

Computing for Medicine: Phase 3, Seminar 5 Project

Jennifer Campbell

Associate Professor, Teaching Stream

campbell@cs.toronto.edu



Seminar 5 Project

- The project handout is posted:
 - <http://c4m.cdf.toronto.edu/cohort1/phase3/>
- Two approaches for doing your work:
 - Use the Computer Science Teaching Labs computing network.
 - Use your personal computer.
- Software to install:
 - tweepy (<http://www.tweepy.org/>)
 - `pip3 install tweepy`
 - textblob (<https://textblob.readthedocs.io/en/dev/>)
 - `pip3 install textblob`

OVERVIEW

Starter code and data

- Example code (SQL and Tweepy tutorials):
 - `config.py`
 - `tutorial_pysqlite.py`
 - `tutorial_tweepy_stream_api.py`
- Starter code:
 - `settings.py`
 - `project-sol.py`
- Data will be collected from Twitter.

Your tasks

Project goal:

Create a database of tweets about medical terms.

- Create a Twitter account (if necessary) and follow the instructions provided to set up Twitter app development.
- Read and understand the example code.
- Complete `settings.py`.
- Complete the TODO items in
 - `project-sol.py`by modeling your solutions after the provided tutorials and starter code.

PYTHON TOOLS

sqlite3

- SQL code can be integrated with Python using the `sqlite3` module.
- See `db_demo.py`

textblob

- A Python natural language processing (NLP) library.
- How you will use it:
 - To perform sentiment analysis on tweets.
 - Two key attributes:
 - *Polarity*: measures the negativity, neutrality, or positivity of the text on a scale of 0 (neg) to 1 (pos).
 - *Subjectivity*: measures the subjectivity of the text as a value from 0 (objective) to 1 (subjective).
- See `textblob_demo.py`

tweepy

- A Python library used to access the Twitter API.
- See `tweepy_demo.py`

Object-oriented programming (OOP)

- Up until now, we have been using classes (e.g., `str` and `list`) and methods (e.g., `str.startswith` and `list.append`), but we haven't written our own classes.
- In `project-sol.py`, a class named `StreamListener` is defined.
- You do not need to know about OOP in order to finish the TODO items in this project.
- However, if you would like to learn more about OOP, you may find [these videos](#) helpful.

Running `project-sol.py`

- Set variable `db_flag` to:
 - 1 to initialize the database
 - 0 if the database has already been initialized

Running `project-sol.py` (2)

- `project-sol.py` imports `settings.py`
- In `settings.py`, you must initialize the values of `ACCESS_KEY`, `ACCESS_SECRET`, `CONSUMER_KEY`, `CONSUMER_SECRET`.
- Also, in `settings.py`, `MAX_TWEETS` is initialized to 500.
 - When you are first running your program, you may want to set `MAX_TWEETS` to a smaller number (e.g., 25) so that the program takes less time to run.
 - Once you are fairly confident that your program is working, you can increase the value that `MAX_TWEETS` refers to.

SUBMITTING PROJECTS

Submitting your first project

- Deadline: Tuesday, December 20, 2016 by 6:00pm
- Submission process:
 - Submit on MarkUs:
 - <https://markus.teach.cs.toronto.edu/c4m-2016-01>
 - Email Jen <campbell@cs.toronto.edu> to specify which project you submitted.
- Getting help:
 - Piazza (<https://piazza.com/utoronto.ca/fall2016/c4mph3>)
 - Email (Jen or the seminar speakers)
 - Note: Jen is travelling Dec 16-23 with very limited email access.

UPCOMING SEMINARS

Seminar 5: Dr. Michael Brudno

- Tuesday January 17 2017 6-8pm
- Location: DCS Innovation Lab
- Topic: Rare disease data capture
- <http://www.cs.toronto.edu/~brudno/>

FEEDBACK

Phase 3, Seminar 5 Survey

- You will receive an email with the subject “C4M: Phase 3, Seminar 5 Feedback Survey”.
- Please complete that survey now:
 - <https://www.surveymonkey.com/r/C4MSeminar5>