* getelementptr inbounds ([6 x double], [6 x double]* @m, i32 0,
%17 = load double, double* getelementptr inbounds ([6 x double], [6 x double]* @m, i32 0,
%18 = call i32 @i8*, ...) @print(i8* getelementptr inbounds ([18 x i8], [18 x i8]* @1, i32
%15, double %16, double %17)
ret i32 0
}

```