

Syllabus for POLI 7002
Text-as-Data
Fall 2016

Location:

M 1:30 – 4:20pm, Pappajohn Business Building (PBB) - C250

Professor: Bryce J. Dietrich

Office: 314 Schaeffer Hall
Email: bryce-dietrich@uiowa.edu
Web Page: <http://www.brycejdietrich.com>
Office Hours: W 1:30 – 4:30pm, and by appointment

Material to be Covered: This course will give students the tools necessary to analyze a variety of documents, both spoken and written. To achieve this end, students will develop a basic understanding of R, Python, and HTML. While programming is an important component of the course, this is not a programming class. Instead, programming is used to achieve a desired end, meaning the course is more applied than theoretic.

By the end of the course students will be able to utilize some of the most useful APIs available to social scientists (including Twitter and Facebook), scrape web pages, use Amazon's Mechanical Turk for large scale data collection, and execute both supervised and unsupervised learning algorithms. These techniques are useful for a variety of text related objectives, such as topic and sentiment classification. They can also be used more broadly for computer assisted content analysis.

Homework: There will be weekly homework assignments. Each assignment is designed to help students develop skills that will be useful for their work. All R assignments will be completed in R Markdown. Python will be completed in markdown, LATEX, or some other format. We will discuss installing R Studio and Python on the first day of class. We will also discuss more about homework and formatting at that time. In total, the homework will be worth 40 percent of your grade.

Final Project: Students will be required to complete a final project. This project can either be an original research paper or can be part of an ongoing research project that has some large text analytic component. On the first day of class we will discuss these options in more detail. After that, students should schedule a time within the first two weeks of class to talk with me during my office hours about what they plan on doing for the class. The sooner we get started the better.

Collaboration is an important part of our discipline. Thus, students will be allowed (and encouraged) to work together on their final project. However, no team should be larger than two people. Additional details about the final project will be discussed within the first two weeks of class. In total, the final project will be worth 40 percent of your grade.

Research Presentation: All research projects will be presented the final week of class. This will provide a great opportunity for students to present their work. Prior to the presentation, students should submit a rough draft of their paper. After they present their work, students should incorporate the comments they receive into their final draft. Additional details about the research presentation will be discussed within the first two weeks of class. In total, the research presentation will be worth 20 percent of your grade.

Grading:

Final Exam	40%
Homework	40%
Research Presentation	20%

Readings: There are no required books for the class. All course readings will be posted on the course website. The following are the books I use the most. I will make additional recommendations as the semester progresses.

Jurafsky, Daniel and James Martin. 2008. *Speech and Language Processing*. Prentice Hall.

Hastie, Tibshirani, and Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Bishop, Christopher. 2006. *Pattern Recognition and Machine Learning*. Springer.

Scale:

A+	98-100%
A	93-97.99%
A-	90-92.99%
B+	87-89.99%
B	83-86.99%
B-	80-82.99%
C+	77-79.99%
C	73-76.99%
C-	70-72.99%
D+	67-69.99%
D	63-66.99%
D-	60-62.99%
F	0-59.99%

Late Work and Make-Up Exams: Late work will not be accepted and no make-up exams will be given. There should be no problems meeting the expectations of the course. All the due dates are given, and I will make sure to give you plenty of notice ahead of time.

Administrative Home: The College of Liberal Arts and Sciences is the administrative home of this course and governs matters such as the add/drop deadlines, the second-grade-only option, and other related issues. Different colleges may have different policies. Questions may be addressed to 120 Schaeffer Hall, or see the CLAS Academic Policies Handbook at <http://clas.uiowa.edu/students/handbook>.

Electronic Communication: University policy specifies that students are responsible for all official correspondences sent to their University of Iowa e-mail address (@uiowa.edu). Faculty and students should use this account for correspondences ([Operations Manual, III.15.2, k.11](#)).

Accommodations for Disabilities: The University of Iowa is committed to providing an educational experience that is accessible to all students. A student may request academic accommodations for a disability (which include but are not limited to mental health, attention, learning, vision, and physical or health-related conditions). A student seeking academic accommodations should first register with Student Disability Services and then meet with the course instructor privately in the instructor's office to make particular arrangements. Rea-

sonable accommodations are established through an interactive process between the student, instructor, and SDS. See <http://sds.studentlife.uiowa.edu/> for information.

Academic Honesty: All CLAS students or students taking classes offered by CLAS have, in essence, agreed to the College's [Code of Academic Honesty](#): "I pledge to do my own academic work and to excel to the best of my abilities, upholding the [IOWA Challenge](#). I promise not to lie about my academic work, to cheat, or to steal the words or ideas of others; nor will I help fellow students to violate the Code of Academic Honesty." Any student committing academic misconduct is reported to the College and placed on disciplinary probation or may be suspended or expelled ([CLAS Academic Policies Handbook](#)).

CLAS Final Examination Policies: The final examination schedule for each class is announced by the Registrar generally by the fifth week of classes. Final exams are offered only during the official final examination period. No exams of any kind are allowed during the last week of classes. All students should plan on being at the UI through the final examination period. Once the Registrar has announced the date, time, and location of each final exam, the complete schedule will be published on the Registrar's web site and will be shared with instructors and students. It is the student's responsibility to know the date, time, and place of a final exam.

Making a Suggestion or a Complaint: Students with a suggestion or complaint should first visit with the instructor (and the course supervisor), and then with the departmental DEO. (**Wenfang Tang, 335-2358**) Complaints must be made within six months of the incident ([CLAS Academic Policies Handbook](#)).

Understanding Sexual Harassment: Sexual harassment subverts the mission of the University and threatens the well-being of students, faculty, and staff. All members of the UI community have a responsibility to uphold this mission and to contribute to a safe environment that enhances learning. Incidents of sexual harassment should be reported immediately. See the UI [Comprehensive Guide on Sexual Harassment](#) for assistance, definitions, and the full University policy.

Reacting Safely to Severe Weather: In severe weather, class members should seek appropriate shelter immediately, leaving the classroom if necessary. The class will continue if possible when the event is over. For more information on Hawk Alert and the siren warning system, visit the [Department of Public Safety website](#).

Calendar: The calendar below gives the dates of exams and other important deadlines for the course. Readings should be completed **prior** to the start of class. This calendar is subject to change. Any changes will be announced and posted on Blackboard.

	<u>Week 1: Text-as-Data</u>
Monday 22 Aug	Getting Started <i>Grimmer and Stewart 2013</i> <i>Monroe and Schrodtt 2008</i>
	<u>Week 2: Acquiring Data I</u>
Monday 29 Aug	Webscraping <i>Handout</i>
	<u>Week 3: Acquiring Data II</u>
Monday 5 Sep	Labor Day <i>Regular Expressions Tutorial</i>
	<u>Week 4: Manipulating Data</u>
Monday 12 Sep	Preprocessing and Stemming <i>Vijayarani et. al 2015</i>
	<u>Week 5: Dictionary Methods</u>
Monday 19 Sep	Using and Creating Dictionaries <i>Black et. al 2011</i> <i>Dodds and Danforth 2010</i> <i>Young and Soroka 2012</i>
	<u>Week 6: Unsupervised Methods I</u>
Monday 26 Sep	PCA versus MDS <i>Hastie, Tibshirani, and Friedman, Chapter 14</i>
	<u>Week 7: Unsupervised Methods II</u>
Monday 3 Oct	K-Means and EM Algorithm <i>Bishop, Chapter 9</i>

	<u>Week 8: Topic Models I</u>
Monday 10 Oct	LDA <i>Blei 2012</i> <i>Blei and Lafferty 2009</i>
	<u>Week 9: Topic Models II</u>
Monday 17 Oct	Structured Topic Models <i>Grimmer 2010</i> <i>Quinn et al. 2010</i> <i>Roberts et al. 2014</i>
	<u>Week 10: Supervised Methods I</u>
Monday 24 Oct	Model Assessment <i>Hastie, Tibshirani, and Friedman, Chapter 7</i>
	<u>Week 11: Supervised Methods II</u>
Monday 31 Oct	SVM and ReadMe <i>Hopkins and King 2010</i> <i>D’orazio et al. 2014</i>
	<u>Week 12: Ensemble Methods</u>
Monday 7 Nov	SuperLearners <i>Hillard, Purpura, and Wilkerson 2007</i> <i>Hastie, Tibshirani, and Friedman, Chapters 15 and 16</i>
	<u>Week 13: Measuring Ideology</u>
Monday 14 Nov	Wordscore and Wordfish <i>Laver, Benoit, and Garry 2003</i> <i>Lowe 2008</i> <i>Slapin and Prokschk 2008</i>
	<u>Week 14: Thanksgiving Break</u>
Monday 21 Nov	No Class <i>No Readings</i>
	<u>Week 15: New Directions</u>
Monday 28 Nov	Incorporating Audio and Video Data <i>Dietrich 2016</i>

Week 16: Research Presentations

Monday 5 Dec Research Presentations
No Readings