Coq projects for type theory 2014

Herman Geuvers, James McKinna, Freek Wiedijk

November 14, 2014

Here are six projects for the type theory course to choose from. Each student has to choose one of these projects. More than one student can choose the same project, but they should not collaborate on it.

At the end of the course each student needs to submit both a Coq formalization, plus a short report (up to ten pages) that describes the formalization and discusses the choices made while formalizing.

For each project we first give a high level description of what should be done, followed by details of one specific way to do that. However, students are free not to follow these specifics.

1 Type check the simply typed lambda calculus

Formalize the typing rules of the simply typed lambda calculus, then formalize a type checker *for this system, and finally prove that the type checker will produce a correct type judgment if it succeeds.*

This project is an instance of *reflection.* Although the type theory of Coq contains simply typed lambda calculus as a subsystem, the terms and types that the formalization talks about will not be *those* Coq terms and types, but syntactic objects *modelled* in Coq.

You do not need to prove the *completeness* of your type checker.

Here are specifics of one possible solution, which takes 97 lines of Coq:

- Define inductive types `type` and `term` for the types and terms of the simply typed lambda calculus. For example the definition of `type` might look like

  ```
  Inductive type : Set :=
  | var_type : string -> type
  | fun_type : type -> type -> type
  ```

  With this definition the type $(A \to B) \to C$ would be represented by

  ```
  fun_type (fun_type (var_type "A") (var_type "B"))
          (var_type "C")
  ```

To get string notation in Coq, put

```
Require Import String.
Open Local Scope string_scope.
```

at the start of your file.

- Define an inductive predicate `has_type`, such that the Coq formula

  ```
  has_type Gamma M A
  ```

  corresponds to the derivability of the judgment

  $$\Gamma \vdash M : A$$

- Write a recursive function `type_check` that (in a given context) returns the type of an element of `term`. A possible Coq type for this function might be

  ```
  type_check
       : list (string * type) -> term -> option type
  ```

  If the input term (the second argument) is not type correct, the function will have to return `None`. For this reason the output type is not `type` but `option type`.

- You will need to look up variables in the context. For this define an inductive predicate `assoc` (to be used in the variable case of `has_type`) and a recursive function `lookup` (to be used in the variable case of `type_check`):

  ```
  assoc
       : forall A B : Set, list (A * B) -> A -> B -> Prop
  lookup
       : forall A B : Set,
         (forall x y : A, {x = y} + {x <> y}) ->
         list (A * B) -> A -> option B
  ```

  The third argument of `lookup` is a decision procedure for equality on `A`. If the keys are `string`s this argument should be `string_dec`.

  The functions `assoc` and `lookup` correspond to each other in exactly the same way that `has_type` and `type_check` do.

- The type checker needs to be able to decide equality of types. (If you apply a function to an argument, the type of the argument needs to match the type of the domain of the function.) For this prove the lemma

```
type_dec
     : forall A B : type, {A = B} + {A <> B}
```

A convenient tactic for proving this is `decide equality.`

- Next we will need to prove our type checker correct. A nice way to do this is by changing the types of `lookup` and `type_check` to have them also return 'proof objects' for the properties of the objects they return:

  ```
  lookup
       : forall A B : Set,
         (forall x y : A, {x = y} + {x <> y}) ->
         forall (l : list (A * B)) (a : A),
         option {b : B | assoc l a b}

  type_check
       : forall (Gamma : list (string * type)) (M : term),
         option {A : type | has_type Gamma M A}
  ```

  To find out about the meaning of the set notation `{ ... | ... }` do `Check exist` or `Print sig`.

  The function `exist` has implicit arguments. If you want to give those arguments explicitly because Coq cannot figure them out, write `@exist`: then all four arguments can and should be given.

- Often Coq will complain if you just use a simple `match ...with`. In that case using `match ...in ...return ...with` can improve things. For example, the `match` the we used in our definition of `lookup` looks like

  ```
  match l return option {b : B | assoc l a b} with ...
  ```

2 The pigeon hole principle

Use Coq to formally prove the pigeon hole principle *which says that if you put n pigeons in m holes, with $m < n$, then at least one hole with have more than one pigeon in it.*

Here are specifics of one possible solution, which takes 61 lines of Coq (but with many tactics per line!):

- A possible way to write the statement is:

  ```
  Lemma pigeon_hole :
    forall m n, m < n ->
      forall f, (forall i, i < n -> f i < m) ->
        exists i, i < n /\
          exists j, j < n /\ i <> j /\ f i = f j.
  ```

Here `f` is the function that maps the number of a pigeon in $\{0 \ldots n-1\}$ to the number of its hole in $\{0 \ldots m-1\}$. The fact that `f` also will map numbers $\geq n$ to something will not hurt.

- A useful tactic to automatically prove equalities and inequalities between natural numbers is

  ```
  omega.
  ```

  To make it available put

  ```
  Require Import Omega.
  ```

  at the start of your file.

- If for natural numbers x and y in a term you want to make a distinction between whether $x \leq y$ or $y < x$, you can write:

  ```
  if le_lt_dec x y then ... else ...
  ```

  Then to do a case split between those two cases in the proof, one can use:

  ```
  elim (le_lt_dec x y).
  ```

3 Proving an expression compiler correct

Formalize both an interpreter and a compiler for a simple language of arithmetical expressions, and show that both give the same results. Compile the expressions to code for a simple stack machine. Use dependent types to make Coq aware of the fact that the compiled code will never lead to a run time error.

Here are specifics of one possible solution, which takes 78 lines of Coq:

- Consider the following expression language:

  $$\begin{array}{rcl} \langle\text{exp}\rangle & ::= & \langle\text{literal}\rangle \mid \langle\text{exp}\rangle + \langle\text{exp}\rangle \mid \ldots \\ \langle\text{literal}\rangle & ::= & 0 \mid 1 \mid 2 \mid \ldots \end{array}$$

  Give an `Inductive` definition of a datatype `Exp` of (the abstract syntax for) $\langle\text{exp}\rangle$s.

- Define a function

  ```
  eval: Exp -> nat
  ```

  giving a semantics for $\langle\text{exp}\rangle$s.

- Give an `Inductive` definition of a datatype `RPN` of Reverse Polish Notation for ⟨exp⟩s.
- Write a compiler

  ```
  rpn : Exp -> RPN
  ```

- Write an evaluator `rpn_eval` for `RPN`, returning an `option nat`.
- Prove that

  ```
  forall e:Exp, Some (eval e) = rpn_eval (rpn e)
  ```

- Generalize the above to `Exp`ressions containing variables, and evaluation with respect to an environment of bindings of variables to `nat`s.
- Discuss how you might avoid explicit consideration of `None` terms in the definition of `rpn_eval`, and explain how you need to modify your formalization in Coq.

4 The unary versus the binary natural numbers

Define both the unary and the binary natural numbers, define addition on both types, define mappings in both directions, show that those mappings are inverse to each other, and finally show that the two addition functions correspond to each other under these mappings.

This exercise is more difficult than it looks, but it is a nice challenge. There are various approaches possible:

- Define the binary numbers with the possibility of leading zeroes. This amounts to 'lists of bits'.

  In the case the mappings between the two types are not simply inverse to each other, as different binary representations might represent the same number. One can define a predicate of two representations being 'equal' (= representing the same number), a predicate of a representation being in normal form (= having no leading zeroes), and a function to normalize a representation (= remove the leading zeroes).

  Our solution using this approach (with many tactics per line!) takes 347 lines and has 34 lemmas.

- Alternatively one can use a binary representation that is unique. There are at least two possibilities for this:

  - Define a type of *positive* binary natural numbers first, and then use that to define a type of *all* binary natural number. This is similar to the way the type of binary integers `Z` is defined in the Coq standard library.

    Our solution using this approach takes 155 lines and has 11 lemmas.

– Define *mutual* types of positive and non-negative binary natural numbers:

```
Inductive bnat : Set :=
| b0 : bnat
| i : pnat -> bnat

with pnat : Set :=
| bit0 : pnat -> pnat
| bit1 : bnat -> pnat.
```

Here the functions `bit0` and `bit1` add a 0 or 1 at the right end of the number, so they amount to doubling the number respectively doubling it and adding one. Note that the type `bnat` has unique representations for all natural numbers.

Our solution using this approach takes 143 lines and has 10 lemmas.

Both of these approaches are quite hairy with most definitions and lemmas appearing twice for the two different types for binary numbers.

5 Integers à la Margaris

Margaris [1] gives a "direct" formalization of the integers, so not as pairs of naturals, or as two copies of the naturals, but by directly defining a language with 0, s (successor) and p (predecessor) with suitable axioms, including an induction scheme. The assignment is to formalize this in Coq, and to prove certain properties about it. The formalization will be slightly different from [1].

Basic definitions

Create a section with the following parameters and hypotheses

```
Parameter ZZ : Set.
Parameter oZ : ZZ.
Parameter pZ sZ : ZZ -> ZZ.

Hypothesis Zps : forall x : ZZ, pZ (sZ x) = x.
Hypothesis Zsp : forall x : ZZ, sZ (pZ x) = x.

Hypothesis ZZ_ind_margaris : forall Q : ZZ-> Prop,
  Q oZ -> (forall y, Q y <-> Q (sZ y)) -> forall x, Q x.
```

So we have a zero and a successor and predecessor that are each others inverses. The final hypothesis is the induction principle for the integers, that Margaris assumes.

Define the predicate N ($N(x)$ says that "x is a natural number") on $\mathbb{Z}$ by

$$N(x) := \forall Q, Q(0) \to (\forall y, Q(y) \to Q(s(y))) \to Q(x)$$

and add as an axiom (using Coq's `Hypothesis` command) the assumption

$$\neg N(p(0))$$

Define the predicate M on $\mathbb{Z}$ by

$$M(x) := \forall Q, Q(0) \to (\forall y, Q(y) \to Q(p(y))) \to Q(x)$$

Prove the following lemmas for $\mathbb{Z}$:

- s and p are injective.
- $N(0)$ and $M(0)$, $\forall x, N(x) \to N(s(x))$ and $\forall x, M(x) \to M(p(x))$.
- $\forall x, N(p(x)) \to N(x)$ and $\forall x, M(s(x)) \to M(x)$.
- $\forall x, p(x) \neq x$ and $\forall x, s(x) \neq x$.
- $\forall x, N(x) \to s(x) \neq 0$ and $\forall x, M(x) \to p(x) \neq 0$.
- $\forall x, M(x) \to \neg N(p(x))$ and $\neg M(s(0))$ and $\forall x, N(x) \to \neg M(s(x))$.

Define the following and prove the given properties

- Define "positive" and "negative" as predicates pos and neg on $\mathbb{Z}$.
- Prove that, for $x : \mathbb{Z}$, *either* $\mathrm{pos}(x)$ *or* $x = 0$ *or* $\mathrm{neg}(x)$.

Possible definitions are $\mathrm{pos}(x) := N(p(x))$ and $\mathrm{neg}(x) := M(s(x))$ and then prove $\forall x : \mathbb{Z}, N(x) \to x = 0 \vee N(p(x))$ and $\forall x : \mathbb{Z}, M(x) \to x = 0 \vee M(s(x))$ to prove $\forall x : \mathbb{Z}, \mathrm{pos}(x) \vee x = 0 \vee \mathrm{neg}(x)$ and $\forall x : \mathbb{Z}, \mathrm{pos}(x) \to x \neq 0 \wedge \neg\mathrm{neg}(x)$ etc.

A full formalization of this assignment can be done in 121 lines of Coq.

Reference. [1] Angelo Margaris – Successor Axioms for the Integers *The American Mathematical Monthly* Vol. 68, No. 5 (May, 1961), pp. 441-444 Published by Mathematical Association of America.
URL: `http://www.jstor.org/stable/2311096`