TEC 154 2014S, Class 22: Digital Humanities (1)

Overview

- Preliminaries.
  - Admin.
- Presentaiton.

Preliminaries

Admin

- This week we have James Lee visiting. MK will take notes on Monday and Wednesday.
- I anticipate getting your graded work back after break.
- The reading for Wednesday was distributed via email.
- Missing: C, A, L, D

Key Points Identified by DK

1. Digital humanities is a concept about interdisciplinary solutions and methods to answering questions. (Typically, the questions are humanities-based at the start)

2. It is encouraged to better use and borrow from a range of people.

   - Examples: Recaptcha, Civil War letter transcribing project

3. Problems can arise in the difference of dialect, such as the scientific and literary method of essay writing.

4. Franco Moretti is an important figure in the digital humanities.

5. He mentions "distant" and close readings of material and situations

6. Other methods to consider when using digital humanities include using statistical or information searching techniques, or increasing the scale of information put into a solution

Prof Lee hopes that digital humanities will move away from just "awesomeness" and become more utilized.

Key Points Identified by DP

1. The humanities are one of the oldest fields of study, and we can use the digital humanities to bring this ancient field into the 21st century. The centuries old techniques may need to be adapted as they are not applicable in today's world. Additionally, the humanities must catch up with modern technology and ask where they fit in.

2. We can ask the big questions about ethics and morals with quantitative tools through the use of digital humanities.

3. At Grinnell there is a phenomenon of "reverse economies of scale" in which the small, porous environment allows cross-discipline communication and collaboration. This is fairly unique to Grinnell and would not occur at a larger research institution.

4. The digital humanities open up a library of literary resources to the public that have never before been available. It also gives individuals access to more literature than they would be able to cover in a lifetime.

5. There is great controversy among scholars about the application of the digital humanities. Some argue that it is a form of "selling out."

Digital Humanities: An Opportunity to Rethink Both Terms

- Side note: He has four MAP students this semester, two CS, two English or Philosophy.
  - It's interdisciplinary
  - Interested in Humanistic questions
  - Also interested in technical tools
  - Linking the two sides of their brains
  - A potentially unique thing to Grinnell
- Wednesday - Let's see what the MAP students are doing and what the English seminar are doing using GIS mapping and text analysis; and how all of this dovetails with faculty research
- Why is Prof. Lee interested in the digital humanities?
  - Most of his college work was in molecular biology, even published a few papers. But didn't want to go into science. Tools and theory were interesting; wet lab was not.
  - Big question: What can he take from the scientific method as he thinks about questions in the humanities.
  - Issue: People get locked into particular methods/camps: Historical, Close reading, etc.
  - What we can important from experimental design - To prove something is plausible, we can't rely on just one method. *You need independent verification.* The more models you can bring to analyze an issue or to support a hypothesis, the better your analysis/support are.

An Overview of Digital Humanities

- Computers have been used to study humanities for a long time. (Visitor later this semester was using punch cards in the 1970's to study Milton.)
- These days, there's now a critical mass of scholars and jobs and students and more.
- Why? In part, because humanities are trying to better fit into society.
- There's an "old school" view that you understand via close reading.
- Pushback: The world is changing rapidly. The humanities need not change, but they should ask whether the age-old methods are still applicable. (Or argue that they are.)
- An opportunity to rethink both terms in new ways.
  - Can digital tools be used to answer historical, cultural, or otherwise humanistic questions.
  - What if your data are emotions or words or ...?
- Side note: Collaborating with his brother, a "quant" on Wall street.
  - Quants build tools for others who use the tools to analyze data.
  - His brother finds the pipeline exciting. Things are not in silos, but are connected.
- How do you answer big cultural questions about society and ethics with digital tools? There's not one answer.
- This is a rare opportunity to define what the humanities do.
  - Current version comes from the early 20th century. Breaking things off into disciplines/areas.
  - Rather than assuming that the humanities mean a certain thing or assume a different method.

Three Important Aspects of the Digital Humanities

Intrinsically, one key issue is to make the work *public*.

- A criticism, particularly in the UK. It was the domain of elite scholars who had the resources (including the knowledge resources, such as knowledge of classical languages) to study these things.
- A related criticism: Most of the work is behind a paywall.
- One goal: Make things public. While digital humanists are interested in using traditional mechanisms, they also want to make sure that their work is also available to others.
- A sense of broadening the *audience*.
- There are many versions of digital humanities.
  - Simple: Take traditional stuff and put it online. For example, a Website that provides an online art exhibition. The original is analog, but the delivery is digital.
- Another important issue of digital humanities is *crowdsourcing*, a way to harness a newly global and interested audience to contribute data points and more.
  - UIowa Civil War Letters project. UIowa has an incredible archive of these letters. OCR will never work on these letters. Is there a community out there who will manually transcribe the texts? There's certainly a vast community of Civil War buffs who might contribute.
  - A history scholar at Carleton looks at cancer in Minnesota. Plotting where people with cancer live. Started as a small map and is gaining a lot of momentum.
    - An opportunity to look at where clusters are formed.
    - An empirical study.
    - Works with epidemiologists.

      ○ Recaptchas as a way of getting people to transcribe text that is difficult to OCR. [Sam notes that this work is done by Luis Van Ahn, who is really clever about such things.]

Most digital projects are *collaborative*

- Digital humanities projects require knowledge that most humanists have not developed
    - Even the simple Web site projects
    - But certainly the big data projects
- So there's a need to work with others.
- And, certainly, the computer scientists and mathematicians they work with care about these issues, too, although may lake the theoretical tools to ask these questions.
- And, even if we've gathered the data, who has the tools to analyze the data?
- Problems with silos, and the lack of incentive to reach across disciplinary lines.
- There's also an issue of language - What terms do we use to talk about things?
- The size of Grinnell makes it easier to reach across disciplines, particularly as we look

There's also a lot of *controversy* about the work.

- Are we moving away from the core mission of the humanities?
- Are we selling out?
- Are we sucking away the soul of the humanities?
- Lee suggests that if we can answer questions from both a quantitative and qualitative, we provide more compelling arguments.
- In most of the public press, the digital humanities sound exciting, but they are very much contested.

Franco Moretti and Stanford's Literary Lab

http://litlab.stanford.edu/

- Was a traditional humanistic scholar.
- In the late 1990's, started to think about statistics and more as ways to think about literary style.
- Fast forward 10 years, founded the "literary lab"
    - Well funded
    - Lots of computers
    - Pull together different people
    - Don't publish traditional literary work - Publish pamphlets
- Very controversial
- Moretti likes to push people's buttons.
    - He speaks and writes in fragments of language.
    - He says thinks like "We don't need books. We don't need another close reading of Hamlet."
- "The slaughterhouse of the British marketplace."
    - Only about ten canonical figures
    - But lots of writers
    - Why do these few authors get treated as canonical?
- The middle part of Liu's article uses one of these pamphlets to talk about the approaches of the lab.

- "Distant" vs. close reading

New Approaches

It's difficult to think about trends across time. What do 50K words look like to you, or words from 50K texts?

- *Visualization* - Finding ways to think about those data.
  - Example: Simpson's work on Ulysses. Illustrate what the careful physical analysis yields.
- Natural language processing - Search vs. *Statistics*
  - Transform text into a corpus
  - Use statistical software to analyze trends
  - An example: Principal Component Analysis
    - Identify principal components
    - Look at relationships
    - Requires a lot of massaging and simplifying data
  - Lee is trying to move away from these statistical approaches
    - Why translate language to numbers so early in the process?
- Really smart people are dealing with language as language in the issue of *search*. Why not work more there?
  - There has to be a hybrid approach that works with both
- The main advantage of the digital humanities: *Scale*
  - He's studying 50K texts (the corpus is expanding to 127K texts)
  - A single human probably can't read that many texts
  - Gives us access to things that an individual human cannot read
  - For a single author, we can read all of the texts
  - Lets us take what we've seen about that author (or about any small set of tests) and see if conclusions we've come up from the sample can be applied to the wider variety of texts.
- We need to move beyond the "that is so awesome" aesthetic response and challenge things at a more substantive level

Questions

Could we apply this to the literature of a discipline, such as Psychology, and look for trends across time.

Certainly. Of course, one nice thing about literary texts is that they have structure. E.g., in plays, character names precede the line. For disciplines, an opportunity to see interesting clustering and relationships. For example, what are the common terms at various points? What if we look at history as a set of discourse, rather than as a series of events?

What are some of the statistical tools?

A basic tool is word count. For example, this word is most prevalent. But words matter in context. But the PCA is more common these days and transforms the words into some set of numbers. Want it to be a means to an end, rather than the end in itself. Statistics need to be woven in strategically.

What are the humanities?

Historically, they are a method. The oldest field of study. The art of rhetoric. Back to Aristotle. Everything derives from Aristotle's definition of rhetoric. Has branched off. At the most basic level, disciplines that focus on questions that are true to life. There are models that the STEM fields give us, but do they correspond to what we experience in life. Are there things that we can't account for in the theories and models of the STEM fields? The study of culture and language. How do we understand the experience of individuals. Reveal the processes by which we get to a truth. Don't just accept truths, but ask about their history. Issues that have happened over time.

How do we use the digital humanities to study non-literary artifacts? Such as visual images and sounds?

It's difficult. Right now, it's building an audience.

Comparing quantiative and NLP methods is interesting for finance.

Humans are irrational to a large extent. Language signifies something