



iscte

INSTITUTO
UNIVERSITÁRIO
DE LISBOA



emprego
digital

 UP**skill**

```
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(x in self.names for x in 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

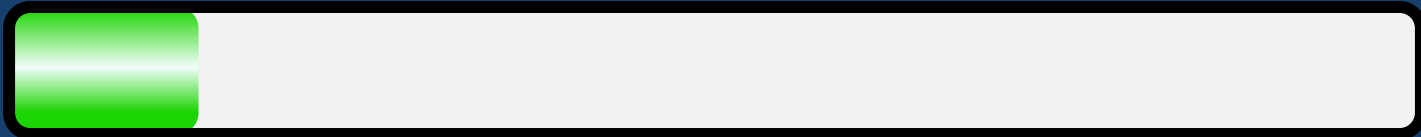
1.

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

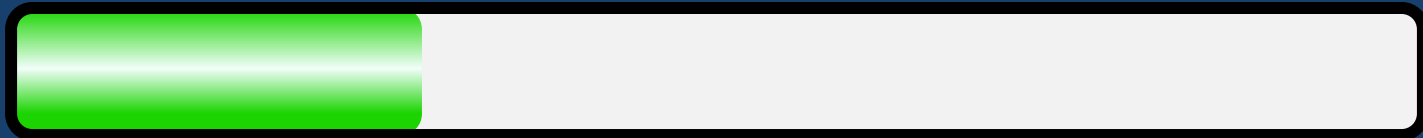
Python programming Introduction Content

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



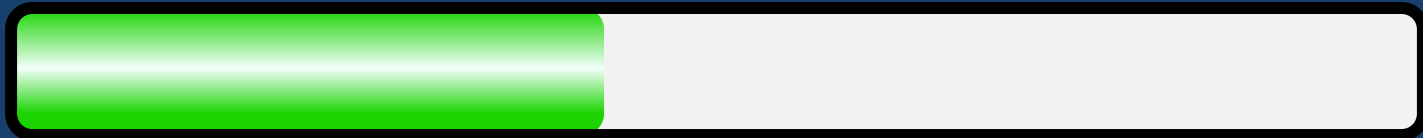
Python programming Introduction Content

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. Set

Set

Define Set

use [] to define a List

use (,) to define a Tuple

use { : } to define a Dictionary

use { , } to define a Set

f = {'apple', 'orange', 'banana'}

D = set() # Empty set

Example

```
f = {'apple', 'orange' , 'banana'}

# f[0]
# Error: 'set' object is not subscriptable

print(type(f))    # <class 'set'>

print(len(f))     # 3

print(f)          # {'orange', 'banana', 'apple'}

for i in f:
    print(i)

m = set(('orange' , 'banana' , 'apple'))

print(f == m)      # True

print('cherry' in f)  # False
```


add()
update()
remove()

```
f = {'apple', 'orange' , 'banana'}
```

```
f.add('cherry')
```

```
print(f)
```

```
# {'orange', 'banana', 'cherry', 'apple'}
```

```
f.update(['mango' , 'grapes'])
```

```
print(f)
```

```
# {'cherry', 'orange', 'banana', 'apple', 'mango', 'grapes'}
```

```
f.remove('apple')
```

```
print(f)
```

```
# {'cherry', 'orange', 'banana', 'mango', 'grapes'}
```

add() update()

```
f = {'a', 'b', 'c'}  
f.add('d')  
# f.add('e', 'f')  
f.add(('e2', 'f2')) # {'e2', 'f2', 'a', 'b', 'c', 'd'}  
f.add(['e2', 'f2']) # Error  
# A = {'a', ['b', 'c']} # Error
```

```
f = {'a', 'b', 'c'}  
f.update('g') # {'a', 'b', 'c', 'g'}  
f.update('h', 'i') # {'a', 'b', 'c', 'g', 'h', 'i'}
```

```
f = {'a', 'b', 'c'}  
f.update(['h', 'i']) # {'a', 'b', 'c', 'h', 'i'}
```

Remove()
Discard()
Copy()
pop()
Clear()

```
vowels = {'a','e','o','i','u'}
```

```
# 'k' is not in vowels
```

```
vowels.remove('k') # Error: KeyError: 'k' ← Error
```

```
vowels.discard('k')
```

```
# if 'k' exist in vowels remove it, else No error
```

```
v2 = vowels # V2 and vowels are dependent
```

```
c = vowels.copy()
```

```
print(vowels) # {'a', 'u', 'o', 'i', 'e'}
```

```
x = vowels.pop() # randomly remove one the members in set
```

```
print(x) # a
```

```
print(vowels) # {'u', 'o', 'i', 'e'}
```

```
print(v2) # {'u', 'o', 'i', 'e'}
```

```
print(c) # {'u', 'o', 'i', 'e', 'a'}
```

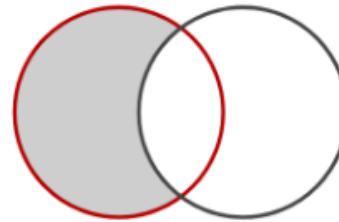
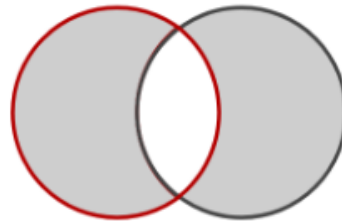
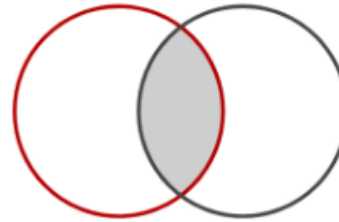
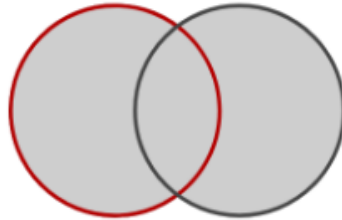
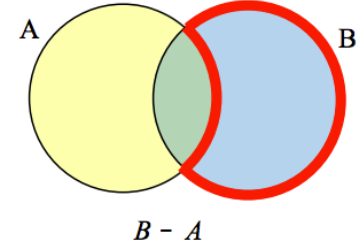
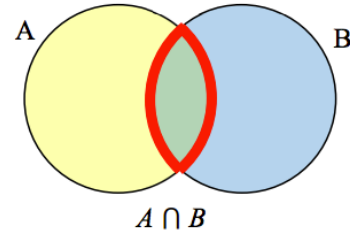
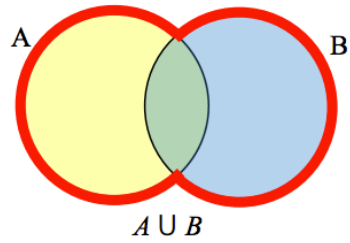
```
vowels.clear()
```

```
print(vowels) # set()
```

```
print(len(vowels)) # 0
```

```
del c
```

review



Union Intersection Update

```
X = {1, 2, 3}
Y = {2, 3, 4}
print(X.union(Y))      # {1, 2, 3, 4}
print(X | Y)           # {1, 2, 3, 4}, | is OR
```

```
X = {1, 2, 3}
Y = {2, 3, 4}
print(X.intersection(Y)) # {2, 3}
print(X & Y)              # {2, 3}, & is and
```

```
X = {1, 2, 3}
Y = {2, 3, 4}
X.update(Y)             # same as union
print(X)                 # {1, 2, 3, 4}
```

Difference

Difference_update

Symmetric_difference

^

```
A = {1, 2, 3, 4, 5}
```

```
B = {2, 4, 7}
```

```
print(A-B)    # {1, 3, 5}
```

```
print(B-A)    # {7}
```

```
r = A.difference(B)
```

```
print(r)      # {1, 3, 5}
```

```
print(A)      # {1, 2, 3, 4, 5}
```

```
print(B)      # {2, 4, 7}
```

```
X = {1, 2, 3}
```

```
Y = {2, 3, 4}
```

```
print(X.symmetric_difference(Y))    # {1, 4}
```

```
print(X ^ Y)                        # {1, 4}, ^ is XOR
```

```
print(X.union(Y) - X.intersection(Y)) # {1, 4}
```

```
print(X.union(Y) - Y.intersection(X)) # {1, 4}
```

```
r = A.difference_update(B)    # nothing return in output
```

```
print(r)    # None
```

```
print(A)    # {1, 3, 5}
```

```
print(B)    # {2, 4, 7}
```

```
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
```

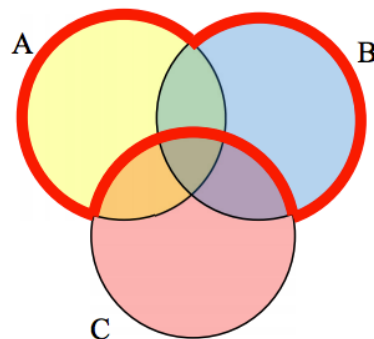
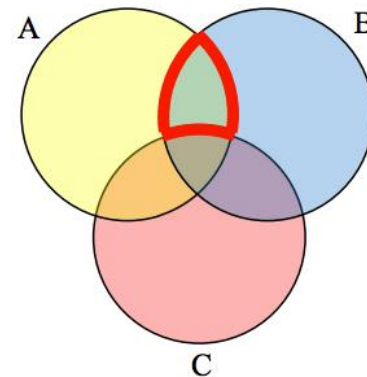
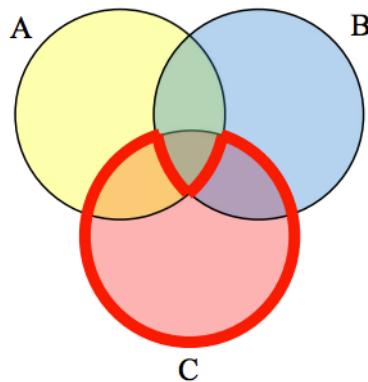
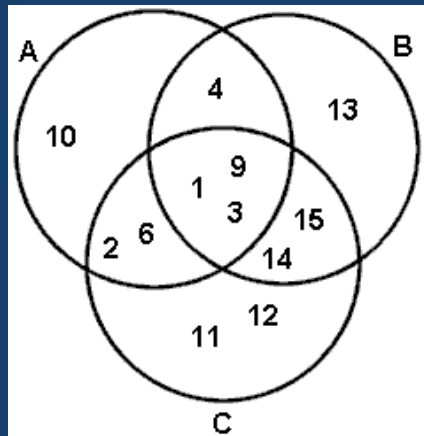
```
    def read(self, requested, isname=True):
        if isname:
            index_name_array = [(self.name2index[x], x) for x in requested if x in self.names]
        else:
            assert(min(requested) >= 0)
            assert(max(requested) < len(self.names))
            index_name_array = [(x,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

Could you answer?



Exercise

Answer

$A = \{1, 2, 3, 4, 6, 9, 10\}$

$B = \{1, 3, 4, 9, 13, 14, 15\}$

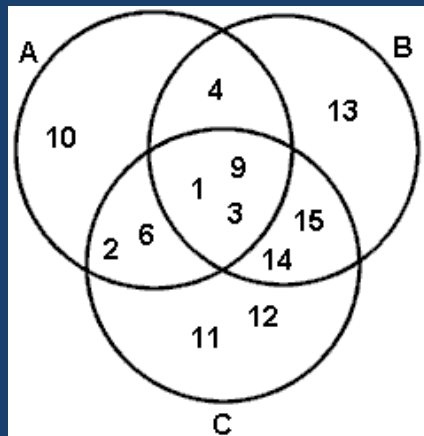
$C = \{1, 2, 3, 6, 9, 11, 12, 14, 15\}$

part 1 is $\{2, 6, 11, 12, 14, 15\}$

```
answer1_1 = C - A.intersection(B).intersection(C)
```

```
answer1_2 = C - A.intersection(B, C)
```

```
answer1_3 = C.difference(A & B & C)
```



Part 2 is $\{4\}$

```
answer2_1 = (A & B) - C
```

```
answer2_2 = (A.intersection(B)) - C
```

Part 3 is $\{4, 10, 13\}$

```
answer3_1 = A.union(B) - C
```

```
answer3_2 = (A | B) - C
```

update

```
s = 'Hamed'           # string
a = [13,25]           # list
t = (7 , 8 )          # tuple
d = {'one':1 , 'two':2} # dictionary
X = {56 , 98}          # set
```

```
X.update(s,a,t,d)
```

```
print(X)  # {'one', 98, 'two', 7, 8, 13, 'a', 'H', 56, 25, 'd', 'm', 'e' }
```

split string to characters

only keys of dict is considered

isdisjoint

X and Y have intersection or Not?

```
X = {1, 2}
Y = {1, 2, 3}
print(X.isdisjoint(Y))    # False
```

```
X = {1, 2}
Y = {3, 7, 8}
print(X.isdisjoint(Y))    # True
```

X and Y have no intersection

issubset

Subset

*We say that A is a subset of B
since every element of A is also in B .*

$A = \{1, 2, 4\}$

$B = \{1, 2, 3, 4, 5\}$

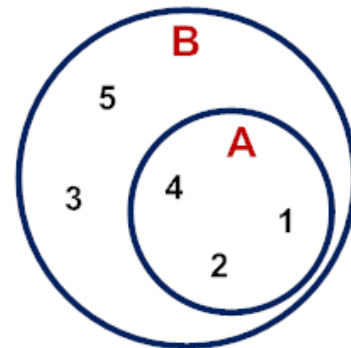
```
print(A.issubset(B)) # True
```

```
print(B.issubset(A)) # False
```

```
C = set() # C is an Empty set
```

```
print(C.issubset(A)) # True
```

```
print(A.issubset(C)) # False
```



$A \subset B$

```
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

*Which characters of
'a' , 'y' , 'c' , 'o' , 'z' are in
'Python Course' ?*

Output: 'o' , 'y'

Solution

```
w = 'Python Course'
```

```
char = {'a' , 'y', 'c', 'o', 'z'}
```

```
w_set = set(w)  ← Convert string to set
```

```
print(char.intersection(w_set))  # {'o', 'y'}
```

```
print(w_set.intersection(char))  # {'o', 'y'}
```

```
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 1>

Homework

Find match key:value in 2 dictionaries

```
d1 = {'a':1 , 'b':3 , 'c':2}
```

```
d2 = {'a':2 , 'b':3 , 'c':1}
```

Try to use set!

Output: {'b':3}

```
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 2>

Review

Interval, mathematics

$$a \leq X \leq b$$



closed interval $[a, b]$

$$a < X < b$$



open interval (a, b)

$$a \leq X < b$$



half-closed interval $[a, b)$

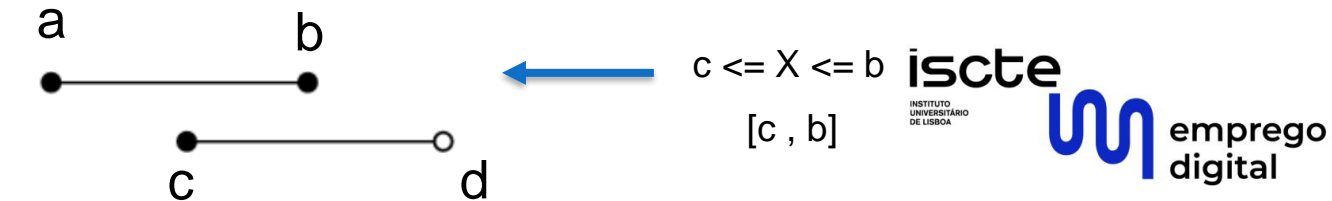
$$a < X \leq b$$



half-closed interval $(a, b]$

Homework

Find the intersection of 2 intervals



Summary

List

define a list with [] list()

ordered

mutable , changeable

length of list can be changed

can be contain of different obj types

indexed, index start from Zero

methods in List:

*'append', 'clear', 'copy', 'count',
'extend', 'index', 'insert', 'pop',
'remove', 'reverse', 'sort'*

Tuple

define a tuple with () tuple()

ordered

immutable , unchangeable

*length of tuple can **Not** be changed*

can be contain of different obj types

indexed, index start from Zero

methods in tuple:

'count', 'index'

Dictionary

define a dictionary with { : }

dict([('a', 1), ('b', 2)])

dict(a = 1, b = 2, c = 3)

mutable , changeable

not ordered until python version 3.7+

can be contain of different obj types

methods in dict:

'clear' , 'copy' , 'fromkeys' ,

'get' , 'items' , 'keys' , 'pop' ,

'popitem' , 'setdefault' ,

'update' , 'values'

Set

define a set with { }

set()

no duplicates ==> good for remove duplicates

not ordered ==> 'set' object is not subscriptable

mutable , changeable

methods in set:

*'add', 'clear', 'copy', 'difference', 'difference_update',
'discard', 'intersection', 'intersection_update',
'isdisjoint', 'issubset', 'issuperset', 'pop',
'remove', 'symmetric_difference',
'symmetric_difference_update',
'union', 'update'*

“

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

O futuro profissional começa aqui

iscte
INSTITUTO
UNIVERSITÁRIO
DE LISBOA

 **emprego
digital**

 **UPskill**