YALE DIGITAL HUMANITIES LAB

- Organizationally & physically in main library
- Space, Community, Resources
- Staff:
  - Director
  - Software Developer
  - User Experience Designer
  - Outreach & Engagement Manager
- Internal faculty grants
- Two post-docs per year, working with Yale humanities faculty
DIGITAL HUMANITIES + LIBRARIES

• Libraries hold two unique kinds of “data”
• Can DH methods be useful to libraries in their stewardship of these?
  • Special collections
    • Enhancing archives through user engagement
  • Electronic data licensed from commercial vendors
    • Ensuring robust researcher access to Text & Data Mining
PART 1: SPECIAL COLLECTIONS & DH

ENHANCING ARCHIVES THROUGH USER ENGAGEMENT
SPECIAL COLLECTIONS & DIGITAL HUMANITIES
Enhancing archives through user engagement

- Transcription
  - Text Mining
  - Handwriting Recognition
- Semantic Relationship Extraction
  - Linked Data
TRANSCRIPTION

1820s
Pages to Words.

Help transcribe Native language manuscripts to build a corpus of text that is legible, searchable, and open to support education and research.

Start Transcribing
Kilpatrick Collection of Cherokee Manuscripts

The Kilpatrick Collection of Cherokee Manuscripts, housed at Yale University's Beinecke Rare Book and Manuscript Library, consists of material created and accumulated by Jack Kilpatrick and Anna Gritts Kilpatrick, dating from the 1890s to the 1960s. The material, entirely in the Cherokee syllabary, documents vernacular literacy in the Cherokee language, the practice of traditional medicine, social aspects of Christian religion and church organizations, dates and circumstances of death, funerary practices, and other topics relating to the history and culture of the Oklahoma Cherokee in the late 19th and early 20th centuries.

The Kilpatrick Collection is cataloged as WA MSS S-2707, and can be requested for research use at the Beinecke Rare Book and Manuscript Library, Yale University. To see a full description and list of contents, view the catalog record in Yale's online catalog, Orbis.
List of food products and prices
Kilpatrick Collection of Cherokee Manuscripts
Box 6, Folder 508

<table>
<thead>
<tr>
<th>KG</th>
<th>ĆGYW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1915</td>
<td>$100</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>AT</td>
<td>50</td>
</tr>
<tr>
<td>T&amp;T</td>
<td>$150</td>
</tr>
<tr>
<td>DSWZ</td>
<td>.ąąą́́́</td>
</tr>
<tr>
<td>$500</td>
<td>50</td>
</tr>
<tr>
<td>Shb</td>
<td>KG</td>
</tr>
<tr>
<td>DSW</td>
<td>hšł</td>
</tr>
<tr>
<td>ĆBD T&amp;T hšřəłčř caveat hšł ĆROT</td>
<td></td>
</tr>
<tr>
<td>ĆAW</td>
<td>$90</td>
</tr>
<tr>
<td>T&amp;T</td>
<td>hřD</td>
</tr>
<tr>
<td>SAWOW</td>
<td>8</td>
</tr>
<tr>
<td>KG</td>
<td>ĆGYW</td>
</tr>
<tr>
<td>SG?EGJ</td>
<td>SC??</td>
</tr>
</tbody>
</table>
TRANSCRIPTION CAN ENABLE...

• Text Mining
  • Machine-actionable corpus of Cherokee
  • Corpus Linguistic approaches

• Handwriting Recognition
  • Matching pixels of drawn glyph to unicode code point
  • Provides “ground truth”
  • Train neural network to recognize glyphs
SEMANTIC RELATIONSHIP EXTRACTION
CITIZEN SCIENCE MEETS THE HUMANITIES
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam eget porttitor felis. Proin sollicitudin turpis at tortor maximus, quis egestas odio gravida. Phasellus nec scelerisque est. Curabitur gravida felis sed. (30 words)

Performance Dates

2002 2016

Select Institution

Select Playwright

Overall completion: 98%

444 88
In-Progress Completed

Yale UNIVERSITY LIBRARY
Digital Humanities Laboratory
twelfth night

by William Shakespeare

Presented by Yale School of Drama

Experimental Theatre

October 23 - 26, 8 pm
October 27, 2 pm & 8 pm
The Department of Drama
of
Yale University
presents

Moliere's Comedy

School for Wives

In an Apartment in Rome

Arthur: Gustavo Franchini

stage by
Robert Hartung

A Drama 140 Production
Experimental Theatre

November 25 and 27, 1942

CAST
(In order of appearance)

Sganarelle
Ergaste
Lisette
Ariste
Valere
Leonor
Isabelle
First Lackey
Second Lackey
Pierrot
Third Lackey
Magistrate
Notary
First Flunkey
Second Flunkey

Robert Stone
Louis Migliorini
Madeleine Thornton
George Latshaw
Bernard Malak
Mary Walters
Ruth Handel
Joe Driscoll
Edward Barber
William Eboro
Robert Seymour
John Fletcher
William Shoidy
Edward Butterfield
Robert Monahan

Act I
YALE UNIVERSITY · SCHOOL OF THE FINE ARTS
The Department of Drama Presents

The King of Nowhere
BY JAMES BRIDIE
DIRECTED BY CONSTANCE WELCH

The Characters

Mrs. Vivaldi · Mary Jane Chiles
Dr. Wamstead · Norman Holland
Dr. McGilp · Lawrence Dobkin
Sam · George Penny
Nurse Appleby · Mary Dickson
Captain Buller · Arthur Sachs
Annie · Betty Lou Evans
Miss Slimmon · Barbara Regola
Lovers · George Wood
Old Man · Robert Stone
Old Lady · Jean Ramsay
Eri-Man · George Schaere
Muller · Donald Muller
Professor · Charles Baker
Attendants · Archibald Douglas, Jr.
Donald Muller
William Laubenstein
Myrmidons · Jack McCusker
Donald Muller
Archibald Douglas, Jr.
Soldiers and Townpeople

Type the name of the actor and the name of the character. For roles with multiple actors and only one character name, draw a box around each actor.

Character

Mrs. Vivaldi

Character Description (Optional)

Actor

Mary Jane Chiles

Bad region
Illegible?
Next Entry
LINKED DATA: WHO ACTED IN WHICH PLAYS?
PART 2: VENDOR DATA
ENSURING ROBUST RESEARCHER ACCESS TO TEXT & DATA MINING
LICENSED VENDOR DATA

• Opportunities
  • Pre-Digitized
  • Markup and/or Metadata Often Present

• Challenges
  • Copyright
  • Licensing

• Vogue Archive as Case Study
  • Text Mining
  • Image Analysis
PROQUEST VOGUE ARCHIVE

• Vogue Magazine
  • 1892-present
  • 2,798 issues
  • 400,000 pages

• ProQuest Raw Data:
  • 6 terabytes
    • JPEG files
    • XML files w/ article segmentation
MEETING THE COPYRIGHT + LICENSING CHALLENGE
A model based on work with Vogue

- Yale Library servers **do store:**
  - Metadata, Citations, Permalinks
  - Dimensionality Reductions (term frequency counts, color extraction etc)

- Yale Library servers **do not store:**
  - Full text
  - Whole images

- Access to full data is based on license
  - Permalinks resolve for any institutional subscriber to ProQuest’s product
VENDOR DATA: TEXT MINING

• Looking for something you think is there (domain expertise)
  • N-gram search

• Letting data organize itself (machine learning)
  • Topic models
TEXT MINING: N-GRAM SEARCH
Looking for something you think is there
TEXT MINING: TOPIC MODELING

Letting data organize itself
IMAGE ANALYSIS
IMAGE ANALYSIS

Vogue Covers 1901-2011
Cut into 277,500 slices
Technique: Damon Crockett / UCSD Software Studies
X: Hue  Y: Brightness
IMAGE ANALYSIS 1901

X: Saturation  Y: Brightness
CONCLUSION
Digital Humanities IN and OF the Library

• What opportunities are there?
  • Special & Rare Collections
  • Vendor-Digitized Material

• How can DH fulfil library objectives?
  • Enhancing archives through user engagement
  • Securing the rights of scholars to TDM on licensed digital materials