Catherine DeRose, PhD
catherine.derose@yale.edu

From Text Mining to Mapping: Digital Humanities Methods

Course Description and Learning Objectives

This week-long course offers an introduction to the field of digital humanities, broadly considered. We will begin with an overview of the field and a general discussion of data in the humanities, looking at different ways to collect, clean, and visualize it. The remaining days of the course will consist of deep dives into specific areas of digital humanities research, including text mining, network analysis, and mapping. Throughout, we will move between aggregate, distant readings of texts and traditional close readings to consider what is gained and lost in the transition. We will conclude the week by discussing options for sharing your work online.

Digital humanities is an expansive, rapidly evolving area of research. This course is designed to give you a foundation, introducing you to approaches and resources you can continue with beyond the seminar. Over the week, you will gain hands-on experience working with popular out-of-the-box, open source tools used by digital humanities practitioners. You will also learn best practices for developing your own digital humanities projects, from how to curate a dataset and identify a tool or approach that is suited to your research question(s) to creating effective and compelling data visualizations you can share.

Texts and Software

All readings for this class are open access and can be read online or printed. Along with readings, we will analyze existing projects that typify different digital humanities approaches. Links to the readings and projects can be found in the Schedule section below.

You are expected to bring a laptop over which you have administrative control to each class. The following software should be downloaded to it. It may seem like an overwhelming amount of software, but we will move slowly with it, and some of it we will only touch on briefly. The range of software is intended to introduce you to a sampling of approaches and tools that are out there.

All of this software is free. Please attempt to download it in advance of the seminar. If you have any trouble with the installation (software can be quirky!) or with locating a laptop you can use for the seminar, please email me so we can try other options. We will set aside class time on Monday for troubleshooting installations.

Mac users, please note that for some software—OpenRefine and Stanford's Named Entity Recognition GUI—you may encounter a message during the installation that says they cannot be opened because they are from an unidentified developer. To override that, you will have to open System Preferences, select Security & Privacy, and click the permission to allow them to open. There is an added security feature in Macs that requires the step of clicking allow for software from unidentified developers (which is a good feature in general, but there isn't a concern with the particular tools we will be using).

Installation is required for:
1. Anaconda's Python 3 https://www.anaconda.com/distribution/#macos
   - The install may take 15-20 minutes, and toward the end you'll receive a prompt to install Microsoft VS Code, which isn't necessary but which you can install if you'd like.
2. Gephi https://gephi.org/
3. JDK 8 https://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html
4. OpenRefine http://openrefine.org/download.html
5. Stanford's Named Entity Recognition GUI bit.ly/stanford-ner-gui
6. Sublime https://www.sublimetext.com/3
   - If you already have a plain text editor you like (such as Notepad or BBEdit), you can skip this installation.
7. Tableau Public https://public.tableau.com/en-us/s/download
8. For Windows-users only, Git for Windows https://gitforwindows.org/
   - During the installation process, look for the option to select Git BASH (if you have a Mac, you will use Terminal, which is already installed by default).

No installation is required for:
1. ArcGIS Online https://www.arcgis.com/index.html (sign up for a free account)
2. Databasic.io https://databasic.io/en/
3. Story Maps https://storymaps.arcgis.com/en/ (use the free account from ArcGIS Online)
4. Voyant https://voyant-tools.org/

For experimenting with these tools, sample datasets will be provided, but there will also be opportunities for you to build you own datasets.

Grading

Special Seminar courses are graded 'Completed Satisfactorily' or 'Completed Unsatisfactorily' (CS/CU). 1 modular credit (1-MC) will be awarded for special seminars that are completed satisfactorily.

40%   IN-CLASS EXERCISES WITH DIGITAL HUMANITIES TOOLS
       One primary goal of the class is to increase your familiarity and comfort with digital humanities methods. Each day, I will introduce different digital humanities tools and approaches that you will be asked to try out. You will *not* be graded on your mastery of those tools or approaches, but rather on your engagement with them, your willingness to try them and experiment.

30%   SHORT WRITING ASSIGNMENTS (3 total, each worth 10%)
       Class will start with a discussion based on a reading or sample projects. In preparation, please write a one-page response on the assigned reading or project that addresses:
       - What is the data that is being used?
       - What method(s) does the project employ?
       - What kinds of questions does that method help you ask of the data?
       - What is the visualization trying to communicate, and why is or isn't it effective?
       - Are there any terms used in the reading or by the project that are unclear?

All short writing assignments should be double spaced and typed in Times New Roman with 12-point font and 1-inch margins. In the upper-left corner of your paper, include a single-spaced header with the following information on separate lines:

- ☐ Your first and last name
- ☐ DH Special Seminar
- ☐ The date
- ☐ The title of the assigned reading or sample project

Short writing assignments should be submitted to our Canvas site before class starts on 12 March, 13 March, and 14 March. Assignments submitted late will lose one point for each day.

20%   PARTICIPATION
Digital humanities is a collaborative area of research that brings together scholars from across disciplines to drive the field forward. We learn from each other. We engage with each other's ideas, and we help to problem solve when we encounter a technical, methodological, or institutional issue. Participation in this seminar includes coming to class prepared and contributing to class discussions by asking questions, sharing ideas, listening attentively, and offering support. I understand that talking in class is not always easy. If you are uncomfortable or unsure of how to enter class discussions, email me or talk to me after class and we can discuss strategies and alternative ways to engage with observations raised in class.

10%   PRESENTATION
As a capstone to the seminar, you will be asked to give a 3-minute presentation on a digital humanities project on Friday, 15 March. You can present on a project you have found online that you think does a good job integrating text mining, network, or spatial techniques, or you can discuss a project of your own design. The aim is to gain experience finding, evaluating, and presenting digital humanities research and, if you present on your own project, to showcase your work and receive input. By the end of the presentations, we will have a class-curated repository of digital humanities projects that you can return to after the seminar to continue thinking about the value of these approaches. To ensure we have enough time for everyone's presentations, please practice and time yourself so that you do not go over 3 minutes.

Classroom Protocol
Your active participation and respect for everyone in the class are key to the success of the course. Please silence your cell phone before class begins and do not take it out during class as it can be disruptive. If you have extenuating circumstances that require you to have a cell phone out, just let me know before class, otherwise your participation grade may be penalized. Please also avoid the temptation to check email or social media during class (unless instructed to do so).

Contacting Me Before, During, and After the Seminar
Please let me know if you have questions or concerns about assignments, course materials, policies, or ways of making the class fully accessible to you. Email is the best way to contact me to set up an appointment or to ask a quick question. I check my email account regularly and will do my best to respond to you promptly. A week is a short time to cover digital humanities. If you have questions after the course ends, would like recommendations about next steps, or ever need someone to discuss ideas with, please feel free to email! I will be happy to stay in touch.

Schedule

11 March 2019: Data in the Humanities
6:00 – 8:30 p.m. | 18:00 – 20:30

      INTRO TO DIGITAL HUMANITIES
            Overview of digital humanities and the class
            Discussion of data in the humanities – what is it and how do we find it
            Data cleaning with OpenRefine
            Data visualization fundamentals with Tableau Public

      READING DUE TODAY
            "Big? Smart? Clean? Messy? Data in the Humanities"
            (http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/)

12 March 2019: Texts
6:00 – 8:30 p.m. | 18:00 – 20:30

      INTRO TO TEXT ANALYSIS
            Overview of text mining approaches
            Discussion of sample projects
            Out-of-the-box tool exploration with Voyant and Databasic.io
            Named Entity Recognition (NER) with Stanford's NER GUI

      PROJECT EXPLORATION DUE TODAY (choose one for the short writing assignment)
            Robots Reading Vogue http://dh.library.yale.edu/projects/vogue/
            Mining the Dispatch http://dsl.richmond.edu/dispatch/
            Early Modern Print https://earlyprint.wustl.edu/

13 March 2019: Networks
6:00 – 8:30 p.m. | 18:00 – 20:30

      INTRO TO NETWORK ANALYSIS
            Overview of network analysis fundamentals
            Discussion of reading
            Creating node and edge lists by hand and with Python
            Generating a network graph in Gephi

      READING DUE TODAY (use this reading for the short writing assignment)
            "Network Theory, Plot Analysis"
            (https://litlab.stanford.edu/LiteraryLabPamphlet2.pdf)

14 March 2019: Maps
6:00 – 8:30 p.m. | 18:00 – 20:30

INTRO TO SPATIAL ANALYSIS
        Overview of geographic information systems (GIS)
        Discussion of sample projects
        Collecting geospatial data by hand and with Python
        Creating a map in ArcGIS online
        Presenting your map in Esri's Story Maps

PROJECT EXPLORATION DUE TODAY (choose one for the short writing assignment)
        The Atlas of Early Printing: http://atlas.lib.uiowa.edu/
        Photogrammar: http://photogrammar.yale.edu/
        Charles Booth's London: https://booth.lse.ac.uk/

15 March 2019: Sharing Digital Humanities Research
6:00 – 8:30 p.m. | 18:00 – 20:30

ENGAGING WIDER AUDIENCES
        Project presentations
        Building a website with GitHub pages
        Crowdsourcing with Zooniverse
        Next steps for continuing with digital humanities

3-MINUTE PROJECT PRESENTATIONS DUE TODAY