Agenda & outcomes

- Introductions
- Overview of the Digital Scholar Lab
- Workflow highlights
- Supporting faculty and students
- Teaching with the DS Lab
- Digital Humanities Specialist Role
- Questions
The Digital Scholar Lab: Removing Key Barriers to Entry
Research tools vary by levels of complexity, usability, access, and support.

This is often a barrier to entry for students new to digital humanities.
THE DIGITAL SCHOLAR LAB

TDM RESEARCH ENVIRONMENT

- Cloud-based Digital Humanities platform
- Access to a broad range of texts from Gale primary source collections
- Construct custom content sets across Gale’s collections
- Access to powerful text mining tools
- Text Cleaning
- Organization of research
- Integrated help and instructional materials
- Export content sets, statistical data, and visualizations in standard formats
The Digital Scholar Lab: Removing Key Barriers to Entry

Access to relevant data in an optimised format

Somewhere to host that data

Familiar, Powerful tools
Parts of Speech uses natural language processing of syntax to recognize, and tag parts of speech. In this implementation of Parts of Speech Tagger you may review how authors use of speech varies over time.

Open Source developer: spaCy

Named Entity Recognition (NER) recognizes and extracts proper and common nouns from documents using a Parts of Speech tagging method, and outputs them as lists of grouped by entity “type”. Some “entity types” available for extraction are: people (including fictional); groups (nationalities, religious, or political); organizations; locations, products, works of art, dates, among others. This implementation uses spaCy’s Named Entity Recognition model. Learn more here>

An Ngram is a term, or collocation of terms, found in your content set. You set the range or number of terms (‘N’) you wish to consider in your analysis. Then, the frequency of those Ngrams is counted and displayed for analysis.

Ngram examples:
N=1: Unigram “a”, “the”, “turtle”, “frankenstein”
N=2: Bigram “on the”, “turtle dove”, “mary shelley”

Learn more here>

Sentiment analysis determines a tally of the positive or negative words within each document of a content set. It uses the AFINN lexicon (dictionary of words and their sentiment value) to compile sentiment scores for each phrase, which are then compiled to produce a document-level sentiment value. By establishing polarity within the texts (i.e. positive/negative word association), this tool can classify the documents in your content set between positive to negative sentiment.

Clustering analyzes the documents from a content set using statistical measures and methods to group them around particular features or attributes. This implementation of clustering leverages the k-means algorithm to create clusters of documents according to similar words contained within each document of your content set.

Open Source developer: Scikit-Learn

Learn more here>

Topic modelling allows users to analyze a large corpus of unstructured (OCR) text. A “topic,” often referred to as a “bag of words,” is a collection of terms that frequently co-occur in your collection of documents. Mallet uses Latent Dirichlet allocation (LDA) models to extract contextual clues in order to connect words with similar meanings, as well as differentiate between words that are spelled similarly but have differing meanings.

Learn more here>
Digital Scholar Lab Workflow Highlights
Digital Scholar Lab Highlights

Digital Scholar Lab
Explore your institution's holdings from Gale Primary Sources using Digital Humanities tools. Rediscover and interpret the past through analysis and visualization of historical texts.

Build your corpus
Search across your institution's Gale Primary Sources holdings and build content sets for use in your research.

Analyze your documents
Analyze your corpus of documents with our visualization tools, experiment and tune your results.

Manage and share
Manage your documents and analyses through Content Sets, export documents and download visualizations.

Explore millions of primary sources
Discover our analysis tools
Manage your Content Sets
Search: Personalized Archive Building

All Content (578)

Search Terms: Keyword (guernica) LIMITS: Content Type ("Newspaper") AND Document Type ("Article") AND Publication Date (> 1936)

Sort by: Relevance

Limit Search by:

- Archives
  - The Times Digital Archive (160)
  - The Telegraph Historical News (117)
  - British Library News (89)
- Content Type
  - Newspaper (578)
- Document Type
  - Article (578)
  - Travel narrative (1)
- Publication Title
  - Times (160)

1. **German guilt over Guernica**
   - BY Reuters
   - OCR CONFIDENCE 61%
   - German government Guernica German guilt over Guernica Germany, in a message from President Roman Herzog, acknowledged the guilt of its schools in the destruction of the Spanish town of Guernica 60 years ago, but stopped...

2. **Bombing Of Guernica**
   - BY From Our Special Correspondent
   - OCR CONFIDENCE 96%
   - Imperial and Foreign Imperial Foreign BOMBING OF GUERNICA FRESH EVIDENCE BOMBING OF GUERNICA FRESH EVIDENCE From Our Special Correspondent BILBAO, MAY 5 Ths statement issued from Salamanca that Guernica was destroyed...

3. **Britain and Bombing of Guernica**
   - OCR CONFIDENCE 96%
   - Britain and Bombing of Guernica Britain and Bombing of Guernica A note from Senor Del Vayo, the Spanish Foreign Minister, to the British Government regarding the destruction of Guernica is being considered by the British Government.

4. **The Independent Digital Archive**
   - Source Library: The Independent
   - Content Type: Newspaper
   - Document Type: Article
   - Publication Date: April 28, 1997
   - Publisher: The Independent

5. **Times Newspapers Limited**
   - Source Library: Times Newspapers Limited
   - Content Type: Newspaper
   - Document Type: Article
   - Publication Date: May 6, 1937
   - Publisher: The Times

6. **British Library Newspapers**
   - Source Library: British Library
   - Content Type: Newspaper
   - Document Type: Article

NEXT
Text Cleaning

CLEAN

Cleaning Configurations are used for:
1. Configuring analysis tools to improve results in Analyze
2. Downloading content sets in My Content Sets

NEW CONFIGURATION

YOUR CONFIGURATIONS

Default Cleaning Configuration
- feminism
- Guernica cleaning config
- Remove special characters and spacing
- Parts of Speech config
- Heavy cleaning

CLEANING CONFIGURATION

Default Cleaning Configuration

This is Gale’s default cleaning configuration. Changes to this configuration can only be saved as a new configuration.

CONTENT SET DOWNLOAD READY

TEXT CORRECTION
- All lowercase

CHARACTERS
- Remove all extended ASCII characters
- Remove all number characters

SPECIAL CHARACTERS
- Remove all special characters
- Set specific special characters to remove

PUNCTUATION
- Remove all punctuation
- Set specific punctuation to remove

SPACING
- Remove all tabs
- Remove all line breaks
- Reduce multiple spaces to one space (ex. “hello there” becomes “hello there”)

REPLACEMENTS

<table>
<thead>
<tr>
<th>Replace this</th>
<th>With this</th>
</tr>
</thead>
</table>

STOP WORDS

Choose a Starter List Clear All
- a
- about
- above
- across
- after
- afterwards
- again
- against
- all
- almost
- alone
- along
- already
- also
- although
- always
- am
- among
- amongst
- amount
- an
- and
- another
- any
- anyhow
- anyone
- anything
- anyway
Analyze: Project Building

CONTENT SET
Guernica in the News

Get started by adding an Analysis Tool

ADD TOOL
Visualizing Your Result

- **Ngram**
  - TOOL SETUP
    - revised stop words (Thu May 06 14:06:35 EDT)
  - RUN STATUS: **COMPLETED**
  - RESULTS: Word Cloud, Bar Chart

- **Parts of Speech Tagger**
  - TOOL SETUP
    - Unnamed
  - RUN STATUS: **READY TO RUN**
  - EXPECTED RESULTS: Searched View

- **Sentiment Analysis**
  - TOOL SETUP
    - New cleaning config (Thu May 02 19:15:49 EC)
  - RUN STATUS: **COMPLETED**
  - RESULTS: Time Series

- **Topic Modelling**
  - TOOL SETUP
    - Guernica (new stop words) (Tue Apr 02 19:12)
  - RUN STATUS: **COMPLETED**
  - RESULTS: Topics, Topic Proportion
My Content Sets: Research Organization

CONTENT SETS

- Catalan Separatism  Edit  Copy
  This project attempts to situate the Catalan separatist movement within a historical context.
  QUICK STATS  Documents: 19  Archives Used: 8

- Family Planning  Edit  Copy

- Feminism in the News  Edit  Copy
  How are feminists depicted in newspapers? What topics are generally discussed by feminists in newspaper articles? What is the language used by newspaper staff when reporting on feminists and their perspectives?

- Guernica during Francoism (April 1, 1939- Nov 20, 1975)  Edit  Copy
  This project deals with the representations of the bombing of the Basque town of Guernica (April 26, 1937) during the Spanish Civil War (1936-1939). This content set focuses on newspaper articles published during the Franco dictatorship which spanned from 1930-1975.
  QUICK STATS  Documents: 184  Archives Used: 11  Timeframe: 1965 - 1975

- Guernica in International Mass Media  Edit  Copy
  This project deals with the representations of the bombing of the Basque town of Guernica (April 26, 1937) during the Spanish Civil War (1936-1939). The content set consists of the text of Francoist España's official news and propaganda to other countries.
Using the Digital Scholar Lab to support faculty & students
Demystifying Digital Collections Content

American Historical Periodicals from the American Antiquarian Society
More than 200 years ago, the American Antiquarian Society was founded and has been recognized as a major source for U.S. primary sources published. Now, Gale and the Society have entered an exclusive partnership to bring digitized periodical holdings to life.

Archives of Sexuality & Gender
Archives of Sexuality & Gender, the largest collection available in support of the study of gender and sexuality, enables scholars to make new connections in LGBTQ+ history and activism, cultural studies, psychology, health, political science, policy studies, and other related areas of research.

Archives Unbound
Since its inception in 2009 the Archives Unbound program has published more than 230 titles. The roots of the program are in microfilm, and the collection makes targeted collections of interest available to scholars engaged in serious research.

Associated Press Collection Online
For more than 170 years, the Associated Press (AP) has delivered the news when and where it occurs. Together, the AP and Gale, a Cengage company, are making previously unseen news copy and footage available digitally to tell the stories of our past in vivid detail.

Brazilian and Portuguese History and Culture
Originating the personal library of the Brazilian diplomat, historian, and journalist Manoel de Oliveira Lima has long been regarded as one of the finest collections of the Luso-Brazilian materials available to scholars. It’s now accessible for students, educators, and researchers alike to delve deeper into Brazilian and Portuguese history and culture from the sixteenth to the twentieth century.

British Literary Manuscripts Online
This unique collection provides an intimate look into the lives and works of more than one thousand authors and delivers insights into the culture and context surrounding centuries of British literary achievement.
Key DSL Value Points

**CONTENT VALUE**
- Create a dataset – solves largest obstacle in the DH workflow
- Provides organization and indexing that TDM drives do not.
- Export content sets for use in open-source tools

**TOOL VALUE**
- Proximity to dataset – bringing content and tools together is especially valuable in a classroom setting where students access the same content.
- Organization – à la carte approach allows researchers to tailor their analysis to a specific project or content set and for specific outcomes.

**INSTRUCTIONAL VALUE**
- DH pedagogy and teaching is scattered. Aligning content and tools through curriculum and use-cases provides value to both students and faculty.
THE DIGITAL SCHOLAR LAB

FEATURES TO SUPPORT CLASSROOM USE

- Cloud-based Digital Humanities platform.
- Access to a broad range of texts from Gale primary source collections
- Construct custom content sets across Gale’s collections
- Access to powerful text mining tools
- Text Cleaning integrated in platform
- Organization of research
- Integrated help and instructional materials
- Export content sets, statistical data, and visualizations in standard formats
EXEMPLARY CLASSROOM USE CASES

- Focus on a specific text mining tool or method
- Answer a specific research question
- Curate a meaningful content set
- Clean a content set
- Build meaningful visualizations and interpret output
- Understand the OCR process
- Learn about metadata standards
- Distant and close reading of texts
Teaching with the Digital Scholar Lab: Instructional Use Cases
Fall 2018
Introduction to Digital Humanities
Declared Majors

35 students from 21 departments across campus
• Goals: foster core computing competencies for humanities and social science students

• Use digital tools to investigate humanities data

• Promote interdisciplinarity, collaboration
## FALL QUARTER 2018 SYLLABUS - INFO 498 C: AN INTRODUCTION TO DIGITAL HUMANITIES

<table>
<thead>
<tr>
<th>DATE</th>
<th>THEME</th>
<th>TOPIC</th>
<th>NOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thursday September 27</td>
<td></td>
<td>Welcome! Syllabus Review</td>
<td></td>
</tr>
<tr>
<td>Tuesday October 2</td>
<td></td>
<td>What is a Digital Project? Evaluating Digital Projects</td>
<td></td>
</tr>
<tr>
<td>Thursday October 4</td>
<td></td>
<td>Copyright &amp; Open Source Material</td>
<td></td>
</tr>
<tr>
<td>Tuesday October 9</td>
<td>PLANNING</td>
<td>Final Presentation Planning &amp; Omeka Overview</td>
<td></td>
</tr>
<tr>
<td>Thursday October 11</td>
<td></td>
<td>Digital Scholar Lab: Intro &amp; Demo</td>
<td>Helene Williams</td>
</tr>
<tr>
<td>Tuesday October 16</td>
<td></td>
<td>LAB: Project Planning and Collaborative Working</td>
<td></td>
</tr>
<tr>
<td>Thursday October 18</td>
<td></td>
<td>Risk Assessment &amp; Data Management Plans</td>
<td></td>
</tr>
<tr>
<td>Tuesday October 23</td>
<td></td>
<td>Metadata Lab</td>
<td></td>
</tr>
<tr>
<td>Thursday October 25</td>
<td>COLLECTING &amp; CURATING</td>
<td>Creating a Digital Archive; What is OCR?</td>
<td>Guest speaker</td>
</tr>
<tr>
<td>Tuesday October 20</td>
<td></td>
<td>Corpus Creation &amp; Curation; Cleaning Data with OpenRefine</td>
<td></td>
</tr>
<tr>
<td>Thursday November 1</td>
<td></td>
<td>Text Cleaning with RegEx, Lexos &amp; Digital Scholar Lab</td>
<td></td>
</tr>
<tr>
<td>Tuesday November 6</td>
<td></td>
<td>Libraries &amp; Digital Scholarship; Text cleaning in Hathi Trust</td>
<td>Helene Williams</td>
</tr>
<tr>
<td>Thursday November 8</td>
<td>WORKING WITH DATA</td>
<td>LAB: Cleaning Content Sets</td>
<td></td>
</tr>
<tr>
<td>Tuesday November 13</td>
<td></td>
<td>Text &amp; Corpus Analysis with Voyant &amp; Digital Scholar Lab</td>
<td></td>
</tr>
<tr>
<td>Thursday November 15</td>
<td></td>
<td>XML &amp; TEI</td>
<td></td>
</tr>
<tr>
<td>Tuesday November 20</td>
<td></td>
<td>Geoparsing with Cleefin, Mapping</td>
<td></td>
</tr>
<tr>
<td>Thursday November 22</td>
<td></td>
<td>THANKSGIVING HOLIDAY: NO CLASS</td>
<td></td>
</tr>
<tr>
<td>Tuesday November 27</td>
<td></td>
<td>Named Entity Recognition, Sentiment Analysis &amp; Ngrams</td>
<td></td>
</tr>
<tr>
<td>Thursday November 29</td>
<td></td>
<td>LAB WORK</td>
<td></td>
</tr>
<tr>
<td>Tuesday December 4</td>
<td>DIGITAL SHOWCASE</td>
<td>Final Group Presentations &amp; Digital Showcase</td>
<td></td>
</tr>
<tr>
<td>Thursday December 6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Class Research Focus: Historical Menus

Based on

NYPL Labs 🎉 What's on the menu? 🎉

Est. 2011
Menu Data in the Digital Scholar Lab

Chop Suey, Garlic and Rice
By Patricia Gallagher

Document Text

Centre Cuisine (Chop Suey, Garlic and Rice) p By Patricia Gallagher pARIS, Nov. 8 — If you didn’t know that dogs like salad, need green beans, and quite often have a taste for chop-suey, your pet is not receiving the right nourishment, but he may be misunderstood. At the 18th annual dinner of caviar-medicina (doctor-a-la-caviar), in Paris last week, veterinary surgeon Philippe de Wally of Boulogne announced that response to his recipe for chop-suey has been staggering, especially among Yorkshire terriers, one of the most difficult to breed in matters of cuisine. The reason, he claims, is simple: The "pinch" of aji: no-moto contains a high per cent of gluteamic acid, a scientifically proven stimulant to canine appetites. The recipe is included in Dr. de Wally’s latest book "The Art of Living With Your Dog." Not all of the menu suggestions in "The Art of Living With Your Dog" are as exotic as chop-suey, nor is Dr. de Wally’s book devoted entirely to cuisine. However, the time involved in preparing even simple fare would seem a minor boggling impossibility to dog owners in the canned food community. Raw ground meat, rice, a green vegetable and 200 grams y...
GROUP 1
THANKSGIVING ACROSS THE YEARS

GROUP 2
FOOD PATTERNS IN BOOM & BUST

GROUP 3
THE TERRACE ROOM IN WWII

GROUP 4
BY THE GLASS

GROUP 5
THE JOURNEY OF EGGS BENEDICT

GROUP 6
EVOLVING ATTITUDES TOWARD CHINESE FOOD IN THE U.S.

GROUP 7
VEGETARIANISM IN WWI

Intro to DH Student Group Projects
Creating & Curating Content Sets

“I think the conclusion I reached is that hand curating a dataset instead of just taking the top of search results will make for a much more compelling analysis.”
Text Cleaning

“Using these tools really helped me understand how necessary it is to clean up your dataset, because if it’s uncleaned, then these analysis tools and visualizations will not be as effectively meaningful and the analysis itself is inherently weaker compared to if the dataset was more refined.”
Term Frequency Analysis:

Bar chart and word cloud visualizations

This visualization is a cirrus/word cloud derived from datasets collected from Digital Scholar Lab. The terms appearing on this visualization are the most used terms in our dataset, with font size representing which are used more than the others. Being a collection of vegetarian dishes, we can see that the most used term is "vegetables", which we believe is the centerpiece of many of the dishes. "Meat" is also frequently used in our data, which we observe is most likely due to the fact that vegetarian (meat alternative) dishes are offered simultaneously with meat dishes, so that guests with different food preferences can enjoy their meals at the restaurants. It is also displayed in the visualization that many vegetarian dishes in this time feature "potatoes", "butter", "soup", "bread", and "pudding", and that many of the dishes were served in the UK. "War" is also a common term from these menus, as it was wartime and many of these menus were found in British and American newspapers.

This bar chart of term frequencies reaffirms the trends discussed above, and it is interesting to note that potatoes and beans are the most commonly mentioned vegetables across the 34 menus from the content set used to create this analysis. Even though these menus were pulled from Digital Scholar Lab, it upholds the commonality of beans and potatoes as popular vegetarian options that were seen in the New York Public Library menus displayed on our Gallery page.
“The sentiment analysis was by far the most useful because it really captured what happened in 1934. However, we learn that as people got used to the conditions of the Great Depression, they start being more positive again even in the midst of the time period.”
Exporting texts and data to use with external tools
Required components for each digital project:

• Research question and engaging narrative.

• Project Planning Documentation: One-pager, Group Project Charter, weekly worklogs.

• Data Management Plan.

• Standardized and consistent Dublin Core metadata.

• Data Visualizations including maps, term frequency analysis, sentiment analysis, topic modeling, timeline and/or story map.

• ‘About the Project’ and the team.
<table>
<thead>
<tr>
<th>MODULE</th>
<th>THEME</th>
<th>TOPIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SETTING THE SCENE</td>
<td>Welcome!</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Syllabus review, assessment, expectations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>What is digital humanities?</td>
</tr>
<tr>
<td>2</td>
<td>PLANNING</td>
<td>Evaluating Digital Projects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Copyright &amp; Open Source Material</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Risk Assessment &amp; Data Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Digital Scholar Lab: Intro &amp; Demo</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Introduction to the Primary Source archives &amp; topics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Building a narrative with StorymapJS</td>
</tr>
<tr>
<td>3</td>
<td>COLLECTING, CURATING &amp; PREPARING DATA FOR ANALYSIS</td>
<td>Creating a Digital Archive; What is OCR?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Effective Searching</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Creating and Curating Content Sets</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cleaning text in the Digital Scholar Lab</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exporting OCR text</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Other tools for cleaning text: OpenRefine, Lexos, RegEx</td>
</tr>
<tr>
<td>4</td>
<td>ANALYSIS AND VISUALIZATION</td>
<td>Survey of text mining tools</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ngrams/term frequency</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Named Entity Recognition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Topic Modeling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sentiment Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clustering and Parts of Speech Tagging</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Compare with external tools: Voyant</td>
</tr>
<tr>
<td>5</td>
<td>PUTTING IT ALL TOGETHER: FINAL PROJECTS</td>
<td>Exporting analyses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mapping</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Developing a compelling narrative</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hands-on work and support</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Final project URLs!</td>
</tr>
</tbody>
</table>
Digital Scholar Lab Implementation

The Digital Scholar Lab Implementation tool is based on the open-source spaCy model. When you run the tool, it will parse (work through) your context set, and identify words classified as 'entities' by the model, which has been trained using the OntoNotes 5e corpus. For your purposes, here is the list of entities which will be recognized, along with their abbreviations:

- **PERSON**: People, including fictional.
- **NORP**: Nationalities or religions or political groups.
- **ORG**: Organizations, including companies, agencies, institutions, etc.
- **GPE**: Countries, cities, states.
- **LOC**: Non-GPE locations, mountain ranges, etc.
- **PRODUCT**: Objects, vehicles, foods, etc. (Web services)

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.

EXEMPLARY PROJECTS USING NAMED ENTITY RECOGNITION:

- Exemplar projects show named entities recognized correctly, in the corresponding place locations in text. Often, this process is used as the basis for knowledgequisition, or knowledge harvest.
Student Analysis Output Examples
Introduction to Digital Humanities Summer 2019

A no-prerequisite online survey course introducing students to the concepts and methodologies of working with, and analyzing primary source texts using digital tools.

The topics chosen by students were wide-ranging, and include: sea serpents, Hitler in the Press, the Chicago World’s Fair, organic food and agriculture, Iraq in European newspaper editorials (1991-2013), comparisons of author style, Civil Rights events, the Battle of Bull Run as reported in North and South newspapers, and sentiments about Kim Jong-Un. With student permission, projects are displayed below, each one divided by a separator.
DSL Workflow:
This tool simplifies so much of the DH process because it has features that cover every point of the DH lifecycle. My search of primary sources went very quickly when I found the correct limiters needed to create a content set that would address my research question. Adding documents to the sets was very easy and with the right naming convention, working with them was simple as well. It got to the point where I could quickly search, create a new content set, add the files I needed, and search again in under 30 seconds.
The cleaning tool was easily configurable and I could switch between tabs running analyses and adjusting the configuration until I got it just right. Once you find out where the controls are, it is pretty easy to knock out in an hour what would usually take a day or two of work. Once my text was clean, running analysis was just a few clicks away, and I could run multiple analyses at the same time so I never had to wait.
The n-gram analysis provided the best exploratory analysis. I could refine my stop-word list, identify outlier terms, and get an overview of the content sets. The clustering tool fascinated me, however, I had to leave that alone due to our time limitations. Given proper time and preparation, I want to see how k-means clustering could reveal unseen patterns. Sentiment analysis is my jam for this project. This will be front and center in my final project, and it was great to see how cleaning configurations changed the sentiment curves. I spent considerable time preparing the data to not just show what I wanted, but show the best representation of sentiment. The named entity recognition and parts of speech tagger were both cool to mess with, but were not as relevant to my research question.

Student Feedback, August 2019
Gale DH Specialist Support

- Support digital scholarship on campus
- Trainings
- Product Q&A
- Direct feedback loop to Gale
- Consultation on DH projects
- DH-centric workshops
<table>
<thead>
<tr>
<th>Available Archives in the Digital Scholar Lab</th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ 17th &amp; 18th Century Burney Collection</td>
</tr>
<tr>
<td>▪ 17th &amp; 18th Century Nichols Collection</td>
</tr>
<tr>
<td>▪ 19th Century UK Periodicals</td>
</tr>
<tr>
<td>▪ American Amateur Newspapers</td>
</tr>
<tr>
<td>▪ American Civil Liberties Union Papers, 1912-1990</td>
</tr>
<tr>
<td>▪ American Fiction</td>
</tr>
<tr>
<td>▪ Archives Unbound</td>
</tr>
<tr>
<td>▪ Archives of Sexuality &amp; Gender</td>
</tr>
<tr>
<td>▪ Associated Press Collections Online</td>
</tr>
<tr>
<td>▪ Brazilian and Portuguese History and Culture</td>
</tr>
<tr>
<td>▪ British Library Newspapers</td>
</tr>
<tr>
<td>▪ China from Empire to Republic</td>
</tr>
<tr>
<td>▪ Crime, Punishment, and Popular culture 1790-1920</td>
</tr>
<tr>
<td>▪ Daily Mail Historical Archive, 1896-2004</td>
</tr>
<tr>
<td>▪ The Economist Historical Archive, 1843-2014</td>
</tr>
<tr>
<td>▪ The Illustrated London News Historical Archive 1842-2003</td>
</tr>
<tr>
<td>▪ The Independent Digital Archive</td>
</tr>
<tr>
<td>▪ Eighteenth Century Collections Online</td>
</tr>
<tr>
<td>▪ Nineteenth Century Collections Online</td>
</tr>
<tr>
<td>▪ Indigenous Peoples of North America</td>
</tr>
<tr>
<td>▪ International Herald Tribune Historical Archive 1887-2013</td>
</tr>
<tr>
<td>▪ Liberty Magazine Historical Archive 1924-1950</td>
</tr>
<tr>
<td>▪ The Listener Historical Archive 1929-1991</td>
</tr>
<tr>
<td>▪ Making of the Modern Law: Foreign Primary Sources</td>
</tr>
<tr>
<td>▪ Making of the Modern Law: Foreign, Comparative and International Law</td>
</tr>
<tr>
<td>▪ Making of the Modern Law: Legal Treatises</td>
</tr>
<tr>
<td>▪ Making of the Modern Law: Primary Sources</td>
</tr>
<tr>
<td>▪ Making of the Modern Law: Trials</td>
</tr>
<tr>
<td>▪ Making of the Modern World</td>
</tr>
<tr>
<td>▪ Nineteenth Century U.S. Newspapers</td>
</tr>
<tr>
<td>▪ Picture Post Historical Archive</td>
</tr>
<tr>
<td>▪ Political Extremism &amp; Radicalism</td>
</tr>
<tr>
<td>▪ Punch Historical Archive, 1841-1992</td>
</tr>
<tr>
<td>▪ Sabin Americana, 1500-1926</td>
</tr>
<tr>
<td>▪ Smithsonian Collections Online</td>
</tr>
<tr>
<td>▪ The Sunday Times Digital Archive</td>
</tr>
<tr>
<td>▪ The Telegraph Historical Archive</td>
</tr>
<tr>
<td>▪ The Times Digital Archive</td>
</tr>
<tr>
<td>▪ Times Literary Supplement Historical Archive</td>
</tr>
<tr>
<td>▪ U.S. Declassified Documents Online</td>
</tr>
<tr>
<td>▪ U.S. Supreme Court Records and Briefs</td>
</tr>
<tr>
<td>▪ Women’s Studies Archive</td>
</tr>
</tbody>
</table>
What will *not* be available in the Digital Scholar Lab in September

- British Literary Manuscripts Online
- Chatham House Archive
- Early Arabic Printed Books
- Financial Times Historical Archive
- National Geographic Magazine Archive
- State Papers Online
- Slavery & Anti-Slavery: A Transnational Archive
- World Scholar: Latin America & the Caribbean

- OCR confidence levels for manuscript content are historically low – not just at Gale, but across most academic publishers and institutions.
- Hand-written texts are much more difficult to process through currently available tools…but we’re working on it!
- Some content providers/publishers do not grant TDM rights at this time…but we’re working on this, too!
Thank you!

Questions?