

HathiTrust Research Center



HATHITRUST
research center

Announcements

[HTRC Announcements](#)

Read about the important events, grants, and projects HTRC is involved in.

Planned Outages

[2022 Outages and Downtime](#)

See the scheduled dates for HTRC Analytics maintenance periods, so you can anticipate these planned outages.

Thinking about text as data

[Getting started](#)

New to the HathiTrust Research Center? This page breaks down HTRC, its relationship to the HathiTrust Digital Library, and provides brief introductions to the tools and resources available on the HTRC Analytics website.

[HTRC Analytics Overview](#)

General overview of all you can do on HTRC Analytics.

[About the Collection](#)

Learn how to understand the data that you will be working with.

[HTRC Policies](#)

Learn our policies from account creation, non-consumptive use, and more.

[HTRC Data Access](#)

Learn about the different access points and formats to the data HTRC provides, as well as the various affordances and limitations of each method. Your research project will largely dictate which method is best suited for your needs.

[MARC Fields](#)

Interested in viewing the MARC fields used by the HathiTrust catalog records? We provide a table here.

[HTRC on the HathiTrust website](#)

Where the data comes from: find our HT documentation pages on the HathiTrust Digital Library site.

Create Worksets

[HTRC Worksets](#)

[HTRC Workset Builder 2.0 \(Beta\) for Extracted Features 2.0](#)

HTRC's Workset Builder tool offers advanced search and result filtering functionality in order to facilitate the creation HTRC worksets. Learn how about and how to use the tool here.

[HTRC Workset Tutorials](#)

Learn the three different ways you can create worksets in HTRC Analytics, as well as how to validate and download a workset to your personal machine.

Run Algorithms

[HTRC Algorithms](#)

Read about the different algorithms HTRC offers, and the kinds of data each algorithm provides.

[HTRC Algorithms tutorial](#)

See step-by-step instructions for running an HTRC algorithm.

Visualize Word Trends with HathiTrust+Bookworm

[HathiTrust+Bookworm Tool](#)

This easy-to-use tool creates a visual output to show word trends across the HT corpus.

[HathiTrust+Bookworm step-by-step tutorial](#)

Follow this tutorial in order to learn how to run the HathiTrust+Bookworm visualization tool.

Do Text Analysis in a Data Capsule

[HTRC Data Capsule Environment](#)

Understand the basics of using a data capsule.

[HTRC Data Capsule Specifications and Usage Guide](#)

Dive deeper into understanding the specs provided in each data capsule environment. Here we break down the different types of capsules, modes, and preloaded libraries, data, and tools we include in each capsule.

[HTRC Data Capsule Step-by-Step Tutorials](#)

View this page as a comprehensive source of all data capsule-specific tutorials.

Download and Use HTRC Derived Datasets

[HTRC Derived Datasets](#)

HTRC Derived Datasets are structured sets of metadata representing a curated collection of HathiTrust volumes. Read about the basics of our Extracted Features and partner-created datasets here.

[Request Data from HathiTrust](#)

Read about the different dataset formats the HathiTrust Digital Library provides their users.

Extracted Features

[Extracted Features \[v.2.0\]](#)

Read about HTRC's most recent version of the Extracted Features derived dataset.

[Basic walk-through of an Extracted Features 2.0 file](#)

Learn what is inside an EF file, how the data is structured, and what it looks like.

[Downloading Extracted Features](#)

See the different ways you can download EF files to your local machine or inside a data capsule.

[Additional steps for downloading an extracted features dataset for Windows users only](#)

Use these steps if you work in a Windows system.

[Stubbytree directory structure](#)

See how EF files are stored in a stubbytree structure.

[Extracted Features in the Wild](#)

See how researchers and developers are using the EF dataset in real-world projects.

[Extracted Features Use Cases and Examples](#)

See two examples of how EF files can be used for accomplishing text analysis research goals.

Word Frequencies in English-Language Literature

[Word Frequencies in English-Language Literature, 1700-1922](#)

Researcher Ted Underwood created a dataset to list word frequencies and help identify genre within a subset of HT volumes. Read about all the details of this dataset here.

Geographic Locations in English-Language Literature

[Geographic Locations in English-Language Literature, 1701-2011](#)

Read about and learn how to download this geospatial dataset created by researcher Matthew Wilkens.

For help with any of our products, services, or documentation, please send an email to htrc-help@hathitrust.org.

General Help Info

[Troubleshooting and FAQs](#)

When something goes wrong, this page is here to help you identify and hopefully remedy the problem. Please contact HTRC support at htrc-help@hathitrust.org with any questions or bugs or peculiarities you would like to report.

[HTRC, Help!](#)

Here are a few examples of some common ways our users receive help from HTRC staff.

[All HTRC Tutorials](#)

A comprehensive list of all HTRC tutorials to walk you through the steps for using HTRC tools and data.

[Glossary](#)

This glossary contains definitions and explanations for certain key terms and tools used in the Digging Deeper, Reaching Further (DDRF) Workshops.

Outreach

[Workshops and Educational Materials](#)

HTRC teaches several workshop series every semester. During the COVID-19 pandemic we switched much of our workshop curriculum to a remote, online format. Please send inquiries based on the information you read here.

[Tips for hosting a workshop at your institution](#)

Are you interested in leading an HTRC workshop at your own institution? Here are some tips to get you started.

[Attend a workshop](#)

Visit the HathiTrust Digital Library's page about attending an HTRC workshop.

[Further reading](#)

Here is a list of both popular and scholarly articles portraying a variety of digital humanities/cultural analytics research. Reading articles is a great way to see how other researchers have approached text analysis-based projects in the past.

[HTRC Trainers Community](#)

Librarians who attend all four days of the virtual series or who attend both days of our in-person training are invited to join the HTRC Trainers Community. The community is composed of librarians from HathiTrust member organizations who are interested in teaching with and about HTRC tools and services.

[HTRC User Group](#)

The HTRC User Group is a community of scholars and librarians who utilize HTRC tools and data in their research and teaching.

HTRC Research Projects

[Examples and Use Cases](#)

Curious what kind of data-driven research and teaching others have done? Here are some examples of what is possible.

[How they did it - In Search of Zora](#)

Read how the research team of the "In Search of Zora" project devised and executed their analysis.

[How they did it - Textual Geographies](#)

Read how the research team for the "Textual Geographies" project used text analysis methods to create an interactive visualization based on geographic locations.

[HTRC Publications and Presentations](#)

View a list of HTRC and HathiTrust Digital Library publications and presentations here.

HTRC Advanced Collaborative Awards and Projects

[Advanced Collaborative Support \(ACS\) Awards](#)

Advanced Collaborative Support (ACS) is a scholarly service at HTRC offering collaboration between external scholars and HTRC staff to solve challenging problems related to HTRC tools and services.

[2019-2020 ACS Project updates](#)

ACS awardees are sharing updates about their work to date roughly midway through their project cycles. Check them out here!

[2020 ACS Project updates](#)

ACS awardees are sharing updates about their work to date roughly midway through their project cycles. Check them out below!

HTRC's Grant-Funded Projects

[Grant-funded projects](#)

Read about HTRC's grant-funded projects here.

[Scholar-Curated Worksets for Analysis, Reuse & Dissemination \(SCWAReD\)](#)

Read about our most recently funded project, SWAReD, and its impact on the HathiTrust's digital collections.

Development Documentation

[HTRC Releases](#)

Read about our past software releases here.

[HTRC Software Development](#)

View HTRC's advanced documentation for developers interested in contributing to, understanding, or viewing certain aspects of our backend development.

[HTRC Technical Documents](#)

View the key technical documents for understanding the main HTRC services. These pages are under continuous development.

[HTRC Software Development Process for Arriving at Technical Decisions](#)

Read about the steps HTRC staff takes for making technical decisions.

User Contributions

[Developing Algorithms for the HTRC Framework](#)

Learn how to develop your own algorithm code for HTRC's front-end.

[Contributing your Code to the HTRC Community](#)

Learn how to contribute your own code to HTRC Analytics.

Other links

- [HathiTrust Research Center Official Logo](#)
- [HathiTrust Research Center Project Website at Indiana University](#)