Meaning and Mining: Method and Interpretation in the Digital Humanities

Time/Location: Wed, 1:15-4:05, Stanford Literary Lab (401, Building 460)

Instructor: Mark Algee-Hewitt
Email: mark.algee-hewitt@stanford.edu
Office: 321, Margaret Jacks Hall, Building 460
Office Hours: Tuesday, 1:00-2:30; Wednesday 10:30 – 12:30

Required Books:
Franco Moretti, *Distant Reading*
Edward Tufte, *The Visual Display of Quantitative Information*
David M Barry (ed) *Understanding Digital Humanities* (optional)

All other readings will be made available on the class CourseWork site.

Description:

In this class, we will explore ways to apply the methods of the digital, or computational, humanities to questions of literary theory or criticism. Our goals will be twofold: first, you will gain hands-on experience in the techniques and strategies of the digital humanities, particularly what has been called by Moretti et al “quantitative formalism.” Through this course you will learn the basic techniques of how to design and implement a digital humanities project, perform analyses based on standard “text-mining” techniques, and present the results of these analyses within an appropriate visualization. Far more important, however, is understanding how these methods can be integrated within the practices of literary criticism. How can we quantify, or operationalize, questions that are meaningful to our understanding of literary theory or history? Similarly, how can we interpret the results of a quantitative analysis and mobilize the information we learn using the methods of the digital humanities with humanistic modes of inquiry? These are the primary questions that we will seek to tackle together through our practical explorations of quantitative and computational methodologies. Accordingly, we will take both a theoretical and practical approach to the Digital Humanities, seeking to understand the implications of digital methodologies for literary studies, while remaining sensitive to the kinds of data that such methods subordinate, or exclude. Through discussions and hands-on original research within the Stanford Literary Lab, we will work to understand how to use the digital humanities as a new, essential, tool for literary criticism in the twenty-first century.

Course Layout

Classes will be divided between discussion and lab-time, with the expectation that this line will blur as the quarter progresses. For class, each student should have a laptop onto which they can install the statistical software environment “R” (we will do this together on the first day). In keeping with the aims of the course, classes will combine a theoretical discussion of the method (and its application to literature) with some time spent together learning how to implement the
method within a lab environment. Classes will end with a short exercise designed to be completed in the lab, working together or alone.

**Students with Documented Disabilities**

Students who may need an academic accommodation based on the impact of a disability must initiate the request with the Office of Accessible Education (OAE). Professional staff will evaluate the request with required documentation, recommend reasonable accommodations, and prepare an Accommodation Letter for faculty dated in the current quarter in which the request is being made. Students should contact the OAE as soon as possible since timely notice is needed to coordinate accommodations. The OAE is located at 563 Salvatierra Walk (phone: 723-1066, URL: [http://studentaffairs.stanford.edu/oae](http://studentaffairs.stanford.edu/oae)).

**Honor Code**

The Honor Code is the University's statement on academic integrity written by students in 1921. It articulates University expectations of students and faculty in establishing and maintaining the highest standards in academic work:

The Honor Code is an undertaking of the students, individually and collectively:

1. that they will not give or receive aid in examinations; that they will not give or receive unpermitted aid in class work, in the preparation of reports, or in any other work that is to be used by the instructor as the basis of grading;
2. that they will do their share and take an active part in seeing to it that others as well as themselves uphold the spirit and letter of the Honor Code.
3. The faculty on its part manifests its confidence in the honor of its students by refraining from proctoring examinations and from taking unusual and unreasonable precautions to prevent the forms of dishonesty mentioned above. The faculty will also avoid, as far as practicable, academic procedures that create temptations to violate the Honor Code.
4. While the faculty alone has the right and obligation to set academic requirements, the students and faculty will work together to establish optimal conditions for honorable academic work.

**Course Requirements**

**Participation (20%)**: 

As this class is split between discussions of the methodologies and hands-on explorations of these methods, you are all tasked with keeping the spirit of experimentation alive. This is another way of saying that participation is mandatory: your voice must be heard in class contributing, questioning or challenging or in the lab as we work together or separately to learn the techniques of literary quantitative analysis.
Group Project (40%):

As we will discuss this quarter, one of the most important aspects of the Digital Humanities is its collaborative nature. For your major project, therefore, you will work together within four teams, of similar size, to apply the techniques that we will be learning together to one or more subsets of the Literary Lab’s corpus. You will design a project, formulate meaningful questions, assemble your corpus and perform multiple analyses on the data in keeping with the techniques that we learn in class. Finally, on one of the two last weeks, you will present your findings, complete with visualizations, to the class and engage the group as a whole in a discussion of your methods and results. In this way, the class itself will simulate the process of a literary lab, as your individual project teams will be presenting to the lab as a whole. Remember, although your group must work together, and therefore agree on the corpus and the overall research design, it is perfectly possible, and, in fact, encouraged, that you all have your own individual goals for what you want to learn through your analysis.

Final Paper (40%)

Building on your work in the group project (or, in special cases, on projects that you have conceived of alone), you will write a paper on a topic of literary theory or history that uses your quantitative work to formulate an argument of your choosing. Here too, experimentation is encouraged: those of you who would like to continue working collaboratively, or who want to venture outside of the boundaries of the traditional research paper are welcome to do so: remember though that your project must combine a quantitative approach with a critical or theoretical idea.

Course Schedule:

1. January 8th
   Intro to DH, the Literary Lab, and Getting to know R
   Introduction to R

2. January 15th
   Why do we mine?
   James Hughes et al. "Quantitative patterns of stylistic influence in the evolution of literature"
   Deleuze and Guattari “A Thousand Plateaus: Capitalism and Schizophrenia: Introduction”

3. January 22nd
   GIGO: Corpus, Metadata and Research Question design
   Brad Pasanek and D. Scully, “Meaning and Mining: The Impact of Implicit Assumptions in Data-Mining for the Humanities.”
   Stephen Ramsay, “Algorithmic Criticism”
4. January 29th  
Who Wrote What? Authorship Attribution  
  J.F. Burrows, “Questions of Authorship: attribution and beyond”  
  David Hoover, “Testing Burrows’ Delta”  
  Matt Jockers and Daniela Witten, “A Comparative Study of Machine Learning Methods for Authorship Attribution”  
  Jorge Borges, “Pierre Menard, Author of the Quixote”  
  Michel Foucault, “What is an Author?”

5. February 5th  
Information Design: Aesthetics in Action  
  Edward Tufte, *The Visual Display of Quantitative Information*  
  Franco Moretti, selections from *Graphs, Maps, Trees*  

6. February 12th  
Exploratory Data Analysis: Finding Patterns and Making Sense  
  Dan Dixon, “Analysis Tool or Research Methodology: Is there an epistemology for patterns?”  
  Michel Foucault, “The Archaeology of Knowledge: The Statement and the Archive”

7. February 19th  
Classificatory Data Analysis: Training the Machine  
  Matt Jockers, “Macroanalysis: Style”  
  Gérard Genette, selections from “The Architext: An Introduction”

8. February 26th  
Advanced Mining: Topic Modeling  
  Blei, “Probabilistic Topic Models”  
  Joris van Zundert et al., “Cultures of Formalisation: Towards and Encounter between Humanities and Computing.”  
  Derrida, selections from *Of Grammatology*

9. March 5th  
Discussion of projects and exploration of Results: Groups 1 and 2

10. March 12th  
Discussion of projects and exploration of Results: Groups 3 and 4