Mining Social Media Data
Spring 2014
IST 400/600

Nancy McCracken
Research Associate Professor

Contact Information: Office: 209 Hinds Hall
njmccrac@syr.edu
443-3955 (office)

Class Sessions: Mon/Wed 2:15 – 3:35 027 Hinds Hall

Course Description:

This course focuses on the collection and analysis of Social Media data to understand the properties and behavior of information that pertains to individuals, communities and other organizations and their connections. The course will encompass a process that starts with collecting publically available Social Media data, storing and carrying out several forms of analysis and visualization and making the information accessible for further use.

In this course, the Web APIs of several online Social Media platforms will be used in order to collect social media data. Scripts to use these APIs will be in the programming language Python. The first part of the course will focus on using the Twitter API while exploring the use of Python programming for these techniques, and then will apply those techniques to other sites, which may include Facebook, YouTube comments, FourSquare, blogs and newsfeeds.

The analysis portion of the course will include introductory sessions in data analysis and visualization, Social Network Analysis (SNA) and Natural Language Processing (NLP). These will include techniques to use the data collected from the Web APIs to answer questions such as:

What do the people using social media have in common? Who are the most influential/popular people in the network?
What are they chatting about and is it interesting? How are people responding to events? How do they feel about other communities, companies and organizations?

In the class, we will connect the results of our data collection with other software packages to manipulate and view the data. These will include

- packages for building and visualizing networks,
- microformats such as JSon and geo for representing data,
- NoSQL databases,
• authorization for access to online resources, and
• the use of regular expressions to find people and topics in text.

Textbook
There will be no textbook for the course, but several programs for accessing web APIs will be used from the open source repository of Matthew Russell, as described in:


Online resources for learning Python programming and using other software packages will be introduced as needed.

Pre-requisite:
Students will be expected to know some programming before taking the course; IST 256 is an appropriate pre-requisite for undergrads or equivalent programming knowledge gained through another avenue such as self-study. Please send email to the instructor if you have any questions about meeting the pre-requisite.

Course Organization:
The organization of the course will be largely informal, viewed as an exploration of the tools in order to achieve the goals. The format will be a combined lecture/lab format with some short presentations of materials, followed by a set of examples to explain and work through those parts of Python that result in using Python scripts to access the Web APIs and to manipulate the collected data.

As the Social Media platforms change fairly rapidly, we will adapt our use of the software as we go through the course. Students will be encouraged to pose problems and share their solutions and discoveries for all to use.

Coursework:
Most of the work in this class will involve a mix of running existing Python scripts for the Web APIs and doing original programming to connect the data with other software to achieve analysis and visualization. Since most of the coursework involves writing and running programs, students should be prepared to learn Python and to work in a complex software environment.

In addition to working through examples in class, weekly homework assignments will be given in the first half of the class for doing additional examples or adapting the Python scripts to other uses. The second half will be spent in working on a project where students will design a data collection and analysis problem to solve. Students may work together in small groups or may work independently on the project.

As this is a combined undergrad/grad class, graduate students will be required to do an additional report based on the project.

Learning Outcomes
After completing this course, students will understand a variety of problems that use information from the connections and content of social media data. They will be able to
use interfaces to a number of Social Media platforms to collect and store social media data. They will be able to use the Python programming language to manipulate this data into other formats, using additional Python tools. They will be familiar with several ways to visualize and analyze the data.

**Grading**

Attendance will be required for this course as students will be expected to participate in and contribute to techniques to analyze and visualize data. A second component of participation is contributing to in-class discussions or on-line discussion threads.

<table>
<thead>
<tr>
<th>Participation</th>
<th>20%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework assignments</td>
<td>40%</td>
</tr>
<tr>
<td>Project</td>
<td>40%</td>
</tr>
</tbody>
</table>

**Academic Integrity**

Syracuse University sets high standards for academic integrity. Those standards are supported and enforced by students, including those who serve as academic integrity hearing panel members and hearing officers. The presumptive sanction for a first offense is course failure, accompanied by the transcript notation “Violation of the Academic Integrity Policy.” The standard sanction for a first offense by graduate students is suspension or expulsion. Students should review the Office of Academic Integrity online resource “Twenty Questions and Answers About the Syracuse University Academic Integrity Policy” and confer with instructors about course-specific citation methods, permitted collaboration (if any), and rules for examinations. The Policy also governs the veracity of signatures on attendance sheets and other verification of participation in class activities. Additional guidance for students can be found in the Office of Academic Integrity resource: What does academic integrity mean?

**Disabilities**

If you believe that you need accommodations for a disability, please contact the Office of Disability Services (ODS), [http://disabilityservices.syr.edu](http://disabilityservices.syr.edu), located at 804 University Avenue, room 309, or call 315-443-4498 for an appointment to discuss your needs and the process for requesting accommodations. ODS is responsible for coordinating disability-related accommodations and will issue students with documented disabilities “Accommodation Authorization Letters,” as appropriate. Since accommodations may require early planning and generally are not provided retroactively, please contact ODS as soon as possible.

**Educational Use of Student Work**

I intend to use academic work that you complete this semester in subsequent semesters for educational purposes. Before using your work for that purpose, I will either get your
written permission or render the work anonymous by removing all your personal identification.