Week 6

Slides & Exercises

Review on exercise 3-2(again!)

HOMEWORK Hw 6-1 Need more practice/example on dictionary!(list & tuples are easier)!

mygrades={‘Python’:’A’,’Java’:’B’,’JavaScript’:’A’,’English’:’C’,’Japanese’:’B’, ‘PE’:’D’}

gradevsvalue={‘A’:100,’B’:80,’C’:65,’D’:20}

creditsnsemester={‘Seminar’:(2,’fall’), ‘JavaScript’:(2,’spring’),’Java’:(4,’fall’), ‘Algebra’(3:’fall’)

,’Python’:(2,’spring’),’Japanese’:(4,’fall’),’English’:(4,’spring’),‘PE’:(2,’spring’)}

Calculate myGPA and count how many course taken in spring and fall!
Print myGPA and course numbers taken in spring&fall!

Review on exercise 4 (OOP)??

Steps/tips

-

Check your system

- if you don’t have mac laptop then you may need to use CNS machine (ssh from your laptop using any terminal program)
- check your virtual environment installation (week 1 slides)

TODAY

We will start writing some “real”/ practical use Python! Time to put together what you’ve learnt in action.

Automatically “browse” a web page, extract informations and save it to a file

(“web scraping”)

Required knowledge as the basis for web scraping today

- Basic File I/O
- Regular Expression/Pattern Matching
- Some useful modules

yes we have learnt these:

how to use & make modules, OOP(modules are usually written in OOP style: have methods, etc), data structures(list,tuples, dictionary for data processing)

Basic File I/O

OPENING AND WRITING INTO A FILE

```
#!/usr/bin/python
# Open a file
fo = open("log.txt", "w")
fo.write( "Python is a.....\nYeah its fun!!\n")
# Close opend file
fo.close()
```

READING AND PRINTING THE CONTENT OF A TEXT FILE

```
#!/usr/bin/python
# Open a file
fo = open("log.txt", "r")
str = fo.read();
print "Read from file: ", str
# Close opend file
fo.close()
```

Exercise 6-1

- Open this URL: http://www.sfc.wide.ad.jp/IRL/contactus.html, save it as <scrape.html>
- Write a script that open/read “scrape.html” and print the content (to the screen)
- Add more lines/codes to open/create another file <log.txt>, define a string hey=”This is a log file”. Write that string into <log.txt>. Check/verify <log.txt> by opening in text editor

Regular expression / Pattern Matching

Very powerful and useful for text processing!

Let’s use one of the good materials on this topic:

https://developers.google.com/edu/python/regular-expressions

A good tools for practising:

http://regexr.com/

Exercise 6-2

Modify Exercise 6-1 , add lines to (dont forget to do import re)

- Grab/Match phone number based on this (japanese) pattern/format 0466-49-3529 and email addresses

Hints: experiment using the tools suggested on the previous slide. Use “findall” not search. To make it easier specify 2 (email, phone) separate match and use to separate variable for the result

- Write the phone numbers (and email too if you want)into <log. txt>

Hints: need to find a way to convert list into string before writing it into file(for example ('\n'.join(match)) will convert a list called “match” into a string , members of the list will be separated by new line (\n) A reminder for “\n” : https://docs.python.org/2.0/ref/strings.html)

- Check/verify <log.txt> by opening in text editor

Web scraping (“browsing”)

install these modules(using “easy_install”):

- urllib & urllib2 if not exist (try with import)-reference URL on SFS
- mechanize-reference URL on SFS
- BeautifulSoup-reference URL on SFS

Let’s have a look at these examples:(download them from SFS) **ex51.py,ex52.py,***ex54.py,***ex55.py,ex56.py**

Exercise 6-3

Modify Exercise 6-2 so that it access this URL
http://www.sfc.wide.ad.jp/IRL/contactus.html
instead of reading “scrape.html”

(refer to the examples provided above)

Optional today Exercise (homework?) 6-4

Write a script that use “mechanize” and “BeautifulSoup” to get a list of university in Japan. The script should save the result to a file, call it “japanuniv.txt”