Digital Humanities Lab Fellow

Hannah Rafkin, Class of 2017

During the month of June, I worked alongside Professor Crystal Hall, Sabina Hartnett ’18, Gabby Papper ’18, and Phoebe Bumsted ’17 to develop the curriculum of Introduction to Digital and Computational Studies, a course that will be taught in Fall 2015. I also performed independent research on Twitter’s role in modern protest, particularly the #BlackLivesMatter movement.

Twitter is becoming an increasingly relevant tool for academic study. As political events and protests unfold, millions of citizens take to Twitter to express their opinions. Collected en masse, this data serves as a modern-day form of primary document. Datasets of tweets can be powerful indicators of public sentiment, prevalent themes, patterns of communication, formation of online communities, and group identities.

This summer, I created a detailed tutorial walking researchers through the process of collecting and analyzing Twitter data. The collection process begins with choosing a hashtag to analyze. (#BlackLivesMatter, #AskRachel, #IfIMWereGunnedDown, and #Charleston were relevant hashtags during the time of my research). Next, the researcher scrapes tweets. This entails collecting thousands of individual tweets per day along with the related metadata (date, time, username, geolocation, etc.) – using a service called Hawksey. The data is most powerful if tweets are scraped every day over an extended period of time.

The analysis process involves topic modeling, a tactic used by digital humanists to identify overarching patterns across a large corpus by identifying the most commonly recurring patterns of word clusters, or topics. When working with a set of texts too large for close reading, topic modeling allows researchers to uncover prevalent themes as well as subtler trends.

I wrote a script in R (a programming language popular among digital humanists) to apply this analytical approach to Twitter. This script inputs the collected Twitter data in spreadsheet form, breaks up the larger dataset into chunks containing a particular number of tweets, and outputs a set of text files to be used later on in topic modeling. The topic modeling is done using Google’s Topic Modeling Tool. After using the R script, the researcher uploads the outputted text files into the Topic Modeling Tool. From there, the tool will run through the files and return a list of topics along with more specific metadata. Analyzed in context of the issues and dynamics attached to the particular hashtag, these results can lead to eye-opening discoveries.

I hope that this tutorial will be helpful to students and professors as Twitter becomes increasingly applicable in the classroom and in academia more broadly. I express my gratitude to Professor Hall, the Mellon Humanities Initiative, and Bowdoin’s Digital and Computational Studies Initiative for the opportunity to contribute to such an exciting and relevant field.

Faculty Mentor: Professor Crystal Hall

Funded by the Mellon Humanities Initiative