Hands-on Intermediate Python

@__mharrison__
http://hairysun.com

About Me

Co-chair Utah Python. Consultant with 14 years Python experience across ML, Data Science, BI, Web, Open Source Stack Management, and Search.

http://hairysun.com/ http://metasnake.com/

TREADING ON PYTHON SERIES

Intermediate Python Programming

Learn Decorators, Generators, Functional Programming & more

Matt Harrison

METASNAKE

@__mharrison__

Begin Intermediate

Impetus

You can get by in Python with basic constructs ...

Impetus (2)

But you might:

- get bored
- be confused by others' code
- be less efficient

Warning

- Starting from basic Python knowledge
- Hands on
  - (short) lecture
  - (short) code
  - repeat

Python 2 or 3?

Most of this is agnostic. I'll note the differences.
Labs work with either.

Outline

- Testing
- Functional Programming
- Functions
- Decorators
- Class Decorators
- Properties
- Iteration
- Generators
- Context Managers

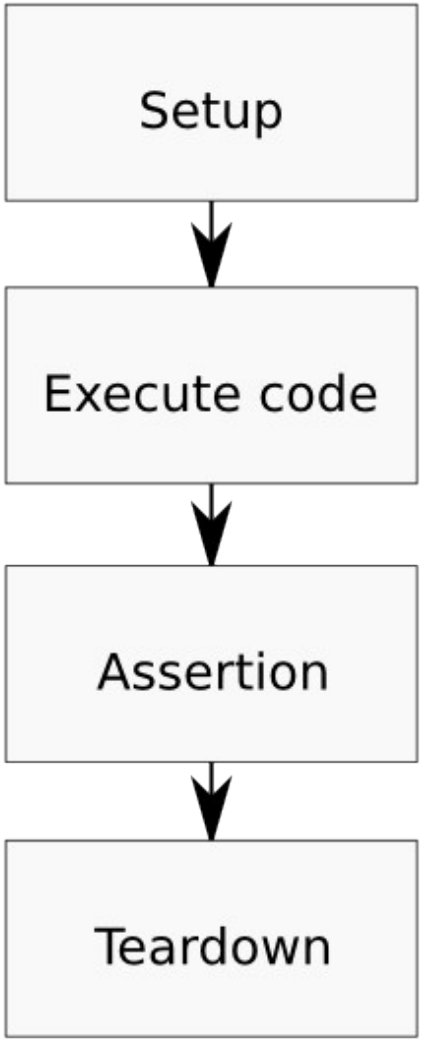
unittest

(Python 2.1)

`unittest`

Implements Kent Beck's *xUnit* paradigm

xUnit workflow



```
import unittest
import integr
class TestIntegr(unittest.TestCase):
    def setup(self):
        # setup is called *before* each test is run
        # if I need to adjust to a well known state before starting
        # I can do that here
        pass
    def teardown(self):
        # teardown is called *after* the each test is run
        pass
    def test_basic(self):
        # any method beginning with "test" is a test
        results = integr.parse('1,3,4')
        self.assertEqual(results, [1,3,4])
if __name__ == '__main__':
    unittest.main()
```

```
class TestIntegr(unittest.TestCase):
    ...
if __name__ == '__main__':
    unittest.main()
```

```
def setup(self):
    # setup is called *before*
    # each test is run.
    # Adjust to a well known
    # state before each test.
    pass
```

```
def test_basic(self):
    # any method beginning with
    # "test" is a test
    results = integr.parse('1,3,4')
    self.assertEqual(results, [1,3,4])
```

```
def teardown(self):
    # teardown is called *after*
    # the each test is run
    pass
```

Assertion Methods

Method signature	Explanation
`assertTrue(expression, [message])`	Complains if expression is False
`assertEqual(this, that, [message])`	Complains if this != that
`assertNotEqual(this, that, [message])`	Complains if this == that
`assertRaises(exception, callable, *args, **kwargs)`	Complains if `callable(*args, **kwargs)` does not raise `exception`

assertRaises example

```
def test_spaces(self):
    with self.assertRaises(ZeroDivisionError) as cm:
        result = 1/0
    # py3 division by zero
    # py2 integer division or modulo by zero
    self.assertTrue('by zero' in cm.exception.args[0])
```

Critique of `unittest`

- **Cons**
  - Modeled after java, why classes? (inheritance/abstraction bad)
- **Pros**
  - In the standard library
  - Straightforward

Multi-paradigmatic

Imperative Programming

Using statements to affect a program's state

Imperative Programming

```
>>> total = 0
>>> for i in range(10):
...     total += i


>>> total
45
```

Object Oriented Programming

Using objects and methods to affect a program's state

Object Oriented Programming

```
>>> class Summer:
...     def __init__(self):
...         self.sum = 0
...     def add(self, num):
...         self.sum = self.sum + num

>>> s = Summer()
>>> for num in range(10):
...     s.add(num)

>>> s.sum
45
```

Declarative Programming

```sql
SELECT *
FROM sales
WHERE store_id = 5;
```

Functional Programming

Change state by applying functions, avoiding state, side effects and mutable data

Functional Programming

```
>>> import operator
>>> reduce(operator.add, range(10))
45
```

Functional Programming

```
>>> sum(range(10))
45
```

Functional Programming

(Python 1.4)

Functional Programming

Change state by applying functions, avoiding state, side effects and mutable data:

```
>>> sum(range(10))
45
```

Imperative Programming

Using statements to affect a program's state:

```
>>> total = 0
>>> for i in range(10):
...     total += i

>>> total
45
```

First-class functions

Functions are treated as data. They can be passed around, not just invoked.

Higher-order functions

Function that accept functions as parameters or returns a function.

Pure functions

- Always produces the same result (ie not accessing global state)
- No side effects (writing to disk, mutating global state, etc)

Pure functions (2)

Pure: `math.cos`

Impure: `print`, `random.random`

Tail call optimization

Optimization for recursion to not create a new stack. Python does not have it (Guido says no).

lambda

Create simple functions in a single line:

```
>>> def mul(a, b):
...     return a * b

>>> mul_2 = lambda a, b: a*b

>>> mul_2(4, 5) == mul(4,5)
True
```

`lambda` examples

Useful for `key` and `cmp` when sorting

lambda key example

```
>>> data = [dict(number=x) for x in '019234']

>>> data.sort(key=lambda x: float(x['number']))

>>> data   # doctest: +NORMALIZE_WHITESPACE
[{'number': '0'}, {'number': '1'}, {'number': '2'},
{'number': '3'}, {'number': '4'}, {'number': '9'}]
```

lambda cmp example

```
>>> data = [dict(number=x) for x in '019234']

>>> data.sort(cmp=lambda x,y: cmp(x['number'], y['number']))

>>> data   # doctest: +NORMALIZE_WHITESPACE
[{'number': '0'}, {'number': '1'}, {'number': '2'},
{'number': '3'}, {'number': '4'}, {'number': '9'}]
```

Hint: Use `key` not `cmp`

`lambda` parameters

Supports:

- normal
- named
- `*args`
- `**kwargs`

`lambda` *expressions*

Statements cause problems:

```
>>> is_pos = lambda x: if x >=0: 'pos'
Traceback (most recent call last):
  ...
    is_pos = lambda x: if x >=0: 'pos'
                       ^
SyntaxError: invalid syntax
```

`lambda` *expressions* (2)

(Conditional) expressions don't:

```
>>> is_pos = lambda x: 'pos' if x >= 0 else 'neg'
>>> is_pos(3)
'pos'
```

See PEP 308

`lambda` *expressions* (3)

Simple rule for *expressions*: Something that could be returned from a function:

```
def func(args):
    return expression
```

Std lib example

from `cookielib.py`

```
# add cookies in order of most specific
# (ie. longest) path first
cookies.sort(key=lambda arg: len(arg.path),
    reverse=True)
```

`lambda` *expressions* (5)

Good for one-liners

map

Higher-order function that applies a function to items of a sequence:

```
>>> map(str, [0, 1, 2])
['0', '1', '2']
```

map (2)

With a `lambda`:

```
>>> pos = lambda x: x >= 0
>>> map(pos, [-1, 0, 1, 2])
[False, True, True, True]
```

Std lib example

from `tarfile.py`:

```
def namelist(self):
  return map(lambda m: m.name,
self.infolist())
```

`map` (3)

In Python 3, `map` is not a function but a lazy class:

```
>>> map(str, range(10))
<map object at 0x7fa285727b90>
>>> next(_)
'0'
```

`map` (4)

Use `itertools.imap` in Python 2 to apply to an infinite sequence (generator)

reduce

Apply a function to pairs of the sequence:

```
>>> import operator
>>> reduce(operator.mul, [1,2,3,4])
24 # ((1 * 2) * 3) * 4
```

`reduce` (2)

Reduce moved to `functools` module in Python 3.
Unlike `map`, still a function and not lazy.

Std lib example

from `csv.py`. Guessing the quote character:

```
quotechar = reduce(lambda a, b, quotes=quotes:
                   (quotes[a] > quotes[b]) and
                    a or b, quotes.keys())
```

Lambda equivalent:

```
if quotes[a] > quotes[b]:
    return a
return b
```

Std lib example

from Python 3 `csv.py`. Guessing the quote character:

```
quotechar = max(quotes, key=quotes.get)
```

`reduce` (4)

Note the `lambda` uses a trick. Named parameter to pass in `quotes`.

filter

Takes a function and a sequence. Return a sequence items for which `function(item)` is `True`:

```
>>> filter(lambda x:x >= 0, [0, -1, 3, 4, -2])
[0, 3, 4]
```

`filter` (2)

Lazy in Python 3. Use `itertools.ifilter` in Python 2 for infinite sequences.

Std lib example

from `tarfile.py`:

```
def infolist(self):
    return filter(
        lambda m: m.type in REGULAR_TYPES,
        self.tarfile.getmembers())
```

Notes about “functional” programming in *Python*

- `sum` or `for` loop can replace `reduce`
- List comprehensions replace `map` and `filter`
- No tail call optimization (means limit on recursion depth)

Example Assignment

`sample.py`

Assignment Notes

- Use spaces not tabs (PEP 8)
- define functions as globals

Assignment

`functional.py`

More about functions

a function is an instance of a `function`

```
>>> def foo():
...     'docstring for foo'
...     print 'invoked foo'

>>> foo #doctest: +ELLIPSIS
<function foo at ...>
```

a function is `callable`

```
>>> callable(foo)
True
```

function invocation

Just add `()`:

```
>>> foo()
invoked foo
```

a function has attributes

```
>>> foo.__name__
'foo'

>>> foo.__doc__
'docstring for foo'
```

(PEP 234 Python 2.1)

function scope

A function knows about itself:

```
>>> def foo2():
...     print "NAME", foo2.__name__

>>> foo2()
NAME foo2
```

function attributes

Can attach data to function prior to invocation:

```
>>> def foo3():
...     print foo3.stuff

>>> foo3.stuff = "Data"

>>> foo3()
Data
```

function definition

```
def func_name(arg1, arg2=value,
              *args, **kwargs):
    """docstring"""
    # implementation
```

Parameter types

- No parameters
- standard parameters (many)
- default/keyword/named parameters (many)
- variable parameters (one), preceded by *
- variable keyword parameters (one), preceded by **

Standard/named parameters

```
>>> def param_func(a, b=2, c=5):
...     print [a, b, c]

>>> param_func(2)
[2, 2, 5]

>>> param_func(3, 4, 5)          # override defaults
[3, 4, 5]

>>> param_func(c=4, b=5, a=6)    # order changed if name
[6, 5, 4]                        # provided
```

a gotcha

When the function is created (usually module import time), the named/default parameters values are assigned to the function (`.func_defaults`, `__defaults__` in Python 3)

named parameters

Don't default to mutable types.

```
>>> def named_param(a, foo=[]):
...     if not foo:
...         foo.append(a)

>>> named_param.func_defaults
([],)

>>> named_param(1)
>>> named_param.func_defaults
([1],)
```

mutable types

lists and *dicts* are mutable. When you modify them you don't create a new list (or dict). *Strings* and *ints* are immutable.

Parameters are evaluated when the `def` they belong to is evaluated. This usually happens at module import.

named parameters (2)

Don't default to mutable types.

```
>>> def named_param(a, foo=None):
...     foo = foo or []
...     if not foo:
...         foo.append(a)
```

`*args` and `**kwargs`

`*args` (variable parameters) is a *tuple* of parameters values.

`**kwargs` (keyword parameters) is a *dictionary* of name/value pairs.

Only one of each type. Naming above is standard convention

*args

```
>>> def demo_args(*args):
...     print type(args), args

>>> demo_args()
<type 'tuple'> ()

>>> demo_args(1)          # Note type
<type 'tuple'> (1,)

>>> demo_args(3, 'foo')
<type 'tuple'> (3, 'foo')
```

`*args` (2)

The `*` before a sequency *parameter* in an invocation “unpacks” (or splats) the sequence

*args (3)

```
>>> args = [1, 2, 3]

>>> demo_args(args[0], args[1], args[2])
<type 'tuple'> (1, 2, 3)

>>> demo_args(*args)          # same as above
<type 'tuple'> (1, 2, 3)

>>> demo_args(args)           # only 1 arg passed in
<type 'tuple'> ([1, 2, 3],)   # List(!) in a tuple
```

*args (4)

```
>>> def add3(a, b, c):     # No *args! yet...
...     return a + b + c

>>> add3(4, 5, 6)
15

>>> add3(*[4, 5, 6])   # unpack list
15
```

**kwargs

```
>>> def demo_kwargs(**kwargs):
...     print type(kwargs), kwargs

>>> demo_kwargs()
<type 'dict'> {}

>>> demo_kwargs(one=1)
<type 'dict'> {'one': 1}

>>> demo_kwargs(one=1, two=2)
<type 'dict'> {'two': 2, 'one': 1}
```

`**kwargs` (2)

The ** before a dict *parameter* in an invocation "unpacks" (or splats) the dict

**kwargs (3)

```
>>> from math import pow
>>> def distance(x1, y1, x2, y2):
...     return pow(pow(x1-x2, 2) +
...                pow(y1-y2, 2), .5)
```

**kwargs (3)

```
>>> points = {'x1':1, 'y1':1,
...           'x2':4, 'y2':5}

>>> # these calls are the same
>>> distance(**points)
5.0

>>> distance(x1=1, y1=1, x2=4, y2=5)
5.0
```

*args and **kwargs

```
>>> def demo_params(normal, kw="Test", *args, **kwargs):
...     print normal, kw, args, kwargs
>>> args = (0, 1, 2)
>>> kw = {'foo': 3, 'bar': 4}

>>> demo_params(*args, **kw)   # args can migrate
0 1 (2,) {'foo': 3, 'bar': 4}

>>> demo_params(args[0], args[1], args[2],
...                foo=3, bar=4)
0 1 (2,) {'foo': 3, 'bar': 4}
```

`*args` and `**kwargs` (2)

Python website [1] has gory details

[1] See http://docs.python.org/reference/expressions.html#calls

Assignment

`funcargs.py`

Closures

(PEP 227 Python 2.1)

Closure Definitions

First-class function with free variables that are bound by the lexical environment

Wikipedia

In Python functions can return new functions. The inner function is a *closure* and any variable it accesses that are defined outside of that function are *free variables*.

me

Closures (2)

Useful as function generators:

```
>>> def add_x(x):
...     def adder(num):
...         # we have read access to x
...         return num + x
...     return adder

>>> add_5 = add_x(5)
>>> add_5 #doctest: +ELLIPSIS
<function adder at ...>
>>> add_5(10)
15

>>> add_bang = add_x("!")
>>> add_bang("hello world")
'hello world!'
```

Closures (3)

Notice the function attributes:

```
>>> add_5.__name__
'adder'
```

Closures (4)

Nested functions only have write access to global and local scope (Python 2.x):

```
>>> X = 3
>>> def outer():
...     X = 4    # now local
...     y = 2
...     def inner():
...         global X
...         X = 5
...     print X
...     inner()
...     print X

>>> outer()
4
4
>>> X
5
```

Closures (5)

Python 3.x has `nonlocal` scope to access variables in outer functions

```
>>> X = 3         # run with python3
>>> def outer():
...     X = 4     # now local
...     y = 2
...     def inner():
...         nonlocal X  # 4
...         X = 5
...     print X
...     inner()
...     print X

>>> outer()
4
5
>>> X
3
```

Assignment

`closures.py`

Decorators

(PEP 318, 3129, Python 2.4)

Decorators

Functions are first class objects! You can wrap them to alter behavior

Decorators (2)

Allow you to

- modify arguments
- modify function
- modify results

Uses for decorators

- caching
- monkey patching stdio
- jsonify
- logging time in function call

Decorator Definition

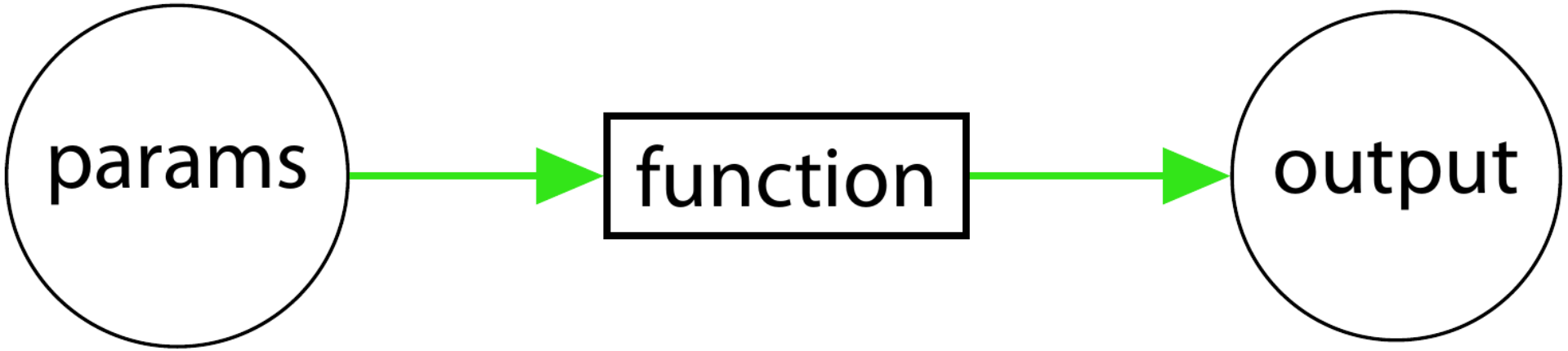
[A]llows behavior to be added to an individual object, either statically or dynamically, without affecting the behavior of other objects from the same class.

Wikipedia

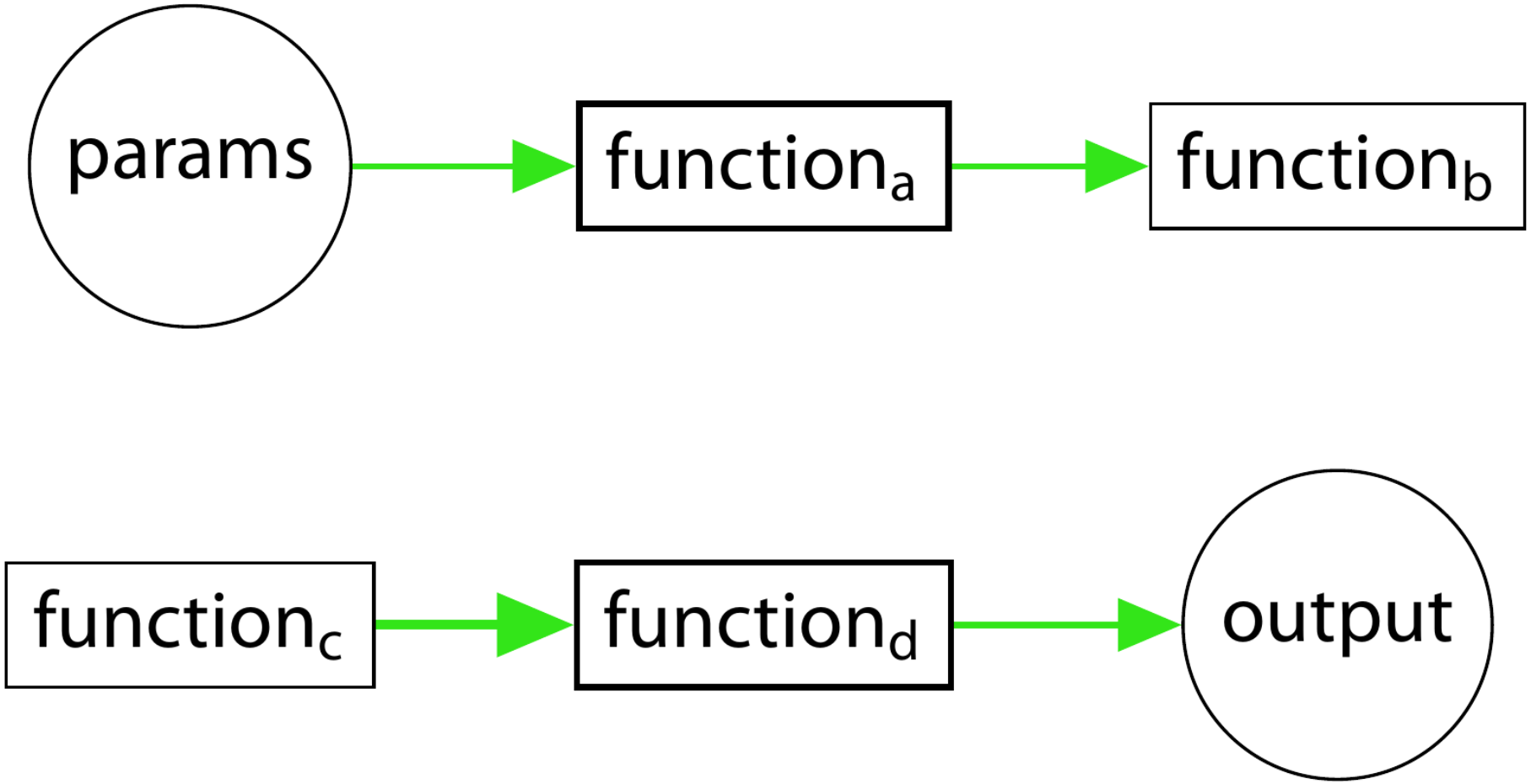
A *callable* that accepts a *callable* and returns a *callable*

me

Function Takes parameters as input returns output

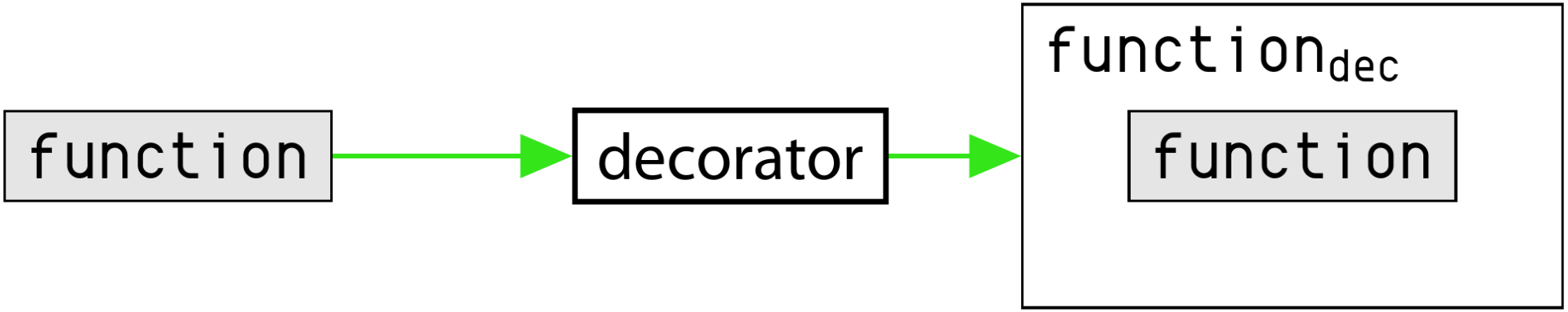


Higher-Order Function **Accepts or returns functions**

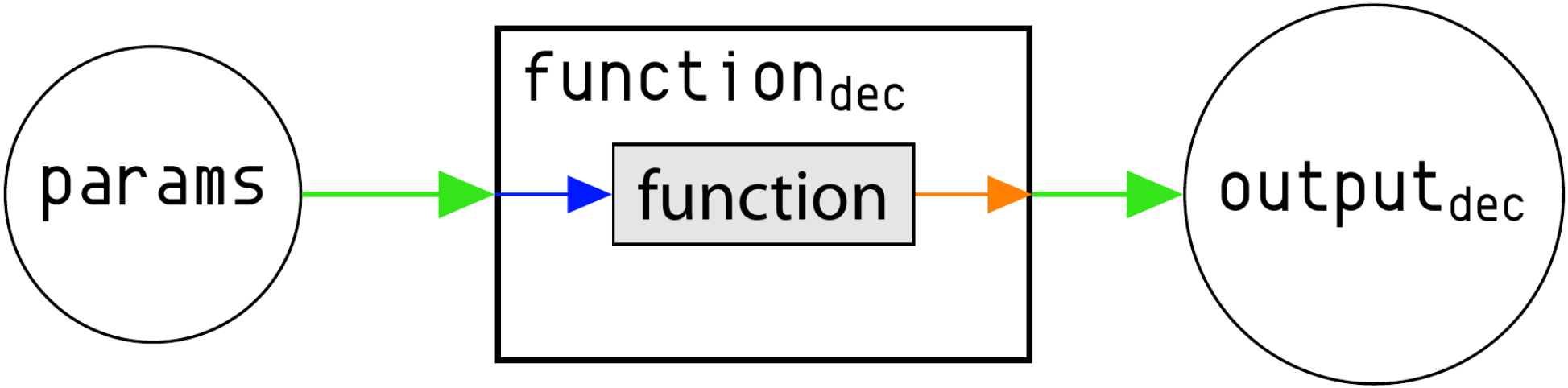


Decorator Takes function as input, returns a function

Decorator Takes function as input, returns a function



Decorated Function can do something before or after



Identity Decorator

```
>>> def iden(func):
...     return func

>>> def add(x, y):
...     return x + y

>>> add = iden(add)

>>> add(3, 4)
7
```

Identity Decorator (2)

```
>>> def iden(func):
...     def wrapper(*args, **kwargs):
...         # before
...         res = func(*args, **kwargs)
...         # after
...         return res
...     return wrapper

>>> def add(x, y):
...     return x + y

>>> add = iden(add)

>>> add(3, 4)
7
```

Decorators

Count how many times a function is called.
Create a decorator—`count`:

```
>>> call_count = 0
>>> def count(func):
...     def wrapper(*args, **kwargs):
...         global call_count
...         call_count += 1
...         return func(*args, **kwargs)
...     return wrapper
```

Decorators (4)

Create a function:

```
>>> def hello():
...     print 'invoked hello'
```

"Decorate" the `hello` function:

```
>>> hello = count(hello)
```

Decorators (5)

Test it:

```
>>> hello()
invoked hello
>>> call_count
1

>>> hello()
invoked hello
>>> call_count
2
```

Syntactic Sugar

```
>>> @count
... def hello():
...     print 'hello'
```

equivalent to:

```
>>> hello = count(hello)
```

Syntactic Sugar(2)

Don't add parens to decorator:

```
>>> @count()   # notice parens
... def hello():
...     print 'hello'
Traceback (most recent call last):
  ...
TypeError: count() takes exactly 1 argument (0
given)
```

Same as `hello = count()(hello)`

There was a problem with `count`

Better decorator

Attach data to `wrapper`:

```
>>> def count2(func):
...     def wrapper(*args, **kwargs):
...         wrapper.call_count += 1
...         return func(*args, **kwargs)
...     wrapper.call_count = 0
...     return wrapper
```

Better decorator(2)

```
>>> @count2
... def bar():
...     "my docstring"
...     pass

>>> bar(); bar()
>>> print bar.call_count
2

>>> @count2
... def snoz():
...     pass

>>> snoz()
>>> print snoz.call_count
1
```

Another problem

```
>>> bar.__name__
'wrapper'
>>> bar.__doc__
```

Better decorator (2)

Update `__name__` and `__doc__` (or use `@functools.wraps`):

```
>>> def count3(func):
...     def wrapper(*args, **kwargs):
...         wrapper.call_count += 1
...         return func(*args, **kwargs)
...     wrapper.call_count = 0
...     wrapper.__name__ = func.__name__
...     wrapper.__doc__ = func.__doc__
...     return wrapper
```

Decorator Template

```
>>> import functools
>>> def decorator(func_to_decorate):
...     @functools.wraps(func_to_decorate)
...     def wrapper(*args, **kwargs):
...         # do something before invocation
...         result = func_to_decorate(*args,
**kwargs)
...         # do something after
...         return result
...     return wrapper
```

Std lib Example

From `contextlib.py`:

```
def contextmanager(func):
    @wraps(func)
    def helper(*args, **kwds):
        return GeneratorContextManager(
            func(*args, **kwds))
    return helper
```

How would we parameterize a decorator?

Parameterized decorators (need 2 closures)

```
>>> def limit(length):
...     def decorator(function):
...         def wrapper(*args, **kwargs):
...             result = function(*args, **kwargs)
...             result = result[:length]
...             return result
...         return wrapper
...     return decorator

>>> @limit(5) # notice parens
... def echo(foo): return foo

>>> echo('123456')
'12345'
```

Parameterized decorators

```
>>> @limit(5)
... def echo(foo): return foo
```

syntactic sugar for:

```
>>> echo = limit(5)(echo)
```

decorator tidying

function attributes get mangled:

```
>>> def echo2(input):
...     """return input"""
...     return input
>>> echo2.__doc__
'return input'
>>> echo2.__name__
'echo2'

>>> echo3 = limit(3)(echo2)
>>> echo3.__doc__    # empty!!!
>>> echo3.__name__
'wrapper'
```

decorator tidying (2)

```
>>> def limit(length):
...     def decorator(function):
...         def wrapper(*args, **kwargs):
...             result = function(*args, **kwargs)
...             result = result[:length]
...             return result
...         wrapper.__doc__ = function.__doc__
...         wrapper.__name__ = function.__name__
...         return wrapper
...     return decorator

>>> echo4 = limit(3)(echo2)
>>> echo4.__doc__
'return input'
>>> echo4.__name__
'echo2'
```

decorator tidying (3)

```
>>> import functools
>>> def limit(length):
...     def decorator(function):
...         @functools.wraps(function)
...         def wrapper(*args, **kwargs):
...             result = function(*args, **kwargs)
...             result = result[:length]
...             return result
...         return wrapper
...     return decorator

>>> echo5 = limit(3)(echo2)
>>> echo5.__doc__
'return input'
>>> echo5.__name__
'echo2'
```

Django Example

From `django/views/decorators/http.py`:

```
def require_http_methods(request_method_list):
    """
Decorator to make a view only accept particular request methods. Usage::

@require_http_methods(["GET", "POST"])
def my_view(request):
# I can assume now that only GET or POST requests make it this far
# ...

Note that request methods should be in uppercase.
"""
    def decorator(func):
        @wraps(func, assigned=available_attrs(func))
        def inner(request, *args, **kwargs):
            if request.method not in request_method_list:
                logger.warning('Method Not Allowed (%s): %s',
                    request.method, request.path,
                    extra={
                        'status_code': 405,
                        'request': request
                    }
                )
                return HttpResponseNotAllowed(request_method_list)
            return func(request, *args, **kwargs)
        return inner
    return decorator
```

Multiple decorators

```
>>> @count
... @limit(4)
... def long_word():
...     return "supercalafrag"
>>> long_word()
'supe'
```

equivalent to:

```
>>> long_word = count(limit(4)(long_word))
```

Decorator rehash

Allows you to:

- Before function invocation
  - modify arguments
  - modify function
- After function invocation
  - modify results

What if I want to tweak decoration parameters at runtime?

(ie `@limit(4)` instead of `@limit(3)`)

Tweak Parameters

- Use class instance decorator
- Tweak `wrapper` attributes
- or

Don't decorate

Since a decorator is just a closure, you can invoke at run time. Like this:

```
result = limit(4)(echo2)('input')
```

Another option

Context managers let you dictate before and after conditions of execution. It is possible to create decorators that serve as context managers.

Assignment

`decorators.py`

Class Decorators

(PEP 3129, Python 2.6)

Class Decorators

A callable that takes a class and returns a class

Class Decorators (2)

```
>>> def shoutclass(cls):
...     def shout(self):
...         print self.__class__.__name__.upper()
...     cls.shout = shout
...     return cls

>>> @shoutclass
... class Loud: pass

>>> loud = Loud()
>>> loud.shout()
LOUD
```

Class Decorators (3)

Occurs during class definition time (not instance creation):

```
>>> def time_cls_dec(cls):
...     print "BEFORE"
...     def new_method(self):
...         print "NEW METHOD"
...     cls.new_method = new_method
...     return cls

>>> @time_cls_dec   # definition time
... class Timing(object): pass
BEFORE

>>> t = Timing()    # instance creation time
```

Class Decorators (4)

Works with subclasses:

```
>>> class SubTiming(Timing): pass

>>> s = SubTiming()
>>> s.new_method()
NEW METHOD
```

Std lib example

`functools.total_ordering` in Python3.2 adds `__le__`, `__gt__`, and `__ge__` if `__lt__` and `__eq__` are defined.

List comprehensions

(PEP 202, Python 2.0)

Looping

Common to loop over and accumulate:

```
>>> seq = range(-10, 10)
>>> results = []
>>> for x in seq:
...     if x >= 0:
...         results.append(2*x)
```

List comprehensions

```
>>> results = [ 2*x for x in seq if x >= 0 ]
```

Shorthand for accumulation:

```
>>> results = []
>>> for x in seq:
...     if x >= 0:
...         results.append(2*x)
```

Construction

- Assign result:

```
results = []
```

- Insert for loop:

```
results = [for x in seq]
```

- Add filter (if any):

```
results = [for x in seq if x >= 0]
```

- Put accumulated object in front:

```
results = [2**x for x in seq if x >= 0]
```

List comprehensions (2)

`if` statement optional:

```
>>> results = [ 2*x for x in xrange(9)]
>>> results
[0, 2, 4, 6, 8, 10, 12, 14, 16]
```

List comprehensions (3)

Can be nested:

```
>>> nested = [ (x, y) for x in xrange(3) \
...             for y in xrange(4) ]
>>> nested
[(0, 0), (0, 1), (0, 2), (0, 3), (1, 0), (1, 1), (1, 2), (1, 3), (2,
0), (2, 1), (2, 2), (2, 3)]
```

Same as:

```
>>> nested = []
>>> for x in xrange(3):
...     for y in xrange(4):
...         nested.append((x,y))
```

List comprehensions (4)

Acting like `map` (apply `str` to a sequence):

```
>>> [str(x) for x in range(5)]
['0', '1', '2', '3', '4']
```

List comprehensions (5)

Acting like `filter` (get positive numbers):

```
>>> [x for x in range(-5, 5) if x >= 0]
[0, 1, 2, 3, 4]
```

Std lib example

From `csv.py`:

```
ascii = [chr(c) for c in range(127)]  # 7-bit ASCII
```

Assignment

`listcomprehensions.py`

Iterators

(PEP 234)

Iterators

Sequences in *Python* follow the iterator pattern (PEP 234):

```
>>> sequence = [ 'foo', 'bar', 'baz']
>>> for x in sequence:
...     # body of loop
```

equivalent to:

```
>>> iterator = iter(sequence)
>>> while True:
...     try:
...         x = iterator.next()  # py3 .__next__()
...     except StopIteration, e:
...         break
...     # body of loop
```

Iterators (2)

```
>>> sequence = [ 'foo', 'bar']
>>> seq_iter = iter(sequence)
>>> seq_iter.next()
'foo'
>>> seq_iter.next()
'bar'
>>> seq_iter.next()
Traceback (most recent call last):
  ...
StopIteration
```

Making objects iterable

```
>>> class Foo(object):
...     def __iter__(self):
...         return self
...     def next(self):    # py3 __next__
...         # logic
...         return next_item
```

Object example

```
>>> class RangeObject(object):
...     def __init__(self, end):
...         self.end = end
...         self.start = 0
...     def __iter__(self): return self
...     def next(self):
...         if self.start < self.end:
...             value = self.start
...             self.start += 1
...             return value
...         raise StopIteration

>>> [x for x in RangeObject(4)]
[0, 1, 2, 3]
```

Std lib example

From `csv.py`:

```
class DictReader:

    def __iter__(self):
        return self

    def next(self):
        if self.line_num == 0:
            # Used only for its side effect.
            self.fieldnames  # property: calls .next()
        row = self.reader.next()
        self.line_num = self.reader.line_num
```

Std lib example (2)

```
# unlike the basic reader, we prefer not to return blanks,
# because we will typically wind up with a dict full of None
# values
while row == []:
    row = self.reader.next()
d = dict(zip(self.fieldnames, row))
lf = len(self.fieldnames)
lr = len(row)
if lf < lr:
    d[self.restkey] = row[lf:]
elif lf > lr:
    for key in self.fieldnames[lr:]:
        d[key] = self.restval
return d
```

Assignment

`iterators.py`

Generators

(PEP 255, 342, Python 2.3)

generators

Functions with the `yield` keyword remember their state and return to it when iterating over them

generators (2)

Can be used to easily “generate” sequences

generators (3)

Can be useful for lowering memory usage (ie `range(1000000)` vs `xrange(1000000)`)

Note `xrange` is *not* a generator

generators (4)

```
>>> def gen_range(end):
...     cur = 0
...     while cur < end:
...         yield cur
...         # returns here next
...         cur += 1
```

generators (5)

Generators return a generator instance. Iterate over them for values:

```
>>> gen = gen_range(4)
>>> gen #doctest: +ELLIPSIS
<generator object gen_range at ...>
```

generators (6)

Follow the iteration protocol. A generator is iterable!

```
>>> nums = gen_range(2)
>>> nums.next()
0
>>> nums.next()
1
>>> nums.next()
Traceback (most recent call last):
  ...
StopIteration
```

Generators (7)

Generator in for loop or list comprehension:

```
>>> for num in gen_range(2):
...     print num
0
1

>>> print [x for x in gen_range(2)]
[0, 1]
```

Generators (8)

Re-using generators may be confusing:

```
>>> gen = gen_range(2)
>>> [x for x in gen]
[0, 1]

>>> # gen in now exhausted!
>>> [x for x in gen]
[]
```

generators (9)

Can be chained:

```
>>> def positive(seq):
...     for x in seq:
...         if x >= 0:
...             yield x
>>> def every_other(seq):
...     for i, x in enumerate(seq):
...         if i % 2 == 0:
...             yield x
>>> nums = xrange(-5, 5)
>>> pos = positive(nums)
>>> skip = every_other(pos)
>>> [x for x in skip]
[0, 2, 4]
```

generators (10)

Generators can be tricky to debug. Can't step into them when invoked, only when *iterated over*.

Objects as generators

```
>>> class Generate(object):
...     def __iter__(self):
...         # just use a
...         # generator here
...         yield result
```

list or generator?

List:

- Need to use data repeatedly
- Enough memory to hold data
- Negative slicing

Generator Hints

- Make it “peekable”
- Generators always return `True`, [] (empty list) is `False`
- Might be useful to cache results
- If recursive, make sure to iterate over results

Generator Hints (2)

- Rather than making a complicated generator, consider making simple ones that chain together (Unix philosophy)
- Sometimes one at a time is slow (db) - wrap with “fetchmany” generator
- `itertools` is helpful (`islice`)

`xrange`

`xrange` doesn't really behave as an generator.

- you can index it directly (but not slice)
- it has no `.next()` method
- it doesn't exhaust

Generator example

```
def fetch_many_wrapper(result, count=20000):
    """
    In an effort to speed up queries, this wrapper
    fetches count objects at a time. Otherwise our
    implementation has sqlalchemy fetching 1 row
    at a time (~30% slower).
    """
    done = False
    while not done:
        items = result.fetchmany(count)
        done = len(items) == 0
        if not done:
            for item in items:
                yield item
```

Recursive generator example

```
def find_files(base_dir, recurse=True):
    """
    yield files found in base_dir
    """
    for name in os.listdir(base_dir):
        filepath = os.path.join(base_dir, name)
        if os.path.isdir(filepath) and recurse:
            # make sure to iterate when recursing!
            for child in find_files(filepath, recurse):
                yield child
        else:
            yield filepath
```

Std lib example

From `collections.py`

```
class OrderedDict(dict):
    ...

    def iteritems(self):
        'od.iteritems -> an iterator over the (key, value)
pairs in od'
        for k in self:
            yield (k, self[k])
```

Peeker

```
class Peeker(object):
    def __init__(self, seq, ignore_done=False):
        self.seq = iter(seq)
        self.buffer = []
        self.ignore_done = ignore_done

    def pop(self):
        if self.buffer:
            return self.buffer.pop(0)

    def peek(self, n=0):
        """ this can raise an exception if peeking off the end. be
        aware and handle PeekDone appropriately"""
        try:
            if n == len(self.buffer):
                self.buffer.append(self.seq.next())
        except StopIteration as e:
            if not self.ignore_done:
                raise PeekDone('Exhausted')
            else:
                return
        return self.buffer[n]
```

Peeker (2)

```
def __iter__(self):
     return self

def next(self):
    if self.buffer:
        return self.buffer.pop(0)
    else:
        return self.seq.next()
```

Peeker (3)

```
lines = Peeker(data)
slide = []
for line in lines:
    try:
        line1 = lines.peek()
        line2 = lines.peek(1)
        if line.startswith('===') and \
            line2.startswith('==='):
            ...
```

Assignment

`generators.py`

Generator Expressions

(PEP 289 Python 2.4)

Generator expressions

Like list comprehensions. Except results are generated on the fly. Use (and) instead of [and] (or omit if expecting a sequence)

Generator expressions (2)

```
>>> [x*x for x in xrange(5)]
[0, 1, 4, 9, 16]

>>> (x*x for x in xrange(5)) # doctest: +ELLIPSIS,
<generator object <genexpr> at ...>
>>> list(x*x for x in xrange(5))
[0, 1, 4, 9, 16]
```

Generator expressions (3)

```
>>> nums = xrange(-5, 5)
>>> pos = (x for x in nums if x >= 0)
>>> skip = (x for i, x in enumerate(pos) if i % 2 == 0)
>>> list(skip)   # materialize
[0, 2, 4]
```

Generator expressions (4)

If Generators are confusing, but List Comprehensions make sense, you can simulate some of the behavior of generators as follows....

Generator expressions (5)

```
>>> def pos_generator(seq):
...     for x in seq:
...         if x >= 0:
...             yield x

>>> def pos_gen_exp(seq):
...     return (x for x in seq if x >= 0)

>>> list(pos_generator(range(-5, 5))) == \
...   list(pos_gen_exp(range(-5, 5)))
True
```

Std lib example

from `string.py`

```
def capwords(s, sep=None):
    """capwords(s [,sep]) -> string

    Split the argument into words using split, capitalize
    each word using capitalize, and join the capitalized
    words using join. If the optional second argument sep
    is absent or None, runs of whitespace characters are
    replaced by a single space and leading and trailing
    whitespace are removed, otherwise sep is used to split
    and join the words.

    """
    return (sep or ' ').join(x.capitalize() for x in s.split(sep))
```

Assignment

`genexp.py`

Dict Comprehensions

(PEP 274 Python 2.7)

Dict comprehensions

Similar to list comprehensions.

Dict comprehensions (2)

This:

```
>>> result = {x:x*x for x in range(5)}
```

Instead of:

```
>>> result = dict((x,x*x) for x in range(5))
```

Dict comprehensions (3)

- More legible
- No list created first (when `dict` combined with LC)

Std lib example

None found in 2.7 and 3.2

Set Comprehensions

(PEP 274 Python 2.7)

Set comprehensions

Similar to list comprehensions. But with { and }.

Set comprehensions (2)

This:

```
>>> result = {x for x in range(5)}
>>> result
set([0, 1, 2, 3, 4])
```

Instead of:

```
>>> result = set(x for x in range(5))
```

(`range(5)` is lousy here)

Set comprehensions (3)

- More legible
- No list created first (when `set` combined with LC)

Std lib example

None found in 2.7 and 3.2

Context Managers

(PEP 343 Python 2.5)

Context Mgr

Shortcut for “try/finally” statements

Context Mgr (2)

Makes it easy to write

```
# setup
try:
    variable = value
    # body
finally:
    # cleanup
```

as

```
with some_generator() as variable:
    # body
```

Try/Finally Diversion

What does this do?

```
>>> def foo():
...     try:
...         return 1
...     finally:
...         return 2
```

Try/Finally Diversion

What does this do?

```
>>> foo()
2
```

Try/Finally Diversion

What does this do?

```
>>> def foo2():
...     try:
...         1/0
...     finally:
...         return 2
```

Try/Finally Diversion

What does this do?

```
>>> foo2()
2
```

Try/Finally Diversion

What does this do?

```
>>> def foo3():
...     try:
...         raise KeyError("Bad!")
...     finally:
...         return 2
```

Try/Finally Diversion

What does this do?

```
>>> foo3()
2
```

Try/Finally Diversion

Takeaway - `finally` always runs

Context Mgr (3)

Seen in files:

```
fin = open('/tmp/foo')
# do something with  fin
fin.close()
```

Context Mgr (4)

Seen in files:

```
with open('/tmp/foo') as fin:
    # do something with fin
# fin is automatically closed here
```

Context Mgr (5)

Two ways to create:

- class
- decorated generator

Context Mgr (5)

Context managers can optionally return an item with `as`

Lock example (PEP 343)

```
lock.acquire()
# run some code while locked
lock.release()
```

Lock example (PEP 343)

Buggy, should be:

```
lock.acquire()
try:
    # run some code while locked
finally:
    lock.release()
```

Lock example (PEP 343)

```
with locked(myLock):
    # Code here executes with
    # myLock held. The lock is
    # guaranteed to be released
    # when the block is left
    # (even if via return or
    # by an uncaught exception).
```

Lock example (PEP 343) (2)

Class style:

```
class locked:
    def __init__(self, lock):
        self.lock = lock
    def __enter__(self):
        self.lock.acquire()
    def __exit__(self, type, value, tb):
        # if error in block, t, v, & tb
        # have non None values
        # return True to hide exception
        self.lock.release()
```

Lock example (PEP 343) (3)

Generator style:

```
from contextlib import contextmanager

@contextmanager
def locked(lock):
    lock.acquire()
    try:
        yield
    finally:
        lock.release()
```

Context Manager with `as`

Seen in files:

```
with open('/tmp/foo') as fin:
    # do something with fin
# fin is automatically closed here
```

Context Manager with `as` (2)

Class style:

```
class a_cm:
    def __init__(self):
        # init
    def __enter__(self):
        # enter logic
        return self
    def __exit__(self, type, value, tb):
        # exit logic
```

Context Manager with `as` (3)

Generator style yield object:

```
from contextlib import contextmanager

@contextmanager
def a_cm():
    # enter logic
    try:
        yield object
    finally:
        # exit logic
```

Error Handling

- In generators can use bare `raise` from `finally`
- In class, return `True` to swallow error. Can inspect error if needed. Arguments to `__exit__` correspond to results of `sys.exc_info()` (class, instance, traceback)

Error Handling

In class:

```
def __exit__(self, type, value, tb):
    # type is class of exception
    # value is instance
    # tb is traceback
    # return True to swallow exception
```

Rollback Code

From PEP

```
@contextmanager
def transaction(db):
    db.begin()
    try:
        yield None
    except:
        db.rollback()
        raise
    else:
        db.commit()
    # don't have to have finally
```

Rollback Code

Not in PEP

```
class transaction(object):
    def __init__(self, db):
        self.db = db

    def __enter__(self):
        self.db.begin()

    def __exit__(self, type, value, tb):
        if type:
            self.db.rollback()
            return False
        self.db.commit()
        return True
```

Uses for Context Managers

- Managing external resources (socket, file, connection)
- Transactions
- Acquiring locks
- closing/cleaning up
- nesting for generating html/xml

Std lib example

from `tempfile.py`

```
class SpooledTemporaryFile:
    """Temporary file wrapper, specialized to switch from
    StringIO to a real file when it exceeds a certain size or
    when a fileno is needed.
    """

    # Context management protocol
    def __enter__(self):
        if self._file.closed:
            raise ValueError("Cannot enter context with closed file")
        return self

    def __exit__(self, exc, value, tb):
        self._file.close()
```

Decorator/Context Manager

```
>>> from contextlib import contextmanager
>>> def verbose(what=None):
...     @contextmanager
...     def verbose_cm():
...         print "BEFORE"
...         yield
...         print "AFTER"
...     if hasattr(what, '__call__'):
...         def wrapper(*args, **kwargs):
...             with verbose_cm():
...                 return what(*args, **kwargs)
...         return wrapper
...     else:
...         return verbose_cm()
```

http://dabeaz.blogspot.com/2010/02/function-that-works-as-context-manager.html

Invoking Combo

```
>>> @verbose
... def middle():
...     print "MIDDLE"

>>> middle()
BEFORE
MIDDLE
AFTER

>>> with verbose():
...     print "MIDDLE"
BEFORE
MIDDLE
AFTER
```

Decorator/Context Manager (2)

```
>>> import sys
>>> class verbose(object):
...     def __init__(self, func=None):
...         self.func = func
...     def __enter__(self):
...         print "BEFORE"
...     def __exit__(self, type, value, tb):
...         print "AFTER"
```

Decorator/Context Manager (3)

```
...     def __call__(self, *args, **kwargs):
...         self.__enter__()
...         exc = None, None, None
...         try:
...             result = self.func(*args, **kwargs)
...         except Exception:
...             exc = sys.exc_info()
...         catch = self.__exit__(*exc)
...         if not catch and catch is not None:
...             cls, val, tb = exc
...             raise cls, val, tb
...         return result
```

http://code.activestate.com/recipes/577273-decorator-and-context-manager-from-a-single-api/
(Michael Foord)

Invoking Combo

```
>>> @verbose
... def middle():
...     print "MIDDLE"
>>> middle()
BEFORE
MIDDLE
AFTER

>>> with verbose():
...     print "MIDDLE"
BEFORE
MIDDLE
AFTER
```

Assignment

`ctxmgr.py`

Properties

(Python 2.2, 2.6 added `getter`, `setter`, `deleter`)

Properties

Utilize *descriptors* to allow attribute access to invoke methods. If you have an attribute you can later add an underlying method to do setting, getting, deleting.

Properties (2)

```
>>> class Person(object):
...     def __init__(self):
...         self.name = None
```

later need to alter attribute value w/o changing interface

```
>>> class Person(object):
...     def __init__(self):
...         self._name = None
...     def get_name(self):
...         return self._name
...     def set_name(self, name):
...         self._name = name.replace(';', '')
...     name = property(get_name, set_name)
```

Properties (3)

```
>>> p = Person()
>>> p.name = 'Fred; Drop TABLE people;'
>>> print p.name
Fred Drop TABLE people
```

Properties (4)

2.6 style:

```
>>> class Person2(object):
...     def __init__(self):
...         self._name = None
...
...     @property
...     def name(self):
...         return self._name
...
...     @name.setter
...     def name(self, value):
...         self._name = value.replace(';', '')
...
...     @name.deleter
...     def name(self):
...         del self._name
```

Properties (5)

```
>>> p = Person2()
>>> p.name = 'Fred; Drop TABLE people;'
>>> print p.name
Fred Drop TABLE people
```

Std lib example

From `csv.py`

```
class DictReader:
    ...
    @property
    def fieldnames(self):
        if self._fieldnames is None:
            try:
                self._fieldnames = self.reader.next()
            except StopIteration:
                pass
        self.line_num = self.reader.line_num
        return self._fieldnames

    @fieldnames.setter
    def fieldnames(self, value):
        self._fieldnames = value
```

That's all

- Volunteer at PyCon!

matthewharrison@gmail.com

@__mharrison__
http://hairysun.com