# Digital Humanities

## Contact information

| | |
|---|---|
| Instructor | Matthew Wilkens |
| Course | English 90127: Digital Humanities |
| Meetings | M 3:00–5:30ish |
| Location | 200 O'Shaughnessy |
| Office hours | 320 Decio; W 3:00–5:00 and by appointment. |
| | Reserve office hour slots at bit.ly/wilkens_appointments |
| E-mail | mwilkens@nd.edu |
| Phone | 574.631.257 |

## Synopsis

A graduate-level introduction to problems and methods in digital humanities with an emphasis on computational and quantitative literary studies.

## Description

Contemporary criticism has a problem. We long ago gave up the idea that our task was to appreciate and explain a handful of great texts, replacing that goal with a much more important and ambitious one: to understand cultural production as a whole by way of the aesthetic objects it creates. But we have continued to practice our craft as if the methods developed in pursuit of the old project were the only ones suited to the new task. In consequence, we have remained less successful and less persuasive concerning the operations of large-scale cultural production than we would like to be.

This course is devoted to new methods and new objects in cultural and literary studies, specifically those enabled by digital media. It is not, however, primarily a course in media studies (though we'll do a bit of that at the outset). We'll spend most of our time covering both what kinds of criticism are made possible by the availability of digital cultural objects (especially digitized texts), whether those objects are born digital or are *post facto* electronic surrogates, and how to perform the technical operations necessary to carry out such criticism. *The course thus has a substantial technical component, one no more difficult than — but substantially different from — most of your existing experience in literary studies.* That said, the idea is certainly not to replace the methods you've previously mastered, but to supplement them with new approaches, issues, and questions that will allow you to do better the kinds of cultural and literary criticism you've already begun to practice.

As I say, we'll begin via the briefest encounter with media theory (McLuhan, Galloway), primarily as a way to remind us that digital and physical texts do potentially different things and occupy different positions in their various historical and cultural matrices. There's a whole branch of digital humanities that's essentially rebranded media studies and considers these issues in depth. Noble work, but not really ours; we'll move quickly to first quantitative and then computationally assisted criticism, generally performed on large bodies of text. Much of this work falls under the broad heading of data mining; we'll

study what that term means and how to do it. We'll aim for a mix of theoretical elaborations concerning what is and is not implied by quantitative methods and how those methods integrate with conventional humanities approaches to interpretation, and (the second part of the mix) specific examples of achieved work in the field, plus technical exercises that will help you carry out similar work on your own.

We'll read as widely as possible in both the major works of digital humanities theory and practice and in the often-divisive debates about the value of that work. Much of the schedule is set out below, but be aware that DH is a rapidly evolving field; we'll adjust our readings in response to new developments that emerge over the course of the semester and to our collective interests or hangups.

Four notes:

1. I'll repeat that this is a course with significant technical components. There's no assumption that you'll enter with any particular computational or mathematical expertise and there will be plenty of help (along with prebuilt tools) available along the way, but you *must* be willing to work outside your presumptive comfort zone as literary critics to develop the skills necessary to conduct the kinds of research we'll explore. This is really exciting stuff and it's not tremendously difficult, but if you shut down early in the face of the command line or a list of numbers, it'll be impossible to do well.

2. Our week-to-week assignments and work will be somewhat different from what you're probably accustomed to. Most notably, there will be — in addition to the usual books and articles to read — graded problem sets or other exercises to complete. We'll talk more about the form these will take as the semester progresses.

3. Your final project will also take an unconventional form. You will work in groups of three or four students to produce either a piece of quantitatively informed literary scholarship or a grant proposal to perform the same. All members of the group will receive the same grade for the project (though not necessarily for the course).

4. When we perform our own analyses, we'll work primarily with a prepared corpus of nineteenth-century American fiction. There are legal *cum* technical reasons we can't do much with the twentieth century, but we can explore other corpora as our needs, desires, and technical abilities dictate.

## Objectives

This course has three primary objectives, all implicit in the synopsis above. These objectives can be arranged in order of increasing generality. First, it aims to provide a specific body of knowledge about recent work in the digital humanities. Students who complete this course will be prepared to undertake more advanced coursework in the field, to begin making their own contributions to it, and to enter debates concerning new results in the professional literature. Second, the course provides a basis for more robust work in a methodologically expanded literary humanities and cultural studies. It introduces students to methods and critical vocabularies that are portable outside DH proper and familiarizes them with the resources used in conducting computationally informed literary research. Third, it seeks to build students' professional skills in critical and metacritical

analysis, including their ability to evaluate competing forms of argument, evidence, and justification beyond those conventionally associated with literary studies.

## Written work and grading

The largest single assignment will be a final group project informed in some way by quantitative and computational methods. In addition, short responses or other exercises will be required for each meeting over the course of the semester. Overall grades will be based on the final project (60%), weekly responses and exercises (20% in sum), and class participation (including a project presentation, 20%). *You must satisfactorily complete all assignments to pass the course.*

## Required texts

Alexander Galloway. *Protocol: How Control Exists after Decentralization* (2006).
Matthew K Gold. ed. *Debates in the Digital Humanities* (2012).
Franco Moretti. *Graphs, Maps, Trees* (2005).
Stephen Ramsay. *Reading Machines: Toward an Algorithmic Criticism* (2011).

In addition, essays from the scholarly DH literature will be assigned and available via Sakai.

## Schedule

**NB. All dates and assignments subject to change.** You'll also note that details get thinner toward the end of the semester; this is by design, to allow for adjustments depending on progress, interests, and new developments in the field. All readings except those from the four required texts will be available on Sakai. Problem sets and related assignments will be posted to Sakai the week before they're due.

| | |
|---|---|
| WEEK 1 (8/27) | Welcome and introduction. Media theory. |
| | Marshal McLuhan, from *Understanding Media*. |
| | Matthew Kirschenbaum, "What is Digital Humanities?" |
| | Set up accounts and tools. |
| | |
| WEEK 2 (9/3) | Network media theory. |
| | Alex Galloway, *Protocol*. |
| | UVa's "Command Line Boot Camp" (if necessary) |
| | |
| WEEK 3 (9/10) | Intro to quantitative methods. |
| | Franco Moretti, *Graphs, Maps, Trees*. |
| | Unsworth et al. "How Not to Read a Million Books." |
| | Exercises 1–3 (through "Working with Files and Web Pages") from *The Programming Historian*. |
| | |
| WEEK 4 (9/17) | Theory of computational literary analysis. |
| | Stephen Ramsay, *Reading Machines*. |

Elson, David K., Nicholas Dames, and Kathleen R. McKeown. "Extracting Social Networks from Literary Fiction."

Exercises 4–5 ("From HTML …" and "Computing Frequencies") from *The Programming Historian*.

WEEK 5 (9/24)     Packaged tools.

Google *n*-gram viewer, Voyant, WordHoard, MONK, etc.

Michel, Jean-Baptiste et al. "Quantitative Analysis of Culture Using Millions of Digitized Books."

Cohen, Dan. "Searching for the Victorians."

Clement, Tanya. "'A Thing Not Beginning and Not Ending': Using Digital Tools to Distant-Read Gertrude Stein's *The Making of Americans*."

Exercises 6–8 (all remaining) from *The Programming Historian*.

WEEK 6 (10/1)     Data sources and text-processing basics.

Programming exercise 1: Parsing XML.

Fyfe, Paul. "Electronic Errata" in *DDH*.

Grimmelmann, James. "The Elephantine Google Books Settlement."

Legal documents related to *Authors Guild v. Google*.

WEEK 7 (10/8)     Entity extraction.

Programming exercise 2: Extracting entities.

Matthew Jockers, *Macroanalysis* (chapters 4, 6, and 10).

Dunning, Ted. "Accurate Methods for the Statistics of Surprise and Coincidence."

WEEK 8 (10/15)     **Fall break, no class meeting.**

WEEK 9 (10/22)     Databases.

Programming exercise 3: MySQL.

Google Code University, "Introduction to Databases and MySQL."

Lee, Maurice S. "Searching the Archives with Dickens and Hawthorne."

WEEK 10 (10/29)     GIS.

Exercise 4: GIS.

Lyndegaard, Kate. "Online GIS Using ArcGIS.com."

Cooper, David, and Ian N Gregory. "Mapping the English Lake District: A Literary GIS."

Wilkens, Matthew. "The Geographic Imagination of Nineteenth-Century American Fiction."

WEEK 11 (11/5)     Topic modeling.

Exercise 5: MALLET

"Getting Started with Topic Modeling and MALLET" from *The Programming Historian*

Blei, David M. "Probabilistic Topic Models."

Underwood, Ted, and Jordan Sellers. "The Emergence of Literary Diction."

Blevins, Cameron. "Topic Modeling Martha Ballard's Diary."

WEEK 12 (11/12)    Advanced topic models.

Exercise 6: Analyzing topic models.

Wang, X., and A. McCallum. "Topics over Time."

Chaney, A. J. B., and D. M. Blei. "Visualizing Topic Models."

Mimno, David. "Computational Historiography: Data Mining in a Century of Classics Journals."

Nelson, Robert K. "Mining the *Dispatch*."

Schmidt, Ben. "Sapping Attention: When You Have a MALLET, Everything Looks Like a Nail."

WEEK 13 (11/19)    **Class cancelled** for the Chicago Colloquium on Digital Humanities and Computer Science at the University of Chicago, 11/18–19

WEEK 14 (11/26)    Visualization.

Yau, Nathan. *Visualize This*, Chapter 6.

Manovich, Lev. "Media Visualization: Visual Techniques for Exploring Large Media Collections."

"Data Visualization." Stanford University *Tooling Up for the Digital Humanities* series.

WEEK 15 (12/3)    Presentations.

FINALS WEEK    **Final papers or projects due 12/14 by 5:00 pm**