HathiTrust and collaborative collection services

Weltin, Heather
2019

Dostupný z http://www.nusl.cz/ntk/nusl-407856

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.
Licence Creative Commons Uveďte původ-Neužívejte komerčně-Nezpracovávejte 4.0

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).
Datum stažení: 03.11.2023

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz.
HathiTrust Collections & Collaborative Collection Services

16th IFLA ILDS Conference
Heather Weltin, HathiTrust Shared Print Program Officer
October 11, 2019
Today’s Plan

1. HathiTrust
   - History
   - Mission and Vision
   - Membership

2. Digital Collection

3. Collaborative Programs

4. Access and Use
   ... or Beyond the Paywall in the HathiTrust Realm
Google Is Adding Major Libraries to Its Databases

By JOHN MARKOFF and EDWARD WYATT  DEC. 14, 2004

Google, the operator of the world's most popular Internet search service, plans to announce an agreement today with some of the nation's leading research libraries and Oxford University to begin converting their holdings into digital files that would be freely searchable over the Web.

It may be only a step on a long road toward the long-predicted global virtual library. But the collaboration of Google and research institutions that also include Harvard, the University of Michigan, Stanford and the New York Public Library is a major stride in an ambitious Internet effort by various parties. The goal is to expand the Web beyond its current valuable, if eclectic, body of material and create a digital card catalog and searchable library for the world's books, scholarly papers and special collections.

October 13, 2008

HathiTrust launches with 12 founding members from the Committee on Institutional Cooperation and the 11 libraries of the University of California.
Mission and Vision

Mission
The mission of HathiTrust is to contribute to research, scholarship, and the common good by collaboratively collecting, organizing, preserving, communicating, and sharing the record of human knowledge.

Vision
HathiTrust will be a vital catalyst for emerging forms of research, teaching, and learning that engage the scholarly and cultural record.
Membership

Eligibility
Academic and research institutions worldwide.

Requirements

● Required
  ○ Submission of holdings data on physical collections.
  ○ Implementation of SAML-based authentication system for access services.
  ○ Annual fees.

● Not required, but encouraged
  ○ Deposit of collections
  ○ Participations in working groups, governance
Membership Across the Globe
HathiTrust Infrastructure
HathiTrust Collection
Collections - Currently Digitized

- 16,987,842 total volumes
- 8,237,760 book titles
- 454,195 serial titles
- 5,945,744,700 pages
- 761 terabytes
- 201 miles
- 13,803 tons
- 6,460,113 volumes (~38% of total) in the public domain
Top 10 Contributors to HathiTrust, July 2019

<table>
<thead>
<tr>
<th>Library</th>
<th>Deposited Volumes</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of California Libraries</td>
<td>235,605</td>
</tr>
<tr>
<td>University of Illinois</td>
<td>97,793</td>
</tr>
<tr>
<td>University of Michigan</td>
<td>56,468</td>
</tr>
<tr>
<td>Ohio State</td>
<td>54,688</td>
</tr>
<tr>
<td>Michigan State University</td>
<td>43,334</td>
</tr>
<tr>
<td>University of Minnesota</td>
<td>22,365</td>
</tr>
<tr>
<td>Indiana University</td>
<td>14,402</td>
</tr>
<tr>
<td>University of Virginia</td>
<td>13,012</td>
</tr>
<tr>
<td>Cornell</td>
<td>10,258</td>
</tr>
<tr>
<td>University of Texas</td>
<td>7,807</td>
</tr>
</tbody>
</table>
Collections by Publication Date

Publication Dates as of October 2018
Collections by Language (Top Ten)
Subjects of HathiTrust Content - as derived from Library of Congress Classification classes

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
<th># of Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>GENERAL WORKS</td>
<td>52,925</td>
</tr>
<tr>
<td>B</td>
<td>PHILOSOPHY. PSYCHOLOGY. RELIGION</td>
<td>333,191</td>
</tr>
<tr>
<td>C</td>
<td>AUXILIARY SCIENCES OF HISTORY</td>
<td>59,964</td>
</tr>
<tr>
<td></td>
<td>WORLD HISTORY AND HISTORY OF EUROPE, ASIA, AFRICA, AUSTRALIA, NEW ZEALAND, ETC.</td>
<td>540,697</td>
</tr>
<tr>
<td>D</td>
<td>HISTORY OF THE AMERICAS</td>
<td>96,330</td>
</tr>
<tr>
<td>E</td>
<td>HISTORY OF THE AMERICAS</td>
<td>278,174</td>
</tr>
<tr>
<td></td>
<td>GEOGRAPHY. ANTHROPOLOGY.</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>RECREATION</td>
<td>111,267</td>
</tr>
<tr>
<td>G</td>
<td>SOCIAL SCIENCES</td>
<td>613,535</td>
</tr>
<tr>
<td>H</td>
<td>POLITICAL SCIENCE</td>
<td>141,800</td>
</tr>
<tr>
<td>J</td>
<td>LAW</td>
<td>108,474</td>
</tr>
<tr>
<td>K</td>
<td>EDUCATION</td>
<td>120,108</td>
</tr>
<tr>
<td>L</td>
<td>MUSIC AND BOOKS ON MUSIC</td>
<td>115,180</td>
</tr>
<tr>
<td>M</td>
<td>FINE ARTS</td>
<td>165,334</td>
</tr>
<tr>
<td>N</td>
<td>LANGUAGE AND LITERATURE</td>
<td>1,105,730</td>
</tr>
<tr>
<td>P</td>
<td>SCIENCE</td>
<td>314,816</td>
</tr>
<tr>
<td>Q</td>
<td>MEDICINE</td>
<td>129,411</td>
</tr>
<tr>
<td>R</td>
<td>AGRICULTURE</td>
<td>113,591</td>
</tr>
<tr>
<td>S</td>
<td>TECHNOLOGY</td>
<td>217,405</td>
</tr>
<tr>
<td>T</td>
<td>MILITARY SCIENCE</td>
<td>26,615</td>
</tr>
<tr>
<td>U</td>
<td>NAVAL SCIENCE</td>
<td>10,330</td>
</tr>
<tr>
<td>V</td>
<td>BIBLIOGRAPHY. LIBRARY SCIENCE.</td>
<td></td>
</tr>
<tr>
<td>Z</td>
<td>INFORMATION RESOURCES (GENERAL)</td>
<td>94,230</td>
</tr>
</tbody>
</table>

Note: LCC classes are only present in 60% of HathiTrust records.
Collaborative Programs
Collaborative Programs - User Support

In order to efficiently and effectively serve the needs of users, and ensure the highest level of function of HathiTrust services, HathiTrust has charged a multi-institution user support team.
Collaborative Programs - Copyright Review Program

The Copyright Review Management System (CRMS) has enabled large-scale, distributed investigations into the copyright status of works in the HathiTrust collection.
Collaborative Programs - U.S. Federal Documents Program

Mission:
Through coordinated and collective action, expand and enhance digital access to U.S. federal publications including those issued by GPO and other federal agencies.

As of July 1, 2019

- 542,731 bibliographic records
- 1,353,343 digital objects
- 509,600 monographs
- 32,619 serial titles
Collaborative Programs - Shared Print

Goal
The primary goals of the HathiTrust Shared Print Program are to ensure preservation of print and digital collections by linking the two, to reduce overall costs of collection management for HathiTrust members, and to catalyze national/continental collective management of collections.

As of May 2019:

- 79 retention libraries participating in HathiTrust Shared Print Program
- 17,889,303 single-part monographs
- 5,476,598 individual titles
Collaborative Programs - HathiTrust Research Center

Featured Services

Extracted Features
An unrestricted dataset of metadata and word counts for each page in the HathiTrust Digital Library. Download and explore on your own machine.

Text Analysis Algorithms
Web-based, click-and-run tools that perform computational text analysis on worksets, which are user-created collections of volumes. No programming required.

Data Capsules
Secure virtual environments for non-consumptive text analysis, where researchers can implement their own data analysis and visualization tools.
Collaborative Programs - Projects and Grants

Global Registry of Digitized Texts Project

Project to develop an international registry of digitized texts enabling discovery and use.

Federated Module for Accessible Material for Education (FRAME)/Accessibility Grant

Project looking at facilitating access to education materials for students with print disabilities
Access and Use or Beyond the Paywall
Viewing and Downloading

Anyone Anywhere Can:

- Read public domain and open access works (via web)
- Search across the full text of the entire collection (via web)
- Receive a dataset of public domain content (under certain conditions)
- Mine text and data (via HTRC)

Members Can:

- Download public domain and open access works
- Provide access to their users who are blind or with print disabilities
- Get access to replacement copies for lost and damaged print copies (under certain conditions)
- Mine text of copyrighted content (in pilot mode via HTRC)
Providing Access to All Users

● Accessibility
  ○ HathiTrust Accessible Text Request Service
  ○ National Federation of the Blind (NFB) User Access to HathiTrust Copyrighted Collection
  ○ Marrakesh Treaty
International Access and Use

- Membership for International Libraries
- International Access
- Full PDF Downloads
Future Things

- Data Analysis and Collection Management Support
- Investigate Discovery to Delivery for Print Materials
- Policy on Metadata Sharing and Use
- Expanding Lawful Access to Works
Mission and Vision

**Mission**
The mission of HathiTrust is to contribute to research, scholarship, and the common good by collaboratively collecting, organizing, preserving, communicating, and sharing the record of human knowledge.

**Vision**
HathiTrust will be a vital catalyst for emerging forms of research, teaching, and learning that engage the scholarly and cultural record.
Questions & Thank You!

Heather Weltin
HathiTrust Shared Print Program Officer
weltin@hathitrust.org