

# HathiTrust Research Center Workshop Materials Re-use Guide



Up-to-date HTRC workshop materials exist for two workshop versions that can be re-mixed for different workshop lengths and audiences.

1. Full day workshop: *Text mining with HathiTrust: An Introduction for Librarians*
2. ½ day workshop: *HathiTrust's Data and Analysis Tools for Text Mining Research*

All workshop materials are intended to be reused as desired, without permission, though we do appreciate hearing how they have been used.

## Available materials:

Materials are currently stored in Box (<https://uofi.box.com/v/HTRC-fall2019>). They include:

- Slides (for both workshops) – in PDF and Word formats
- Handouts (for both workshops) – in PDF and Word formats
- Jupyter notebook activities (for both workshops)
- 3 case studies

## Slides & content

- The slides for both workshops are provided with an optional presenter's script in the notes field.
- There is overlap between the content of the two workshops. Here are key ways in which they are different:
  - ½ day workshop includes additional “tracking one item across HathiTrust” slides
  - ½ day workshop uses only 1 case study
  - Full day workshop includes more information about text analysis methods and approaches
  - Full day workshop includes readings and guided discussion questions

## Activities

Activity	Full day workshop	½ day workshop
HTRC Named Entity Recognizer Algorithm	x	
Plotting locations from NER algorithm (notebook)	x	
Learning the Jupyter environment (notebook)	x	x
Analyzing Extracted Features (notebook)	x	x
Research with the HTRC Data Capsules		x

The Jupyter notebooks can be revised and reused. They can also be run locally. In the workshops, we interact with them using a web-viewer called Binder, which takes the GitHub repository—including the .ipynb files, data files, and an environment file specifying the libraries needed for the code used in the notebooks to execute—and then renders it in the interactive Binder environment.

- To run locally (and with different data):
  - Download the GitHub repository at <https://go.illinois.edu/htrc-workshop>.
  - Make sure that the dependencies listed in the environment.yml file are installed on your computer.
  - Run Jupyter Notebooks from the command line on your computer by typing and entering `jupyter notebook`.
  - Change out the data, for example to another set of Extracted Features files or a different Named Entity Recognizer Algorithm output file. Note that you may have to make sure that the file paths are pointing to the correct places on your local machine.
- To reuse the GitHub repository for another Binder:
  - Clone the GitHub repository at <https://go.illinois.edu/htrc-workshop>.
  - Modify the files in the repository, such as using a different set of Extracted Features files or a different Named Entity Recognizer Algorithm output file.
  - Edit the notebooks with different explanatory text or add or remove parts of the activities (making sure that you don't break the code if you can't fix it!).
  - Follow the steps to initialize Binder using the documentation here: <https://mybinder.readthedocs.io/en>

# Preparing to teach

## Further reading

- Please see this page on the HTRC wiki for recommended readings:  
<https://wiki.htrc.illinois.edu/display/COM/Further+reading>.

## Additional materials

- Glossary: <https://wiki.htrc.illinois.edu/display/COM/Glossary>.
- HTRC introductory video: <https://www.youtube.com/watch?v=Vq6E9u67GnM>.
- Many additional materials can be found on the DDRF project website including short video tutorials, webinars, and teaching materials that use slightly different examples and activities. They are available at: <https://teach.htrc.illinois.edu/>.

## Scheduling

- Don't schedule a workshop during HTRC's monthly maintenance (every 2<sup>nd</sup> Sunday and Monday of the month).
- Please feel free to let us know when you have scheduled a training event so we can keep an eye on user tickets or account-creation issues.