FAQ

These are Frequently Asked Questions (FAQ) about using HTRC Analytics.

See here for the general HTRC FAQ.

For detailed instructions, follow the step-by-step guide.

Q: How do I get to HTRC Analytics?

A: The front page of HTRC Analytics is available at https://analytics.hathitrust.org/.

Q: What is HTRC Analytics?

A: HTRC Analytics is a set of complementary tools studying sub-collections of volumes from the HathiTrust Digital Library, called worksets, using computational text analysis. Services offered through HTRC Analytics include a suite of algorithms, Extracted Features Dataset, and the interface to create and manage a Capsule in the HTRC Data Capsule secure computing environment.

Q: How do I create an account to log in to HTRC Analytics?

A: You can create an account by going to HTRC Analytics and clicking "Sign Up" in the top right corner. Anyone possessing an email address from a nonprofit institution of higher education is allowed to register, including those whose institutions are not HathiTrust members.

Q: What terms do I need to know to get started using HTRC Analytics?

A: HTRC Analytics has several overarching paradigms: algorithms, jobs,results, and worksets.

1. Text analysis algorithms are programs that will run one or more function against your workset in order to understand the patterns, themes, or content of a workset. You can choose from a set of algorithms that have been integrated into the HTRC.
2. Jobs: When you hit submit, you are submitting a job. A job is a set of instructions that are executed by one of the computing resources available to the HTRC. You can view the status of the jobs that you have submitted. You can also delete jobs. If you find that you have made an error in your set-up, you can delete the job.
3. Results: When your job has completed, you can view the results of the job. The results can be viewed in the HTRC. You can also download the results.
4. Worksets are subc-ollections of HathiTrust content that you can create using the HathiTrust Digital Library and then download for subsequent use with the tools provided at HTRC Analytics.

Q: What are worksets and what do I do with them?

A: Worksets are sub-collections of HathiTrust volumes created by researchers. You can run HTRC algorithms against worksets in order to analyze them or download their Extracted Features. Worksets can be cited, and researchers can choose to make their worksets public or private.

Q. How do I create a workset?

You create a workset by uploading a list of volume IDs to HTRC Analytics. The easiest way to generate a list of volume IDs for analysis is to first create a collection in the HathiTrust Digital Library. The file you upload must be in TXT or CSV format, and while it can contain other data, the only requirement is that the first column of data list the volume IDs for your workset.

Here are the basic steps:

1. Create a collection in the HathiTrust Digital Library. For now, HTRC Algorithms can only access a snapshot of public domain volumes, so be sure to only select volumes that are in "Full View" in the digital library.

2. Download the metadata for your collection. (You can also browse public collections and download their metadata.)
3. Upload the resulting file to HTRC Analytics.

Note: HTRC Analytics may display the order of your volumes differently than they appeared on the input file.

Q. What happened to the Workset Builder?

A: As HTRC upgrades its services and builds a new Workset Builder, the retired Workset Builder has been taken offline. The new system of creating a collection in the HathiTrust Digital Library better aligns workset-building with the HathiTrust and offers improved search and selection.

Q: What kinds of text analysis services and/or algorithms are in HTRC Analytics?

A: HTRC Analytics includes access to 9 off-the-shelf HTRC algorithms that facilitate text analysis. These algorithms help you extract, refine, analyze, and visualize the content of a workset. Please see this description of the algorithms provided via HTRC Analytics for more information.

Q. What is the login timeout?

A: The current login timeout is 1 hour. However, your submitted job won't be affected by this logout time. It will still run even if you logout or if the system logs you out.