

# Homework 4

Due April 10, 2017

Submissions are due by 11:59PM on the specified due date. Submissions may be made on the Blackboard course site under the Assignments tab. Late submissions will be accepted for a 10% penalty each day (up to 48 hours). Make sure your name and FSUID are in a comment at the top of the file.

## 1 faculty\_directory.py (50 points)

Suggested modules: requests, lxml or re

In this part of the assignment, your job is to create a script that will navigate to the FSU CS department website and print out a directory containing the telephone number, office location, email, and webpage of every faculty member in the department. The root page for the faculty listings is <http://www.cs.fsu.edu/department/faculty/>.

From this page, you can find links to every individual faculty page. Each individual faculty page lists the links that you need. *However, the only link that you may hardcode into your file is the one above. You also may not hardcode faculty names.* All other information from this point must be retrieved using crawling and scraping methods. Here is some guidance:

1. First, try to gather all of the links to the individual faculty pages.
2. Navigate to the first page and scrape the data to be output to the user.
3. Repeat step 2 until all faculty information has been output to the user. An example of what your output should look like is shown below. Missing information should be indicated with an "N/A" entry. Be careful of edge cases or inconsistencies, you may need to be a little creative – double check your output against the web pages!

```
caitlin@pymachine$ python faculty_directory.py
Name:  Sudhir Aggarwal
Office:  263 Love Building
Telephone:  (850) 644 0164
E-Mail:  sudhir [ at cs dot fsu dot edu ]
*****
Name:  Theodore Baker
Office:  N/A
Telephone:  N/A
E-Mail:  baker [ at cs dot fsu dot edu ]
*****
Name:  Mike Burmester
Office:  268 Love Building
Telephone:  (850) 644-6410
E-Mail:  burmeste [ at cs dot fsu dot edu ]
...

```

## 2 imgur\_info.py (50 points)

Suggested modules: requests, json

Your job in this part of the assignment is to sort through Imgur user comment data. Your program should begin by prompting the user to enter a username of an Imgur account.

An Imgur user page can be found at <http://imgur.com/user/<username>>, but the comments on this page are loaded dynamically which can make it very hard to pull data. The dynamically loaded content, however, is systematic and easy to retrieve. All of the user's comments can be found at

<http://imgur.com/user/<username>/index/newest/page/<num>/hit.json?scrolling>

where <num> is a counter that starts at 0 and increases as needed. When there are no more comments to receive, the next page in the sequence will simply contain an empty string. For example, navigate to the page

<http://imgur.com/user/LastAtlas/index/newest/page/0/hit.json?scrolling>

Now, increase the page counter in the address by 1. If we set the page counter to 100, for instance, we'll see that the page is empty. However, there's no way to know beforehand how many pages each user requires.

If the username does not exist or has not been used to post any comments, you may simply print a message and end the program. Otherwise, you should sort through the user's comment data to find the top 5 comments (i.e. the comments with the most points). Your output should list these comments in descending order of points, detailing the post identifier (the "hash"), title of the post commented on, number of points received, and timestamp of the comment. In the case of a tie between points, break the tie by comparing "hash" values.

```
caitlin@pymachine$ python imgur_info.py
Enter username: LastAtlas
1. XJ7xbSk
Points: 19
Title: This man pulled, pushed and lifted his disabled twin brother
through an IronMan. Here is a touching picture of them at the finish
line.
Date: 2014-08-24 15:36:54

2. wEF0R
Points: 11
```

Title: What a guy  
Date: 2015-08-02 04:00:30

3. ZsqSJ  
Points: 7  
Title: MRW I find out I am unknowingly sharing a BF  
Date: 2015-01-25 15:43:58

4. HfrBZSJ  
Points: 5  
Title: This is just magnificent  
Date: 2014-02-17 03:42:57

5. NLiuVXC  
Points: 5  
Title: My wife...  
Date: 2014-02-17 09:58:07