



iscte

INSTITUTO
UNIVERSITÁRIO
DE LISBOA



emprego
digital



```
class BigFile:
```

```
    def __init__(self, datadir, ndims):
        idfile = os.path.join(datadir, "id.txt")
        self.names = [x.strip() for x in str.split(open(idfile).read()) if x.strip()]
        self.name2index = dict(zip(self.names, range(len(self.names))))
        self.ndims = ndims
        self.featurefile = os.path.join(datadir, "feature.bin")
        print "[BigFile] %d features, %d dimensions" % (len(self.names), self.ndims)
        print "        binary: %s" % self.featurefile
        print "        txt: %s" % idfile
```

```
    def read(self, requested, isname=True):
        if isname:
            index_name_array = [(self.name2index[x], x) for x in requested if x in self.names]
        else:
            assert len(requested) > 0
            assert all((requested[i] in self.names) for i in range(len(requested)))
            index_name_array = [(x, self.names[x]) for x in requested]
            index_name_array.sort()
            vecs = seq_read(self.featurefile, self.ndims, [x[0] for x in index_name_array])
            return [x[1] for x in index_name_array], vecs

    def shape(self):
        return (len(self.names), self.ndims)
```



pythonTM

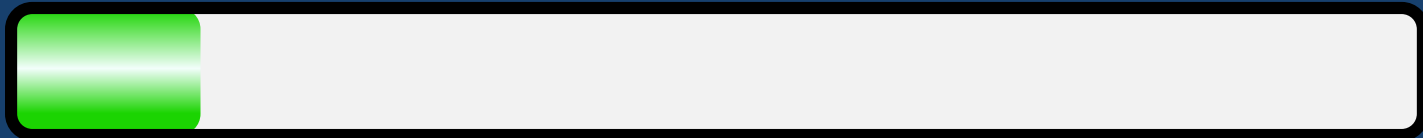
1.

Overall Program Content

Web development with Python	Hours
Work skills development	50
Python Programming Introduction	150
Web Programming Introduction (html/css)	100
Databases Concepts and Structures	50
Web Servers Programming	150
Web services development	150
Total	650

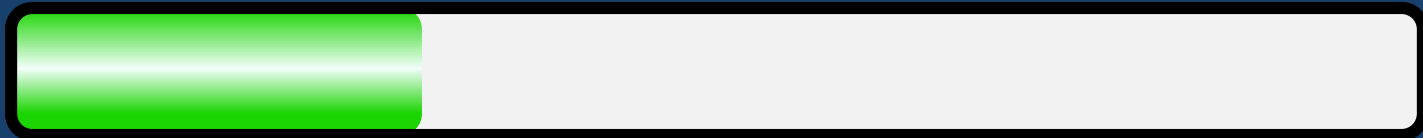
Python programming Introduction Content

1. Course Introduction
 - Why Python?
 - Python Applications
 - Installation Tools
 - Building your code catalog
 - Useful websites



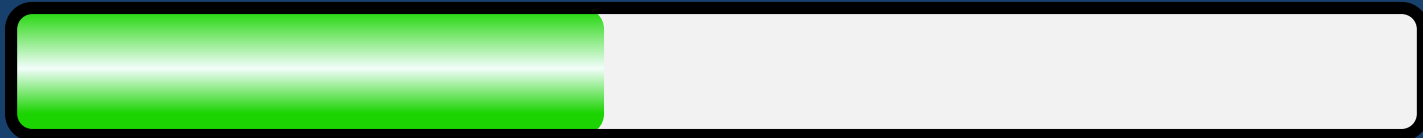
Python programming Introduction Content

2. Data types/outputs/inputs
3. Operators
4. Functions and Modules



Python programming Introduction Content

- 5. Conditional statements and expression
- 6. Loops
- 7. Work with standard Library and Modules



Python programming Introduction Content

- 8. Data structure in python
- 9. List,
- 10. Tuple,
- 11. Dictionaries,
- 12. Set



Python programming Introduction Content

- 13. Files
- 14. Functions and Modules
- 15. Classes
- 16. Introduction to Numpy
- 17. Introduction to Pandas



Python programming Introduction Content

- 18. Introduction to matplotlib for data visualization
- 19. Data Preprocessing

100% Loaded

Our Teachers:



Joseanne Viana (Josi)

Email: jcova1@iscte-iul.pt



Stefan Postolache

Email: stefanpostolache@edu.ulisboa.pt



Hamed Farkhari

Email: Hamed_Farkhari@iscte-iul.pt


```
class BigFile:
```

```
    def __init__(self, datadir, ndims):
        idfile = os.path.join(datadir, "id.txt")
        self.names = [x.strip() for x in str.split(open(idfile).read()) if x.strip()]
        self.name2index = dict(zip(self.names, range(len(self.names))))
        self.ndims = ndims
        self.featurefile = os.path.join(datadir, "feature.bin")
        print "[BigFile] %d features, %d dimensions" % (len(self.names), self.ndims)
        print "        binary: %s" % self.featurefile
        print "        txt: %s" % idfile
```

```
    def read(self, requested, isname=True):
        if isname:
            index_name_array = [self.name2index[x], x] for x in requested if x in self.names
        else:
            assert(min(requested) >= 0)
            assert(max(requested) < len(self.names))
            index_name_array = [(x, self.names[x]) for x in requested]
            index_name_array.sort()
            vecs = seq_read(self.featurefile, self.ndims, [x[0] for x in index_name_array])
            return [x[1] for x in index_name_array], vecs

    def shape(self):
        return [len(self.names), self.ndims]
```

<Let's get started >

Contents

1. Tuple

Tuple

Define tuple

use [] to define a list

use () to define a tuple

```
t = ('English', 'History', 'Mathematics')  
  
print(t)  
  
print(type(t))      # <class 'tuple'>  
  
print(len(t))       # 3
```

Define tuple

To define a tuple with single obj we should use ',' after that obj

```
t1 = (3)           # t1 is integer
t2 = (3,)          # t2 is tuple
s1 = ('a')         # s1 is string
s2 = ('a',)        # s2 is tuple
```


Access members Index Slicing

*access the members in tuple we use [],
same as list*

```
t = ('English', 'History', 'Mathematics')  
  
print(t[0])          # English  
  
print(t[1:3])        # ('History', 'Mathematics')  
  
print(t.index('English')) # 0  
  
print('English' in t) # True
```

Access members

tuples are immutable, same as strings

```
t = ('English', 'History', 'Mathematics')
```

```
t[0] = 'art'    # Error
```

```
for i in t:  
    print(f'I like to read {i}')
```

```
'''
```

```
I like to read English
```

```
I like to read History
```

```
I like to read Mathematics
```

```
'''
```

Review

'i in a'

'if' vs 'for'

```
"""
what is different of

' i in a '

when we use it in 'for' and 'if' ?

"""
a = (1, 2, 3)

if i in a:
# that means check this compares:
if ( i == a[0] or i == a[1] or i == a[2] ) :
    ...

for i in a:
# that means for every loop run do these assignment:
i = a[0]   for first run
i = a[1]   for second run
i = a[2]   for third run
    ...
```

Max
Min
Sum
Count

```
t = (1, 9, 2)

print(sum(t))           # 12

print(max(t))           # 9

print(min(t))           # 1

print(t.count(9))       # 1
```

+ *
reversed

** , + on tuples*

```
t = (1, 9, 2)

print(t*2)           # (1, 9, 2, 1, 9, 2)

print(t + t + t)     # (1, 9, 2, 1, 9, 2, 1, 9, 2)

print((3, 6) + (9,))  # (3, 6, 9)
print((1, 2) + (9, 6)) # (1, 2, 9, 6)
```

Reversed()

```
print(tuple(reversed(t))) # (2, 9, 1)
```

Ordered

tuples are ordered

```
t1 = (1,2)
```

```
t2 = (2,1)
```

```
print(t1 == t2)    # False
```

Append

tuples are not changeable
Here we overwrite t with new value!

```
t = (4 , 6)
```

```
t = t + (9,)
```

```
print(t)          # (4, 6, 9)
```

```
class BigFile:
```

```
    def __init__(self, datadir, ndims):
        idfile = os.path.join(datadir, "id.txt")
        self.names = [x.strip() for x in str.split(open(idfile).read()) if x.strip()]
        self.name2index = dict(zip(self.names, range(len(self.names))))
        self.ndims = ndims
        self.featurefile = os.path.join(datadir, "feature.bin")
        print "[BigFile] %d features, %d dimensions" % (len(self.names), self.ndims)
        print "        binary: %s" % self.featurefile
        print "        txt: %s" % idfile
```

```
    def read(self, requested, isname=True):
        if isname:
            index_name_array = [(self.name2index[x], x) for x in requested if x in self.names]
        else:
            assert(min(requested) >= 0)
            assert(max(requested) < len(self.names))
            index_name_array = [(x, self.names[x]) for x in requested]
            index_name_array.sort()
            vecs = seq_read(self.featurefile, self.ndims, [x[0] for x in index_name_array])
            return [x[1] for x in index_name_array], vecs

    def shape(self):
        return [len(self.names), self.ndims]
```

<Exercise 1>

Exercise

*Try append tuple with another way!
By Converting tuple to another data type*

Input: $t = (4, 6)$

Output: t be $(4, 6, 9)$

What is my Input ? It is tuple

How to convert Tuple → list ?

What should I use to Add/append/extend to a list?

How I can convert List → tuple ?

Exercise

How to solve a
Problem by
splitting to small
questions?

```
Input:  t = (4 ,6)
```

```
Output:      t be (4, 6, 9)
```

What is my Input ? It is tuple

How to convert Tuple ==> list ? list()

What should I use to Add/append/extend to a list?

How I can convert List ==> tuple ? tuple()

```
"""
```

```
t = (4 ,6)
```

```
a = list(t)
```

```
print("what is the value of a? ", a)
```

```
a.extend([9]) # a.append(9)
```

```
t = tuple(a)
```

Check what happen to 'a' until this line



Exercise solution

*Try append tuple with another way!
By Converting tuple to another data type*

Input: t = (4 ,6)

Output: t be (4, 6, 9)

```
t = (4 ,6)
```

```
a = list(t)
```

```
a.append(9)
```

```
t = tuple(a)
```

```
print(t)        # (4, 6, 9)
```

```
class BigFile:
```

```
    def __init__(self, datadir, ndims):  
        idfile = os.path.join(datadir, "id.txt")  
        self.names = [x.strip() for x in str.split(open(idfile).read()) if x.strip()]  
        self.name2index = dict(zip(self.names, range(len(self.names))))  
        self.ndims = ndims  
        self.featurefile = os.path.join(datadir, "feature.bin")  
        print "[BigFile] %d features, %d dimensions" % (len(self.names), self.ndims)  
        print "        binary: %s" % self.featurefile  
        print "        txt: %s" % idfile
```

```
    def read(self, requested, isname=True):  
        if isname:  
            index_name_array = [(self.name2index[x], x) for x in requested if x in self.names]  
        else:  
            assert(min(requested) >= 0)  
            assert(max(requested) < len(self.names))  
            index_name_array = [(x, self.names[x]) for x in requested]  
            index_name_array.sort()  
            vecs = seq_read(self.featurefile, self.ndims, [x[0] for x in index_name_array])  
            return [x[1] for x in index_name_array], vecs  
  
    def shape(self):  
        return [len(self.names), self.ndims]
```

<Exercise 2>

Exercise

Try change string?

you need to Convert string to another data type then overwrite it with new value!

1-Remove '\$', '#' and spaces in t

2-Add space between 'Python' and 'Course'

Input: t = "\$ \$ Python\$#Course \$"

Output: t = "Python Course"

How to solve a Problem by splitting to small questions(Parts)?

"""

- 1-Remove '\$', '#' and spaces in t
- 2-Add space between 'Python' and 'Course'

Input: t = "\$ \$ Python\$#Course \$"

Output: t = "Python Course"

- 1 - what is my input? string, unchangeable
- 2 - Remove '\$', '#' and spaces?
is there any method to do that? yes, t.strip('# \$')
- 3 - is it solved? I need to remove '\$#' from the middle of string! ==> convert string to list by list()
- 4 - I need a loop to check all list members and remove unwanted members
I found there is no need to strip() and developed my code!
- 5 - I need to insert space between "n" and "C"

"""

Exercise Solution 1

```
t = "$ $ Python$#Course $"
#a = t.strip('# $')
a = list(t)
b = a.copy()
for i in a:
    # i in ["#", "$", "\n", "\t", " "]
    if i == "#" or i == "$" or i == " " :
        b.remove(i)

C_index = b.index("C")
b.insert(C_index, " ") # b.insert(b.index("C"), " ")

t = ''.join(b)
```

Exercise Solution 2

```
t = "$ $ Python$#Course $"

# remove '#' and '$' from the first and end of t
a = list(t.strip("# $"))    # remove '#', '$' and spaces

b = a.copy()

# remove '#' and '$' from the middle of t
for i in a:
    if i == "#":
        b.remove("#")
    if i == "$":
        b.remove("$")

# add space between 'Python' and 'Course'
b.insert( b.index("C") , " " )

# convert List to String
t = ''.join(b)
```


Remove

*Remove a members from tuple
only possible by converting to another
data type!*

```
t = (4, 7, 2, 9, 8)
```

```
x = list(t)
```

```
x.remove(2)
```

```
t = tuple(x)
```

```
print(t)          # (4, 7, 9, 8)
```

unpack

Example 1

```
t = (4, 8)
a, b = t
print(a)      # 4
print(b)      # 8
```

Example 2

```
car = ('blue', 'auto', 7)
color, _, a = car
print(color)   # blue
print(_)       # auto
print(a)       # 7
```

Zip Zip (*)

zip, zip()*

```
a = (1, 2)
b = (3, 4)
c = zip(a,b)
x = list(c)
print(x)                # [(1, 3), (2, 4)]
print(x[0])              # (1,3)
print(type(x[0]))        # <class 'tuple'>
```

z can be tuple or list

```
z = ((1, 3), (2, 4))    # OR z = [(1, 3), (2, 4)]
u = zip(*z)
print(list(u))           # [(1, 2), (3, 4)]
```

*Try this example
when z is a dictionary!
what is output?*


Zip Zip (*) Example

Example 1

You can use tuple or list

```
a = (1, 2, 'A')           # a = [1, 2, 'A']
b = (3, 4, 8)             # b = [3, 4, 8]
c = zip(a,b)
x = list(c)
print(x)                  # [(1, 3), (2, 4), ('A', 8)]

print(list(zip(*x)))      # [(1, 2, 'A'), (3, 4, 8)]
```



minimum length between 'a' and 'range(2)' is 2

```
a = [11, 22, 33]
b = zip(a, range(2))
print(list(b))           # [(11, 0), (22, 1)]
```

```
class BigFile:
```

```
    def __init__(self, datadir, ndims):  
        idfile = os.path.join(datadir, "id.txt")  
        self.names = [x.strip() for x in str.split(open(idfile).read()) if x.strip()]  
        self.name2index = dict(zip(self.names, range(len(self.names))))  
        self.ndims = ndims  
        self.featurefile = os.path.join(datadir, "feature.bin")  
        print "[BigFile] %d features, %d dimensions" % (len(self.names), self.ndims)  
        print "        binary: %s" % self.featurefile  
        print "        txt: %s" % idfile
```

```
    def read(self, requested, isname=True):  
        if isname:  
            index_name_array = [(self.name2index[x], x) for x in requested if x in self.names]  
        else:  
            assert(min(requested) >= 0)  
            assert(max(requested) < len(self.names))  
            index_name_array = [(x, self.names[x]) for x in requested]  
            index_name_array.sort()  
            vecs = seq_read(self.featurefile, self.ndims, [x[0] for x in index_name_array])  
            return [x[1] for x in index_name_array], vecs  
  
    def shape(self):  
        return [len(self.names), self.ndims]
```

<Exercise 3>

Exercise

Try Zip with 3 input variables?

A = [1, 2, 'A']

B = ('Python', 161.8, 0, 5)

C = {10, 12, 14, 16, 18, 20}

Output ==> print(list(zip(A, B, C)))

[(1, 'Python', 10), (2, 161.8, 12), ('A', 0, 14)]

*Write program with the same input and output
without using Zip() function?*

Exercise Solution Step by step

```
A = [1, 2, "A"]  
B = ("Python", 161.8, 0, 5)  
C = {10, 12, 14, 16, 18, 20} ←
```

Hint: C is a set type and is not ordered
That means each time you run the code,
The order of C items maybe changed!
This example with set is just for practicing!

```
# print(list(zip(A, B, C)))  
# output :  
# [(1, 'Python', 10), (2, 161.8, 12), ('A', 0, 14)]  
  
min_ABC = min(len(A), len(B), len(C))  
C = list(C)  
  
i = 0  
# step1: change output for better understanding  
d1 = [(A[0], B[0], C[0]), (A[1], B[1], C[1]), (A[2], B[2], C[2])]  
# step2: change indexes with i  
d2 = [(A[i], B[i], C[i]),  
      (A[i + 1], B[i + 1], C[i + 1]),  
      (A[i + 2], B[i + 2], C[i + 2])]  
# step3: its like a loop with 'i' iteration  
e = []  
for i in range(min_ABC):  
    my_tuple = (A[i], B[i], C[i])  
    e.append(my_tuple)  
  
output = e  
print(output)
```

isinstance

how many tuples are in 'num' list

```
num = [8, 2, (9,3), 4, (1,6,7), 34]
c = 0
for i in num:
    if isinstance(i, tuple):
        continue
    c += 1
print(c)                # 4
print(len(num) - c)     # 2
```



```
class BigFile:
```

```
    def __init__(self, datadir, ndims):
        idfile = os.path.join(datadir, "id.txt")
        self.names = [x.strip() for x in str.split(open(idfile).read()) if x.strip()]
        self.name2index = dict(zip(self.names, range(len(self.names))))
        self.ndims = ndims
        self.featurefile = os.path.join(datadir, "feature.bin")
        print "[BigFile] %d features, %d dimensions" % (len(self.names), self.ndims)
        print "        binary: %s" % self.featurefile
        print "        txt: %s" % idfile
```

```
    def read(self, requested, isname=True):
        if isname:
            index_name_array = [(self.name2index[x], x) for x in requested if x in self.names]
        else:
            assert(min(requested) >= 0)
            assert(max(requested) < len(self.names))
            index_name_array = [(x, self.names[x]) for x in requested]
            index_name_array.sort()
            vecs = seq_read(self.featurefile, self.ndims, [x[0] for x in index_name_array])
            return [x[1] for x in index_name_array], vecs

    def shape(self):
        return [len(self.names), self.ndims]
```

<Exercise 4>

Exercise

how many tuples are in 'num' list

Try this example again

without using 'isinstance()' function?

```
num = [8, 2, (9,3), 4, (1,6,7), 34]
```

Exercise Solution

```
"""  
How many tuples are in Num?  
"""  
  
Num = [8, 2, (9, 3), 4, (1, 6, 7), 34]  
  
c = 0  
for i in Num:  
    if type(i) == tuple:  
        c += 1  
print("There are ", c, "tuples in Num")
```

Example

Input : [(1,2,3) , (4,5,6)]

Output : [(1,2,9) , (4,5,9)]

```
a = [(1,2,3) , (4,5,6)]
```

```
b = [i[:-1]+(9,) for i in a]
```

```
print(b)      # [(1, 2, 9), (4, 5, 9)]
```

```
class BigFile:
```

```
    def __init__(self, datadir, ndims):
        idfile = os.path.join(datadir, "id.txt")
        self.names = [x.strip() for x in str.split(open(idfile).read()) if x.strip()]
        self.name2index = dict(zip(self.names, range(len(self.names))))
        self.ndims = ndims
        self.featurefile = os.path.join(datadir, "feature.bin")
        print "[BigFile] %d features, %d dimensions" % (len(self.names), self.ndims)
        print "        binary: %s" % self.featurefile
        print "        txt: %s" % idfile
```

```
    def read(self, requested, isname=True):
        if isname:
            index_name_array = [(self.names.index[x], x) for x in requested if x in self.names]
        else:
            assert(min(requested) >= 0)
            assert(max(requested) < len(self.names))
            index_name_array = [(x, self.names[x]) for x in requested]
            index_name_array.sort()
            vecs = seq_read(self.featurefile, self.ndims, [x[0] for x in index_name_array])
            return [x[1] for x in index_name_array], vecs

    def shape(self):
        return [len(self.names), self.ndims]
```

<Homework 1>

Homework 1

Input : $[(1,2,3) , (4,5,6)]$

Output : $[(1,2,9) , (4,5,9)]$

Try last example again with another way?

```
class BigFile:
```

```
    def __init__(self, datadir, ndims):
        idfile = os.path.join(datadir, "id.txt")
        self.names = [x.strip() for x in str.split(open(idfile).read()) if x.strip()]
        self.name2index = dict(zip(self.names, range(len(self.names))))
        self.ndims = ndims
        self.featurefile = os.path.join(datadir, "feature.bin")
        print "[BigFile] %d features, %d dimensions" % (len(self.names), self.ndims)
        print "        binary: %s" % self.featurefile
        print "        txt: %s" % idfile
```

```
    def read(self, requested, isname=True):
        if isname:
            index_name_array = [(self.names.index[x], x) for x in requested if x in self.names]
        else:
            assert(min(requested) >= 0)
            assert(max(requested) < len(self.names))
            index_name_array = [(x, self.names[x]) for x in requested]
            index_name_array.sort()
            vecs = seq_read(self.featurefile, self.ndims, [x[0] for x in index_name_array])
            return [x[1] for x in index_name_array], vecs

    def shape(self):
        return [len(self.names), self.ndims]
```

<Homework 2>

Homework 2

By using unpacking method define `my_obj` to avoid raising error?

```
my_obj = ...  
for i, j in my_obj :  
    print("i is: ", i)  
    print("j is: ", j)
```

*When can we use “for i, j in my_obj”?
For example by defining `my_obj = (1, 3)`
We get this Error:*

```
TypeError: cannot unpack non-iterable int object
```



```
class BigFile:
```

```
    def __init__(self, datadir, ndims):
        idfile = os.path.join(datadir, "id.txt")
        self.names = [x.strip() for x in str.split(open(idfile).read()) if x.strip()]
        self.name2index = dict(zip(self.names, range(len(self.names))))
        self.ndims = ndims
        self.featurefile = os.path.join(datadir, "feature.bin")
        print "[BigFile] %d features, %d dimensions" % (len(self.names), self.ndims)
        print "        binary: %s" % self.featurefile
        print "        txt: %s" % idfile
```

```
    def read(self, requested, isname=True):
        if isname:
            index_name_array = [(self.name2index[x], x) for x in requested if x in self.names]
        else:
            assert(min(requested) >= 0)
            assert(max(requested) < len(self.names))
            index_name_array = [(x, self.names[x]) for x in requested]
            index_name_array.sort()
            vecs = seq_read(self.featurefile, self.ndims, [x[0] for x in index_name_array])
            return [x[1] for x in index_name_array], vecs

    def shape(self):
        return [len(self.names), self.ndims]
```

<Homework 3>

Homework 3

Input `[(1, 3), (2, 4), ('A', 8)]`

Output `[(1, 2, 'A'), (3, 4, 8)]`

Do Not use zip() function

“

- *Make it work*
- *Make it Right*
- *Make it Fast*

O futuro profissional começa aqui

iscte
INSTITUTO
UNIVERSITÁRIO
DE LISBOA



emprego
digital



UPskill