Wiley Webinar Series

Strategies for the Digital Humanities

Dr. Miguel Escobar Varela, Ph.D
Assistant Professor, National University of Singapore (NUS) and Academic Advisor on Digital Scholarship, NUS Libraries
miguelescobar.com | @migueljogja
This talk

What are the digital humanities?

What this means for researchers and institutions: challenges and opportunities

Some ideas for how to tackle these challenges and capitalize on these opportunities
What are the digital humanities (DH)?

The usage of digital data and digital methods for carrying out research in the humanities.
Areas in the DH

Databases

Data-driven analysis

Digital publishing

The DH are also a community, a space for meeting up and discussing how technology is changing the humanities
Databases
Databases

Data is ‘potential information’. A process needs to be applied to data to become meaningful. This is a challenge to how people usually think about research in the humanities.
Data is context dependent

- The question should not be what are data but "when are data" (Borgman, 2015)

That which is data in one context might be information or knowledge in another.
Sources of data in the humanities

Digitized books, films, artworks, video recordings

Motion capture

3d models

Surveys

Production details (film, theatre)

Historical records
Problems of DH data

Standardization

Incompleteness

Inaccuracy

These problems are important for historical data.

Annotation and discussion are the best way to deal with this problems.
Limited data in the humanities

- 20-30,000 novels published
- 6,000 novels still extant
- 3,500 novels available in full text

Digitization and availability of British Nineteenth-Century Novels. Image published under a CC-BY license (Schöch).
The future of data in DH

From Schöch (2013)
Smart data

Markup, annotations and metadata

Clear data models

Clear relations to external entities

“Smart data to be semi-structured or structured, clean and explicit, as well as relatively small in volume and of limited heterogeneity” Schöch
Welcome to the Chinese Text Project homepage. The Chinese Text Project is an online open-access digital library that makes pre-modern Chinese texts available to readers and researchers all around the world. The site attempts to make use of the digital medium to explore new ways of interacting with these texts that are not possible in print. With over thirty thousand titles and more than five billion characters, the Chinese Text Project is also the largest database of pre-modern Chinese texts in existence.

You may wish to read more about the project; view the “How to Use” and “User Guide” (or “Wiki” tables of contents, or consult the instructions, FAQs, or list of texts). If you're looking for a particular Chinese text, you can search for texts by title across the main textual sections of the site.

This site is provided as an open-access resource. If you use this site in your teaching or research, please ask your university library to subscribe. Donations are also greatly appreciated.

Latest additions

<table>
<thead>
<tr>
<th>Date</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016-12-30</td>
<td>Thanks to the support of Harvard Yenching Library, over 5 million pages of scanned materials from the Yenching Library collection have been added to the Library section of the site, including high quality images from the Chinese Rare Books Collection. Approximate transcriptions created using the open source OCR software have also been added to the YLS, making these materials fully searchable. In future we hope to collaborate with other libraries to include materials from their Chinese language collections.</td>
</tr>
<tr>
<td>2015-07-02</td>
<td>Support for Unicode 8.0 adds new characters: A new version of the Unicode standard has been released, defining thousands of additional newly used and variant Chinese characters. Support for these has been added to the dictionary section of the site. For these characters, please [install the latest version of the Tangut font]. Many new characters belong to &quot;CJK Extension E&quot; - you can confirm system support for these from the [Font Test page].</td>
</tr>
</tbody>
</table>
The Comédie-Française Registers Project
Data-driven research

Network analysis, geospatial visualizations, time series analysis and textual analysis
Networks

A network consists of:

Nodes (things that are connected).

Edges (connections between those things). Specific, explicit connections between things. They can be directed or undirected.

Examples: social networks, communication networks, citation networks, collaboration networks.
Network measurements

**Degree.** Total number of edges it has to other nodes.

**Density.** The portion of the potential connections in a network that are actual connections.
Network analysis of wayang kulit characters

# Network analysis of Javanese wayang kulit characters

<table>
<thead>
<tr>
<th>Characters</th>
<th>Percentage</th>
<th>Average Weighted Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indian Characters</td>
<td>53%</td>
<td>153.47</td>
</tr>
<tr>
<td>Javanese Characters</td>
<td>47%</td>
<td>73.36</td>
</tr>
</tbody>
</table>
Textual analytics

Many different approaches (most developed area of DH).

- Corpus linguistics
- Topic modelling
- Sentiment analysis
Voyant Tools (voyant-tools.org)

Analysis of letters by Vincent Van Gogh (1853-1890)

https://voyant-tools.org/?corpus=2a9aa299a95d7eca47cf68d25f0382e7
Robots reading vogue

Bookworm

http://bookworm.library.yale.edu/

Topic modelling

http://dh.library.yale.edu/projects/vogue/topics/
Geospatial analysis

Increasingly common approach

- Geospatial visualizations
- Geostatistics
Visualization of wayang kulit performances (Java, Indonesia)
Spatial distribution of Chinese Culture in Singapore

http://shgis.nus.edu.sg/
Time-series analysis

https://www.cfregisters.org/en/the-data/basic-tool
Javanese Wayang Kulit Performances Per Year
Titus Andronicus, Hong Kong Arts Festival and No Man's Land
Director: TANG Shu-wing, Date: 2008
Pathfinders (project built with Scalar)

http://scalar.usc.edu/works/pathfinders/traversals-and-interviews-documentation
The Chinese Deathscape

Edited by Thomas S. Mullaney, Stanford University Press
http://chinesedeathscape.org/
Challenges and suggestions
Challenges for researchers

DH changes to how we think about sources, methods and evidence

Requirements for new skills, forms of review and collaborations
Two provocative ideas

- We need to critically assess the potential of DH for the many different areas of the humanities
- We all need better training in statistics and computational thinking...even if we don’t want to use this methods
- Why? Because this methods are becoming increasingly important. We need to be able to critique them in nuanced ways.
Thank you

Comments and questions

m.escobar@nus.edu.sg
miguelescobar.com
@miguelJogja
Bibliography
