

# The Genealogy of Texts and Ideas

Looking Back and Forth through Early English Books Online

## Workshop Guide

A Digital Humanities Start Up Project funded by:



NATIONAL ENDOWMENT FOR THE  
**Humanities**  
FIFTY YEARS

**A brainstorming workshop hosted by Rice University's Humanities Research Center to envision a DTA tool so intuitive that any humanities researcher can take a "fishing expedition" through the EEBO corpus with very little effort.**

February 8 to 10, 2017 in Houston, Texas.



**Computational analysis of texts is a paradigm shift in our ability to interpret and understand written history.**

---

# Workshop Goals

1. Publish a general description of the current DTA tool environment.
2. Develop a list of DTA techniques that humanities researchers want.
3. Outline parameters for an *intuitive* DTA tool, tailored to the EEBO corpus:
  1. Data preparation and structure (POS/Variants/Geo-Tagging, etc)
  2. Analytical Techniques (Ngram/Frequency/Sentiment/Network, etc.)
  3. Visualization (Network Mapping/Block Graph/Line Graph, etc.).
4. Consider package adaptations for R and Python that will help satisfy the demands/desires we have identified .
5. Plan next steps towards the realization of the intuitive EEBO DTA tool.

Our **BHAG** is to create a DTA tool so intuitive that any humanities researcher can take a "fishing expedition" through the EEBO corpus with very little effort.

# Competitions & Prizes

Token prizes will be awarded for the following:

- ❖ Best name for the new BHAG tool
- ❖ Best pithy tag line for our project
- ❖ Best suggested new tag(s) for enhancing the EEBO corpus
- ❖ Best visualization idea or concept
- ❖ Best freehand sketch (must be on sketchbook paper)
- ❖ Best colored OWL
- ❖ Most congenial participant
- ❖ Best emulation of an emoji on pillow 

Participants will determine winners by vote or consensus.

# Workshop Topics (to be prioritized)

- ❖ Innovations or novelty, word clouds and other useless objects
- ❖ The NEH and the Future
- ❖ Spelling Variation in Early Modern English
- ❖ Current Online DTA tools, Bookworm, CPQweb, and others
- ❖ What Humanities Scholars want to Know
- ❖ Sentiment Analysis
- ❖ The TEI XML Schema - Tags for Everything
- ❖ Copyrights and Wrongs - Understanding the Legal Implications of DTA
- ❖ EEBO and Metadata - Sorting things Out
- ❖ Practicality - Developing an Intuitive User Interface
- ❖ Bookworm, Ngrams, and Frequency Analysis
- ❖ Defining Parameters for a New Analytic Framework

**Tuesday February 7**  
**Check-in time - 4pm.**

---

**Hotel ZaZa**

5701 Main Street  
Houston, TX 77005

[Google Maps Link to Hotel ZaZa](#)

2pm - 12am **Arrivals, Hotel Check-in. Workshop Staff available to help with logistics.**  
**Call Benjamin 713-252-1952 or Sylvia 713-705-6051.**

**Hospitality Suite Open - Snacks, bottled water, and lite bites.**

**Wednesday February 8**  
**8:00am – 9:00pm**

---

**Rice Shuttle leaves ZaZa for BRC at 8:00 am, 8:40 am.**

*It is a 30 minute walk, 15 minute bus/tram ride (\$1.25 fare) 5-10 minute Uber.*

10th Floor Boardroom

**Bioscience Research Collaborative**

6500 Main St  
Houston, TX 77030

[Google Maps Link to BRC](#)

Participants check in at the security desk to gain access to the elevator.

8:15 - 9:30 **Coffee & Breakfast Buffet Available**

**Register and select Thursday box lunch.**  
Mix. Mill. Meet.

**9:00 -10:30 Photographer on-site taking headshots and candid shots. Portraits are for the participant directory.**

9:30 - 9:55 **Opening Ritual: “Sacrificing the Word Cloud” Thought Experiment**

Innovation over novelty. Utility over entertainment.  
Read Workshop Goals  
Welcome with gratitude, without introductions (yet)  
The story of a sad humanities researcher

10:00 - 12:00 **Spelling Variation in Early Modern English**

*Objectives:* Understand the complexities of the spelling variation issue in EEBO and how VARD addresses it.  
*Activities:* Understand the complexities of the spelling variation issue in EEBO and how VARD addresses it. Whiteboarding.

- 12:00 - 1:15 **Lunch - 5411 Empanadas**  
Biological relief and ritual hand washing. Prayers optional.
- 1:15 - 2:00 **Rules of Engagement - Tools of Engagement - Competitions and Prizes**  
*Objectives:* Discuss the workshop methodology and our emphasis on collaborative brainstorming. Right brain activation & stimulation - scenic venues, relaxed schedule, art supplies, contests. Soft Feedback Mechanism = emoji pillows.  
*Activities:* Pass out the tools of engagement, introduce the pillows, toss ideas and pillows around. **Prioritize workshop topics.**
- 2:00 - 2:30 **Tour National Space Biomedical Research Institute**
- 2:45 - 4:00 **Goals and Introductions**  
*Objectives:* Get the participants to know and become comfortable with each other. Focus into and align with the workshop goals. Distinguish novel from innovative.  
*Activities:* Participants introduce themselves to complement the bio handout. Discuss goals for the workshop in conjunction with the introductions. Divide into 3 or 4 groups; each group comes up with the most workshop important goal. Reconvene and discuss the goals as a group. (See goals above). Demonstration of Sentiment Analysis and Network Tool.  
*Materials:* Participant directory handout with photos - Computer Projector, Whiteboard
- 4:15 - 5:00 **The Online DTA Landscape**  
*Objectives:* Begin the conversation about mapping the online DTA terrain.  
*Activities:* Open discussion, Demonstration of Sentiment Analysis and Network Tool.  
*Materials:* Computers, Whiteboard
- 5:30pm **Adjourn to hotel or walk Rice Campus and/or Hermann Park**
- 6:45pm **Leave Hotel ZaZa for America's Restaurant**
- 7:00 - 9:00 **Dinner - private room - America's River Oaks**  
2040 West Gray, Houston, TX 77019 (832) 200-1492

**Thursday, February 9**  
**8:00am – 10:30pm**

---

Huff House  
Rice Campus  
6100 Main Street [Google Maps link to Huff House](#)

It is a 10 minute walk, 3-5 minute Uber from ZaZa.

8:30 - 9:30 **Breakfast, Coffee, & Casual Conversation**

9:30 - 10:15 **CQPweb - Advanced Linguistic Analytical Tools - Andrew Hardie**

*Objective:* Participants learn about a web-based corpus analysis system, which provides a user-friendly interface to the Corpus Workbench (CWB) system. This interface is compatible with any corpus, but is especially useful for large corpora, corpora with word-level annotation (such as part-of-speech tagging), and corpora with rich text-level metadata.

*Materials:* CQPweb Handout

10:30 - 11:15 **Mapping The Online DTA Universe - Whiteboard Session**

*Objective:* Develop a topological map of the current DTA environment

*Activity:* Establish a list of strengths and weaknesses to be applied to the universe of DTA tools and evaluate online DTA tools for their strengths and weaknesses.

*Materials:* whiteboard, computer and screen.

11:15 - 12:00 **Brainstorm: What Humanities Questions are Possible with DTA**

*Objectives:* Create a list of ALL the statistical analyses that can be applied to text data, without regard to their utility to scholars. Imagine the possible uses of these techniques.

*Activities:* Whiteboarding questions.

*Materials:* Whiteboard/ PPT/ Computers and Internet.

12:00 - 1:30 **Box Lunch from Salento**

Garden Walk, Draw, Scribble, Build, Play.

1:30 - 3:00 **Delineate the DTA Environment and Tools - Whiteboard Session - Part 2**

*Objective:* Develop a topological map of the current DTA environment

*Activity:* Establish a list of strengths and weaknesses to be applied to the universe of DTA tools and evaluate all DTA tools for their strengths and weaknesses.

*Materials:* whiteboard, computer and screen.

3:00 - 3:30 **Coffee, Calories & Casual Conversation**

3:45 - 5:45 **Corpus Data: The TEI XML Schema & Relating Data to Queries**

*Objective:* Participants discuss the TEI XML Schema and which “new” tags might usefully or fruitfully be implemented that have yet to be included into the EEBO corpus. Generally, we will discuss which tags should be implemented in every humanities purposed corpus.

*Materials:* TEI XML Schema, Whiteboard

6:15 **Dinner at Pepper Twins Organic Sichuan Cuisine**  
1915 W Gray, 77019

7:00 - 10:00 **Beetle Happy Hour at the Continental Club**  
3700 Main Street - Houston, Texas (713) 529-9899  
It is a 30 minute walk and 5 minute Uber back to the ZaZa.

---

**Friday 8:00 – 9:30**

---

Huff House  
Rice Campus  
6100 Main Street [Google Maps link to Huff House](#)

It is a 10 minute walk, 3-5 minute Uber from ZaZa.

8:30 – 9:30 **Coffee, Breakfast, & Casual Conversation**

9:30 - 10:00 **Copyrights and Wrongs - Understanding the Legal Implications of DTA**  
*Objective:* Participants learn about the limits copyright status imposes on DTA tools and the benefits of small corpora with well-defined copyright permissions like the EEBO phase 1.  
*Materials:* Copyright Handout, whiteboard, and recorders.

10:00 - 11:00 **Spelling Variation in Early Modern English**  
*Objectives:* Take the information from the previous sessions and use it to understand to present  
*Activities:* Understand the complexities of the spelling variation issue in EEBO and how VARD addresses it. Whiteboarding.

11:00- 11:30 **Finding Structure in EEBO**  
*Objective:* Participants consider spreadsheet vs. database for selecting subsets of the corpus to analyze. Paul Schaffner, Andrew Hardie, Anupam Basu relate their experiences with EEBO.  
*Materials:* Whiteboard and recorders

11:30 - 11:45 **Defining Parameters**

*Objective:* In this session participants synthesize the information gleaned from the workshop into a set of specifications and requirements for DTA analytics to address humanities researchers' needs. This synthesis will form the basis for an implementation grant request. It will also provide the structure for a DTA syllabus to be distributed to conference participants and on the web.

*Materials:* Whiteboard and recorders

11:45 – 12:15 **Practicality - Developing an Intuitive User Interface**

*Objective:* In this session participants synthesize the information gleaned from the workshop into a set of specifications and requirements for implementing an intuitive user interface for the analytics tools discussed in the previous session. This synthesis will form the basis for an implementation grant request. It will also provide the structure for a DTA syllabus to be distributed to conference participants and on the web.

12:30 – 2:00 **Lunch Cohen House (Rice Faculty Club)**

**Bookworm - Erez Lieberman-Aiden**

*Objective:* Participants learn about the functionality of Bookworm and the next phase of Bookworm implementation - the NEH-Hathi Trust Project. Participants will consider approaches to using Bookworm and other DTA tools to address weaknesses identified in the morning.

*Materials:* Erez Lieberman-Aiden Handout (if applicable)

2:30 - 3:00 **NETFLIX - ABTRACT: THE ART OF DESIGN (This could postponed to the evening or Saturday)**

3:15 - 5:00 **Conclusion - Planning the next Phase**

*Objective:* Build a team for writing the next phase grant, either NEH Start-Up Phase 2 or Implementation.

*Materials:* Group web search for calls for proposals?

*Activity:* Consider future plans and responsibilities going forward.

7:00 **River Oaks Country Club** Conference participant Anne Chao has invited us to the club's acclaimed Friday night seafood buffet. Our plan is to dine on the terrace where jackets are not required. Other parts of the club have a stricter dress code.