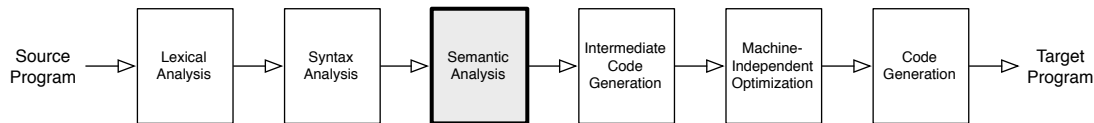


HW 5: Scopes and Type Systems

CSCI 434T
Fall, 2011

Overview



This week, we will examine the symbol table management required for the IC language and explore type systems as a foundation for describing and reasoning about type checking. The first reading is an article from the *Encyclopedia of Computer Science* on type systems written by Luca Cardelli, a very influential researcher. (Cardelli designed and implemented the first ML compiler and helped to develop much of the formal theory of object-oriented programming.) This article covers the basic theory and formalism used to describe type systems.

I have also included a paper on type-based data race detection to illustrate how type systems can capture additional program properties. The last reading, from Kim Bruce's book, touches on many problems specific to designing and type checking object-oriented languages. It should be mostly background reading and will serve as a OOP refresher as you implement IC.

Readings

- “Type Systems,” Luca Cardelli, 1997.
 1. Sections 1–3 (stop after Table 9), and Section 9.
 2. Even though you will not need to use it in any detail, this paper does assume familiarity with the λ -calculus as a basic syntax for a programming language. You may wish to refer back to your CS 334 notes or Mitchell book for a quick refresher as you begin to read this. The Wikipedia entry for lambda calculus may also be sufficient.
- Dragon, Ch 6.3–6.3.2, 6.5.

Skim.
- “Type-Based Race Detection for Java,” Cormac Flanagan and Stephen Freund, *Proceedings of the ACM Conference on Programming Language Design and Implementation*, 2000.

As when reading any research paper, focus on the big ideas and do not get lost trying to learn every detail.
- *Foundations of Object-Oriented Languages*, Kim Bruce, 2002. Chapters 2–3.

This is background reading — You can skim it quickly, or use it as a reference for OOP-specific items as they come up in later problem sets and the programming projects.

Exercises

1. This question applies the symbol table discussion from last week's problem set to write the design of the `syntab` package for your IC compiler. This package will contain two main classes: a `SymbolTable` class that encodes all symbol information for a scope, and a `Visitor` that will traverse the AST, constructing the symbol tables and annotating each AST node with the symbol table corresponding to its scope, as described in the PA 2 handout.

Your design need not be long or contain every detail, but it should address the following items:

- Define the general interface to the `SymbolTable` class. Describe what each method does, and sketch the internal representation details. Include a description of what information is stored along-side each symbol in the tables (ie, your `SymbolTable` is really a map from identifier to what?)
- Describe the operations performed by the scope-building visitor upon entry and exit to each form of scope (e.g., global, class, method, block).
- How will the visitor create a new symbol entry when it encounters a declaration?
- When, how, and where will the scoping rules be enforced?
- Sketch the contents of your `SymbolTable` structure after processing the following program.

```

1:   class A {
2:       int x,u;
3:       void f(int y, int z) {
4:           int x;
5:           y = x + this.x;
6:           g(y-u);
7:       }
8:       void g(int y) { }
9:   }
10:
11:   class B extends A {
12:       int a,b,c;
13:       void g(int y) {
14:           if (y > 0) {
15:               int a = 2;
16:               f(y, x);
17:           }
18:       }
19:   }

```

- For each occurrence of `CLASSID` and `ID` in the IC grammar, indicate whether or not it is a defining occurrence that introduces a new name into the current scope. For those that are not, indicate the scope in which the defining occurrence of that name would occur. Use this characterization to indicate how your compiler will resolve the following symbols:
 - the first `x` on line 5.
 - `this` on line 5.
 - the second `x` on line 5.
 - `u` on line 6.
 - `x` on line 16.
- In what ways, if any, will you change the AST package to support construction of the symbol tables?

2. Using the type system of Cardelli's paper, prove that

- (a) $\emptyset, x : \text{Nat} \vdash \text{succ}(\text{succ } x) : \text{Nat}$
- (b) $\emptyset, x : \text{Nat} \vdash \text{if true then } x \text{ else } 0 : \text{Nat}$

3. For each of the following IC constructs, state whether it is well-typed in some well-formed typing context. If so, give the most general typing context in which the construct is well-typed and write the corresponding proof tree. If the construct is not well-typed in any type context, explain why. (Note that, unlike in the Cardelli system, you do not need to prove that environments and types are well-formed. You should be able to convince yourself that any type or environment that you mention is well-formed, however.)

- (a) `(new int[x.length])[x[2]]`
- (b) `if (x == v[x] && y == "true") x = y;`
- (c) `((a == b) == c) && (a == (b + "c"))`
- (d) `f(x)[x.length] = y[2]`
- (e) `if (x == a[b[x]] && y) y = b[c[x]];`

4. This problem concerns the typing and the translation of `for` loop constructs.

- (a) Suppose we extend IC to support `for` loops that operate over ranges of integer values:

`for (x from e1 to e2) s`

where `x` is an integer field or local variable (that has been declared before), and `e1`, `e2` are the loop bounds. The loop body is executed for all values of `x` in the range from `e1` to `e2`. Write a typing rule that ensures the safe execution of this loop.

- (b) It is difficult to reason about `for` loops when the execution of the loop body might change the iteration variable or the loop bounds. Describe a semantic check that would ensure this never happens.
- (c) Suppose we further augment IC with extended `for` loops, in a fashion similar to those in Java 1.5. An extended `for` loop in IC will have the following form:

`for (T x : e) s`

Here, `x` is a newly declared local variable of type `T` that iterates over all of the elements of the array `e`, and `s` is the loop body. Write an appropriate typing rule for this construct.

5. Suppose we extend IC with tuples of the following form. A tuple type is written as a sequence of types in parentheses. For example, the type `(int, bool, string)` represents a 3-tuple. The individual elements of the tuple can be accessed (i.e., read or written) in a manner similar to array elements. For example, if `x` has type `(int, bool, string)`, the expression `x[0]` has type `int`, `x[1]` has type `bool`, and `x[2]` has type `string`. Tuples are unlike arrays in that the index expression must be a constant. For simplicity, we assume that tuples don't contain class types; this ensures that different tuples cannot be subtypes of each other.

- (a) Explain why is it necessary to require that the index of a type expression must be a constant.
- (b) Write additional typing rules in the static semantics of IC for expressions and statements to support tuples.
- (c) Consider the types `T1 = (int, (int, int)[])` and `T2 = (int, (int, int))[]`. Consider a variable `x` having either type `T1` or type `T2`. Write an expression that type-checks and has the same type in both cases. Write an expression that type-checks if `x` has type `T1`, but doesn't if `x` has type `T2`. We require that `x`, `0`, and `1` are the only variables and constants in your expressions.
- (d) Syntactically, the tuple element access expression looks like an array element access expression. Will this create problems for type checking? Explain briefly.

6. Read "Type-Based Race Detection for Java" by Flanagan and Freund.

- The system extends the Java syntax with three items:
 - `guarded_by` modifiers on fields;
 - `requires` modifiers on methods; and
 - classes parameterized by locks.

What does each syntactic form capture?

- What guarantee does the type system make about well-typed programs?
- The main typing judgment is $P; E; ls \vdash e : t$. This is a fairly standard judgment form, except for the inclusion of `ls`. What is `ls`? How is it computed, and how is it used?
- Can you identify any limitations of the approach taken in the paper?