

Workshop Outline

HathiTrust Digital Library

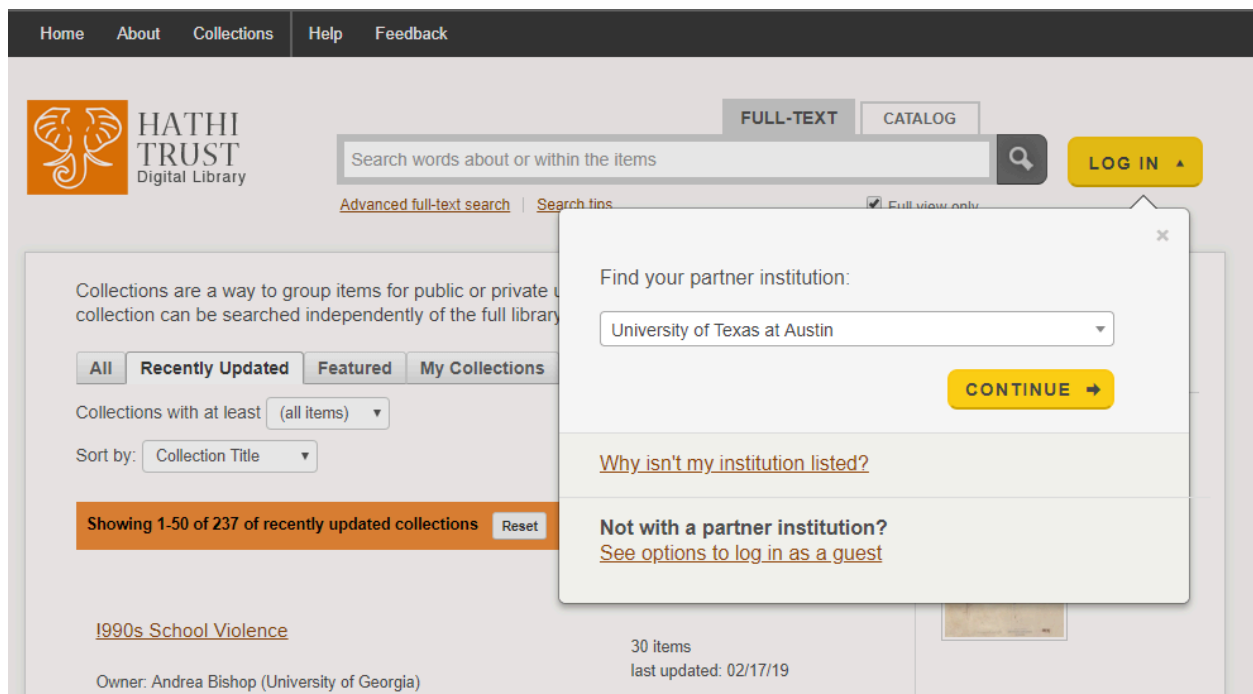
1. What's in the HathiTrust?

[Statistics and Visualizations](#)

2. Creating Collections

[View Collections](#)

To log in, Select “University of Texas at Austin” from the list of partner institutions and then log in with your EID.



HathiTrust Research Center Analytics

1. Create an Account

This is a different log in from the main HathiTrust site.

<https://analytics.hathitrust.org/>

2. Ready to use tools

Import a Workset

The screenshot shows the HTRC Analytics interface. The top navigation bar includes 'HTRC Analytics', 'Algorithms', 'Data Capsules', 'Worksets', 'Datasets', and 'Explore'. The 'Worksets' menu item is circled in blue with a '1' above it. On the right side of the navigation bar, 'Help', 'About', and 'aaguzman' are listed. Below the navigation bar, the breadcrumb 'Home / Worksets' is visible. Two buttons, 'Validate A Workset' and 'Create A Workset', are circled in blue. The main content area is titled 'Worksets' and contains a search filter '(all items)', a search input 'Filter by name/description...', a dropdown for 'My Worksets', and a 'Sort by: Workset Name' dropdown. A table lists worksets with columns: Name, Author, Description, Volume Count, Last Modified Date, Availability, and Actions. One workset is listed: '1990s School Violence' by 'researcher10' with a volume count of 30, last modified in February 2019, and private availability.

Validate a Workset

This screenshot is identical to the previous one, showing the HTRC Analytics interface. The 'Validate A Workset' button is circled in blue. The rest of the page content, including the navigation bar, breadcrumb, search filters, and the workset table, remains the same.

Using Algorithms

HTRC Analytics Algorithms Data Capsules Worksets Datasets Explore Help About aaguzman

Home / Algorithms Jobs

Algorithms

Extracted Features Download Helper (v3.0.2)

Generate a script that allows you to download extracted features data for your workset of choice. The script is a file containing a list of the rsync commands to access the volumes of the workset. After you download the script from HTRC Analytics, it can be run locally (from your computer), which will then download the extracted features data to your computer via rsync. For more information on the extracted features data see the [documentation](#).

Execute

InPhO Topic Model Explorer (v1.0)

The InPho Topic Explorer trains multiple LDA topic models and allows you to export files containing the word-topic and topic-document distributions, along with an interactive visualization. For full detailed description, please review the [documentation](#). Can be run on worksets of fewer than 3000 volumes, as long as the total size of the workset is less than 3 GB.

Execute

Named Entity Recognizer (v2.0)

Generate a list of all of the names of people and places, as well as dates, times, percentages, and monetary terms, found in a workset. You can choose which entities you would like to extract. Can be run on worksets of fewer than 3000 volumes, as long as the total size of the workset is less than 3 GB.

Execute

3. Data Capsule

Due to the slow speed of the data capsules, this will be a demo, not a hands on part of the workshop. You're welcome to try it out on your own, but there are only 15 capsules available at a given time.

[Data Capsule documentation and specs](#)

4. Additional Tools to Explore

Explore

Have fun experimenting with HTRC in-development and beta tools. They are still being tested and may one day be fully incorporated into HTRC's suite of tools, or they may be taken down. Some of these tools require experience with programming to use. Here's where you can get a feel for the HTRC's exploratory work.

HathiTrust+Bookworm

Create a line graph showing word use trends in 13.7 million HathiTrust volumes.

[Try it now](#)

HathiTrust+Bookworm advanced interface

Create a heatmap of word use trends in 13.7 million HathiTrust volumes.

[Try it now](#)

HathiTrust+Bookworm Playground

Create a bar chart, map, or heatmap of word use trends in 13.7 million HathiTrust volumes.

[Try it now](#)

HTRC Feature Reader

Python library for interacting with HTRC Extracted Features.

[Try it now](#)