Exploring HathiTrust

2023 GIL USERS GROUP MEETING
May 18, 2023

Chris Stotelmyer, Head of Electronic Resources
Clayton State University

Jim Rickerson, Electronic Resources Librarian
Clayton State University

Chandra Jackson, Electronic & Continuing Resources Librarian
University of Georgia

Anu Moorthy, Electronic Resources Librarian
Georgia Institute of Technology
Scope of Presentation

• Introduction
• Origin and overview of HathiTrust Digital Library
• Accessing and searching HathiTrust
• Viewing search results and viewing options
• Research Center and other tools and services
• Future plans
• Questions and answers
Introduction

• Chris Stotelmyer, Clayton State University
Basic Facts

- Hathi is pronounced “hah-tee” and is the Hindi word for elephant
- Launched in 2008 as a partnership of University of California and Big Ten library systems
- Known for its Digital Library
- Members get much more
HathiTrust Digital Library

Currently Digitized

- 17,645,725 total volumes
- 8,475,062 book titles
- 469,424 serial titles
- 6,176,003,750 pages
- 791 terabytes
- 7,048,477 volumes (~40% of total) in the public domain
HathiTrust Digital Library continued

Materials

- Digitized books and book-like materials
  - manuscripts
  - Pamphlets
- Bound and unbound serials
- No unbound maps or large-format items
- No audio or video
- No items digitized from microfilm/fiche
- No born-digital materials (PDF, e-book)
Background

• The foundation for the HathiTrust library began in 2005

• Digitization projects before HathiTrust
  • Google’s Library Project
  • The Internet Archive
  • The Open Content Alliance
  • Microsoft’s “Live Search” service

• University of California began digitizing as early as 2005

• The Big Ten Academic Alliance began digitizing in 2007
Membership

- Over 170 members in US and Canada
- International participation includes members in Australia, Spain and England
- Early members include University of Virginia and Harvard
- Research focus
Categories of membership

- Academic and research institutions
- University Systems of libraries
- Consortia of libraries
- No libraries associated with for-profit organizations
- No public libraries without a clear research mission
- No public or private K-12 or secondary school libraries
- No non-libraries, including publishers
- No individuals
Introduction to

There's an Elephant in the Library; Organizers Promise It Will Never Forget

Programs

• Emergency Temporary Access Service
• Shared Print Program
• Digital Library benefits
• Accessible Text Request Service
• Many others
Final Thoughts

Member Toolkit (link in bibliography)

Materials for students and faculty

Links to informational videos
Searching HathiTrust

• Jim Rickerson, Clayton State University
It is recommended that you log in using your institutional profile. This will enable you to access additional features such as:

- Downloading books
- Creating/Managing Collections
From the homepage you will find four search options to choose from.

- Basic Full-Text searching
- Basic Catalog searching
- Advanced full-text searching
- Advanced catalog searching
Indicates that you can access the full text of the book in the library and as a member you can download the full-text of the book as well.
(1) **Basic Full-text Search**

All Items will return both full view and limited search only items.
(1) BASIC FULL-TEXT SEARCH

Limited (search-only)

Search in this text

Find

#22 (p. 18) - 2 matching terms
#23 (p. 19) - 2 matching terms
#24 (p. 20) - 3 matching terms
#25 (p. 21) - 1 matching term
#41 (p. 37) - 1 matching term
#44 (p. 40) - 1 matching term
Click Get This Item to search WorldCat or Google Books for this Limited Search item.
of human nature, has shown me some verses; but alas! how came such charming poetry to be thrown away on so unmeritorious a topic? I don't know whether I ought to praise the lines most or censure the object most. Voltaire makes the excellence of French poetry consist in the number of difficulties it vanquishes. Pope, who celebrated Lord Bolingbroke, could not have succeeded, did not succeed, better; and yet I hope that, though a meaner subject, I am not so bad a one! Well! with all my humility, I cannot but be greatly flattered. Madame de Sévigné spread her leaf gold over all her acquaintance and made them shine; I should not doubt of the same glory when Lady Nuneham's poetry shall come to light, if my own works were but burnt at the same time; but alas! Coulanges's verses were preserved, and so may my writings, too. Ah! my lord, I have a new volume of that divine woman's letters. Two are entertaining; the rest not very divine. But there is an application, the happiest, the most exquisite that even she herself ever made! She is joking with a Président de Provence, who was hurt at becoming a grandfather. She assures him there is no such great misfortune in it; 'I have experienced the case,' says she, 'and believe me, papa, now do not.' If you are not both transported with this, ye are not the Lord and Lady Nuneham I take ye to be. 

MORACE WALPOLE
(2) BASIC CATALOG SEARCH

Subject

Author

ISBN
Clicking Catalog Record in the results page displays a limited number of metadata fields:

- Author
- Language
- Publisher / Publication Date
- Subject Headings
- Physical Description
Scrolling down the page - below Description you will see a section called Viewability.

This provides you with links to volumes 1 and 2 which are held by different institutions.
Clicking the button:

**View HathiTrust MARC record**

displays the MARC record with additional metadata for the item
**Visions and beliefs in the west of Ireland, collected and arranged by Lady Gregory: with two essays and notes by W.B. Yeats ...**

<table>
<thead>
<tr>
<th>Description</th>
<th>Tools</th>
<th>Cite this</th>
<th>Export citation file</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Main Author:</th>
<th>Gregory, Lady, 1852-1932.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Related Names:</td>
<td>Yeats, W. B. 1865-1939.</td>
</tr>
<tr>
<td>Language(s):</td>
<td>English</td>
</tr>
</tbody>
</table>

**Subjects:**
- Folklore > Ireland
- Manners and customs
- Folklore
- Folklore > Ireland
- Atlantic Coast (Ireland) > Social life and customs
- Ireland > Atlantic Coast
- Ireland
- Atlantic Coast (Ireland) > Social life and customs

**Physical Description:**
- 2 volumes frontispieces 20 cm

**Locate a Print Version:**
- Find in a library
Advanced Full-text Search

Search information within the item (Search Tips).

Prefer to search about the item in an Advanced Catalog search?

Search by field

[Dropdown menu with options: Full-Text + All Fields, Title, Only Full-Text, Full-Text + All Fields, Title, Author, Subject, Publisher, Series Title, ISBN/ISSN]

- Full-Text + All Fields
- Title
- Only Full-Text
- Full-Text + All Fields
- Title
- Author
- Subject
- Publisher
- Series Title
- ISBN/ISSN

- all of these words
- Search Term 1

- this exact phrase
- Search Term 2

AND  OR

Add another group of search fields
Advanced Full-text Search

Search information within the item (Search Tips).

Prefer to search about the item in an Advanced Catalog search?

Search by field

- Full-Text + All Fields
- Title

.search dropdown:
- all of these words
- any of these words
- this exact phrase

Search Term 1

OR

Search Term 2

Add another group of search fields

Advanced Search
(3) ADVANCED FULL TEXT SEARCH

[any of these words] formorians formori

Will search both full-text and all bibliographic fields for: formorians OR formori and retrieve works containing either term.

[all of these words] "fairy tales" legends

Will search subject headings of works for: fairy tales AND legends and retrieve works containing both terms.
Additional search options provides the option to filter only those items to which you have full view access and if you want which items are available to you as physical copies you can borrow within the USG.
(3) ADVANCED FULL TEXT SEARCH

Because we checked our Institution Name we know that these are available to borrow as physical copies in the USG.
Because we checked our Institution Name under additional search options it will only display those items that are available to borrow as physical copies in the USG.

These are examples of the links to hatho Trust from the brief record and the PRIMO Services page.
(3) ADVANCED FULL TEXT SEARCH

Language
Select one or more checkboxes to narrow your results to items that match all of your language selections.

- Abkhazian
- Achi
- Acoli
- Adangme
- Adyghe
- Afrikaans
- Afro-Asiatic
- Akan
- Akkadian
- Akha
- Alain

Original Format
Select one or more checkboxes to narrow your results to items that match all of your format selections.

- Archive
- Audio
- Audio (music)
- Audio (spoken word)
- Audio CD
- Audio LP
- Biography
- Book
- Computer File
- Conference

Search
Reset Form
(4) ADVANCED CATALOG SEARCH

Search by field drop down menu

Advanced Catalog Search
Search information about the item (Search Tips).
Prefer to search within the item in an Advanced Full-text search?

Search by field

- Author
- All Fields
- Title
- Subject
- Publisher
- Series Title
- Year of Publication
- ISBN/ISSN

Search Term 1
AND OR
Search Term 2
AND OR
Search Term 3
AND OR
Search Term 4

Search Advanced Search
Search by field

- **Author**: William Butler Yeats
- **Title**: The land of heart's desire
- **Year of Publication**: Search Term 3
- **Subject**: Search Term 4

Additional search options

- View Options: Full view only
- Date of Publication: Before, After, Between, Only during 1900
On the catalog record page, you can easily navigate to related items.
(5) NULL SEARCH

Entering an asterisk * into the search box and selecting Full-Text will retrieve all records in Hathi Trust. The default filter is Full View. Here you can see there are currently over seven million records.
(5) NULL SEARCH

Entering an asterisk * into the search box and selecting Full-Text will retrieve all records in Hathi Trust. The default filter is Full View. Here you can see there are currently over seven million records.
Hathi Trust supports both Trunaction and Wildcard searching symbols: 

? (question mark) and * (asterisk)

Wildcards are a technique for helping you look for terms with different spelling such as between British and American English. A search for organi?ation will retrieve both spellings of the word with a Z and S.
(6) WILDCARD & TRUNCATION SEARCHING

Hathi Trust supports both Trunaction and Wildcard searching symbols:

? (question mark) and * (asterisk)

Truncation is a technique for helping you broaden your search for various word spellings and endings. A search for music* will retrieve various words such as musician and musical.
If the work is in a Latin alphabet with diacritical marks, it is not necessary to enter the character with diacritical marks.
Viewing Results

- Chandra Jackson, University of Georgia
LIFE IN UTOPIA

BEING
A FAITHFUL AND ACCURATE DESCRIPTION OF THE INSTITUTIONS THAT AROSE OUT OF LABOUR, ART, SCIENCE, AGRICULTURE, HABITATION, MATRIMONY, LAW,}

AN
LIFE IN UTILITY

BEING
A FAITHFUL AND ACCURATE DESCRIPTION
OF THE INSTITUTIONS THAT ARE
INVOLVED IN LABOUR, ART, SCIENCE, AGRICULTURE,
HABITATION, MATRIMONY, LAW, AND RELIGION.
LIFE IN U...
LIFE IN UTILITY AND CONVENIENCE

A FAITHFUL AND ACCURATE DESCRIPTION OF THE INSTITUTIONS THAT RESULT FROM LABOUR, ART, SCIENCE, AGRICULTURE, HABITATION, MATRIMONY, LAW, AND RELIGION,
LIFE IN THE COUNTRY

BEING
A FAITHFUL AND ACCURATE
REPRESENTATION
OF THE INSTITUTIONS
OF LABOUR, ART, SCIENCE, AGRICULTURE,
HABITATION, MATRIMONY, AND RELIGION
IN THIS DELIGHTFUL INHABITANCY.
LIFE IN 

A FAITHFUL AND ACCURATE ACCOUNT OF THE INSTITUTIONS TO WHICH LABOUR, ART, SCIENCE, AGRICULTURE, HABITATION, MATRIMONY, AND RELIGION HAVE GIVEN FORM 

BEHIND THE SCENES
LIFE IN UTOPIA.

BEING
A FAITHFUL AND ACCURATE DESCRIPTION OF THE INSTITUTIONS THAT REGULATE LABOUR, ART, SCIENCE, AGRICULTURE, HABITATION, MATRIMONY, LAW, GOVERNMENT, AND RELIGION, IN THIS DELIGHTFUL REGION OF HUMAN INHABITATION.
LIFE IN U

BEING

A FAITHFUL AND ACCURATE

THE INSTITUTIONS THAT

LABOUR, ART, SCIENCE, AGRICULTURE,

HABITATION, MATRIMONY, LAW,

AND RELIGIOUS

IN THIS DELIGHTFUL REGION OF E
Accessibility

• The HathiTrust statement of accessibility can be found here: https://www.hathitrust.org/accessibility

• The digital library platform meets the following web accessibility guidelines:
  • United States Federal Section 508 Standards
  • WCAG 2.1 at level A and AA
  • European Union “Accessibility requirements suitable for public procurement of ICT products and services in Europe” EN 301 549
HathiTrust Research Center

• Anu Moorthy, Georgia Institute of Technology
Discovery of HathiTrust Materials

• Widgets

• Discovery on ILS (Primo, Summon)
  • Authentication
    • Shibboleth (preferred), IP offered
    • User access controls
    • Metadata rigor

• Accessibility
  • Compliance of 508
  • Other services
Discovery - Embed Widgets

Catalog Search With Multiple Options

Find

All Fields

Full view only

All Libraries
Discovery (Primo VE)
Embed Hathitrust Books

**Flip view**

To embed the flip book, simply make the iframe wider than it is tall, e.g.,

**HTML Code**

```
<iframe src="https://hdl.handle.net/2027/wu.89004292751?urlappend=%3Bui=embed" height="450" width="700"></iframe>
```

**Display**

![Flip view image](https://hdl.handle.net/2027/wu.89004292751?urlappend=%3Bui=embed)
HathiTrust's Access and User Control

• HathiTrust follows rigorous protocol for access rights
• Copyright status for the items
• Giving attributes to the digitizing agents
• Tracking changes to rights (continuously update public domain)
• User Control – Correct user to access the material without violating copyright
• Metadata control

Source: https://www.hathitrust.org/access_use
Metadata quality control

Zephir uses a list of volume identifiers from the University of Michigan representing digital objects ingested to the HathiTrust repository the previous day to determine which bibliographic records are included in its daily export to the University of Michigan for further processing and use in HathiTrust's access systems.

[Diagram of metadata quality control process]

https://www.hathitrust.org/zephir
Accessible Text Request Service Configuration

Set up this service at your university or library

• Designated contact
• Signed Agreement
• Configure secure access (IP or Proxy)
• Annual Renewal – to maintain this service

This service can help minimize the work that your disability office does to scan and convert print books into accessible formats.
HathiTrust's Accessibility Control

• Accessible Text Request Service Configuration
• ATRS is available to any HathiTrust member
• Patron eligibility is determined by the institution
• Copyright laws must be adhered
• Personal use only and may not be shared or copied for anything other than personal use
• Institutions are responsible for misuse

Source: https://www.hathitrust.org/accessibility#accessible-texts
HathiTrust Research Center (HTRC)

• HTRC enables computational analysis of the HathiTrust collection
  • Worksets (user created HathiTrust collection treated as data)
    • Text Analysis Algorithms
  • Datasets
    • Can be used for Data Capsules
  • Data Capsules
    • Secure data analysis and visualization tools

![HTRC Dataset](image)
HTRC Worksets

- Curated Worksets created by real researchers on the subject
  - Recommended Worksets to explore
  - User created Collection treated as Data
  - Worksets may contain In-copyright OR public domain
  - Shareable and citable for reproducibility

- Scholar-Curated Worksets for Analysis, Reuse & Dissemination (SCWAReD)
  - The Black Book Interactive Project in HathiTrust

- Workset Algorithms
  - HTRC Algorithms on Worksets
  - Text Analysis
    - No programming required
HathiTrust Workset Builder
Visualizations of Datasets

- Downloadable
- Non-consumptive book data
- Copyright protected
- Volume level metadata
- Page level metadata
Visualization of Datsets

This visualization tool graphs the usage of words (whether in description or dialogue) over time, distinguishing that usage both by the gender of the fictional characters the terms are associated with, and by the gender of the authors who used them. It draws on text from 110,041 volumes of fiction from the HathiTrust Digital Library published between 1800 and 2007. Publication years are plotted on the x-axis; the y-axis shows the frequency of the searched-for word (e.g., "heart") as a ratio of its occurrences in each year and gender category out of all the words used in either the associated description or associated dialogue for each year. The feminine and masculine gender categories are plotted on separate lines.
Data Availability and Datasets

• HTRC derived Datasets
• Bibliographic API
• Data API
• Open Archive Initiative (OAI feed)
• Hathifiles (metadata in Tab Delimited file)
• HathiTrust and OCLC records
• Renewal ID file (for copyright registration numbers)
Explore – Fun Tools

• HathiTrust+Bookworm
  • Create a line graph showing word use trends in 17.1 million HathiTrust volumes.

• HathiTrust+Bookworm Playground
  • Create a bar chart, map, or heatmap of word use trends in 17.1 million HathiTrust volumes.

• HTRC Feature Reader
  • Python library for interacting with HTRC Extracted Features.
HathiTrust Bookworm

Word use trends from the 17.1 million HathiTrust Volumes
Bookworm Map

See where a word occurs in the 17 million volume HathiTrust collection.

Locations correspond to the places that volumes were published in.

Search For a Term

Cheque

Combine search words with a comma. Only single word queries supported.

Optional: Compare to another term

Check

Update Words

Type of Map

Scatter

Map Scope

World

Example Books

Choose a place on the map to see matching books from there. All search and compare words included in matches.

* Nothing selected

Geographic word use trends from the 17.1 million HathiTrust Volumes
Did you know?

- More than a 17 million items available in various languages (including 17,000 Ukranian)
- More than 6 billion pages
- Cost to Digitize
  - Cost per volume is less than $0.25 for 2023
- 1 million Dollar Grant
  - from Mellon Foundation in Feb 2023
- Other astounding Stats
  - Publication dates from 1500 to 2015
- Save the Date
  - Community Week – July 10 – 14, 2023
<table>
<thead>
<tr>
<th>Collection Features</th>
<th>HathiTrust</th>
<th>Google Books</th>
<th>Internet Archive Texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Volumes</td>
<td>17.4+ million</td>
<td>40 million</td>
<td>33+ million</td>
</tr>
<tr>
<td>Reading Access</td>
<td>39% of volumes</td>
<td>Public domain volumes,</td>
<td>100% of volumes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>snippets, and content</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>permitted by publishers</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>and rights holders</td>
<td></td>
</tr>
<tr>
<td>UC Volumes</td>
<td>4.5 million UC volumes</td>
<td>4.3 million UC volumes</td>
<td>200K UC volumes</td>
</tr>
<tr>
<td></td>
<td>scanned by Google, Internet</td>
<td>scanned by Google</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Archive, and UC campuses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content Types</td>
<td>Books, serials</td>
<td>Books, serials</td>
<td>Books, serials, texts</td>
</tr>
<tr>
<td>Content Origin</td>
<td>From over 60 (mostly U.S.)</td>
<td>From Google's Library Project and its Partner Program (for publishers and authors)</td>
<td>From Internet Archive partners or uploaded by users</td>
</tr>
<tr>
<td></td>
<td>research libraries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highlights</td>
<td>Great for historical and genealogical research. Provides computational research support for UC researchers via HTRC</td>
<td>Greatest breadth of content. Snippet views in search results. At least 20% of text available for reading in Publisher Project partner volumes.</td>
<td>Includes “digitize and lend volumes”. Has read-aloud feature. Great for reuse of book images.</td>
</tr>
<tr>
<td>Download Access</td>
<td>Full view volumes only: UC affiliates may login and download via PDF, EPUB, Text (.txt), Text (.zip), or Image (JPEG). Anyone may download 1 page at a time via PDF, .txt, or image (JPEG)</td>
<td>Everyone may download public domain volumes. Format: PDF, EPUB</td>
<td>Everyone may download full versions of all volumes. Format: Many options available</td>
</tr>
</tbody>
</table>

Did you know?
Future of HathiTrust

HathiTrust in the Future with partnership with Athenaeum21
Strategic Visioning with input from HathiTrust member libraries
Q1 – Discovery and Listening
Q2 & Q3 – Visioning
Q4 – Sharing of information to the community
Useful links

- HathiTrust Homepage: https://www.hathitrust.org
- HathiTrust Help Page: https://www.hathitrust.org/help_digital_library
- HathiTrust 101 Video Series: https://www.youtube.com/playlist?list=PLtYBrz1-79lW0Cf9zZkCYDLXLf6Tu8rDB
- Member Toolkit: https://www.hathitrust.org/member-toolkit
- Search Box Widgets: https://www.hathitrust.org/widgets
Questions?

Chris Stotelmyer, Head of Electronic Resources
Clayton State University

Chandra Jackson, Electronic & Continuing Resources Librarian
University of Georgia

Jim Rickerson, Electronic Resources Librarian
Clayton State University

Anu Moorthy, Electronic Resources Librarian
Georgia Institute of Technology