Compiler construction

Lecture 2: Software Engineering for Compilers

Magnus Myreen
Spring 2018
Chalmers University of Technology — Gothenburg University

Structuring the project

Compiler structure

Passes

- Lexer
- Parser
- Type checker
- Return checking[1]
- Code generator

Structuring passes

- In functional languages, a pass correspond to a function
- In OO languages, a pass corresponds to a visitor method

[1] Can be done as a separate pass or as part of the type checker

What you have to do

- BNFC takes care of lexing and parsing, however, you will have to change the BNFC file for JAVALETTE that we provide for you
- Write typechecker
- Write code generator
- Write a `main` function which connects the above pieces together and invokes the various LLVM tools to generate an executable program (for submissions B and C)

Version control

- It is highly recommend that you use version control software; using version control software is an essential practice when developing code
- For example: git, darcs, subversion, mecurial, …
- However, do not put your code in a public repository, where others can see your code
- Use educational account for GitHub or BitBucket
- Alternative: use a Dropbox folder as a git remote (create a bare repo)

Testing compilers

Trusting the compiler

Bugs
When finding a bug, we go to great lengths to find it in our own code.

- Most programmers trust the compiler to generate correct code
- The most important task of the compiler is to generate correct code

Establishing compiler correctness

Alternatives

- Proving the correctness of a compiler is prohibitively expensive
- Testing is the only viable option

Testing compilers

- Most compilers use unit testing
- They have a big collection of example programs which are used for testing the compiler
- For each program the expected output is stored in the test suite
- Whenever a new bug is found, a new example program is added to the test suite; this is known as regression testing

Random testing

Generating random inputs and check correctness of output.

Property-based testing

- Specify (semi-formal) properties that software should have
- Generate random inputs to validate these properties
- In case of a violation, then we have found a counterexample
- Shrink the counterexample to a minimal failing test case

Example

```
propReverse :: [Int] -> [Int] -> Bool
propReverse xs ys =
  reverse (xs ++ ys) == reverse ys ++ reverse xs

Prelude Test.QuickCheck> quickCheck propReverse
+++ OK, passed 100 tests.
```

Random testing for compilers

- Testing compilers using random testing means generating programs in the source language
- Writing good random test generators for a language is very difficult
- Different parts of the compiler might need different generators
  - The parser needs random strings, but they need to be skewed towards syntactically correct programs in order to be useful
  - The type checker needs a generator which can generate type correct programs (with high probablity)
- It can be hard to know what the correct execution of a program is; we need another compiler or interpreter to test against
- What if the generated program doesn't terminate, or takes a very long time?
- Using random testing for compilers is difficult and a lot of work

Testing your JavaLette compiler

Remember to test your compiler!

- Use the provided test suite!
- Write your own tests!

Compiler Bootstrapping

A real language

Some people say:
A programming language isn't real until it has a self-hosting compiler

A self-hosting compiler
If you're designed an awesome programming language you would probably want to program in it.

In particular, you would want to write the compiler in this language.

The chicken and egg problem

If we want to write a compiler for the language X in the language X, how does the first compiler get written?

Solutions

- Write an interpreter for language X in language Y
- Write another compiler for language X in language Y
- Write the compiler in a subset of X which is possible to compile with an existing compiler
- Hand-compile the first compiler

Porting to new architectures

A related problem
How to port a compiler to a new hardware architecture?

Solution: cross-compilation
Let the compiler emit code for the new architecture while still running on an old architecture.

Writing Makefiles

Make

The build automation tool `make` is handy for compiling large projects. It keeps track of which files need to be recompiled.

A Makefile consists of rules which specifies:

- Which target file will be generated
- How these files are generated

General structure of rules

```
target : dependencies ...
  shell commands specifying how to generate target
```

Concrete example

```
compiler : parser.o typechecker.o
  gcc -o compiler parser.o typechecker.o
parser.o : parser.c
  gcc -c module.c -o module.o
```

Using `make`

Pattern rules

- When having lots of targets it can be inconvenient to list all of them in the in a Makefile
- Then pattern rules come in handy

```
%.o : %.c
    gcc -c $< -o $@
```

Warning

- The space before the shell commands needs to be a tab stop!
- If you just use spaces then the commands will not execute

Using `make`

Invoking `make`

- Invoking `make` without any arguments will make the first target in a Makefile
- When giving `make` a target as an argument it will try to build that target and all of its dependencies if needed

Using `PHONY` rules

- Sometimes it is convenient to have targets which do not produce files
- A common example is `clean` which removes all generated files
- These targets should be declared as `PHONY`

```
.PHONY clean
clean:
  rm -f *.o
```

Outlook

- There is a lot more to `make`, but these basic principles will get you very far
- `make` is not without flaws, but it is very widely available and good to know

Project

- In the project you automatically get a Makefile from the BNFC tool
- Don't forget to `make clean` before packaging your solution for submission
- It can be very convenient to have a target which automatically makes a package for submission

Managing state in the compiler

OO vs functional implementation language

- When writing the type checker and code generator, the compiler needs to carry around symbol tables with information about e.g. the type of a variable
- This is handled differently when implementing the compiler in an object-oriented language or a functional language

Object-oriented
In OO languages it is easy to manage state, simply by using a local variable which is updated, or an object field.

Functional
In functional languages it can be tiresome to carry around state.

Can be made much more convenient by using a state monad.

The state monad

The state monad provides a convenient way to carrying around state in Haskell.

```
data CompileState = ...

type CompileMonad a = State CompileState a
```

State transformer

For debugging purposes it is often convenient to use the state monad transformer on top of the IO monad.

This allows for easily printing debug-information.

```
data CompileState = ...

type CompileMonad a = StateT CompileState IO a
```

State monad demo

Live coding

The lens package

The package `lens` provides functions which makes it more convenient to use the state monad.

Suppose we wish to use the following state in our state monad:

```
data FState = FState
  { _consts  :: [Int]
  , _subst   :: [(V,V)]
  , _nameGen :: Int
  }

makeLenses ''FState
```

This produces lenses named `const`, `subst` and `nameGen`.

Note the underscores in the names!

Requires language extension `TemplateHaskell`.

State monad and lenses

Getting and setting a field in the state:

Without lenses

```
st <- get
let cs = consts st

set (st {consts = []})
```

With lenses

```
cs <- use consts

consts .= []
```

State monad and lenses: Updating

Updating a field in the state:

Without lenses

```
set (st {const = c : const st)})
```

With lenses

```
const %= (c:)
```

Uniplate

Uniplate is library for writing simple and concise generic operations.

Queries

```
[v | Let v _ _ <- universe ast]
```

Transformations

```
let r x = case x of Neg (Const n) -> Const (-n); _ -> x
in  transform r ast
```

State monad, lenses, and uniplate

- The `lens` library is a huge library with lots of convenient functionality, but use with care
- We have only scratched the surface here
- Uniplate is a handy library for queries and traversals
- It is not mandatory to use the state monad, unipate, or the `lens` library in the project
- Use the tools you feel are helpful