Introduction to Python

October 8, 2013

hack@uchicago
http://hack.uchicago.edu/

hack@uchicago

- RSO for students interested in technology
- We organize hack nights, hackathons, workshops, etc.
- Join our Facebook group, or visit our website (http://hack.uchicago.edu/)

Upcoming Events

- Hack Night
  Every Friday 5pm-8pm @ Ryerson 4th Floor
- Tech Talk: “An Introduction to Object Relational Impedance”
  10/10 4:30pm @ Ryerson 4th Floor
- Tech Talk: “Analytics Technology in the Obama Campaign and Beyond”
  10/17 4:30pm @ Ryerson 4th Floor
- Google Tech Talk
  10/24 4:30pm @ Kent 107
- Hackathon
  10/26-10/27 http://theforge.eventbrite.com/
- Twitter Tech Talk
  11/7 4:30pm @ Kent 107

For more details, check out the event calendar at
https://studentactivities.cs.uchicago.edu/

What does this do?

```
a = [10, 1, 2, 76, 20, 3, 20, 49, 98, 5]
n = 0

for i in a:
    if i > 5:
        print i
    else:
        n += 1

print "n is " + str(n)
```

```
#include<stdio.h>

int main()
{
    int a[] = {10, 1, 2, 76, 20, 3,
               20, 49, 98, 5};
    int n = 0;

    for(int i=0; i < 10; i++)
    {
        if(a[i] > 5)
            printf("%i\n", a[i]);
        else
            n++;
    }

    printf("n is %i\n", n);
}
```

```
10
76
20
20
49
98
n is 4
```

Submit your output at http://bit.ly/hack-python

Exercise

- Go to workshop website: http://bit.ly/hack-python-workshop
- Go to “Examples” and download the examples.
- Run simple/intro.py from the command line.
- Run the simple/intro.py from the interpreter.

How this is going to work

- We will be alternating between three things:
  - Demonstrating basic examples using an IPython Notebook. You will be able to access this code as I type it (follow the “live examples” link on the website).
  - Live coding of more substantial examples that use a Twitter dataset, alternating between an editor and the interpreter.
  - Exercises using the Twitter dataset.

Lists

Lists example

- I am going to write a program that prints the lengths of the first N tweets from the dataset.
  - examples/twitter/get_tweet_lengths.py
- We have prepared some functions that handle the actual reading of the dataset.
  - examples/twitter/workshop.py

Lists exercise

- Modify my program to compute the *average* length of the tweets.
- If you find that easy, modify it further so it will also compute the standard deviation.
  - Square root:

```
import math

math.sqrt(x)
```

Dictionaries

Dictionaries example

- Instead of load_tweets_text(), we're going to use load_tweets(), which returns a list of dictionaries.
- Each dictionary is a single tweet, with all the information returned by the Twitter API.
  - examples/twitter/print_tweet_info.py
- I am going to write a program that computes the frequency of each length.
  - examples/twitter/get_length_frequencies.py

Dictionaries exercise

- Modify my program to compute the frequency of hashtags.
  - Use the workshop.get_hashtags() function to extract the hashtags from a single tweet.
- Print out only the top 10 hashtags.

Functions

Functions example

- I am going to write a function that returns a list of all the lengths in the dataset.
  - examples/twitter/functions.py
- And, I am going to produce a histogram of those lengths.
  - examples/twitter/lengths_histogram.py
  - You will need matplotlib installed for this to work (http://matplotlib.org/)

Functions exercise

- Modify functions.py to implement these functions:
  - `extract_values(n, tweets_file, field)`
    Similar to extract_lengths, but extracting a specific field from the tweet.
  - `compute_frequencies(l)`
    Given a list of values, returns a dictionary mapping values to the number of times that value appears in the list.
    Similar to how we counted the hashtags
- Once you've done this, you can test it with get_frequencies.py

List Comprehensions

Simple visualizations