Visualizing Texts, Modeling History: Possibilities for Digital Scholarship in the Humanities

Lisa Spiro
Digital Media Center
October 2008
How can computing advance humanities research?

Computing can enable:

- Discovering information, by making it accessible and searchable
- Collecting, organizing and sharing information, as one tags it, sorts it, and makes it available
- Building community
- Revealing patterns in information
- Visualizing movement & change
- Simulating cultural heritage
- Representing & disseminating knowledge, through new models for publication

But we’re still in the early days, so many tools & resources are not fully developed
Exploring digital scholarship in the humanities

Remix dissertation as work of digital scholarship
Rely on digital resources
Experiment with digital tools
Share openly
Blog the process
Explore multimedia publication

http://digitalscholarship.wordpress.com/
Enabling Research by Creating CyberInfrastructure

Cyberinfrastructure: “layer of information, expertise, standards, policies, tools, and services that are shared broadly across communities of inquiry but developed for specific scholarly purposes” (Revolutionizing Science and Engineering through Cyberinfrastructure, 2003)

Example of CI: nanoHUB:

Tools, simulations, data, community
“We have remarkable opportunities to bring new analytic and interpretive power to bear on the materials and the methods of the humanities and the social sciences: by so doing, we can advance our understanding of human cultures past, present, and future.” (ACLS Cyberinfrastructure Commission for Humanities & Social Sciences)
Access: Valley of Shadow

http://valley.vcdh.virginia.edu/choosepart.html
Access: The Whitman Archive

http://www.whitmanarchive.org/
Impact of the Whitman Archive (WWA)

- Whitman scholars on WWA: “indispensable,” “the first place that I go to do research on Whitman,” “the most important development in the history of Whitman studies.”
- Enables scholars to analyze key sources more rapidly & effectively, e.g. Traubel’s *With WW in Camden*, a 9 volume chronicle of Whitman
- Supports manuscript and textual study of the poet
- Enriches understanding of contexts, e.g. periodicals, disciples, visual culture
Collection, Annotation, Collaboration

- NINES (Networked Infrastructure for Nineteenth-century Electronic Scholarship) Collex:
  - “search and browse more than 300,000 peer-reviewed texts, images, and citations in 19th-century studies
  - build your own collections of documents, articles, images, and ephemera;
  - organize, add keywords, and annotate your work;
  - discover lines of critical inquiry related to your own
  - and (coming soon!) create illustrated exhibits about material in NINES.”

http://nines.org/collex
Collex

NINES is a federation of peer-reviewed resources, citation records, and innovative research tools, made freely available to students and scholars of 19th-century culture. :: learn more ::

This exhibit is a shared resource. It is available for general use but has not yet been peer-reviewed by the NINES Editorial Boards. If you would like to recommend this exhibit for peer-review, please contact the Project Manager.

Annotated Bibliography

Pre-Raphaelite Caricature

This is a sample annotated bibliography created using the NINES data set and Collex, a tool for searching, browsing, collecting, and annotating federated digital resources. Collex now allows users to create and share online "exhibits" based on peer-reviewed NINES resources.

These exhibits currently take the form of annotated bibliographies and illustrated essays. Both of those formats are being expanded and fine-tuned in cooperation with the Poetess Archive Journal. Future development will allow for the creation of course syllabi and timelines in the Collex exhibit builder.
Text Visualization

- Represent text graphically
- See **patterns** in data, e.g.
  - Word frequency
  - Semantic structures
  - Relationships among words
- Build interpretations or explanations based on analysis of those patterns
Tag Clouds

“a visualization of word frequencies.” (Many Eyes)

http://chir.ag/phernalia/preztags/
Comparison: Document Cloud Comparison

<table>
<thead>
<tr>
<th>Obama 2004 DNC Keynote Address</th>
<th>Obama 2008 DNC Acceptance Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>america</td>
<td>america</td>
</tr>
<tr>
<td>states believes united</td>
<td>promise keep cannot</td>
</tr>
<tr>
<td>white black latino</td>
<td>time work states blue</td>
</tr>
<tr>
<td>pundits asian</td>
<td>country right served last</td>
</tr>
<tr>
<td>conservative</td>
<td>democrats change</td>
</tr>
<tr>
<td>place liberal</td>
<td>together need turn can</td>
</tr>
<tr>
<td>feel kerry red can name</td>
<td>news make mccain red john</td>
</tr>
<tr>
<td>end john face rich blue child</td>
<td>help part belief life settle</td>
</tr>
<tr>
<td>faith safe hope knows</td>
<td>essence mend safe destiny</td>
</tr>
<tr>
<td>hard country politics across</td>
<td></td>
</tr>
<tr>
<td>keep</td>
<td></td>
</tr>
</tbody>
</table>

ip to study in a magical place. America, that shone a
'O believing that in a tolerant America, your name is
't rich, because in a generous America, you don't hav
That is the true genius of America, a faith -- a faith I
make sure that every child in America has a decent s
John Kerry believes in an America where hard work
John Kerry believes in an America where all America
use our military might to keep America safe and sec
John Kerry believes in America.

tonight, there is not a liberal America and a conserv
There is not a Black America and a White America a
fending the United States of America.

n, but shared a belief that in America, their son coul
America, we are better than these last eight years.
It's time for us to change America.
That's the promise of America - the idea that we are
create good jobs right here in America.
e future are built right here in America.
America, now is not the time for small plans.
And I will not settle for an America where some kids
mily leave, because nobody in America should have
'd tax havens that don't help America grow.
must also admit that fulfilling America's promise wil
sibility - that's the essence of America's promise.
Text Analysis

- Break down texts into component parts, then represent that information to support analysis

What you can do with text analysis:
- Create concordances
- Evaluate tone
- View words in context, word pairs, etc.
- Study authorship, linguistic change, etc.

Example Tools:
- TAPOR: http://portal.tapor.ca/portal/portal
- TokenX at Cather Archive
Both Melville’s *Pierre* (1852) and Mitchell’s *Reveries of a Bachelor* (1850) focus on sentimental, artistic bachelors. What is the relationship between the two works?

I got electronic copies of each text from the Early American Fiction project.

I used Wordle word cloud generator to “see” the works (Available now through Many Eyes, which offers tools for creating visualizations)

I used TAPOR’s Comparator & collation tools to examine two works in relation to each other.
Reveries Word Cloud
(Wordle)
Pierre Word Cloud (Wordle)
Comparing *Reveries & Pierre* with Wordle
TAPoR (Text Analysis Portal for Research): Humanities Cyberinfrastructure

Geoffrey Rockwell, “Cyberinfrastructure: Reflections from TAPoR to Tools”
Comparing *Reveries & Pierre* with TAPOR Comparator

<table>
<thead>
<tr>
<th>Words</th>
<th>Rev counts</th>
<th>Rev relative</th>
<th>Pier relative</th>
<th>Pier counts</th>
<th>Rel ratio (R/P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mother</td>
<td>58</td>
<td>0.0009</td>
<td>0.0015</td>
<td>237</td>
<td>0.5953</td>
</tr>
<tr>
<td>father</td>
<td>39</td>
<td>0.0006</td>
<td>0.0009</td>
<td>138</td>
<td>0.6875</td>
</tr>
<tr>
<td>sweet</td>
<td>73</td>
<td>0.0011</td>
<td>0.0008</td>
<td>125</td>
<td>1.4206</td>
</tr>
<tr>
<td>light</td>
<td>45</td>
<td>0.0007</td>
<td>0.0007</td>
<td>106</td>
<td>1.0327</td>
</tr>
<tr>
<td>morning</td>
<td>56</td>
<td>0.0009</td>
<td>0.0005</td>
<td>86</td>
<td>1.584</td>
</tr>
<tr>
<td>night</td>
<td>68</td>
<td>0.001</td>
<td>0.0007</td>
<td>110</td>
<td>1.5037</td>
</tr>
<tr>
<td>time</td>
<td>106</td>
<td>0.0016</td>
<td>0.0014</td>
<td>227</td>
<td>1.1359</td>
</tr>
<tr>
<td>heart</td>
<td>199</td>
<td>0.003</td>
<td>0.0012</td>
<td>186</td>
<td>2.6026</td>
</tr>
<tr>
<td>hand</td>
<td>102</td>
<td>0.0016</td>
<td>0.001</td>
<td>163</td>
<td>1.5222</td>
</tr>
<tr>
<td>face</td>
<td>62</td>
<td>0.0009</td>
<td>0.001</td>
<td>162</td>
<td>0.931</td>
</tr>
<tr>
<td>eye</td>
<td>71</td>
<td>0.0011</td>
<td>0.0004</td>
<td>67</td>
<td>2.5778</td>
</tr>
<tr>
<td>love</td>
<td>134</td>
<td>0.002</td>
<td>0.0012</td>
<td>192</td>
<td>1.6977</td>
</tr>
<tr>
<td>feel</td>
<td>32</td>
<td>0.0005</td>
<td>0.0006</td>
<td>90</td>
<td>0.8649</td>
</tr>
<tr>
<td>feeling</td>
<td>34</td>
<td>0.0005</td>
<td>0.0003</td>
<td>48</td>
<td>1.723</td>
</tr>
<tr>
<td>think</td>
<td>70</td>
<td>0.0011</td>
<td>0.0005</td>
<td>85</td>
<td>2.0033</td>
</tr>
</tbody>
</table>
### Some Unique Words That Occur Frequently in Reveries or Pierre

<table>
<thead>
<tr>
<th>Reveries</th>
<th>Pierre</th>
</tr>
</thead>
<tbody>
<tr>
<td>coal</td>
<td>portrait</td>
</tr>
<tr>
<td>flirt</td>
<td>ambiguous</td>
</tr>
<tr>
<td>sympathies</td>
<td>marvelous</td>
</tr>
<tr>
<td>sparkles</td>
<td>original</td>
</tr>
<tr>
<td>sensibility</td>
<td>visible</td>
</tr>
<tr>
<td>mused</td>
<td>seized</td>
</tr>
<tr>
<td>prettier</td>
<td>miserable</td>
</tr>
</tbody>
</table>
Words co-occurring with “mother”:

<table>
<thead>
<tr>
<th>Reveries</th>
<th>Pierre</th>
</tr>
</thead>
<tbody>
<tr>
<td>heart</td>
<td>dear</td>
</tr>
<tr>
<td>kiss</td>
<td>conceal</td>
</tr>
<tr>
<td>lap</td>
<td>torture</td>
</tr>
</tbody>
</table>
Impact of Experiments with Comparison

- Allows you to extract out key features of texts and observe patterns
- You can recontextualize those features by using concordancing tools
- Establish a “linguistic profile”: see how Melville appropriates and remixes language of sentimentality
- Text analysis tools open up new questions more than they reveal answers--stimulus to interpretation
Text Mining

- Problem: experts can only read so much in own field, and little in other fields.
- Use computer to extract new information & patterns from large textbase (vs. search, where you’re looking for something specific)
- Applications:
  - Biomedicine: text mining led to discovering link between magnesium deficiency & migraines
  - National security
  - Humanities: conduct queries across, say, Google Books
Text Mining: “Hot” or “Not”

- Use NORA to examine erotic language in corpus of 300 letters between Emily Dickinson & Susan Dickinson

- Text mining enabled Martha Nell Smith to “plumb much more deeply into little four- and five-letter words, the function of which I thought I was already sure, and has also enabled me to expand and deepen some critical connections I've been making for the last 20 years.”

Plaisant, et al “Exploring Erotics in Emily Dickinson's Correspondence with Text Mining and Visual Interfaces.” JCDL '06.
Data Visualization: “Historical Weather Maps”

“We know where the hurricanes ending up hitting and can retrospectively understand how these complex processes unfolded. In a similar process, we can comprehend the historical weather, tracing where the currents led, how the storms brewed, and how the unpredictable somehow came to pass.” (Ed Ayers, “Mapping Freedom”)
Valley & Historical Visualization: The Emancipation Project

Augusta County
Real estate wealth, according to census:
- > $15,000
- $9,000-$15,000
- $3,000-$9,000
- $300-$3,000
- < $300

Other Views:
- Race
- Biography
- Literacy

http://www.vcdh.virginia.edu/emancipation/index.html
Cinematic Maps: Voting America

- Visualize evolution of American presidential politics by transforming data into dynamic images “to reveal patterns across time”

http://americanpast.richmond.edu/voting/
Growth of Neighborhoods over Time: Trulia Hindsight

http://hindsight.trulia.com/
Challenges thesis that from 1900-1930, social forces stifled Irish-American literature

Jockers’ data reveals while this is true in the Eastern US, Irish-American lit expanded in the West

Example of “literary geospaces”: mapping literature, experience of space

https://www.stanford.edu/~mjockers/cgi-bin/drupal/node/19
Interactive Simulations

- The World's Columbian Exposition of 1893
- Developed by UCLA’s Urban Simulation Team
- Goals:
  - “interactive exploration of the modeled site, thereby creating unprecedented opportunities for experiential interpretation and innovative pedagogy”
  - “Real-time visual simulation technology allows us to reclaim the lost experience of navigating through the White City”

http://www.ust.ucla.edu/ustweb/Projects/columbian_expo.htm
Winner of the Bancroft Prize

http://www.vcdh.virginia.edu/AHR/
Open Archives: arXiv

http://arxiv.org/
The New Metrics of Scholarly Authority?

According to Michael Jensen, ranking of scholarship will be based on automatically-determined criteria such as:

- Prestige of commenters and other participants.
- Obvious attention: discussions in blogspace, comments in posts, etc.
- Nature of the language in comments: positive, negative, etc.
- Percentage of phrases that are valued by a disciplinary community.
- Quality of author's institutional affiliation(s).
- Amount of author's participation in other projects, as commenter, editor, etc.
- Inclusion of a document in lists of "best of," in syllabi, indexes, etc.
- Types of tags assigned to it, the terms used, the authority of the taggers, the authority of the tagging system.

http://chronicle.com/free/v53/i41/41b00601.htm
Challenges facing digital scholarship

- **Access to data**: Locating or creating data, then getting it into standard format.
- **Copyright**: Copyright prohibits access to much scholarly info.
- **Funding**: More money is needed to support digital humanities projects.
- **Academic culture**: Many departments don’t know how to evaluate digital scholarship; often it is not rewarded.
- **Training**: Technical skills are needed to develop digital humanities projects.
More challenges

- Tool development: Researchers need powerful, easy-to-use tools for analyzing & organizing information.
- Publishing: Outlets for publishing digital scholarship.
- Preservation: Providing long-term access to digital humanities resources.
- Proving itself: Is digital humanities a fad? Is it good for humanities scholarship?
Why should humanities engage with computing?

- Jerome McGann: “Because we have no choice.” The archive is becoming digital.
- Cathy Davidson: “Hybridity, exchange, flow, and cultural transaction are all explored more responsibly and adventurously when the resources of many nations, in many languages, have been digitized, made interoperable, and offered for research by scholars around the world.”
- Brett Bobley of the NEH: “it is about getting things done that couldn’t be done before.”
Opportunities for digital scholarship

- What’s difficult about the way we currently work?
- What are key questions and challenges that can benefit from the use of digital resources & tools?
- How can we imagine the future use of digital tools and resources?
Maps (Space, Geometry)

Example: changing map of village novels--from circle centered on village (1824) to scattered, distributed
Bonus Slides
Trees (morphology, form)

Examples:
- evolution of genres
- Features that enable genres to survive (e.g., use of clues in detective fiction)
- changing nature of free indirect discourse
Why should the humanities experiment with visualization?

- “Exposing my language-based discipline of history to the visual, the graphic, and the dynamic throws it off balance. And that's exactly what the humanities, like all academic disciplines, need to do if they are to stay alive and vital.” (Ed Ayers, “Mapping Freedom,” 2007)

- “The great fun of information visualization is that it gives you answers to questions you didn’t know you had.” (Ben Shneiderman, NYT, 2008)
Research Approaches: “Distant Reading” (Franco Morretti)

- How do we study world literature and understand the literary system?
- Close reading = small canon
- “Distant reading: where distance… is a condition of knowledge: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes—or genres and systems.”
- Microscope vs. telescope
- Elaborated in *Graphs, Maps, Trees* (2005)

http://www.newleftreview.org/A2094
Graphs (Time)

Examples: rise and fall of genres, national literatures