Queries With For

CS-214 Software Construction

Queries with `for`

The for notation is essentially equivalent to the common operations of query languages for databases (e.g. SQL's `SELECT ... WHERE`)

Example: Suppose that we have a database `books`, represented as a list of books.

```
case class Book(title: String, authors: List[String])
```

A Mini-Database

```
val books: List[Book] = List(
  Book(title   = "Structure and Interpretation of Computer Programs",
       authors = List("Abelson, Harald", "Sussman, Gerald J.")),
  Book(title   = "Introduction to Functional Programming",
       authors = List("Bird, Richard", "Wadler, Phil")),
  Book(title   = "Effective Java",
       authors = List("Bloch, Joshua")),
  Book(title   = "Java Puzzlers",
       authors = List("Bloch, Joshua", "Gafter, Neal")),
  Book(title   = "Programming in Scala",
       authors = List("Odersky, Martin", "Spoon, Lex", "Venners, Bill")))
```

Some Queries

To find the titles of books whose author's name is "Bird":

```
for
  b <- books
  a <- b.authors
  if a.startsWith("Bird,")
yield b.title
```

To find all the books which have the word "Program'' in the title:

```
for b <- books if b.title.indexOf("Program") >= 0
yield b.title
```

Another Query

To find the names of all authors who have written at least two books present in the database.

```
for
  b1 <- books
  b2 <- books
  if b1 != b2
  a1 <- b1.authors
  a2 <- b2.authors
  if a1 == a2
yield a1
```

Another Query

To find the names of all authors who have written at least two books present in the database.

```
for
  b1 <- books
  b2 <- books
  if b1 != b2
  a1 <- b1.authors
  a2 <- b2.authors
  if a1 == a2
yield a1
```

Why do solutions show up twice?

How can we avoid this?

Modified Query

To find the names of all authors who have written at least two books present in the database.

```
for
  b1 <- books
  b2 <- books
  if b1.title < b2.title
  a1 <- b1.authors
  a2 <- b2.authors
  if a1 == a2
yield a1
```

Problem

What happens if an author has published three books?

- The author is printed once
- The author is printed twice
- The author is printed three times
- The author is not printed at all

Problem

What happens if an author has published three books?

- O The author is printed once
- O The author is printed twice
- X The author is printed three times
- O The author is not printed at all

Modified Query (2)

Solution: We must remove duplicate authors who are in the results list twice.

This is achieved using the `distinct` method on sequences:

```
val repeated =
  for
    b1 <- books
    b2 <- books
    if b1.title < b2.title
    a1 <- b1.authors
    a2 <- b2.authors
    if a1 == a2
  yield a1
repeated.distinct
```

Modified Query

Better alternative: Compute with sets instead of sequences:

```
val bookSet = books.toSet
for
  b1 <- bookSet
  b2 <- bookSet
  if b1 != b2
  a1 <- b1.authors
  a2 <- b2.authors
  if a1 == a2
yield a1
```

Recall Translation of For (1)

The Scala compiler expresses for-expressions in terms of `map`, `flatMap` and a lazy variant of `filter`.

Here is the translation scheme used by the compiler (we limit ourselves here to simple variables in generators)

1. A simple for-expression

```
for x <- e1 yield e2
```

is translated to

```
e1.map(x => e2)
```

Translation of For (2)

2. A for-expression

```
for x <- e1 if f; s yield e2
```

where `f` is a filter and `s` is a (potentially empty) sequence of generators and filters, is translated to

```
for x <- e1.withFilter(x => f); s yield e2
```

(and the translation continues with the new expression)

You can think of `withFilter` as a variant of `filter` that does not produce an intermediate list, but instead applies the following `map` or `flatMap` function application only to those elements that passed the test.

Translation of For (3)

3. A for-expression

```
for x <- e1; y <- e2; s yield e3
```

where `s` is a (potentially empty) sequence of generators and filters, is translated into

```
e1.flatMap(x => for y <- e2; s yield e3)
```

(and the translation continues with the new expression)

Generalization of `for`

Interestingly, the translation of for is not limited to lists or sequences, or even collections;

It is based solely on the presence of the methods `map`, `flatMap` and `withFilter`.

This lets you use the for syntax for your own types as well – you must only define `map`, `flatMap` and `withFilter` for these types.

There are many types for which this is useful: arrays, iterators, databases, XML data, optional values, parsers, etc.

For and Databases

For example, `books` might not be a list, but a database stored on some server.

As long as the client interface to the database defines the methods `map`, `flatMap` and `withFilter`, we can use the `for` syntax for querying the database.

This is the basis of data base connection frameworks such as *ScalaSql*, or big data platforms such as *Spark*.

Functional Random Generators

CS-214 Software Construction

Martin Odersky

Other Uses of For-Expressions

Question: Are for-expressions tied to collection-like things such as lists, sets, or databases?

Other Uses of For-Expressions

Question: Are for-expressions tied to collection-like things such as lists, sets, or databases?

Answer: No! All that is required is some interpretation of `map`, `flatMap` and `withFilter`.

We can do without `withFilter`, but then for expressions cannot use conditions.

We have already seen these for expressions for `Future`.

There are many other domains that afford such an interpretation.

Example: random value generators.

Random Values

You know about random numbers:

```
import java.util.Random
val rand = new Random
rand.nextInt()
```

Question: What is a systematic way to get random values for other domains, such as

- booleans, strings, pairs and tuples, lists, sets, trees

?

Generators

Let's define a trait `Generator[T]` that generates random values of type `T`:

```
trait Generator[+T]:
  def generate: T
```

Some instances:

```
val integers = new Generator[Int]:
  val rand = new java.util.Random
  def generate = rand.nextInt()
```

Generators

Let's define a trait `Generator[T]` that generates random values of type `T`:

```
trait Generator[+T]:
  def generate: T
```

Some instances:

```
val booleans = new Generator[Boolean]:
  def generate = integers.generate > 0
```

Generators

Let's define a trait `Generator[T]` that generates random values of type `T`:

```
trait Generator[+T]:
  def generate: T
```

Some instances:

```
val pairs = new Generator[(Int, Int)]:
  def generate = (integers.generate, integers.generate)
```

Streamlining It

Can we avoid the `new Generator ...` boilerplate?

Ideally, we would like to write:

```
val booleans = for x <- integers yield x > 0

def pairs[T, U](t: Generator[T], u: Generator[U]) =
  for x <- t; y <- u yield (x, y)
```

What does this expand to?

Streamlining It

Can we avoid the `new Generator ...` boilerplate?

Ideally, we would like to write:

```
val booleans = integers.map(x => x > 0)

def pairs[T, U](t: Generator[T], u: Generator[U]) =
  t.flatMap(x => u.map(y => (x, y)))
```

Need `map` and `flatMap` for that!

Generator with `map` and `flatMap`

Here's a more convenient version of `Generator`:

```
trait Generator[+T]:
  def self = this

  def generate: T

  def map[S](f: T => S) = new Generator[S]:
    def generate = f(self.generate)
```

Generator with `map` and `flatMap`

Here's a more convenient version of `Generator`:

```
trait Generator[+T]:
  def self = this

  def generate: T

  def map[S](f: T => S) = new Generator[S]:
    def generate = f(self.generate)

  def flatMap[S](f: T => Generator[S]) = new Generator[S]:
    def generate = f(self.generate).generate
```

The booleans Generator

What does this definition resolve to?

```
val booleans = for x <- integers yield x > 0
```

The booleans Generator

What does this definition resolve to?

```
val booleans = for x <- integers yield x > 0
```

```
val booleans = integers.map(x => x > 0)
```

The booleans Generator

What does this definition resolve to?

```
val booleans = for x <- integers yield x > 0
```

```
val booleans = integers.map(x => x > 0)
```

```
val booleans = new Generator[Boolean]:
  def generate = ((x: Int) => x > 0)(integers.generate)
```

The booleans Generator

What does this definition resolve to?

```
val booleans = for x <- integers yield x > 0
```

```
val booleans = integers.map(x => x > 0)
```

```
val booleans = new Generator[Boolean]:
  def generate = ((x: Int) => x > 0)(integers.generate)
```

```
val booleans = new Generator[Boolean]:
  def generate = integers.generate > 0
```

The pairs Generator

```
def pairs[T, U](t: Generator[T], u: Generator[U]) =
  for x <- t; y <- u yield (x, y)
```

The pairs Generator

```
def pairs[T, U](t: Generator[T], u: Generator[U]) =
  for x <- t; y <- u yield (x, y)

def pairs[T, U](t: Generator[T], u: Generator[U]) = t.flatMap(
  x => u.map(y => (x, y)))
```

The pairs Generator

```
def pairs[T, U](t: Generator[T], u: Generator[U]) =
  for x <- t; y <- u yield (x, y)

def pairs[T, U](t: Generator[T], u: Generator[U]) = t.flatMap(
  x => u.map(y => (x, y)))

def pairs[T, U](t: Generator[T], u: Generator[U]) = t.flatMap(
  x => new Generator[(T, U)] { def generate = (x, u.generate) })
```

The pairs Generator

```
def pairs[T, U](t: Generator[T], u: Generator[U]) =
  for x <- t; y <- u yield (x, y)

def pairs[T, U](t: Generator[T], u: Generator[U]) = t.flatMap(
  x => u.map(y => (x, y)))

def pairs[T, U](t: Generator[T], u: Generator[U]) = t.flatMap(
  x => new Generator[(T, U)] { def generate = (x, u.generate) })

def pairs[T, U](t: Generator[T], u: Generator[U]) = new Generator[(T, U)]:
  def generate = new Generator[(T, U)] { def generate = (t.generate, u.generate) }
    .generate
```

The pairs Generator

```
def pairs[T, U](t: Generator[T], u: Generator[U]) =
  for x <- t; y <- u yield (x, y)

def pairs[T, U](t: Generator[T], u: Generator[U]) = t.flatMap(
  x => u.map(y => (x, y)))

def pairs[T, U](t: Generator[T], u: Generator[U]) = t.flatMap(
  x => new Generator[(T, U)] { def generate = (x, u.generate) })

def pairs[T, U](t: Generator[T], u: Generator[U]) = new Generator[(T, U)]:
  def generate = new Generator[(T, U)] { def generate = (t.generate, u.generate) }
    .generate

def pairs[T, U](t: Generator[T], u: Generator[U]) = new Generator[(T, U)]:
  def generate = (t.generate, u.generate)
```

Generator Examples

```
def single[T](x: T): Generator[T] = new Generator[T]:
  def generate = x

def range(lo: Int, hi: Int): Generator[Int] =
  for x <- integers yield lo + x % (hi - lo)

def oneOf[T](xs: T*): Generator[T] =
  for idx <- range(0, xs.length) yield xs(idx)
```

A `List` Generator

A list is either an empty list or a non-empty list.

```
def lists: Generator[List[Int]] =
  for
    isEmpty <- booleans
    list <- if isEmpty then emptyLists else nonEmptyLists
  yield list
```

A `List` Generator

A list is either an empty list or a non-empty list.

```
def lists: Generator[List[Int]] =
  for
    isEmpty <- booleans
    list <- if isEmpty then emptyLists else nonEmptyLists
  yield list

def emptyLists = single(Nil)
```

A `List` Generator

A list is either an empty list or a non-empty list.

```
def lists: Generator[List[Int]] =
  for
    isEmpty <- booleans
    list <- if isEmpty then emptyLists else nonEmptyLists
  yield list

def emptyLists = single(Nil)

def nonEmptyLists =
  for
    head <- integers
    tail <- lists
  yield head :: tail
```

A `Tree` Generator

Can you implement a generator that creates random `Tree` objects?

```
enum Tree:
  case Inner(left: Tree, right: Tree)
  case Leaf(x: Int)
```

Application: Random Testing

You know about unit tests:

- Come up with some some test inputs to program functions and a *postcondition*.
- The postcondition is a property of the expected result.
- Verify that the program satisfies the postcondition.

Property-based testing does away with the need for test inputs.

Instead, it relies on generating *random test inputs*.

Random Test Function

Using generators, we can write a random test function:

```
def test[T](g: Generator[T], numTimes: Int = 100)
           (test: T => Boolean): Unit =
  for i <- 0 until numTimes do
    val value = g.generate
    assert(test(value), s"test failed for $value")
  println(s"passed $numTimes tests")
```

Random Test Function

Example usage:

```
test(pairs(lists, lists)): (xs, ys) =>
  (xs ++ ys).length > xs.length
```

Question: Does the above property always hold?

0 Yes

0 No

Random Test Function

Example usage:

```
test(pairs(lists, lists)): (xs, ys) =>
  (xs ++ ys).length > xs.length
```

Question: Does the above property always hold?

O Yes

X No

ScalaCheck

Shift in viewpoint: Instead of writing tests, write *properties* that are assumed to hold.

This idea is implemented in the `ScalaCheck` tool.

```
forAll: (l1: List[Int], l2: List[Int]) =>
  l1.size + l2.size == (l1 ++ l2).size
```

It can be used either stand-alone or as part of ScalaTest.

Here, the additional idea is that we make generators given instances, so they are provided *automatically* based on the tested types.

Users can override with explicit generators, but they can also rely on the given defaults.

Monads

CS-214 Software Construction

Martin Odersky

Monads

Data structures with `map` and `flatMap` seem to be quite common.

In fact there's a name that describes this class of a data structures together with some algebraic laws that they should have.

They are called *monads*.

What is a Monad?

A monad `M` is a parametric type M[T] with two operations, `flatMap` and `unit`, that have to satisfy some laws.

```
trait M[T]:
  def flatMap[U](f: T => M[U]): M[U]

def unit[T](x: T): M[T]
```

In the literature, `flatMap` is often called `bind`.

Examples of Monads

- `List` is a monad with `unit(x) = List(x)`
- `Set` is monad with `unit(x) = Set(x)`
- `Option` is a monad with `unit(x) = Some(x)`
- `Generator` is a monad with `unit(x) = single(x)`

`flatMap` is an operation on each of these types, whereas `unit` in Scala is different for each monad.

Monads and map

`map` can be defined for every monad as a combination of `flatMap` and `unit`:

```
m.map(f)  ==  m.flatMap(x => unit(f(x)))
          ==  m.flatMap(f.andThen(unit))
```

Note: `andThen` is defined function composition in the standard library.

Monad Laws

To qualify as a monad, a type has to satisfy three laws:

Associativity:

```
m.flatMap(f).flatMap(g)  ==  m.flatMap(f(_).flatMap(g))
```

Left unit

```
unit(x).flatMap(f)  ==  f(x)
```

Right unit

```
m.flatMap(unit)  ==  m
```

Checking Monad Laws

Let's check the monad laws for Option.

Here's `flatMap` for `Option`:

```
abstract class Option[+T]:
  def flatMap[U](f: T => Option[U]): Option[U] = this match
    case Some(x) => f(x)
    case None => None
```

Checking the Left Unit Law

Need to show: `Some(x).flatMap(f) == f(x)`

```
Some(x).flatMap(f)
```

Checking the Left Unit Law

Need to show: `Some(x).flatMap(f) == f(x)`

```
      Some(x).flatMap(f)

==    Some(x) match
        case Some(x) => f(x)
        case None => None
```

Checking the Left Unit Law

Need to show: `Some(x).flatMap(f) == f(x)`

```
      Some(x).flatMap(f)

==    Some(x) match
        case Some(x) => f(x)
        case None => None

==    f(x)
```

Checking the Right Unit Law

Need to show: `opt.flatMap(Some)  ==  opt`

```
opt.flatMap(Some)
```

Checking the Right Unit Law

Need to show: `opt.flatMap(Some) == opt`

```
      opt.flatMap(Some)

==    opt match
        case Some(x) => Some(x)
        case None => None
```

Checking the Right Unit Law

Need to show: `opt.flatMap(Some) == opt`

```
      opt.flatMap(Some)

==    opt match
        case Some(x) => Some(x)
        case None => None

==    opt
```

Checking the Associative Law

Need to show: `opt.flatMap(f).flatMap(g)` == `opt.flatMap(f(_).flatMap(g))`

`opt.flatMap(f).flatMap(g)`

Checking the Associative Law

Need to show: `opt.flatMap(f).flatMap(g)` == `opt.flatMap(f(_).flatMap(g))`

```
      opt.flatMap(f).flatMap(g)

==   (opt match { case Some(x) => f(x) case None => None })
          match { case Some(y) => g(y) case None => None }
```

Checking the Associative Law

Need to show: `opt.flatMap(f).flatMap(g)` == `opt.flatMap(f(_).flatMap(g))`

```
      opt.flatMap(f).flatMap(g)

==    (opt match { case Some(x) => f(x) case None => None })
           match { case Some(y) => g(y) case None => None }

==    opt match
        case Some(x) =>
          f(x) match { case Some(y) => g(y) case None => None }
        case None =>
          None match { case Some(y) => g(y) case None => None }
```

Checking the Associative Law (2)

```
==   opt match
       case Some(x) =>
         f(x) match { case Some(y) => g(y) case None => None }
       case None => None
```

Checking the Associative Law (2)

```
==   opt match
       case Some(x) =>
         f(x) match { case Some(y) => g(y) case None => None }
       case None => None

==   opt match
       case Some(x) => f(x).flatMap(g)
       case None => None
```

Checking the Associative Law (2)

```
==   opt match
       case Some(x) =>
         f(x) match { case Some(y) => g(y) case None => None }
       case None => None

==   opt match
       case Some(x) => f(x).flatMap(g)
       case None => None

==   opt.flatMap(x => f(x).flatMap(g))
```

Checking the Associative Law (2)

```
==   opt match
       case Some(x) =>
         f(x) match { case Some(y) => g(y) case None => None }
       case None => None

==   opt match
       case Some(x) => f(x).flatMap(g)
       case None => None

==   opt.flatMap(x => f(x).flatMap(g))

==   opt.flatMap(f(_).flatMap(g))
```

Significance of the Laws for For-Expressions

We have seen that monad-typed expressions are typically written as `for` expressions.

What is the significance of the laws with respect to this?

1. Associativity says essentially that one can "inline" nested for expressions:

```
    for
      y <- for x <- m; y <- f(x) yield y
      z <- g(y)
    yield z

==  for
      x <- m
      y <- f(x)
      z <- g(y)
    yield z
```

Significance of the Laws for For-Expressions

2. Right unit says:

```
      for x <- m yield x

==    m
```

3. Left unit does not have an analogue for for-expressions.

Another type: Try

You gave seen `Try`: it resembles `Option`, but instead of `Some/None` there is a `Success` case with a value and a `Failure` case that contains an exception:

```
abstract class Try[+T]
case class Success[+T](x: T)       extends Try[T]
case class Failure(ex: Exception) extends Try[Nothing]
```

`Try` is used to pass results of computations that can fail with an exception between threads and processes.

Creating a Try

You can wrap up an arbitrary computation in a `Try`.

```
Try(expr)     // gives Success(someValue) or Failure(someException)
```

Here's an implementation of `Try.apply`:

```
object Try:
  def apply[T](expr: => T): Try[T] =
    try Success(expr)
    catch case NonFatal(ex) => Failure(ex)
```

`NonFatal` is a condition that catches all exceptions except a few ones that cannot be recovered from, such as `StackOverflowException`.

Composing `Try`

Just like with `Option`, `Try`-valued computations can be composed in for-expressions.

```
for
  x <- computeX
  y <- computeY
yield f(x, y)
```

If `computeX` and `computeY` succeed with results `Success(x)` and `Success(y)`, this will return `Success(f(x, y))`.

If either computation fails with an exception `ex`, this will return `Failure(ex)`.

Definition of `flatMap` and `map` on `Try`

```
abstract class Try[T]:

  def flatMap[U](f: T => Try[U]): Try[U] = this match
    case Success(x) => try f(x) catch case NonFatal(ex) => Failure(ex)
    case fail: Failure => fail

  def map[U](f: T => U): Try[U] = this match
    case Success(x) => Try(f(x))
    case fail: Failure => fail
```

So, for a `Try` value `t`,

```
t.map(f)  ==  t.flatMap(x => Try(f(x)))
          ==  t.flatMap(f.andThen(Try))
```

Exercise

It looks like `Try` might be a monad, with `unit` = `Try`.

Is it?

- `Yes`
- `No, the associative law fails`
- `No, the left unit law fails`
- `No, the right unit law fails`
- `No, two or more monad laws fail.`

Solution

It turns out the left unit law fails.

```
Try(expr).flatMap(f)  !=  f(expr)
```

Indeed the left-hand side will never raise a non-fatal exception whereas the right-hand side will raise any exception thrown by `expr` or `f`.

Hence, `Try` trades one monad law for another law which is more useful in this context:

> *An expression composed from 'Try', 'map', 'flatMap' will never throw a non-fatal exception.*

Call this the "bullet-proof" principle.

Conclusion

We have seen that for-expressions are useful not only for collections.

Many other types also define `map`,`flatMap`, and `withFilter` operations and with them for-expressions.

Examples: `Generator`, `Option`, `Try`.

Many of the types defining `flatMap` are monads.

(If they also define `withFilter`, they are called "monads with zero").

The three monad laws give useful guidance in the design of library APIs.