

## MODULE 2.1 Gathering Textual Data: Finding Text

### KEY TOOLS & PLATFORMS

#### HT Collection Builder

An interface for creating collections via the HathiTrust Digital Library.

#### HTRC Analytics

An interface for working with HTRC worksets, which are collections of text from HathiTrust that can be analyzed using the non-consumptive tools and environments in HTRC Analytics.

### ACTIVITY: Evaluating different sources for textual data

 **Slide M2.1 - 9**

*If we are building a corpus for political history, what are the strengths and weaknesses of each of the following broad sources for textual data? Please discuss and take some notes in the chart below.*

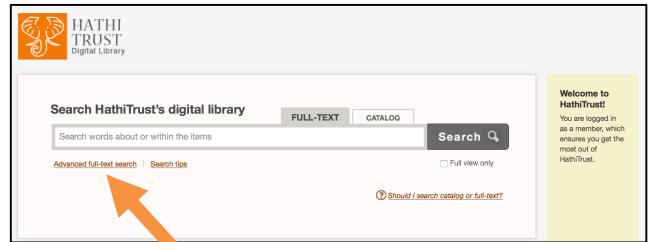
	Strengths	Weaknesses
Vendor database		
Library and archives digital collections		
Social media		

## ACTIVITY: Building and uploading a workset

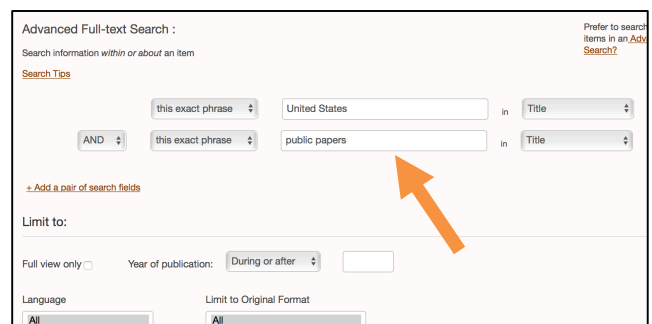
Slide M2.1 - 17

Let's create a workset for a political science student. As an example, we'll do an advanced full-text search for volumes that contain both "public papers" and "United States" in their titles

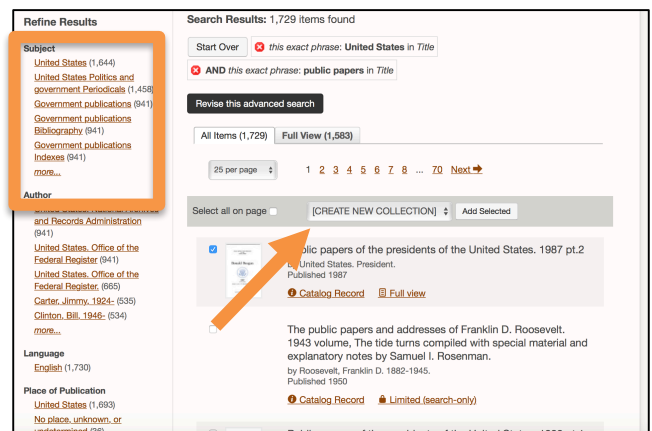
1. Make sure you are still logged in to the HTDL: <https://www.hathitrust.org/>
2. Click the "FULL-TEXT" tab to search in full text.
3. Click "Advanced full-text search" under the search bar.



4. You will see two blank search fields. Since we want to look for volumes that have two specified phrases ("public papers" and "United States") in their titles, the "AND" logic works and we don't need to change it.
5. For both search fields, select "this exact phrase" and "Title" to limit our search. Type *United States* in one search field and *public papers* in the other.
6. Click "Search" at the bottom of the page.



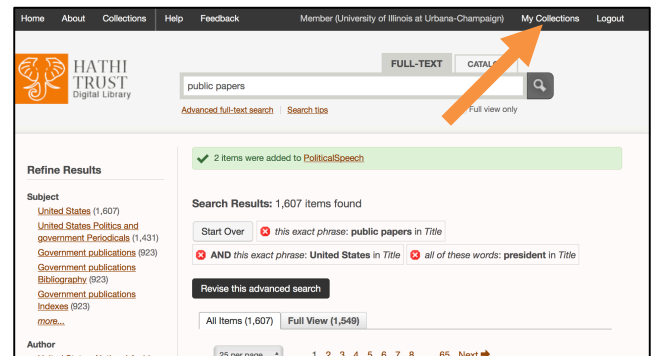
7. Filter your results using the fields on the left.
8. Click the checkbox next to the titles you would like to add to your collection.
9. When ready to create your collection, click on the "Select Collection" bar and choose "[CREATE NEW COLLECTION]" from the drop-down menu. Click "Add Selected".



10. Enter information about your collection in the pop-up window. Click "Save Changes".

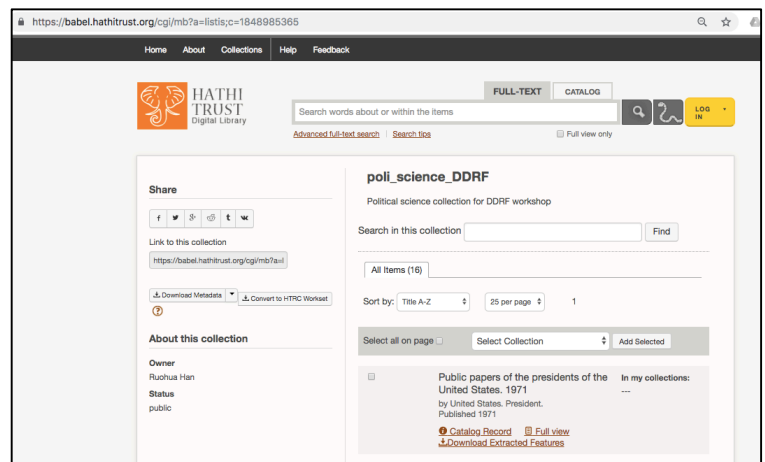


11. You'll see a confirmation that your collection was created. Click "My Collections" in the top right.



12. Click the title of the collection you just created to view it.

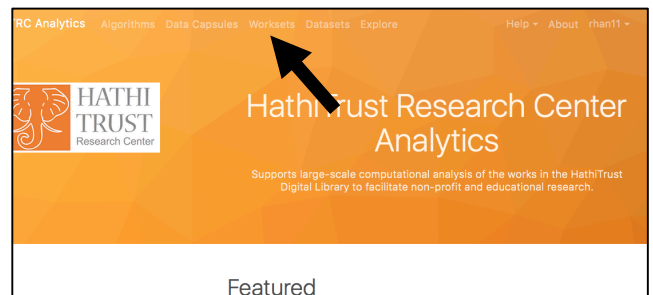
13. Copy the URL for your collection from the address bar.



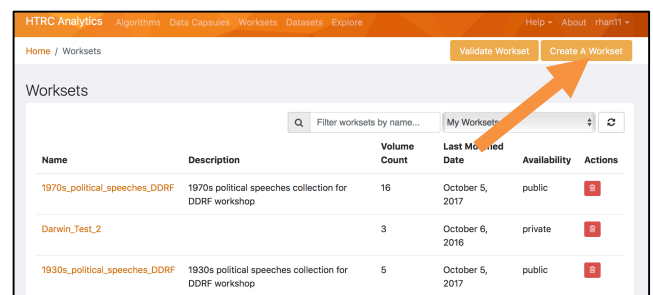
14. Log in to HTRC Analytics:

<https://analytics.hathitrust.org>

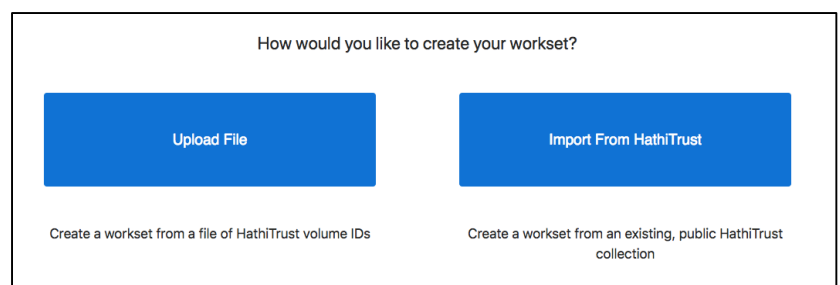
15. Go to "Worksets" in the header menu.



16. Click on "Create a workset" near the top right.



17. Choose "Import from HathiTrust"



18. Paste the URL for your collection in the Collection URL field and click “fetch.”
19. View and/or edit the information for your workset.
20. Click “Create Workset”. You should be able to find your new workset on your Worksets page.

### Create A Workset

Import a collection from HathiTrust using the collection's URL. While HathiTrust grows daily, HTRC syncs data periodically from the HathiTrust Digital Library. Some volumes you would like to include in your workset may not be available. Any volumes in your workset not available through HTRC will be skipped by the algorithm.

**Find collection URL**

When viewing your [collection on HathiTrust](#), simply copy the URL from your browser, or copy the “Link to this collection” found on the left sidebar, and paste the URL below.

**Import**

Hit “Fetch Collection” and your collection will be transformed into an HTRC workset. You may need to edit the default name in order to meet HTRC requirements.

**HathiTrust Collection URL**

**Name**

Disallowed characters: ~ ! @ # % ^ \* + = [ ] | < > , ' " \ /

**Description**

☐ Private Workset  
If checked, your workset will be accessible to only you.