Advances in Programming Languages

APL4: Row variables in OCaml — Structural typing for objects

Ian Stark

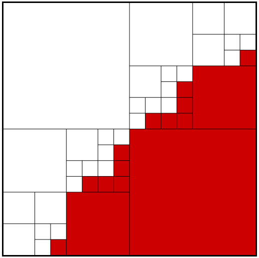
School of Informatics
The University of Edinburgh

Thursday 22 January 2009
Semester 2 Week 2

Outline

Example: Quadtrees

A *region quadtree* represents two-dimensional spatial data, such as images, with variable resolution. Where information density is nonuniform it is more efficient than a simple two-dimensional array.



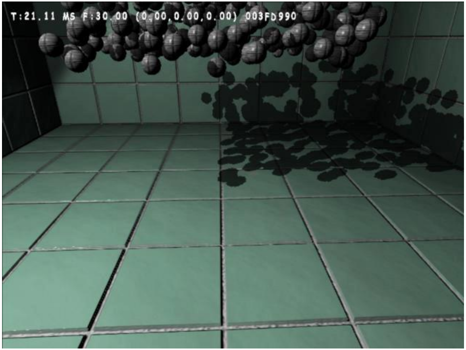
$$
\begin{aligned}
\textbf{type}\ \text{quadtree} \ = \ & \text{Clear} \\
\mid \ & \text{Black} \mid \text{White} \mid \text{Red} \mid \text{Green} \mid \text{Blue} \\
\mid \ & \text{Tree}\ \textbf{of}\ \text{quadtree} * \text{quadtree} * \text{quadtree} * \text{quadtree}
\end{aligned}
$$

$$\textbf{type}\ \text{picture} = \{\ \text{title : string};\ \text{image: quadtree}\ \}$$

Quadtree example?

Video by Handkor
http://handkor.googlepages.com/
Early test sequence for "Hippocrates's Dilemma"

Octree example



Video by Handkor
http://handkor.googlepages.com/
Collision test — 256 marbles & gravity

Outline

Subtyping arrays in Java

Java has subtyping: a value of one type may be used at any more general type. So String $\leqslant$ Object, and every String is an Object.

Not all is well with Java types

```
String[] a = { "Hello" };                // A small string array
Object[] b = a;                          // Now a and b are the same array
b[0] = Boolean.FALSE;                    // Drop in a Boolean object
String s = a[0];                         // Oh, dear
System.out.println(s.toUpperCase());     // This isn't going to be pretty
```

This compiles without error or warning: in Java, if $S \leqslant T$ then $S[] \leqslant T[]$.

Except that it isn't. So every array assignment gets a runtime check.

Typing in OO languages

Ideally, an statically-checked object-oriented language should have a type system that is

Typing in OO languages

Ideally, an statically-checked object-oriented language should have a type system that is

(a) usable, and

Typing in OO languages

Ideally, an statically-checked object-oriented language should have a type system that is

(a) usable, and

(b) correct.

Typing in OO languages

Ideally, an statically-checked object-oriented language should have a type system that is

(a) usable, and
(b) correct.

Building such type systems is a continuing challenge.

Typing in OO languages

Ideally, an statically-checked object-oriented language should have a type system that is

(a) usable, and
(b) correct.

Building such type systems is a continuing challenge.

One problem is that subtyping is crucial to OO programming, but unfortunately:

Typing in OO languages

Ideally, an statically-checked object-oriented language should have a type system that is

(a) usable, and

(b) correct.

Building such type systems is a continuing challenge.

One problem is that subtyping is crucial to OO programming, but unfortunately:

- subtyping is not inheritance; (really, it's not)

Typing in OO languages

Ideally, an statically-checked object-oriented language should have a type system that is

(a) usable, and
(b) correct.

Building such type systems is a continuing challenge.

One problem is that subtyping is crucial to OO programming, but unfortunately:

- subtyping is not inheritance; (really, it's not)
- it's also extremely hard to get right.

How hard?

Fixing object subtyping has been a busy research topic for several years.

You can see this by observing that the type declared for the max method in the Java collections class has gone from: (Java 1.2, 1998)

How hard?

Fixing object subtyping has been a busy research topic for several years.

You can see this by observing that the type declared for the max method in the Java collections class has gone from: (Java 1.2, 1998)

public static Object max(Collection coll)

which always returns an Object, whatever is stored in the collection, to:

How hard?

Fixing object subtyping has been a busy research topic for several years.

You can see this by observing that the type declared for the max method in the Java collections class has gone from: (Java 1.2, 1998)

```
public static Object max(Collection coll)
```

which always returns an Object, whatever is stored in the collection, to:

```
public static <T extends Object & Comparable<? super T>> T
  max(Collection<? extends T> coll)
```

How hard?

Fixing object subtyping has been a busy research topic for several years.

You can see this by observing that the type declared for the max method in the Java collections class has gone from: (Java 1.2, 1998)

```
public static Object max(Collection coll)
```

which always returns an Object, whatever is stored in the collection, to:

```
public static <T extends Object & Comparable<? super T>> T
  max(Collection<? extends T> coll)
```

and it might *still* throw a ClassCastException. (Java 6, 2006)

How hard?

Fixing object subtyping has been a busy research topic for several years.

You can see this by observing that the type declared for the max method in the Java collections class has gone from: (Java 1.2, 1998)

```
public static Object max(Collection coll)
```

which always returns an Object, whatever is stored in the collection, to:

```
public static <T extends Object & Comparable<? super T>> T
  max(Collection<? extends T> coll)
```

and it might *still* throw a ClassCastException. (Java 6, 2006)

This is not a criticism: the new typing is more flexible, it saves on explicit downcasts, and the Java folks do know what they are doing.

Nominal vs. Structural

Java uses predominantly *nominative* or *nominal typing*: the only relations between types are those stated explicitly by the programmer.

```
class pair1 { int x; int y; }        // Pair of integers
class pair2 { int x; int y; }        // Also a pair of integers

pair1 a = new pair1();               // Create one new pair object
pair2 b = a;                         // Assign it to another
                                     // Get an "incompatible types" error
```

This is by design:

- it can help with safe programming; and
- it certainly helps the compiler with typechecking.

Nominal vs. structural

In contrast, OCaml uses *structural typing*: the properties of types can be deduced from their structure.

```
type pair1 = int * int        (* Type abbreviation *)
type pair2 = int * int        (* An identical one *)

let a : pair1 = (5,6)         (* Create a new pair *)
let b : pair2 = a             (* Copy it to another *)
                              (* No error *)
```

If object typing is tough to sort out nominally, then how do we attempt to do it structurally?

Outline

Records and record types

OCaml provides strongly-typed *records*:

```
type picture = { title : string; image : quadtree }
let p = { title = "Look at me"; image = i }

# p.title;;
- : string = "Look at me"
```

This could be the basis for an object system; records can even have *mutable* fields to serve as instance variables.

However, field names are strictly tied to their record:

```
# fun x -> x.title;;
- : picture -> string = <fun>
```

Objects need more flexibility. Subtyping is one possibility, but there is another mechanism already available...

Parametric polymorphism

A simple type system:

$$\tau ::= \alpha \quad | \quad \tau \times \tau \quad | \quad \tau \to \tau$$
$$\sigma ::= \forall \vec{\alpha}.\tau$$

Here τ is a type, α is a *type variable* and σ is a *type scheme*.

Type schemes characterise functions that carry out the same action at a range of types, for example:

$$\lambda x.x : \forall \alpha.\alpha \to \alpha$$

This is *parametric polymorphism*, implemented in Java/C# as *generics*.

OCaml automatically infers polymorphic types where possible:

```
let id x = x;;
val id : 'a -> 'a = <fun>
```

Row variables

Add types for records, where $m_1 \dots m_k$ are labels and ρ is a *row variable*:

$$\tau ::= \alpha \quad | \quad \tau \times \tau \quad | \quad \tau \to \tau \quad | \quad \langle m_1 : \tau_1, \dots, m_k : \tau_k \mid \rho \rangle$$
$$\sigma ::= \forall \vec{\alpha} \vec{\rho}.\tau$$

We can now type functions that carry out the same action at a range of different record types. For example, using # for field selection:

$$\lambda x.(x\#m) : \forall\alpha\forall\rho.\langle m{:}\alpha|\rho\rangle \to \alpha$$

This is *row polymorphism*.

OCaml automatically infers polymorphic row types where possible:

```
let getfield p = p#m
val getfield : < m : 'a; .. > -> 'a = <fun>

let double p = p#height * 2;;
val double : < height : int; .. > -> int = <fun>
```

OCaml uses row types to represent an object as a record of methods.

```
let a =                                   (* Saving account *)
  object
    val mutable balance = 0
    method credit n = balance <- balance + n
    method enquire = balance
  end;;
val a : < credit : int -> unit; enquire : int > = <obj>
```

Automatic type inference gives the most general type for an object.

(OCaml does also have classes, for objects that share method suites.)

Different object types can share methods with the same name.

```
let b =                              (* Spending account *)
  object
    val mutable balance = 0
    method credit n = balance <- balance + n
    method debit n = balance <- balance - n
    method enquire = balance
  end;;
val b : < credit : int -> unit; debit : int -> unit; enquire : int >
    = <obj>
```

Account b has all the methods of a, and more.

(We could also use inheritance to generate one class from another.)

Objects in Ocaml

Define a function to add credit to an account.

```
let boost x = x#credit 20;;
val boost : < credit : int -> 'a; .. > -> 'a = <fun>
```

OCaml infers a very general type, so we can apply this to both existing accounts:

```
boost a; a#enquire;;
- : int = 20

boost b; b#debit 5; b#enquire;;
- : int = 15
```

It is even possible to infer a type for the function that takes a list of any type of accounts and selects the one of greatest value:

```
max : (< enquire : int; .. > as a') list -> 'a
```

Homework

- Test both the bank account objects in OCaml.
- Find out what you can about the *F#* language. This builds on a basis very similar to OCaml, but is also driven by the nature of the .NET platform. What are the differences between the F# and OCaml types for objects?
- Subtyping is also possible for records, without row variables. Find out what the difference is between *width* and *depth* subtyping.

  If you find a good source for this, then post a link on the newsgroup.

Summary

- Static typing for object-oriented programming is tricky.
- Co- and contra-variance is important in checking subtypes.
- Where Java uses *nominal* typing, OCaml uses *structural* typing.
- Row variables, and row polymorphism, allow structural typing of objects.

Benjamin C. Pierce.
Types and Programming Languages.
MIT Press, 2002