DUE: Thursday, Apr 1 (11:59pm)

For this part of the project, you will add dataflow optimizations to your compiler. At the very least, you must implement global common subexpression elimination. *The other optimizations listed below are optional.* We list them here as suggestions since past winners of the compiler derby typically implement each of these optimizations in some form. You are free to implement any other optimizations you wish. Note that you will be implementing register allocation for the next project, so you don't need to concern yourself with it now.

Global CSE (Common Subexpression Elimination): Identification and elimination of redundant expressions using the algorithm described in lecture (based on available-expression analysis). See §8.3 and §13.1 of the Whale book, §10.6 and §10.7 in the Dragon book, and §17.2 in the Tiger book.

Global Constant Propagation and Folding: Compile-time interpretation of expressions whose operands are compile time constants. See the algorithm described in §12.1 of the Whale book.

Global Copy Propagation: Given a "copy" assignment like `x = y`, replace uses of `x` by `y` when legal (the use must be reached by only this def, and there must be no modification of `y` on any path from the def to the use). See §12.5 of the Whale book and §10.7 of the Dragon book.

Loop-invariant Code Motion (code hoisting): Moving invariant code from within a loop to a block prior to that loop. See §13.2 of the Whale book and §10.7 of the Dragon book.

Unreachable Code Elimination: Unreachable code elimination is the conversion of constant conditional branches to equivalent unconditional branches, followed by elimination of unreachable code. See §18.1 of the Whale book and §9.9 of the Dragon book.

Dead Code Elimination: Dead code elimination is the detection and elimination of code which computes values that are not subsequently used. This is especially useful as a post-pass to constant propagation, CSE, and copy propagation. See the §18.10 of the Whale book and §10.2 of the Dragon book.

The following are not generally considered dataflow optimizations. However, they can increase the effectiveness for the transformations listed above.

Algebraic Simplification and Reassociation: Basic algebraic simplifications described in class. This includes simplifying the following rules:

$$a + 0 \Rightarrow a$$

$$a - 0 \Rightarrow a$$

$$a * 1 \Rightarrow a$$
$$b == true \Rightarrow b$$
$$b \mathrel{!=} false \Rightarrow b$$
$$a + a \Rightarrow 2 * a$$
$$\text{etc}$$

You may find it useful to canonicalize expression orders, especially if you choose to implement CSE. See §12.3 of the Whale book and §10.3 of the Dragon book.

"Peephole" optimizations: Local manipulation of generated code. Many possibilities for peephole optimization exist. See chapter 18 of the Whale book and §9.9 of the Dragon book for ideas. Peephole optimization is a broad category. If you decide to attack it, talk with your TA in advance about what you are planning to do.

Inline Expansion of Calls: Replacement of a call to procedure P by inline code derived from the body of p. This can also include the conversion of tail-recursive calls to iterative loops. See §15.1 and §15.2 of the Whale book.

You will want to think carefully about the order in which optimizations are performed. You may want to perform some optimizations more than once.

All optimizations (except inline expansion of calls) should be done at the level of a single procedure. Be careful that your optimizations do not introduce bugs in the generated code or break any previously-implemented phase of your compiler. Needless to say, it would be foolish not to do regression testing using your existing test cases. *Do not underestimate the difficulty of debugging this part of the project.*

As in the code generation project, your generated code must include instructions to perform the runtime checks listed in the language specification. It is desirable to optimize these checks whenever possible (*e.g.,* CSE can be used to eliminate array bounds tests). Note that the optimized program must report a runtime error for exactly those program inputs for which the corresponding unoptimized program would report a runtime error (and the runtime error message must be the same in both cases). However, we allow the optimized program to report a runtime error earlier than it might have been reported in the unoptimized program.

What to Hand In

Follow the directions given in project overview handout when writing up your project. For the written portion, make sure to include a description of every optimization you implement. Show clear *examples* of IR excerpts (and x86-64 excerpts, if relevant) detailing what your optimizations do. You should also include a discussion of how you determined the order in which your optimizations are performed.

Each group must place their completed submission at:

$$\text{/mit/6.035/group/le} \mathbf{NN} \text{/submit/le} \mathbf{NN} \text{-dataflow.tar.gz}$$

Your compiler's command line interface must provide the following interface for turning on each optimization individually. Something similar should be provided for every optimization you implement. Document the names given to each optimization.

- `-opt cse` — turns on common subexpression elimination

- `-opt cp` — turns on copy propagation optimization

- `-opt all` — turns on all optimizations

This is the interface provided by `java6035.tools.CLI.CLI` through the `optnames` facility. No `-target` option should be necessary to run the optimizations and get the resulting `.s` file.

Submitted tarballs should have the following structure:

```
leNN-dataflow.tar.gz
|
'-- leNN-dataflow
    |
    |-- AUTHORS            (list of students in your group, one per line)
    |
    |-- code
    |   |
    |   ...               (full source code, can build by running 'ant')
    |
    |-- doc
    |   |
    |   ...               (write-up, described in project overview handout)
    |
    '-- dist
        |
        '-- Compiler.jar   (compiled output, for automated testing)
```

You should be able to run your compiler from the command line with:

```
java -jar dist/Compiler.jar -target codegen <filename>          # no optimizations
java -jar dist/Compiler.jar -target codegen -opt all <filename> # all optimizations
```

Your compiler should then write a x86-64 assembly listing to: `<filename>.s`

Test Cases

The test cases will be provided in `/mit/6.035/provided/dataflow/`. Your grade will be determined by producing correct code and how well your compiler performs the required dataflow optimization (CSE) for the provided tests and for a set of hidden tests.

Grading Script

As with the previous project, we are providing you with the grading script we will use to test your code. This script only shows you results for the public test cases.

/mit/6.035/provided/gradingscripts/p4grader.py

The usage is the same as with previous projects.

The grading script only tests correctness. Your grade is based on 1) correctness 2) CSE optimization implementation 3) writeup.

6.035 Computer Language Engineering
Spring 2010