

Introduction to Digital Research

English 5100, Temple University, Spring 2016, 3 credits
Class meets Wednesdays, 9:30-11:45 pm
416 Weiss Hall; Digital Scholarship Center, bottom floor of Paley Library

Peter Logan (peter.logan@temple.edu)
Office: 914 Anderson Hall, 215-204-8567
Office Hours: 2:00-3:30 pm, Tuesday and Wednesday, and by appt.

Description

This course introduces students to the concepts and tools needed to conduct digital research in English. During the semester, we'll discuss how the broader field of the Digital humanities (DH) is defined, why humanists are using digital tools to do their research, how the new methods compare with older methods of humanities scholarship, and what are their strengths and weaknesses.

This course gives you a chance to explore these new methods. We begin with a focus on the basic theoretical and technological issues involved in creating and analyzing digital texts, before moving on to a series of hands-on exercises in analyzing words and interrogating the results.

By the end of the semester, students will understand the history, theory, and technology of digital textual analysis and produce a 15-page paper applying these new methods to material relevant to their own interests or analyzing examples of digital-based criticism.

The field of DH values collaborative work far more than most other forms of scholarship in the humanities. This is because every DH project involves a collection of many discrete skills, far more than any one person can generally master. In this course, students will be encouraged to work collaboratively where possible. Part of the course will be devoted to discussing the nature of collaborative work and how it differs from "group work," so that students learn how to work together in productive and positive relationships.

Students should be comfortable using a computer and moving around in the file system. No knowledge of computer programming is needed. If you own a laptop, please bring it to class. If you do not, we'll have laptops available for you to work through the in-class exercises.

Required Texts

Texts marked "Web" are online open-source materials. Some text citations are preceded by an abbreviation used in the syllabus for reading assignments.

Jerome McGann, *Radiant Textuality: Literature After the World Wide Web* (New York: Palgrave, 2001). Print.

Franco Moretti, *Graphs, Maps, Trees: Abstract Models for a Literary History* (New York: Verso, 2005). Print.

Franco Moretti, *Distant Reading* (New York: Verso, 2015). Print.

Stephen Ramsay, *Reading Machines: Toward and Algorithmic Criticism* (Urbana: U of Illinois P, 2011). Print.

(D_H) Anne Burdick, et al, [Digital Humanities](#) (Cambridge: MIT P, 2012). Web.

(DDH) Matthew K. Gold, [*Debates in the Digital Humanities*](#) (Minneapolis: U of Minnesota P, 2012). Web.

(CDH) Susan Schreibman, Ray Siemens, and John Unsworth, [*Companion to Digital Humanities*](#) (Oxford: Blackwell, 2004). Web ().

(LSDH) Price and Seimans, [*Literary Studies in the Digital Age*](#). Web.

Daniel Powell, with Constance Crompton and Ray Siemens, "[Glossary of Terms, Tools, and Methods,](#)" [*Literary Studies in the Digital Age*](#). Web.

Assignments and Grading

Assignment #1 (Feb. 10) 20%

Using at least two tools in AntConc, analyze the language of Charlotte Bronte's *Jane Eyre* and Louisa May Alcott's *Little Women*. What does the language tell us about the themes and novelistic world of each text? What conclusions can you draw about the differences between them? Be careful to document your stopword list and the settings used to get your results. (5 pages/1500 words)

Assignment #2 (March 16) 20%

Write a short paper on any of the theoretical questions raised in the readings so far in this course. Use several scholarly essays, at least one of which is from the syllabus so far.

Assignment #3 (April 20) 40%

The culminating event for the course is a 15-page paper. You can either (1) apply digital research techniques to a corpus of your choice and produce your own analysis; or (2) write a research paper critiquing existing work in the field of digital literary studies.

Presentation 20%

Each student will present an analysis of one essay on the syllabus to the class and lead a brief discussion about it.

Disability Statement

Any student who has a need for accommodation based on the impact of a documented disability, including special accommodations for access to technology resources and electronic instructional materials required for the course, should contact me privately to discuss the specific situation by the end of the second week of classes or as soon as practical. If you have not done so already, please contact Disability Resources and Services (DRS) at 215-204-1280 in 100 Ritter Annex to learn more about the resources available to you. I/we will work with DRS to coordinate reasonable accommodations for all students with documented disabilities.

Compulsory Statement on Freedom

"Freedom to teach and freedom to learn are inseparable facets of academic freedom. The University has a policy on Student and Faculty and Academic Rights and Responsibilities (Policy #03.70.02) which can be accessed through the following link:

http://policies.temple.edu/getdoc.asp?policy_no=03.70.02."

Schedule: Introduction to Digital Research

* Items marked with an asterisk are provided

Introduction

January 13

Introduction
History of DH
Introducing GitHub

Stylistics

January 20

Homework

Open [GitHub student account](#) and review [GitHub Guides--Understanding the GitHub Flow; Hello World; Mastering Markdown](#)

Reading

Moretti, *Graphs, Maps, and Trees* (all)
Matt Kirschenbaum, "[What Is Digital Humanities and What's It Doing in English Departments?](#)" (DDH)
Susan Hockey, "[The History of Humanities Computing](#)" (CDH)

Practicum

Update Javascript
Style Basics: *AntConc* word frequencies, concordance

January 27

Reading

Moretti, *Distant Reading*, ch. 2, 3, 5
[Piper, "The Werther Effect I: Goethe, Objecthood, and Handling of Knowledge"](#)

Practicum

Style Basics: *AntConc* collocates, keywords, ngram

Stylometry

February 3

Reading

Moretti, *Distant Reading*, ch. 7, 9, 10
*Burrows, "Delta': A Measure of Stylistic Difference"
[Lancashire and Hirst, "Vocabulary Changes in Agatha Christie's Mysteries"](#) (Paper)
[Lancashire and Hirst, "Vocabulary Changes in Agatha Christie's Mysteries"](#) (Poster)

Textuality

February 10

Assignment #1 Due (5 pages)

Reading

[Burdick, et al, "Humanities to Digital Humanities" \(D_H\)](#)

McGann, *Radiant Textuality*, Introduction, "The Alice Fallacy," "Rethinking Textuality"

Networks

February 17

Reading

*Long\So, "Literary Pattern Recognition: Modernism between Close Reading and Machine Learning"

*Dewitt, "Advances in the Visualization of Data: The Network of Genre in the Victorian Periodical Press"

[Elson, et al, "Extracting Social Networks from Literary Fiction"](#)

Practicum

Voyant

February 24

Reading

*Wilkens, "The Geographic Imagination of the Civil-War Era American Fiction"

[Kretzschmar, "GIS for Language and Literary Study" \(LSDH\)](#)

Practicum

Voyant

[Spring Break]

Topic Models

March 9

Reading

[Brett, "Topic Modeling: A Basic Introduction"](#)

[Underwood, "Topic Modeling Made Just Simple Enough"](#)

[Nelson, "Mining the Dispatch"](#)

[Schmidt, "Typical TV episodes: Visualizing Topics in Screen Time"](#)

[Schmidt, "Fundamental Plot Arcs ..."](#)

Practicum

Install *Topic Modelling Tool*

March 16

Assignment #2 Due (5 pages)

Reading

[Blei, "Probabilistic Topic Models"](#)

[Goldstone and Underwood, "What Can Topic Models of PMLA Teach Us about the History of Literary Scholarship?"](#)

[Schmidt, "Words Alone: Dismantling Topic Models in the Humanities"](#)

Practicum

Interpreting student topic models

Sentiment Analysis

March 23

Reading

*Levi Strauss, "The Structural Study of Myth"

Jockers, "A Novel Method for Detecting Plot"

Jockers, "Revealing Sentiment and Plot Arcs with the Syuzhet Package"

Jockers, "That Sentimental Feeling"

Practicum

Daniel Soper's *Sentiment Analyzer*

hedonometer.org

Plot

March 30

Reading

*Scholes, et al., "Plot in Narrative," in *The Nature of Narrative* (207-39)

Underwood, "Plot Arcs in the Novel"

*Piper, "Novel Devotions: Conversional Reading, Computational Modeling, and the Modern Novel"

Criticism

April 6

Ramsay, *Reading Machines*, "An Algorithmic Criticism," "The Turning Test"

*Latour and Lowe, "The Migration of the Aura or How to Explore the Original through its fac similes"

[Hauser, et al, "A Quantitative Literary History ... Semantic Cohort Model"](#)

Visualization

April 13

*Tufte, "Data-Ink Maximization and Graphical Design," *The Visual Display of Quantitative Information*

[Sinclair, et al., "Information Visualization for Humanities Scholars" \(LSDH\)](#)

*Klein, "The Image of Absence: Archival Silence, Data Visualization, and James Hemings"

Final Projects

April 20

Final Projects

Discussion

English 5100, Spring 2016
Introduction to Digital Research
Peter Logan

First Essay

Please write a 5-page (1500-word) essay, due at the start of class on February 10

Using at least two Tools in AntConc, compare the language of Charlotte Brontë's *Jane Eyre* and Louisa May Alcott's *Little Women*. What does the language tell us about the themes and novelistic world of each text? What conclusions can you draw about the differences between them? Be careful to document your stopword list and the settings used to get your results.

It is not necessary to have actually *read* either of these novels, in the conventional sense. The point is to see what you can infer about unknown texts based solely on their use of words. Your conclusions need not even be "right," as they long as they are derived logically from the language. Discovering how word counts alone could create misleading impressions is itself a valuable lesson in working with statistical data derived from literary texts.

Plain-text files of both novels are posted on Owlbox for you to download. You will also find there files downloaded from AntConc: a standard stopword list file (stopword-list1.txt), two files of general word frequency for American and British English (AmE06_wordlist.txt and BE06_wordlist.txt), and a lemma list (e_lemma_no_hyphen.txt). Finally, there is a file with information on what "keyness" values mean and how to judge whether they are significant or not (about_keyness.txt).

English 5100, Spring 2016
Introduction to Digital Research
Peter Logan

Second Essay

Please write a 5-page (1500-word) essay, due at the start of class on March 16.

In your last essay, you tried your hand at a *practical* digital research assignment. In this essay, I want you to consider the *theoretical* side of digital research.

So far in this course you have read 11 essays and 3 books on digital research methods, mostly for literary study. Some of these essays raise questions about specific methods of textual analysis. Others address more abstract concepts, like the nature of textuality in the digital world. As a group, they address a wide range of topics: stylistics, stylometry, textuality, networks, and (coming up) topic models.

Take any two of these essays or book chapters and use it as the basis for an essay on any one issue or conceptual problem in applying digital methods to literary texts. Whatever question you choose to write about, I want you to identify and analyze the critical method being used and consider its consequences for literary study. A few possibilities, out of many:

How does the use of digital methods alter the object of literary analysis? Are the “texts” in some way different from the traditional print texts used in a literary course? Does the writer base their argument on important critical assumptions that we should think about more carefully? To what extent might the writer’s choice of method predetermine their conclusions about the text? Are the conclusions themselves useful or not? How are discoveries produced by these methods (like networks or style) different from those we identify in close readings?

Feel free to develop your own question; these are simply illustrative.

As an alternate, you can use one essay from the syllabus and one other of your choosing. But please limit yourself to two—five pages will only get you so far.

English 5100, Spring 2016
Introduction to Digital Research
Peter Logan

Third Essay

Please write a 15-page (4500-word) essay, due at the start of class on April 20.

In your first essay, you tried your hand at a *practical* digital research assignment. In the second, you considered the *theoretical* side of digital research.

For this final essay, you can do either.

1. Analyze a corpus of literary texts that you select, using the tools we have worked with in class. You will need to find and clean the digital texts, determine which textual analysis techniques to use, write up the results of your tests and explain their significance. You will also need to locate articles or book chapters relevant to your project.
 - a. Example: you work on all the writing of Harriet Beecher Stowe and decide that her interests change over the course of her career, based on her use of language. What have others said about the topic? Do your findings support or contradict existing assumptions about Stowe's work?
2. Write a research paper critiquing existing work in the field of digital literary studies. This can either be a critical analysis of specific claims made by digital scholars, or it can be more focused on the use of digital methods within your own area of interest.
 - a. Example: You specialize in 20th-century American poetry. What has digital research revealed about it so far? Has it challenged existing assumptions or found quantitative support for qualitative claims made by others? What are the problems created by copyright restrictions, and how are current scholars addressing them? What specific problems or opportunities might poetry present, rather than prose?

The problem with this assignment is that you have to define your own research project, and there are too many possibilities. I suggest you think about a particular writer or group of writers you are interested in, and then use this as a chance to get to know them better. The project also needs to be a manageable one, given the time and space you have.

To facilitate your efforts, I will schedule conferences with each of you individually during the week of March 28-April 1. Before that conference, think about your topic and do some exploratory research. What questions do you have? Then we will use the conference to get you on track as quickly as possible.