DRAG AND DROP IT, ZIP-UNZIP IT, VIEW IT, CODE IT: WHAT ARE DIGITAL HUMANITIES?

HEATHER FROEHLICH
@HEATHERFRO // #DAFTPUNKDH
WHO AM I?

PhD student at University of Strathclyde, Glasgow, UK

- I study gender in Early Modern London plays using computers

Visualizing English Print, 1470-1800

Strathclyde, Folger Shakespeare Library, UW-Madison

- http://winedarksea.org
- http://graphics.cs.wisc.edu/VEPsite/
BUT WHO ARE YOU?

How many of you ...

• Have a digital analysis project? (or one in mind?)
• Use computers in your research already?

• Know about….
  • Corpus linguistics?
  • A programming language?
  • The phrase “digital humanities”?
DIGITAL... HUMANITIES?

Amanda French
Roy Rosenzweig Center for History and New Media
The use of computers by scholars, librarians, archivists, museum professionals, and others to do and to support humanities research.

Scott B. Weingart
Indiana University
It's what humanists do with computers, and what computers do with humans.

Michael Ullyot
English, University of Calgary
Qualitative queries, quantified.

Ryan Hunt
University of Western Ontario
DH is about using technology to collaborate, build, discuss, and, engage. DH is inclusive, allowing individuals, both inside and outside of the academy, to embrace and define it themselves.

DIGITAL... HUMANITIES?

Philip R. "Pib" Burns
Northwestern University
Humanities is humanities, digital or not -- the study of literature, fine arts, history, language, and philosophy. We don't distinguish digital sociology or digital astronomy, so why digital humanities? Just because computers are involved doesn't mean the basic nature of the subject area is any different than it has been been traditionally. Computers allow for doing things with texts and other cultural artifacts that could not be done feasibly without the computational power and storage modern computers provide. Computers should be considered an extension of the scholar's mind -- very useful tools indeed. -Philip R. Pib Burns, Northwestern University, USA

Lauren Klein
Georgia Tech
Last year, I said that this question had become harder and harder to answer. This year, I think the reverse: Digital Humanities is simply THE humanities, as it is practiced, modeled, and theorized today.

Kimon Keramidas
Bard Graduate Center
ugh? really? maybe later

DIGITAL... HUMANITIES?

http://whatisdigitalhumanities.com/

• “Refresh the page to get a new definition. Quotes were pulled from participants from the Day of DH between 2009-2012. As of December 2012, the database contains over 500 rows and randomly selects a quote each time the page is loaded.”

(Jason Heppler, designer)
HUMANITIES, WITH COMPUTERS!?
5 MAJOR CATEGORIES

- Humanistic scholarship presented in digital form(s)
- Humanistic scholarship enabled by digital methods & tools
- Humanistic scholarship about digital technology & culture
- Humanistic scholarship building and experimenting with digital technology
- Humanistic scholarship critical of its own digital-ness

What counts as “digital humanities” then?

• 5 minutes: what do you think counts as digital humanities? Why?
IS THE VARIENG WEBSITE DIGITAL HUMANITIES?

VARIENG stands for the Research Unit for the Study of Variation, Contacts and Change in English. It also stands for innovative thinking and teamwork in English corpus linguistics and the study of language variation and change.

VARIENG members study the English language, its uses and users, both today and in the past. We are interested in how language is situated in social, cognitive, textual and discourse contexts, and produced in speaker interaction; how language varies and changes in meaning and structure; and how change is connected with language typology.

The VARIENG Research Unit was founded by linguists and philologists in the Department of English, University of Helsinki, in 1995. The Academy of Finland appointed VARIENG as a Finnish Centre of Excellence for a six-year term from 2000 to 2005, and again from 2006 to 2011.

VARIENG continues as a vibrant research community in the new Department of Modern Languages. In May 2012, we were acknowledged in the research evaluation of the University of Helsinki as one of the University’s international cutting-edge research communities. Read more...

Fourteen volumes of VARIENG’s eSeries
WHAT ABOUT VARIENG’S EPUBLICATIONS?

One volume is currently in the production pipeline and scheduled for publication in 2013.

Volume 1 – Annotating Variation and Change
    Editors: Anneli Meurman-Solin & Arja Nummi

Volume 2 – Towards Multimedia in Corpus Studies
    Editors: Päivi Pahta, Irma Taavitsainen, Terttu Nevalainen & Jukka Tyrkkö

Volume 3 – Approaches to Language and Cognition
    Editor: Heli Tissari

Volume 4 – Cambridgeshire Dialect Grammar
    Author: Anna-Lisa Vasko

Volume 5 – National Survey on the English Language in Finland: Uses, Meanings and Attitudes
    Authors: Sirpa Leppänen, Anne Pitkänen-Huhta, Tarja Nikula, Samu Kytölä, Timo Törmäkangas, Kari Nissinen, Leila Kääntä, Tiina Räsänen, Mikko Latinen, Päivi Pahta, Heidi Koskela, Salla Lähdesmäki & Henna Jousmäki

Volume 6 – Methodological and Historical Dimensions of Corpus Linguistics
    Editors: Paul Rayson, Sebastian Hoffmann & Geoffrey Leech
IS THIS SLIDESHOW DIGITAL HUMANITIES?
WHAT IS DH AND WHAT IS IT DOING IN ENGLISH DEPARTMENTS?

Text as manipulatable data

- Text-based data processing (stylistics, linguistics, author attribution)

Computers ↔ Composition

- Microsoft Word (among other word-processing software & the like)

Electronic archives, databases, archives

Hypertext and electronic literature

Cultural studies

e-reading devices

(Kirschenbaum 2012 [2010])
COMMON DIGITAL HUMANITIES METHODS

Creating and maintaining digital archives (Omeka, Scalar)
Text encoding (TEI)
Network Analysis (Gephi)
Text analysis and topic modeling (MALLET)
Data Visualization (graphs and charts; R)
CORPUS LINGUISTICS AND DH

Corpus projects are digital endeavors.

- Preparation: Text compilation & collection
- Preparation: Text encoding
- Analysis: keywords, collocation, clusters, ngrams…
- Analysis: close-reading examples in a concordance line
- Analysis: making intuitive visualizations of data (graphs/tables)
DIGITAL LITERACIES

Not just enough to know the humanistic details of your work (or is it?)

“There’s actually a lot of low-hanging fruit out there still worth picking — big questions that are easy to answer quantitatively and that only require organizing large datasets.” (Underwood, 2014)

Easy vs “easy”
“The whole point of quantification is to get beyond binary categories — to grapple with questions of degree that aren’t well-represented as yes-or-no questions.” (Underwood, 2014)

With the right skills – yes, of course this is easy!

But without them? This is difficult!
EASY VS “EASY”

It’s easy to drop items on a map

It’s less easy to decide what kind of catalogable information is relevant to document and address in what ways
EASY VS “EASY”: NEATLINE & OMEKA

Neatline allows scholars, students, and curators to tell stories with maps and timelines. As a suite of add-on tools for Omeka, it opens new possibilities for hand-crafted, interactive spatial and temporal interpretation.

Image from http://neatline.org :: a suite of add-ons to visualize data using Omeka (http://omeka.net)
It’s easy to make a PDF of a book chapter
It’s easy to make that PDF into a TXT file
It’s easy to catalog each feature of the chapter (words, paragraphs, headers) in the converted text file

It’s harder to do that for an image file of an older book, for example

It’s even harder to do that for a LOT of older books
Hamlet, Second Quarto, as seen in EEBO
From Sarah Werner’s “Where Material Book Culture Meets Digital Humanities”,
Journal of Digital Humanities 1.3, Summer 2012: bit.ly/1gBP86t
EASY VS “EASY”

It’s easy to ask a computer to tell you what the topics of what a corpus contain

It’s less easy to understand what the topics present, and why:

• “Simplifying topic models for humanists who will not (and should not) study the underlying algorithms creates an enormous potential for groundless — or even misleading — ‘insights’”.

-- Benjamin M. Schmidt, JDH 2.1, Winter 2012
http://bit.ly/1aR0TUo
According to the model, the total probability of the model is:

$$P(W, Z, \theta, \varphi; \alpha, \beta) = \prod_{i=1}^{K} P(\varphi_i; \beta) \prod_{j=1}^{M} P(\theta_j; \alpha) \prod_{t=1}^{N} P(Z_{j,t}; \theta_j) P(W_{j,t}; \varphi_{Z_{j,t}}),$$

where the bold-font variables denote the vector version of the variables. First of all, $\varphi$ and $\theta$ need to be integrated out.

$$P(Z, W; \alpha, \beta) = \int_\varphi \int_\theta P(W, Z, \theta, \varphi; \alpha, \beta) \, d\varphi \, d\theta$$

$$= \int_\varphi \prod_{i=1}^{K} P(\varphi_i; \beta) \prod_{j=1}^{M} \prod_{t=1}^{N} P(W_{j,t}; \varphi_{Z_{j,t}}) \, d\varphi \int_\theta \prod_{j=1}^{M} P(\theta_j; \alpha) \prod_{t=1}^{N} P(Z_{j,t}; \theta_j) \, d\theta.$$  

All the $\theta$s are independent to each other and the same to all the $\varphi$s. So we can treat each $\theta$ and each $\varphi$ separately. We now focus only on the $\theta$ part.

$$\int_\theta \prod_{j=1}^{M} P(\theta_j; \alpha) \prod_{t=1}^{N} P(Z_{j,t}; \theta_j) \, d\theta_j = \prod_{j=1}^{M} \int_{\theta_j} P(\theta_j; \alpha) \prod_{t=1}^{N} P(Z_{j,t}; \theta_j) \, d\theta_j.$$  

We can further focus on only one $\theta$ as the following:

$$\int_{\theta_j} P(\theta_j; \alpha) \prod_{t=1}^{N} P(Z_{j,t}; \theta_j) \, d\theta_j.$$  

Actually, it is the hidden part of the model for the $j^{th}$ document. Now we replace the probabilities in the above equation by the true distribution expression to write out the explicit equation.

$$\int_{\theta_j} P(\theta_j; \alpha) \prod_{t=1}^{N} P(Z_{j,t}; \theta_j) \, d\theta_j = \int_{\theta_j} \frac{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} \theta_{j,i}^{-1} \prod_{t=1}^{N} P(Z_{j,t}; \theta_j) \, d\theta_j.$$  

Let $n_{j,t}^i$ be the number of word tokens in the $j^{th}$ document with the same word symbol (the $i^{th}$ word in the vocabulary) assigned to the $i^{th}$ topic. So, $n_{j,t}^i$ is three dimensional. If any of the three dimensions is not limited to a specific value, we use a parenthesized point (·) to denote. For example,
<table>
<thead>
<tr>
<th>Music</th>
<th>Book</th>
<th>Art</th>
<th>Game</th>
<th>Show</th>
<th>Film</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band</td>
<td>Life</td>
<td>Museum</td>
<td>Knicks</td>
<td>Nets</td>
<td>Television</td>
</tr>
<tr>
<td>Songs</td>
<td>Novel</td>
<td>Show</td>
<td>Show</td>
<td>Points</td>
<td>Movie</td>
</tr>
<tr>
<td>Rock</td>
<td>Story</td>
<td>Exhibition</td>
<td>Artist</td>
<td>Team</td>
<td>Series</td>
</tr>
<tr>
<td>Album</td>
<td>Books</td>
<td>Artist</td>
<td>Artists</td>
<td>Season</td>
<td>Says</td>
</tr>
<tr>
<td>Jazz</td>
<td>Man</td>
<td>Paintings</td>
<td>Paintings</td>
<td>Play</td>
<td>Life</td>
</tr>
<tr>
<td>Pop</td>
<td>Stories</td>
<td>Century</td>
<td>Century</td>
<td>Games</td>
<td>Man</td>
</tr>
<tr>
<td>Song</td>
<td>Love</td>
<td>Works</td>
<td>Works</td>
<td>Night</td>
<td>Character</td>
</tr>
<tr>
<td>Singer</td>
<td>Children</td>
<td></td>
<td></td>
<td>Coach</td>
<td>Know</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Theater</th>
<th>Clinton</th>
<th>Stock</th>
<th>Restaurant</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Play</td>
<td>Bush</td>
<td>Market</td>
<td>Sauce</td>
<td>Tax</td>
</tr>
<tr>
<td>Production</td>
<td>Campaign</td>
<td>Percent</td>
<td>Menu</td>
<td>Governor</td>
</tr>
<tr>
<td>Show</td>
<td>Gore</td>
<td>Gore</td>
<td>Fund</td>
<td>County</td>
</tr>
<tr>
<td>Stage</td>
<td>Political</td>
<td>Investors</td>
<td>Food</td>
<td>Mayor</td>
</tr>
<tr>
<td>Street</td>
<td>Republican</td>
<td>Funds</td>
<td>Dishes</td>
<td>Billion</td>
</tr>
<tr>
<td>Broadway</td>
<td>Dole</td>
<td>Companies</td>
<td>Street</td>
<td>Taxes</td>
</tr>
<tr>
<td>Director</td>
<td>Presidential</td>
<td>Stocks</td>
<td>Dining</td>
<td>Plan</td>
</tr>
<tr>
<td>Musical</td>
<td>Senator</td>
<td>Investment</td>
<td>Dinner</td>
<td>Legislature</td>
</tr>
<tr>
<td>Directed</td>
<td>House</td>
<td>Trading</td>
<td>Chicken</td>
<td>Fiscal</td>
</tr>
</tbody>
</table>

A BRIEF INTERLUDE ON TOPIC MODELING

Topic models measure statistical co-occurrence

(often with high-frequency function words removed)

Each document is weighted

• 20% one kind of words (topic)
• 15% another kind of words (topic)
• 10% a third kind of word (topic)

Measures words which are likely to appear together

• Selecting number of topics decides how fine-grained the analysis will be
  • 4 topics will show you something very different than 104 topics
"liteRatuRe. theoRy. macRo. gRaph. (aRt?) Ready. stat." from:
http://culturelab.asc.upenn.edu/2013/01/11/literature-theory-macro-art-ready-stat/
We’ve talked a lot about the COMPUTER side of Computers + Humanities = Digital Humanities

What about the HUMANITIES part?
HUMANISTIC INQUIRY WITH DIGITAL APPROACHES

Using XML markup to determine which characters in a Shakespeare play are on-stage at any given time (http://folgerdigitaltexts.org)

Modeling St Paul's Churchyard to determine the acoustics of John Donne's sermon (http://vpcp.chass.ncsu.edu/)

Using grammatical markup to ask who is writing to whom and in what way? (various CEEC-related projects: http://www.helsinki.fi/varieng/domains/CEEC.html)

(among others!)
“LOW-HANGING FRUIT”

Jockers (2012, 35) calls metadata the “lowest-hanging fruit of literary history”

• Not just ease but accessibility

Asking questions with the information we have, then building up to bigger questions we don’t have information about
WHERE’S CORPUS LINGUISTICS IN ALL OF THIS?

Voyant Tools

- [http://docs.voyant-tools.org/about/examples-gallery/](http://docs.voyant-tools.org/about/examples-gallery/)

Various projects involving R

AntConc makes appearances

Many of the text analysis projects I’ve come across have been attempting to do corpus linguistics or stylistics but without corpus techniques, methods, theories
VOYANT TOOLS
(HTTP://VOYANT-TOOLS.ORG)

Image from H. L. Ulman:
http://ccdigitalpress.org/stories/chapters/ulman/toolkit_multifaceted.html
THE NEXT BIG THING IN LITERARY STUDIES?

Digital Humanities picks up pieces from media studies, book history, linguistics, computer science, sociology, statistics, literature, history, art, graphic design…

Are these all imperatives in literary studies? What about humanities inquiry?
PRACTICAL CONCERNS

“Our discussion in DH so often begin with tools (a workshop on this, a tutorial on that, what can X do for you?)”

(Honn 2013, http://joshhonn.com/?p=1)

Experimental: how can we use this new approach to best suit our immediate needs?
HOW DO WE SYNTHESIZE THESE APPROACHES IN A MEANINGFUL WAY?
COLLABORATION: A FEATURE OF DH

Cross-disciplinary: one person can’t do it all

• Or can they? 3 Examples:
  1. A professional programmer who went back to graduate school for a degree in literature, for example
  2. Or someone moving from an undergraduate degree in the sciences into a history program
  3. Someone with an interest in historical linguistics going into a degree in new media and shifting back into digital materiality

But this isn’t everyone – exceptions, not the norm!
A LEARNING CURVE

MOST DHers are learning the skills as they need to, not walking in with them under their belts already.
HUMANISTS ARE LEARNING ABOUT COMPUTERS

But computationally driven disciplines are learning about the humanities, too
INHERENTLY INTERDISCIPLINARY

Diane Jakacki
Georgia Tech

Engagement with technology (tools, platforms, networks) to develop and pursue modes of research and teaching otherwise unpursuable. This engagement requires the participation of transdisciplinary collaborators and correspondents.

DIGITAL HUMANITIES ENGAGES A WORLD OF LINKED AND LIVED EXPERIENCES.

BecAuse netwOrks connect us, they Are sOCial technoloGies. As sCholarship moves from the lIBRARY and the lecture hALL to digital communication networks, it takes on expanded sOCial roles and raises new questions. nEW moDeS of knoWledge formaTion in the digital humanities are dyNAmically linked to communities vastly larger and more diverse than those to which the academy has been accustomed. these com-munities increasingly demand and delight in sociable intellectual interactions, in which critique manifests as versioning, and thinking, making, and doing form itera-tive feedback loops.
“LINKED & LIVED EXPERIENCES”

Digital Humanities thrives on:

1. A network of people (often knowing other people) who can answer your questions
2. Openness (everyone’s welcome to join us)
3. Experimentation (go play, and see what you can come up with)
“LINKED & LIVED EXPERIENCES”

Digital Humanities requires:

1. Access to the network of people, either directly or indirectly
2. A willingness to accept new & old approaches unequivocally
3. A working knowledge of methods & theories being borrowed, remixed, re-imagined
ON ACCESS

DH-specific centres and workshops exist…

…but you often have to get to another corner of the world

• Victoria, British Columbia, Canada (DHSI/ETCL)
• College Park, Maryland, USA (MITH/HILT)

European DH Groups

• Oxford, England http://digital.humanities.ox.ac.uk/
• Leipzig, Germany http://www.culingtec.uni-leipzig.de/ESU_C_T/
• European Association for DH http://eadh.org/

Annual DH conference: swaps continents annually

• this year in Lausanne, Switzerland
• Last year in Lincoln, Nebraska, USA
ON ACCESS

• A willingness to accept new & old approaches unequivocally…

• … but there has been much debate over “what counts” as DH, contrary to what’s been outlined by others today

• Stephen Ramsay’s “Who's In & Who's Out” essay:

And, of course, all of this presupposes:

- Access to a computer with
  - administrative access
  - a willingness to participate in internet discourse
- Access to funding - someone willing to
  - send you elsewhere in the world to learn a highly specific task, which ultimately may not be relevant for what you end up doing
  - sponsor your project from within your own institution and/or nation
- Confidence to jump in with unfamiliar tools & approaches
- Time and energy to learn the unfamiliar tools & approaches
WHERE TO GO FROM HERE?

A SHORT GUIDE TO THE DIGITAL_HUMANITIES, part 5 of Burdick et al., 2012. Digital_Humanities. MIT Press.

• OA PDF: http://goo.gl/Mu5y1J

Journals

• Literary and Linguistic Computing http://llc.oxfordjournals.org/
• Digital Humanities Quarterly http://digitalhumanities.org/dhq/
• Journal of Digital Humanities http://journalofdigitalhumanities.org/
WHERE TO GO FROM HERE?

DH on Twitter

- https://twitter.com/alanyliu/seed-list-of-dh-scholars/members
- https://twitter.com/amandafrench/lists/digital-humanities-women
- https://twitter.com/dgolumbia/critical-digital-theory/members

Organizations

- ADHO http://adho.org
- European Association for DH http://eadh.org/

Humanist List

- http://digitalhumanities.org/humanist
SELECT WORKS CITED


