A Comparison between Concrete Representations for Bindings

Extended Abstract – June 3rd, 2006

Arthur Charguéraud

Many representations of term syntax with variable bindings have been used to formalize programming language metatheory, but so far there is no clear consensus on which is the best representation. We have undertaken an exhaustive comparison between concrete representations, among which are: names, nominal, de Bruijn indices, de Bruijn levels, locally nameless, and their variations. In the context of formalizing System $F_{<:}$, we compare the possible representations of types with respect to:

- representation of bound variables and of their binders, and representation of free variables;
- structural modifications of types that might be required when weakening the environment, performing a substitution, or updating the environment (for example changing the upper bound of a variable in the narrowing lemma);
- well-formation of both types and environments, either using an auxiliary relation to define for some given context what types are well-formed in it, or directly embedding well-formedness in the fundamental definition of types.

Because we aim at formal proofs, it is important to study how each of the representation behave at every places where types are used. We will consider a diverse set of issues raised in the definition of subtyping and in the proofs of properties about this relation (this corresponds to part 1A plus lemma A.10 of the POPLmark Challenge). The most important issues are:

- how bound variables are turned into free variables when passing through a binder, and, in the particular case of working with named free variables, how is the introduced name quantified;
- how types are extracted from the environment: either using a lookup function, or a "env-has" relation, or even annotating free variables with their upbound type;
- how well-formation appears in the definition of the subtyping relation: either in every single rule, or only at the leaves of the derivation, or only at the root of the derivation.

We argue that the locally nameless encoding dominates all other known approaches, at least when it is used with a relation describing the belonging of a binding to an environment, when the well-formation hypotheses are placed at the roots of derivations, and when names introduced when passing through a binder are quantified universally. Practice confirms these theoretical results: our implementation in COQ of part 1A using this technique is significantly shorter than the solutions previously submitted.

This work is still in progress. We have so far implemented two complete solutions to part 1A of the challenge in COQ, one using de Bruijn indices and the other with the locally nameless encoding as described above. We have studied the other solutions using concrete representations that have been proposed, although we still need to investigate some of them in more details. Future work includes the extension of the locally nameless solution to the rest of the POPLmark challenge.