INLS 890: Digital Humanities:
Explorations and Experimentation

http://inls890spring2014.web.unc.edu/
12:30 – 3:15 pm, Manning 304

Richard Marciano, Professor
School of Information and Library Science
e-mail: richard_marciano@unc.edu
skype: richardjmarciano
Office Hours:
   Drop-in or by appointment

In this interdisciplinary graduate seminar, we will explore concepts and case studies in the digital humanities and experiment with software and tools. The class will consist of four main units, roughly one per month, with a hands-on assignment due at the end of each unit:

• January: *digitization / OCR*
• February: *GIS*
• March: *text analysis*
• April: *crowdsourcing*

Discussions of infrastructure and scalability will be incorporated into each of the units (as opposed to being a standalone theme).

Four sets of spatial and textual historical data will be provided and discussed at the start of the semester and teams of two to three students will be formed, allowing for reuse of data across all four units.

Suggested Textbooks


Class Meetings

In the first part of each class, we will explore and discuss topics in each of the 4 class units (*digitization/OCR* in January, *GIS* in February, *text analysis* in March and *crowdsourcing* in April).

In the second part of each class, we will experiment with software and tools that relate to the current unit.

We will also welcome speakers and conduct field trips during the semester.

Evaluation & Grading

Attendance is expected at all class meetings. Your ability to participate, contribute, influence the class is essential.

Course grades will be determined as follows:

10 points: **Participation**.
10 points: **Presentation & Critique** of a map project.

**Experimentation:**

• 15 points: *Digitization/OCR* assignment, due end of January.
• 15 points: *GIS* assignment, due end of February.
• 25 points: 2 *Text Analysis* assignments, due end of March.
• 15 points: *Crowdsourcing* assignment, due end of April.
10 points: **Putting it all together** assignment, due end of course.
Participation, the critique, and the final essay will be graded individually. For the experimentation assignments, you will be graded as a group.

**Digital Humanities Blogs**
There are many great DH blogs out there, among them:
- Digital Humanities Specialist: [https://dhs.stanford.edu/](https://dhs.stanford.edu/)
- GMU’s Center for History and New Media’s Digital Campus: [http://digitalcampus.tv/](http://digitalcampus.tv/)
- Flowing Data Blog on Data Visualization: [http://flowingdata.com/](http://flowingdata.com/)
Many DH centers maintain their own blogs and are easy to find via Google. The *New York Times'* ongoing Humanities 2.0 series may be of interest to many of you, as well. There is also a social media Linkedin “Digital Humanities / Humanities Computing” group.

**Digital Humanities Syllabi and Tools**

**Digital Humanities Journals**
- DHQ: digital humanities quarterly: [http://digitalhumanities.org/dhq](http://digitalhumanities.org/dhq)

<table>
<thead>
<tr>
<th>Explorations</th>
<th>Experimentation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Digitization / OCR</strong></td>
<td>Monday – Part 1</td>
</tr>
<tr>
<td>JAN.</td>
<td><strong>Week 1:</strong> 1/18: #1 – Introduction</td>
</tr>
<tr>
<td></td>
<td><strong>Week 2:</strong> 1/20: NO CLASS (MLK)</td>
</tr>
<tr>
<td></td>
<td><strong>Week 3:</strong> 1/27: #2 – Digitization / OCR</td>
</tr>
<tr>
<td></td>
<td><strong>Week 4:</strong> 2/03: #3 – Digitization at scale</td>
</tr>
<tr>
<td></td>
<td><strong>Experimentation</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Week 1:</strong> Introduction to VCL software / ABBYY</td>
</tr>
<tr>
<td></td>
<td><strong>Week 2:</strong> NO CLASS (Spring Break)</td>
</tr>
<tr>
<td></td>
<td><strong>Week 3:</strong> Field Trip: digitalINC &amp; Wilson Library</td>
</tr>
<tr>
<td></td>
<td><strong>Week 4:</strong> ABBYY</td>
</tr>
<tr>
<td><strong>GIS</strong></td>
<td>Monday – Part 2</td>
</tr>
<tr>
<td>FEB.</td>
<td><strong>Week 5:</strong> 2/10: #4 – Intro to GIS</td>
</tr>
<tr>
<td></td>
<td><strong>Week 6:</strong> 2/17: #5 – Recreating historical maps</td>
</tr>
<tr>
<td></td>
<td><strong>Week 7:</strong> 2/24: #6 – Presentation &amp; critique of map projects</td>
</tr>
<tr>
<td></td>
<td><strong>Week 8:</strong> 3/03: #7 – Map services in the cloud</td>
</tr>
<tr>
<td></td>
<td><strong>Experimentation</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Week 5:</strong> ArcGIS Intro</td>
</tr>
<tr>
<td></td>
<td><strong>Week 6:</strong> Georeferencing</td>
</tr>
<tr>
<td></td>
<td><strong>Week 7:</strong> Field Trip: digitalINC &amp; Wilson Library</td>
</tr>
<tr>
<td></td>
<td><strong>Week 8:</strong> ABBYY</td>
</tr>
<tr>
<td><strong>Text Analysis</strong></td>
<td>Monday – Part 1</td>
</tr>
<tr>
<td>MAR</td>
<td><strong>Week 9:</strong> 3/10: NO CLASS (Spring Break)</td>
</tr>
<tr>
<td></td>
<td><strong>Week 10:</strong> 3/17: #8 – Intro to text analysis</td>
</tr>
<tr>
<td></td>
<td><strong>Week 11:</strong> 3/24: #9 – Linked open data</td>
</tr>
<tr>
<td></td>
<td><strong>Week 12:</strong> 3/31: #10 – Big data</td>
</tr>
<tr>
<td></td>
<td><strong>Experimentation</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Week 9:</strong> NO CLASS (Spring Break)</td>
</tr>
<tr>
<td></td>
<td><strong>Week 10:</strong> Vogueur</td>
</tr>
<tr>
<td></td>
<td><strong>Week 11:</strong> Text analysis</td>
</tr>
<tr>
<td></td>
<td><strong>Week 12:</strong> Text processing</td>
</tr>
<tr>
<td><strong>Crowdsourcing</strong></td>
<td>Monday – Part 1</td>
</tr>
<tr>
<td>APR.</td>
<td><strong>Week 13:</strong> 4/07: #11 – Introduction to crowdsourcing</td>
</tr>
<tr>
<td></td>
<td><strong>Week 14:</strong> 4/14: Hands-on Workshop</td>
</tr>
<tr>
<td></td>
<td><strong>Week 15:</strong> 4/21: #13 – Crowdsourcing at scale</td>
</tr>
<tr>
<td></td>
<td><strong>Experimentation</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Week 13:</strong> Toolkits</td>
</tr>
<tr>
<td></td>
<td><strong>Week 14:</strong> Hands-on</td>
</tr>
<tr>
<td></td>
<td><strong>Week 15:</strong> More hands-on exercises</td>
</tr>
</tbody>
</table>

3D printing: [http://store.makerbot.com/replicator2](http://store.makerbot.com/replicator2)
Future of Higher Ed: [https://www.coursera.org/course/highered](https://www.coursera.org/course/highered)
**Schedule**

**#1: January 13**

**Part 1: Introduction**

Readings:

**Part 2: Introduction to VCL Software / ABBYY FineReader**

Readings:
- Virtual Computing Lab: [http://sils.unc.edu/it-services/remote-access/vcl](http://sils.unc.edu/it-services/remote-access/vcl) & [http://vcl.ncsu.edu](http://vcl.ncsu.edu)

**Unit 1: Digitization & OCR**

January 20: NO CLASS (MLK)

**#2: January 27**

**Part 1: Digitization & OCR**

Readings:
- Improving Access to Text, IMPACT, [http://www.impact-project.eu/](http://www.impact-project.eu/)

**Part 2: Wilson Library Visit:** Nick Graham, Digital NC, [http://digitalnc.org](http://digitalnc.org) & Fred Stipe (Head of CDLA’s Digital Production Center), overview & Scribe operation (large-format studio & Zeutschel/PhaseOne equipment)
#3: February 3

**Part 1: Digitization at scale**

Readings:
- Digitizing Print Collections with the Internet Archives, [http://www.archive.org/scanning](http://www.archive.org/scanning)

Addressing History:
- [http://addressinghistory.blogs.edina.ac.uk/](http://addressinghistory.blogs.edina.ac.uk/)
- [http://addressinghistory.edina.ac.uk/](http://addressinghistory.edina.ac.uk/)
- [http://www.slideshare.net/LIFESHAREProject/addressing-history](http://www.slideshare.net/LIFESHAREProject/addressing-history)

**Part 2: ABBYY**

---

**Unit 2: GIS**

#4: February 10

**Part 1: An Introduction to GIS**

Readings:

**Part 2: Introduction to ArcGIS**

---

#5: February 17

**Part 1: Recreating historical maps**


**Part 2: Hands-on: Georeferencing**
#6: February 24
Part 1: Presentation and critique of map projects
Student presentations

Part 2: Hands-on: Tracing

#7: March 3
Part 1: Map services in the cloud
Readings:
NYPL Labs:
   NYPL Map Warper: http://maps.nypl.org/warper/
Terrahub exercise: http://geocompute3.renci.org/
ESRI:
   Geocommons: http://geocommons.com, Share open data and maps
   GeoIQ: http://www.geioq.com/, “Provide geospatial data management, visualization and analysis platform providing collaborative, browser-based data analysis tools for use by both technical and non-technical users.”
   ArcGIS Online: http://www.esri.com/software/arcgis/arcgisonline
GIS Cloud: http://www.giscloud.com
Urban Mapping: http://www.urbanmapping.com/

Part 2: Hands-on: Geocoding

Unit 3: Text Analysis
March 10: NO CLASS (Spring Break)

#8: March 17
Part 1: Introduction to text analysis
Readings:
Explore the TEI initiative: http://www.tei-c.org/index.xml
“What is TEI”, Scott Hamlin:  
http://www.colleges.org/techcenter/Archives/workshops/xml05/what_is_tei/slideswhat_is_tei-slides0.html


Explore TAPoR tools: http://portal.tapor.ca/portal/portal and http://entry.tapor.ca/

Part 2: Text analysis using OpenRefine and Voyeur

Readings:
The Stone and the Shell (historical questions raised by a quantitative approach to language), by Ted Underwood,  
http://tedunderwood.com/2012/08/14/where-to-start-with-text-mining/


Exercises: https://docs.google.com/document/d/1p1K-MAndX_trL33Zl4of1m3VXkVhkkmkY10G3wWius/edit#

Open Refine: http://openrefine.org/


GREL functions: https://github.com/OpenRefine/OpenRefine/wiki/GREL-Functions

Book: Using OpenRefine

#9: March 24

Part 1: Linked open data

Readings:
Linked Open Data, What is it? http://vimeo.com/36752317

What is Europeana doing with semantic web and linked open data?, in LODLAM blog, June 18, 2013, Linked Open Data, http://www.diglib.org/community/groups/linkeddata/

A. Renear, “Text Encoding”, Chapter 17, A companion to Digital Humanities, 2004,  
http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-3-5&toc.depth=1&toc.id=ss1-3-5&brand=9781405103213_brand

Part 2: Text Analysis

Voyant Tools: http://voyant-tools.org/
 – http://hermeneuti.ca/voyer
 – https://blogs.nd.edu/emoorgan/2014/01/voyant-tools/
 – http://docs.voyant-tools.org/start/

Stylometry:

Patrick Juola:


• Java Graphical Authorship Attribution Program: http://evllabs.com/jgaap/w/index.php/Main_Page


Jose Nilo Binongo:

#10: March 31

**Part 1: Big data**

*Readings:*


JISC Digging into Data, http://diggingintodata.org

**Part 2: Text processing**

---

**Unit 4: Crowdsourcing**

#11: April 7

**Part 1: Crowdsourcing in the cultural realm**

*Readings:*

Some examples (among many...):

- What’s on the Menu, http://menus.nypl.org/
- FamilySearch.org: https://familysearch.org/indexing/


Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research by Stuart Dunn and Mark Hedges from the Centre for e-Research, Department of Digital Humanities King’s College London. http://stuartdunn.files.wordpress.com/2013/04/crowdsourcing-connected-communities.pdf


AND ALSO:


Part 2: Toolkits

**#12: April 14**


Crowd business models (from [http://www.resultsfromcrowds.com](http://www.resultsfromcrowds.com)):

- Crowd Research, [http://crowdresearch.org](http://crowdresearch.org)
- Crowdsourcing For Dummies, David Alan Grier, [http://www.dummies.com/how-to/content/crowdsourcing-for-dummies-cheat-sheet-uk-edition.html](http://www.dummies.com/how-to/content/crowdsourcing-for-dummies-cheat-sheet-uk-edition.html)

Part 2: Hands-on


Richard’s list: software


Amazon Mechanical Turk: [http://www.mturk.com](http://www.mturk.com)

- “Conducting behavioral research on Amazon’s Mechanical Turk”, Winter Mason, Siddarth Suri, in Behav. Res., 2011, [CLICK HERE](http://www.mturk.com)
- “Demographics of Mechanical Turk”, P. Ipeirotis, [CLICK HERE](http://www.mturk.com)

**#13: April 21 (last class)**

**Part 1: Crowdsourcing at scale**

Readings:


Part 2: Hands-on