Machine Readings: Text Analysis for the Information Age

Winter 2016

ENGL 55 / MATH 5 / QSS 30.02

Time: at the 12 hour

Location: Haldeman Center 28 (lower level of Kemeny Hall)

Instructor: Allen Riddell

Office Hours: Kemeny 333, Tuesdays 2:00-3:00pm

Description

Library digitization has made millions of books, newspapers, and other printed materials accessible to the public. In this course we will learn how to draw on computational resources to analyze a range of materials, including poetry, novels, newspaper articles, and personal diaries. We will explore debates about the representation of literary texts as "data" and consider the challenges "machine reading" poses for research in the humanities and how we think about what it means to read a text. Through case studies we will reflect critically on the history of the digital humanities (formerly known as humanities computing) and will gain practical experience in text analysis.

Prerequisites

This course welcomes students from a diversity of academic backgrounds. Students are expected to be familiar with algebra at the level of Math 1 or Math 2 (http://www.math.dartmouth.edu/courses/by-course/). No prior experience with computer programming is required for this course.

Required Texts


It should be possible to find inexpensive copies of these texts. Other readings will be made available.

Having one of the following texts on hand is recommended:

- The Python Tutorial (http://docs.python.org/3.3/tutorial/) (official documentation). Free, very well written.

You may wish to browse the table of contents for these texts and select the one that is best suited to you.
The following (interactive) online resources are recommended:

- Computer Science Circles (University of Waterloo) ([http://cscircles.cemc.uwaterloo.ca/](http://cscircles.cemc.uwaterloo.ca/)) (Python 3)
- Introduction to Computer Science and Programming Using Python (MIT 6.00.1x) ([https://www.edx.org/course/introduction-computer-science-mitx-6-00-1x-0](https://www.edx.org/course/introduction-computer-science-mitx-6-00-1x-0)) (edX, NB: Python 2!)

Assignments

You will complete three written assignments during the term and a final project. The course also features programming assignments (to be completed outside of class) and in-class lab assignments intended to help you learn the Python programming language and to familiarize you with methods of text analysis discussed during the week.

Course Grade

You will be evaluated according to your work on homework (15%) and lab assignments (10%), on written assignments (35%), and on the final project (40%). In grading written assignments and the final project, I will consider your grasp of and ability to engage analytically with the materials encountered in the course.

You are encouraged to discuss course material, including written assignments and homework. The work that you turn in must reflect their own work. Code and writing submitted may not be copied from other students, books, or websites. Plagiarism, the offense of passing off someone else's work as your own, is a violation of Dartmouth's honor code that the instructor must bring to the attention of the disciplinary committee. Such a violation can result in a student's expulsion from Dartmouth College. Please consult the policies concerning the honor principle ([http://www.dartmouth.edu/~deancoll/student-handbook/principles.html](http://www.dartmouth.edu/~deancoll/student-handbook/principles.html)). If you have any questions regarding plagiarism or the honor principle, please consult your instructor.

Course Policies

You are expected to attend all class sessions. Missing more than three sessions will have consequences for your grade. Students wishing to take part in religious observances that conflict with their participation in the course should meet with me before the end of the second week of the term to discuss arrangements.

The lowest two grades for non-written assignments will be dropped; no credit will be given for late work. If you are unable to finish an assignment by the deadline, submit your work for partial credit.

Written assignments must be typed, double-spaced, with regular font-size (10-12 point) and standard margins (1" top and bottom 1.25" sides). Acknowledge any ideas that are not your own. Students are encouraged to consult Kate L. Turabian's *A Manual for Writers of Research Papers, Theses, and Dissertations* (Chicago Style) or the *MLA Handbook for Writers of Research Papers* (MLA Style). For help with writing assignments, you may wish to consult John R. Trimble's *Writing with Style* (any edition) and/or a writing tutor in the Writing Resource Center RWIT (Baker Berry Library; appointments may be scheduled online ([http://www.dartmouth.edu/~rwit/](http://www.dartmouth.edu/~rwit/))).

Students with learning, physical, or psychiatric disabilities are encouraged to contact me in order to discuss accommodations. If you think you might need special accommodation but are not currently registered for those provisions, please contact Student Accessibility Services ([http://www.dartmouth.edu/~accessibility/](http://www.dartmouth.edu/~accessibility/)), immediately.

Discussion Kickoffs
Beginning Week 2, each member of the class will select a session during which they will inaugurate our
discussion. While considerable flexibility is given to the inagurator, a typical discussion kickoff would include: (1) a
succinct description of the project of (one of) the text(s); (2) a brief critical response to the material; and (3) a
query or claim addressed to the group arising from the text.

"Green" and "White" assignments

Assignments typically contain two sets of exercises. One set is labeled "white" and the other is labeled "green".
By the end of the second week you will need to decide which set of assignments is best suited for you. For
example, if you decide on "white" then I expect you will complete the "white" exercises for the rest of the class.

Calendar

Subject to revision. Readings not linked directly are available on the discussion site (https://piazza.com
/dartmouth/winter2016/engl5502math501qss3002/home) (under the "Resources" tab (https://piazza.com
/dartmouth/winter2016/engl5502math501qss3002/resources)).

Week 1 (January 4, 2016)

Reading (before Wednesday)

- Hockey, Susan, "The History of Humanities Computing (http://www.digitalhumanities.org/companion
  /view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-2-1)" in Susan Schreibman,
- Svensson, Patrik, "Humanities Computing as Digital Humanities" (http://digitalhumanities.org/dhq/vol
- Install Python 3.5. If you're new to Python, installing the distribution of Python provided by Anaconda
  (http://docs.continuum.io/anaconda/install) is easiest. Please be sure to install Python 3.5. Do not install
  Python 2.7.
- Read some or all of "Chapter 1: Creating a Digital Library" (http://nbviewer.ipython.org/github/fbkarsdorp
  /python-course/blob/book/Chapter%201%20-%20Getting%20started.ipynb) from Programming for the
  Humanities.

Session 1: Introduction to the course, grand tour of quantitative text analysis in the human and interpretive social
sciences, history of humanities computing/digital humanities

Session 2: Introduction to the programming language, basic data types, representations of texts.

Material discussed in Session 2 is covered in the Python Tutorial "An Informal Introduction to Python"
(https://docs.python.org/3/tutorial/introduction.html) and in "Chapter 1: Creating a Digital Library"

Lab 1

Assignment 1 (due Thursday, January 14 AoE (https://en.wikipedia.org/wiki/Anywhere_on_Earth)))

If you have any problems installing Python, please ask for help on the discussion site. We'll also have a chance
to install the software in class on Friday.

Week 2 (January 11, 2016)
Reading (before Monday)

How (and why) we read


Session 3: How we read (and write).
Session 4: Loops and other constructs
Lab 2
Assignments on Canvas

Week 3 (January 18, 2016)

NB: January 18, Monday -- Martin Luther King Jr. day - classes moved to x-periods.

Reading (before Tuesday)

- Mining the Dispatch (http://dsl.richmond.edu/dispatch/) (website). Browse the website and read "Introduction" (http://dsl.richmond.edu/dispatch/pages/intro).
- Recommended: Watch PBS program A Midwife’s Tale

Session 5: Time and Text Collections: A Midwife’s Tale (Meet on Tuesday during x-hour)

Session 6: Writing and calling functions
Lab 3
Assignments on Canvas
Week 4 (January 25, 2016)

Reading (before Monday)

- Cosma Shalizi, "Graphs, Maps, Trees, Fishing" (http://www.thevalve.org/go/valve/article /graphs_trees_materialism_fishing/)
- Recommended: Scott McLemee, "Literature to Infinity" (http://www.insidehighered.com/views/mclemee /mcleeme193)

Python resources for this week:

- Matplotlib: plotting (http://scipy-lectures.github.io/intro/matplotlib/matplotlib.html)

Session 7: Sociology of Literature I: Moretti

Session 8: Program design

Lab 4

Assignments on Canvas

Week 5 (February 1, 2016)

NB: Lab moved to Monday, discussion on Wednesday

X-hour review moved to Thursday. E-mail me to schedule a time between 2pm and 5pm.

Reading (before Monday)

- James F. English, "Literary Studies"
- Recommended: The VIDA Count 2013 (http://www.vidaweb.org/the-count-2013/)

Session 9: Sociology of Literature II: Radway, Tuchman and Isaac

Session 10: Manipulating text, bag-of-colors, probability

- Recommended: pandas Tutorials (http://pandas.pydata.org/pandas-docs/stable/tutorials.html) (e.g., "10 minutes into pandas")

Lab 5

Assignments on Canvas
Week 6 (February 8, 2016)

Class on Wednesday is canceled. Lecture will be on Friday. Lab moved to next week.

X-hour review moved to Thursday. E-mail me to schedule a time between 2pm and 5pm.

Readings

- Chapters 4-6 from Ian Hacking. 2001. An Introduction to Probability and Inductive Logic. Cambridge University Press. If this material is new to you, I recommend doing the odd (or even) exercises at the end of each chapter. (Answers are supplied at the end of the book).
- Federalist No. 6 (http://thomas.loc.gov/home/histdox/fed_06.html)
- In lieu of Anti-Federalist essays (recommended):
  - Interview with Sanford Levinson. 2015. Do We Need a New Constitutional Convention?: Reading The Federalist in the twenty-first century (https://bostonreview.net/us/richard-kreitner-sanford-levinson-federalist) Boston Review

Session 11: Mosteller & Wallace: The Federalist Papers

Session 12: Probability

Lab 6

Assignments on Canvas

Week 7 (February 15, 2016)

Lab during x-hour on Tuesday. Lab may also be done outside of class anytime on Tuesday.

Readings

- Chapters 7, 11, 15 from Ian Hacking. 2001. An Introduction to Probability and Inductive Logic. Cambridge University Press. Consider doing a few of the exercises at the end of each chapter. (Answers are in the back of the book.)
- (recommended) D. H. Mellor discusses the meaning of probability (http://philosophybites.com/2014/12/hugh-mellor-on-probability.html).
Session 13: Authorship attribution, adversarial stylometry, Henry James

Session 14: Probability (cont.), Bayesian Inference

Lab 7

Assignments on Canvas

Final project (filename)/pages/hw/final-project/final-project.md (due Sunday, March 13, 2016, AoE)

**Week 8 (February 22, 2016)**

Readings


Session 15: Networks, ‘Time on the Cross’

Session 16: Networks, Topic Models

Lab 8 (filename)/pages/labs/08/lab-08.md

**Week 9 (February 29, 2016)**

Session 17: Lab (Final project work)

Session 18: Lab (Final project work)

Lab 9: Lab (Final project work)

**Week 10 (March 7, 2016)**
Acknowledgments

The course's subtitle is borrowed from Matt Erlin and Anupam Basu's "Introduction to Digital Humanities: Cultural Analysis in the Information Age". Inspiration for the organization of the programming content came originally from Cosma Shalizi's *Statistical Computing* (http://www.stat.cmu.edu/~cshalizi/statcomp/).