QTM 340, "Practical Approaches to Data Science with Text"
Final Project Overview

Your final project assignment is in some ways fairly straightforward: to envision and execute your own data science project involving text. Ultimately, you will be submitting a polished final paper that documents your project. You will have the choice of writing a more general-audience paper in the style of The Pudding features or a more traditional scientific paper. More details on the options for the final paper are to come. This document is intended as an overview of the final project, so that you get a sense of the scope and expectations of the project, the project structure, and your various responsibilities along the way.

Assignments
Datasheet, due 10/20
Project proposal, due 10/29
Exploratory analysis, due 11/10
Presentation, in class on 11/17 and 11/19
Final project due, 12/4

Policy on Groups
You may work alone or in groups of 2-4 people. If you choose to work in a group, the scope of your project should reflect that increased capacity. In addition, each group member will be required to submit a 1-page account of their individual contributions to the group; each group member will be graded individually.

Places to look for interesting data:
- Data is Plural
  https://tinyletter.com/data-is-plural/archive
- Kaggle Datasets:
  https://www.kaggle.com/datasets
- HathiTrust Digital Library:
  https://www.hathitrust.org/
- UCI Machine Learning Repository
  http://mlr.cs.umass.edu/ml/
- This meta-list (look at the sentiment analysis and NLP lists in particular):
  https://lionbridge.ai/datasets/the-50-best-free-datasets-for-machine-learning/
- Twitter API:
  https://developer.twitter.com/en/docs/twitter-api

Emory librarians to consult:
Dr. Erica Bruchko (link to schedule appointment via link below)
http://web.library.emory.edu/about/staff-directory/woodruff/bruchko-erica.html
Dr. Rob O'Reilly
http://web.library.emory.edu/about/staff-directory/woodruff/oreilly-rob.html