Empowering Librarians and Users to Mine the HathiTrust Digital Library Resources
The “Digging Deeper, Reaching Further” Project
ASERL Webinar | July 20, 2017
Funded by 3-year IMLS Laura Bush 21st Century Librarian grant award (award #RE-00-15-0112-15)
5 partner institutions, led by the University of Illinois at Urbana-Champaign

http://teach.htrc.illinois.edu
Speakers

- Harriett Green, Head of Scholarly Communication and Publishing, University of Illinois at Urbana-Champaign
- Eleanor Dickson, Visiting HTRC Digital Humanities Specialist, University of Illinois at Urbana-Champaign
The Landscape

ARL SPEC Kit #350: Supporting Digital Scholarship

Studies and reports from Ithaka S+R, Association for Research Libraries (ARL), and Coalition for Networked Information (CNI)


Hartsell-Gundy et al. (eds.), *Digital Humanities in the Library: Challenges and Opportunities for Subject Specialists* (2014)

ACRL forthcoming Digital Scholarship Section (Digital Humanities IG, Digital Curation IG), ADHO Libraries Special Interest Group
http://www.hathitrust.org/htrc

How do we enable our digital collections for research use?

- Jointly led by the University of Illinois at Urbana-Champaign and Indiana University Bloomington
- Facilitates text analysis of HTDL content
- Focus on large-scale, computational research
- User Engagement and Assessment
- Research & Development
  - Finding technical solutions
  - Building tools and services
Findings: User Personas

- **Digital Project Librarian**
  - Wants flexible, transparent tools
  - Role: Research Support staff
  - Challenges: Inaccessible data, matching tool to researcher

- **Faculty Member**
  - Wants computational resources
  - Role: Experienced Researcher
  - Challenges: Collaboration, Finding texts

- **Graduate Student**
  - Wants examples
  - Role: New Researcher
  - Challenges: Understanding stats, choosing areas of interest

Credits: Alex Kinnaman, Peter Organisciak, Eleanor Dickson
Goals of DDRF

- Arm librarians with instructional content and tool skills in digital scholarship and digital humanities;
- Empower librarians to become active research partners on digital projects at their institutions;
- Enable librarians to build foundations for digital scholarship centers and services

http://teach.htrc.illinois.edu
Principal Investigator: Harriett Green, University of Illinois at Urbana-Champaign

University of Illinois: Beth Sandore Namachchivaya, Stephen Downie

Indiana University Bloomington: Angela Courtney

Lafayette College: Neil McElroy, Terese Heidenwolf

Northwestern University: Geoff Morse

University of North Carolina: Amanda Henley (Stewart Varner)
Developing a suite of workshop materials and training resources for training LIS professionals in approaches and tools frequently used in text analysis research

Use the HTDL collections as the cornerstone dataset and HTRC Portal and services as the core toolkit

Teaching pilot workshops at the partner institutions

Final year: National series of workshops
Project so far

- 4 semesters of pilot workshops at partner libraries
- Improve curriculum between each iteration
  - Test, and sometimes drop, components
  - Example: data management slides were added, then reduced
The Curriculum

Lead Project Staff for Curriculum Development: Eleanor Dickson (UIUC), Nicholae Cline (IU), Leanne Nay (IU)
Learning goals

- Bridge concepts, tools, and skills
- Example:

<table>
<thead>
<tr>
<th>Concepts</th>
<th>Tools</th>
<th>Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Textual corpora</td>
<td>HTRC Workset Builder; wget command line tool</td>
<td>Building text datasets; webscraping</td>
</tr>
<tr>
<td>Assorted text analysis methods</td>
<td>HTRC Algorithms; Python libraries</td>
<td>Create topic models; exploratory data analysis</td>
</tr>
</tbody>
</table>

Module 2

Module 4
Learning goals

- Meet variety of experience levels
- Appropriate for different kinds of librarians
- Transferable skills
- Connect skills across specializations
Teaching strategies

- Frame workshop around a research/reference question
- Interactive content
  - Critical thinking activities
  - Technical tasks
  - Guided discussion
Workshop assessment

- Response form administered at end of each workshop
- Goal is to learn about efficacy of content vs quality of instruction
- Comments incorporated into the curriculum
Responses: Year 1

The good
- Access to workshop materials
- Hands-on activities

Areas for improvement
- More concrete examples
- Pare back: Too much content
The good

- Breadth of content
- Good handouts with directions, information, and scripts

Areas for improvement

- Better explanation of technical Python scripts
- More discussion, more critical/creative thinking activities
Responses from Instructors

- Librarians came from many depts.
- More hands-on time / more open-ended activities
- Prepare for issues with computers
- Frame the unfamiliar as familiar - scaffold skills for text mining, with the context of skillsets and knowledge from traditional librarian work
The next iteration

- Revamped curriculum based on attendee and instructor feedback
- Improving the supplementary materials
  - Participant handouts, instructor guide
- Further refinement of activities
  - Incorporate: play, paired-programming, open-ended exploration
Outcomes

- Conduct a nationwide “roadshow” of text mining workshops for librarians and LIS professionals
- Toolkit of open educational resources and curricula for librarians to teach text mining and digital scholarship methods
- Build a national network of librarians teaching and pursuing text mining methods
- Research study on curriculum development for digital scholarship methods and assessment of strategies for LIS professional development

http://teach.htrc.illinois.edu
What’s ahead

- Revising the final DDRF curriculum
- Year 3 national workshop series runs from September 2017 to August 2018
- Final curriculum of OER materials posted by late fall 2018
Upcoming workshops

- Schedule still tentative
- Look for workshops in Atlanta, New Orleans, and Washington, D.C.
- Keep an eye on the website for the schedule!
Building a Community of Practice...

- What infrastructures do libraries need to build or revise for supporting digital scholarship?
- How do we empower librarians to build the skills needed to respond to new research needs?
- How can librarians share skills, materials, and resources for supporting digital scholarship?

- Always Already Computational: Collections as Data
- Data Science in Libraries
- Digging Into Data

D-Lib special issue: http://www.dlib.org/dlib/may17/05contents.html
Thanks again to the Team!

**PI:** Harriett Green, University of Illinois at Urbana-Champaign

**University of Illinois:** Beth Sandore Namachchivaya, Stephen Downie

**Indiana University Bloomington:** Angela Courtney

**Lafayette College:** Neil McElroy, Terese Heidenwolf

**Northwestern University:** Geoff Morse

**University of North Carolina:** Amanda Henley (R. Stewart Varner)

**Curriculum Working Group:** Eleanor Dickson (UIUC), Ruohua Han (UIUC), Leanne Nay (IU), Ewa Zegler-Poleska (Indiana), Nicholae Cline (Indiana)
Thanks to our Funder and Supporters!

The DDRF project was made possible by the Institute for Museum and Library Services, award #RE-00-15-0112-15
Questions?

Harriett Green
Head of Scholarly Communication & Publishing
University of Illinois at Urbana-Champaign

green19@illinois.edu
@greenharr

Eleanor Dickson
HTRC Digital Humanities Specialist
University of Illinois at Urbana-Champaign

dicksone@illinois.edu
@EllieDickson

http://teach.htrc.illinois.edu