HTRC User Getting Started FAQ

Q1: What is the HTRC?

A: The HTRC is the research arm of the HathiTrust. It facilitates scholarly research using the large-scale HathiTrust Digital Library by providing mechanisms for researchers to access content in the HathiTrust and study it using computational tools for text analysis.

The HTRC is a partnership between Indiana University (IU) Libraries, the Pervasive Technology Institute, and the School of Informatics and Computing at IU, as well as the University of Illinois at Urbana-Champaign (UIUC) Libraries and the Graduate School of Library and Information Science at UIUC.

Q2: What are the HTRC tools and services?

A: The HTRC has created a suite of tools that allow researchers to perform text analysis on content in the HathiTrust Digital Library. Most of these tools are available via HTRC Analytics, and include web-based text analysis algorithms, HathiTrust+Bookworm, and the HTRC Data Capsule. They are intended to meet the needs of various HTRC researchers.

- HTRC Algorithms: a set of tools for assembling collections of digitized text and performing text analysis on them.
- HathiTrust+Bookworm: a tool for visualizing and analyzing word usage trends in the HathiTrust Digital Library.
- HTRC Data Capsule: a secure computing environment for performing researcher-driven text analysis on HathiTrust content.

Q3: How do I use the HTRC?

A: You use the HTRC by interacting with our tools and services. Please refer to the documentation for each tool or service for more specific how-to guides.

- HTRC Analytics guide
- HathiTrust+Bookworm guide
- HTRC Data Capsule guide and Data API guide
- Extracted Features Dataset guide

Q4: Do I need an account to use the HTRC and how do I make one?

A: Most of the HTRC services require an account to log in and interact with the tools, though HathiTrust+Bookworm is available without an account.

Register for an account by going to the main page of the HTRC Analytics and choosing "Sign up" from the menu. Anyone possessing an email address from a nonprofit institution of higher education is allowed to register, including those whose institutions are not HathiTrust members.

Q5: What is the difference between using the HTRC and searching the HathiTrust Digital Library?

A: Using the search on the hathitrust.org site, you can find digitized items in the HathiTrust Digital Library (HTDL) and to read them if they are in the public domain. From the HTDL, you can create collections that you are able upload to HTRC Analytics as a workset. With the HTRC tools you can work with material from the HathiTrust Digital Library at scale, using computational methods to analyze collections of content, called worksets in HTRC, relevant to your research.

Q6: What types of data and metadata does HTRC provide?

A: The availability of data and metadata in HTRC depends on the tool or service.

- HTRC algorithms and HTRC Data Capsules currently provides access to a snapshot of the public domain corpus OCR text from HathiTrust, as well as each volume's MARC bibliographic and METS metadata. Both the HTRC algorithms and Capsule-environments draw from the HTRC Data API described below.
- The HTRC makes available also two datasets, the HTRC Extracted Features Dataset and a dataset of Word Frequencies in English Language Literature, 1700-1922. HTRC Extracted Features includes metadata and extracted page-level data (words and word counts) for 13.7 million volumes.
- HathiTrust+Bookworm visualizes data for 13.7 million volumes.

Q7: What is the difference between the HTRC Data API and HathiTrust datasets and APIs?

A: This table outlines the differences between the HTRC Data API and HathiTrust Data API. The HTRC Data API currently functions within the HTRC Data Capsule.

	HTRC Data API	HT Data API
purpose	to serve high-performance large-scale algorithms and programs	to provide public users some volume retrieval capabilities
throttling enforcement	no	yes
security	JWT	OAuth
bulk retrieval of volumes	yes	no
metadata available	METS	METS, MARC

Q8: What happened to the HTRC Solr Proxy API?

A: As the HTRC moves to update and improve its search and workset-building services, the Solr Proxy API has been retired. For now, you can search for HathiTrust volumes via the HathiTrust Digital Library interface. Look for improved functionality in the near future, and please reach out with your workset-building scenarios that require additional search functionality.

Q9: How do I ask questions or start discussions with other users?

A: Please join the HTRC User Group mailing list.

- Subscribe here: https://list.indiana.edu/sympa/info/htrc-usergroup-l
- For questions that you want to discuss with us privately, please write to htrc[dash]help[at]hathitrust[dot]org, a list subscribed by HTRC internal staff only.
- All users are subscribed to a listserv called HTRC-Announce when they create an HTRC Analytics account. Only approved senders can send mail through this list.

Q10: How do I report issues or give feedback?

A: We welcome your feedback! You can send an email to HTRC Support at htrc[dash]help[at]hathitrust[dot]org. We track support requests in using JIRA, and you can log-in to see your requests and our responses here: https://jira.htrc.illinois.edu/servicedesk/customer.

Q11: Where do I go for more information?

A: If you have not found what you are looking for in our documentation, you might find the material posted to our Publications and Presentations page useful for further reading.

You might also consider attending a workshop. You can find information on future workshops on our calendar.

Or you can ask for further assistance on our mailing lists. See below for more information about signing up.