1 CSC299 - Lab Assignment 5

IMPORTANT INSTRUCTIONS

Use this URL to verify your progress:

`https://mdp.cdm.depaul.edu/csc299`

Login into:

`mdp.cdm.depaul.edu`

Under your csc299 folder create a new folder called `lab05` and, under the latter, create a new file called `README.json`. This file should contain:

```
{"student_id": "<yourstudentid>", "name": "<yourname>", "email": "<youremail>"}
```

After completing each task below remember to do:

```
git add README.json
git add .
git commit -a -m "task completed"
git push
```

else I will not receive your work.
ATTENTION: Files are different for each student.

1.1 Task 1 (3 points)

Create a new folder `/csc299/lab05` and `cd` under that folder.
Download the document at this URL

`http://odata.cdm.depaul.edu/Cdm.svc/Courses?$orderby=CatalogNbr&$filter=EffStatus%20eq%20%27A%27%20and%20SubjectId%20eq%27CSC%27`

into a file called `CSC.xml`. The file has the following structure

```
...
      <m:properties>
        <d:CrseId>001362</d:CrseId>
        <d:SubjectId>CSC</d:SubjectId>
        <d:CatalogNbr>200</d:CatalogNbr>
```

```
      ...
    </m:properties>
    <m:properties>
      ...
    </m:properties>
...
```

Use BeautifulSoup to extract the content of the first `<m:properties>` block and dump it to a file `CSC.properties.1.xml`. Call your program `program51.py`.

1.2 Task 2 (3 points)

Use BeautifulSoup to extract the content of the first `<m:properties>` block and convert it to a python dictionary as follows:

```
<m:properties>                      -> {
  <d:CrseId>001362</d:CrseId>       ->   'CrseId': '001362',
  <d:SubjectId>CSC</d:SubjectId>    ->   'SubjectId': 'CSC',
  <d:CatalogNbr>200</d:CatalogNbr> ->   'CatalogNbr': '200',
  ...                                    ...
</m:properties>                     -> }
```

and use simplejson to store the dictionary into a file `CSC.properties.1.json`. Call your program `program52.py`.
Tip: Consider this code:

```
soup = BeautifulSoup(page)
soup.findAll(name=re.compile('^d\:\w+'))
```

The findall finds all tags with a name starting with `d:`*something*.
Tip: Consider this code:

```
import simplejson
obj = {'hello':'world'}
simplejson.dump(obj, open('obj.json','w'))
copy = simplejson.load(open('obj.json','r'))
print obj == copy
```

The simplejson library allows to dump an object to a file into JSON format and load it back.

1.3 Task 3 (3 points)

Use BeautifulSoup to extract the content of all properties items into a list of dictionaries and store the list into a JSON file `CSC.json`. Call your program `program53.py`.

1.4 Task 4 (2 points)

Write a program called `program54.py` that takes as command line argument some keywords for example:

```
python program54.py Python > python.log
```

Run the above command!
and outputs a list of course names and course descriptions that include the keywords (in the example "Python") in the course name.
Tip: Consider this code:

```
import sys
text = 'this is a test'
keywords = sys.argv[1:]
if keywords and all(k in text for k in keywords):
    print text
```

It reads the command line arguments into keywords and prints text if text contains all the keywords.