What Is Digital Humanities?

Doing research with digital materials and tools in a collaborative and open fashion. Not everyone needs to be a builder, though working with someone who can helps!

James Baker

NB: Refresh the page to get a new definition. Quotes were pulled from participants from the Day of DH between 2009-2014. As of January 2015, the database contains 817 rows and randomly selects a quote each time the page is loaded. If you want to do something cool with the data, I am providing a download for the CSV I compiled here.
What has changed since 2004, when DH comes into the picture?

- Large efforts in digitalization that made many objects created by humans available for software tools.
- Other disciplines being hit by technologies that handle big data.
- The irruption of Geographical Information Systems, GIS, and other visualization tools that have been used to easy ‘interpretation’
Summary of DH research

Research in humanities is often related to
the interpretation of objects created by humans

Such objects are mainly texts for literature and history, for instance

Research has to do with:
• Finding ... Texts
• Finding ... Data in texts
• Finding ... Patterns and relations in data, so that ..
• Making a particular interpretation in a sound and evident way is possible
Natural Language Processing (NLP) are available tools for ...

- Finding data in **large quantities of texts**
- Finding and **extracting patterns and correlations**
- Representing data to enable their analysis
NLP tools find names, relations, opinions ...

Finding names. Examples
Geographical Information Systems (GIS) and NLP
Named Entity Recognition & Desambiguation (NER)

The task is:
• identifying an entity mention (a proper noun) and
• identifying to which entity it refers to.

Geographical Names (assigning coordinates, for instance)
Person Names (linking to VIAF, Wikipedia, for instance)
GIS and NER (1) Pelagios Mediaeval Iberia http://commons.pelagios.org

Annotating texts with Identification of Named Entities

• NER with FreeLing
• Relate and extend with LOD, dbpedia...
• And geolocation with Geographic Information Systems (GIS)
Heritage As a Source of Studies into Industrial History: Using Digital Tools to Explore the Geography of the Industrialization

Guillermo Esteban-Oliver*, Adrià San José* and Jordi Martí-Henneberg*

Department of Geography and Sociology, University of Lleida, Lleida, Spain

The main objective of this article is to explore the possibility of combining two very different sources in order to study the distribution of industrial activity throughout history. The traditional primary sources to use for this purpose are the official censuses on population and economic activity that have been conducted in the majority of countries since the mid-nineteenth century. However, the majority of these lack detail at the regional level and also with respect to the types of professional occupations that they quantify. In order to complement and profile these census data, we propose the use of another type of information which can also be quantified, but whose characteristics are very different. We refer to the industrial heritage sites identified in digital format in a given territory, which in this case is Catalonia, Spain. This innovative dataset was obtained using digital tools such as web scraping and data mining techniques. This type of historical information was used to check whether it is reliable and valid for interpreting the spatial impact of the introduction of industrial activity. The article also shows that the systematic identification of elements of industrial heritage offers a new and very useful source of information for interpreting the history of industrial geography.
Finding patterns. Examples
NLP for finding patterns

The corpus has more than 5,000 sonets (about 71,000 lines)

**ADSO project**

The purpose of this Project is to develop a macroanalysis and distant reading of Spanish Golden Age Sonnets, from the Renaissance Era (Garcilaso de la Vega) to the poetry from the end of the Baroque period (Sor Juana Inés de la Cruz). Computational methods will be used in order to perform the analysis, so that the main recurrent traits—both metrical and semantic—can be detected and singled out.

Unlike previous studies, our aim is not to analyze the limited number of sonnets that have been standardized as ‘canonical’ poetry, but rather to achieve the identification and characterization of those literary traits, both metrical and semantic, that all the sonnets from the 16th and 17th centuries share (Navarro Colorado, 2015, 2016), using computational techniques.
NLP for Sentiment Analysis. Example
It analyzed 40 months of data, containing 1.8 billion comments made by over 100 million users across 36,000 communities.
NLP for Data Analysis. Examples
NLP for data analysis

The curious data here is less what Rose says (“I’m flying”) and more what the screen direction prescribes (“she smiles dreamily,” “he pushes against her”). In the following analysis, we go deep on screen direction to understand gender tropes. We examined 2,000 scripts and broke down every screen direction mapped to the pronouns “she” and “he.”

The R code behind this analysis is publicly available on GitHub.

https://juliasilge.com/about
https://pudding.cool/2017/08/screen-direction/
Word embeddings quantify 100 years of gender and ethnic stereotypes


Table 1. The top 10 occupations most closely associated with each ethnic group in the Google News embedding

<table>
<thead>
<tr>
<th>Hispanic</th>
<th>Asian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Housekeeper</td>
<td>Professor</td>
</tr>
<tr>
<td>Mason</td>
<td>Official</td>
</tr>
<tr>
<td>Artist</td>
<td>Secretary</td>
</tr>
<tr>
<td>Janitor</td>
<td>Conductor</td>
</tr>
<tr>
<td>Dancer</td>
<td>Physicist</td>
</tr>
<tr>
<td>Mechanic</td>
<td>Scientist</td>
</tr>
<tr>
<td>Photographer</td>
<td>Chemist</td>
</tr>
<tr>
<td>Baker</td>
<td>Tailor</td>
</tr>
<tr>
<td>Cashier</td>
<td>Accountant</td>
</tr>
<tr>
<td>Driver</td>
<td>Engineer</td>
</tr>
</tbody>
</table>

Table 2. Top adjectives associated with women in 1910, 1950, and 1990 by relative norm difference in the COHA embedding

<table>
<thead>
<tr>
<th></th>
<th>1910</th>
<th>1950</th>
<th>1990</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charming</td>
<td>Delicate</td>
<td>Sweet</td>
<td>Maternal</td>
</tr>
<tr>
<td>Placid</td>
<td></td>
<td></td>
<td>Morbid</td>
</tr>
<tr>
<td>Delicate</td>
<td></td>
<td>Charming</td>
<td>Artificial</td>
</tr>
<tr>
<td>Passionate</td>
<td>Transparent</td>
<td>Placid</td>
<td>Physical</td>
</tr>
<tr>
<td>Sweet</td>
<td>Dreamy</td>
<td>Childish</td>
<td>Caring</td>
</tr>
<tr>
<td>Dreamy</td>
<td>Indulgent</td>
<td>Soft</td>
<td>Emotional</td>
</tr>
<tr>
<td>Indulgent</td>
<td>Playful</td>
<td>Colorless</td>
<td>Protective</td>
</tr>
<tr>
<td>Playful</td>
<td>Mellow</td>
<td>Tasteless</td>
<td>Attractive</td>
</tr>
<tr>
<td>Mellow</td>
<td>Sentimental</td>
<td>Agreeable</td>
<td>Soft</td>
</tr>
<tr>
<td>Sentimental</td>
<td></td>
<td></td>
<td>Tidy</td>
</tr>
</tbody>
</table>
It's impossible to conduct research without software, say 7 out of 10 UK researchers

The Software Sustainability Institute, University of Edinburgh

(http://www.software.ac.uk/blog/2014-12-04-its-impossible-conduct-research-without-software-say-7-out-10-uk-researchers)

<table>
<thead>
<tr>
<th>2014</th>
<th>Do you use research software?</th>
<th>Do you develop your own research software?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>21</td>
</tr>
<tr>
<td>Humanities &amp; Language Based</td>
<td>No</td>
<td>17</td>
</tr>
<tr>
<td>Studies &amp; archaeology</td>
<td>Yes</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>2</td>
</tr>
<tr>
<td>Medicine, Dentistry &amp; Health</td>
<td>Yes</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>6</td>
</tr>
</tbody>
</table>
Researchers need information about tools ....
About research infrastructures ...

https://www.clarin.eu

By Researchers for Researchers

DARIAH is a pan-European infrastructure for arts and humanities scholars working with computational methods. It supports digital research as well as the teaching of digital research methods.

https://www.dariah.eu
Summary of DH research and NLP

Research in humanities is often related to the interpretation of objects created by humans.

Such objects are mainly texts for literature and history, for instance.

Research has to do with:

- Finding ... Texts ????????
- ✓ Finding ... Data in texts
- ✓ Finding ... Patterns and relations in data, so that ..
- ✓ Making a particular interpretation, sound and evident, is possible
Finding texts: Researchers need downloadable and processable DATA as collections ...