

HTRC Analytics Documentation

HTRC Analytics is the gateway to most HTRC tools and services. It is a set of complementary tools for studying sub-collections of volumes from the HathiTrust Digital Library (called worksets) using computational text analysis. On the Analytics website, you can create worksets for textual analysis, run text analysis algorithms, set-up and run a Data Capsule secure analysis environment, and more.

[Go to HTRC Analytics Follow a Tutorial](#)

The basics

| Offering | Description | Data availability | Account required | Documentation |
|--|---|--|---|----------------------------|
| HTRC Algorithms | A set of tools for assembling collections of digitized text from the HathiTrust corpus and performing text analysis on them. | Including items in copyright for ALL USERS. | HTRC Analytics | Learn more |
| HTRC Worksets | Worksets are sub-collections of HathiTrust volumes that can be analyzed with HTRC algorithms or used to access HTRC Extracted Features. | Including items in copyright for ALL USERS. | HTRC Analytics | Learn more |
| Extracted Features Dataset | A dataset allowing non-consumptive analysis on specific features extracted from the full text of the HathiTrust corpus. | Including items in copyright for ALL USERS. | None | Learn more |
| HathiTrust +Bookworm | A tool for visualizing and analyzing word usage trends in the HathiTrust corpus. | Including items in copyright for ALL USERS. | None | Learn more |
| HTRC Data Capsule | A secure computing environment for researcher-driven text analysis on the HathiTrust corpus. | All users may access public domain items. Access to items in copyright is available ONLY to member-affiliated researchers. | HTRC Analytics ; plus additional restrictions | Learn more |

Want to learn more about data availability? [See this additional table.](#)

Accounts

Account creation

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 an institution of higher education or a non-profit research organization is allowed to register, including those whose institutions are not HathiTrust members.

Many email domains from colleges and universities in the United States are recognized in our system, and you can register automatically with an email address from those institutions.

If you possess an email address that is not yet recognized by HTRC Analytics as belonging to an approved domain, you will be prompted to request an account on the sign up page. Upon review of the request, affiliates of institutions of higher education will be approved to create an account, and that email domain is then registered in our system for the benefit of others at their institution. If you are affiliated with a non-profit research institutions, such as a library, your account will also be approved, though you may be asked to provide information about your organization during the request review process. Only in certain cases are researchers from non-academic or for-profit organizations permitted to create an account.

You cannot create an account with Gmail, Hotmail, Yahoo, or similar email address.

Account deactivation

Accounts that have not been active for more than one year are deactivated. HTRC checks activity monthly to identify inactive accounts. Once an account has been deactivated, a researcher will experience an error if they attempt to log into their account. A deactivated account can be reactivated by contacting HTRC at htrc-help@hathitrust.org.