

HATHI
TRUST



HathiTrust数字图书馆： 发展简介和远景展望

杨继东

密歇根大学亚洲图书馆

(Asia Library, University of Michigan)

缘起与发展

- 2004: “谷歌图书项目”（Google Books Project）
- 2008: 美国中西部大学十所大学组成的CIC联盟和加州大学(UC)系统宣布建立HathiTrust数字图书馆。
- 建立HathiTrust的目的。
- HathiTrust名称的由来。
- 2008-2012: 更多大学图书馆加入。
- 最新统计显示，HathiTrust目前已有的数字化图书达550多万种，1050万册，27万多种期刊，大约30%的内容已经进入公有领域。

成员馆

- 联盟馆：CIC联盟(中西部10所大学加上芝加哥大学)、UC(加州大学)系统各馆、TRLN联盟(杜克、北卡等大学)
- 其他成员馆：波士顿、哥伦比亚、康奈尔、哈佛、达特茅斯、麻省理工、麦积尔、纽约、西北、普林斯顿、斯坦福、约翰·霍普金斯、普渡、亚利桑那、宾夕法尼亚、圣母、耶鲁、匹兹堡、弗吉尼亚、华盛顿、特拉华、马里兰、佛罗里达、密苏里、康涅狄格等大学的图书馆以及国会图书馆和纽约公共图书馆。

内容的来源、储存和管理

- 三大来源：谷歌图书项目、Internet Archive、各校独自开展的数字化项目
- 所有数据均存于密歇根大学图书馆，该校有专门团队负责数据维护工作。另在印第安纳大学有一套备份数据。
- 专业化的数据集中管理和保存是HT吸引越来越多大学图书馆入盟的重要原因。
- HT网站有专门的OPAC目录，包括所有数字内容的元数据（马克数据）。各成员馆的OPAC系统也包含全部或部分HT元数据。

组织与管理

- 2008-2012: 执行委员会和策略顾问组
- 2011年10月: “立宪大会”在华盛顿召开
- 2012年3月: 由12个成员组成的董事会成立。董事会随即又通过选举方式成立了有5个成员的执行委员会。
- 费用分摊方式: 目前每馆每年以\$3.86/GB的比率承担费用, 明年开始采用新的更复杂的计算公式, 考虑多种因素(是否有纸本, 区分公有领域和有版权限制的著作等)。

获取和使用权限

- 一般公众: 可以在HT网站检索OPAC和全文, 在网上浏览已进入共有领域的书刊。
- 成员馆的用户: 除了一般公众的权限外, 还可以下载公有领域内的书刊, 并建立自己的个性化书库(collection)。

主页

The screenshot shows the HathiTrust Digital Library homepage. The browser window title is "HathiTrust Digital Library | Millions of books online - Mozilla Firefox". The address bar shows "www.hathitrust.org". The page features a navigation menu with "Home", "About", "Collections", and "My Collections". The main content area is divided into three columns: "Catalog Search", "Full-text Search", and "Collections". Each column has a search form with a "Find" button and a "Full view only" checkbox. Below the search forms are three sections: "Recent News and Publications", "Currently Digitized", and "Featured Collection". The "Recently Digitized" section lists statistics such as total volumes, book titles, serial titles, pages, and terabytes. The footer contains links for "About", "Help", "Feedback", "Take-Down Policy", "Privacy", and "Contact".

HathiTrust Digital Library | Millions of books online - Mozilla Firefox

File Edit View History Bookmarks Tools Help

HathiTrust Digital Library | Milli... +

www.hathitrust.org

HATHI TRUST Digital Library

Home About Collections My Collections

Catalog Search

Search information **about** the items.

All Fields Full view only

[Advanced Catalog Search](#) | [Search Tips](#)

Full-text Search

Search words that occur **within** the items.

Full view only

[Advanced Full-text Search](#) | [Search Tips](#)

Collections

Browse, search, or make HathiTrust collections.

Collections are a way to group items for private or public use. The full-text of items within a collection can be searched independently of the full digital library. Read more about [creating collections](#).

Recent News and Publications

[Board of Governors Election Results](#)

[Try our new mobile website!](#)

[Information about the Authors Guild Lawsuit](#)

- [Update on July 2012 Activities](#)
- [Getting everything you want from HathiTrust - The Stone and the Shell](#)
- [HathiTrust And TRAC](#)
- [HathiTrust: On TRAC](#)
- [Update on June 2012 Activities](#)

[view more news >>](#)

Currently Digitized

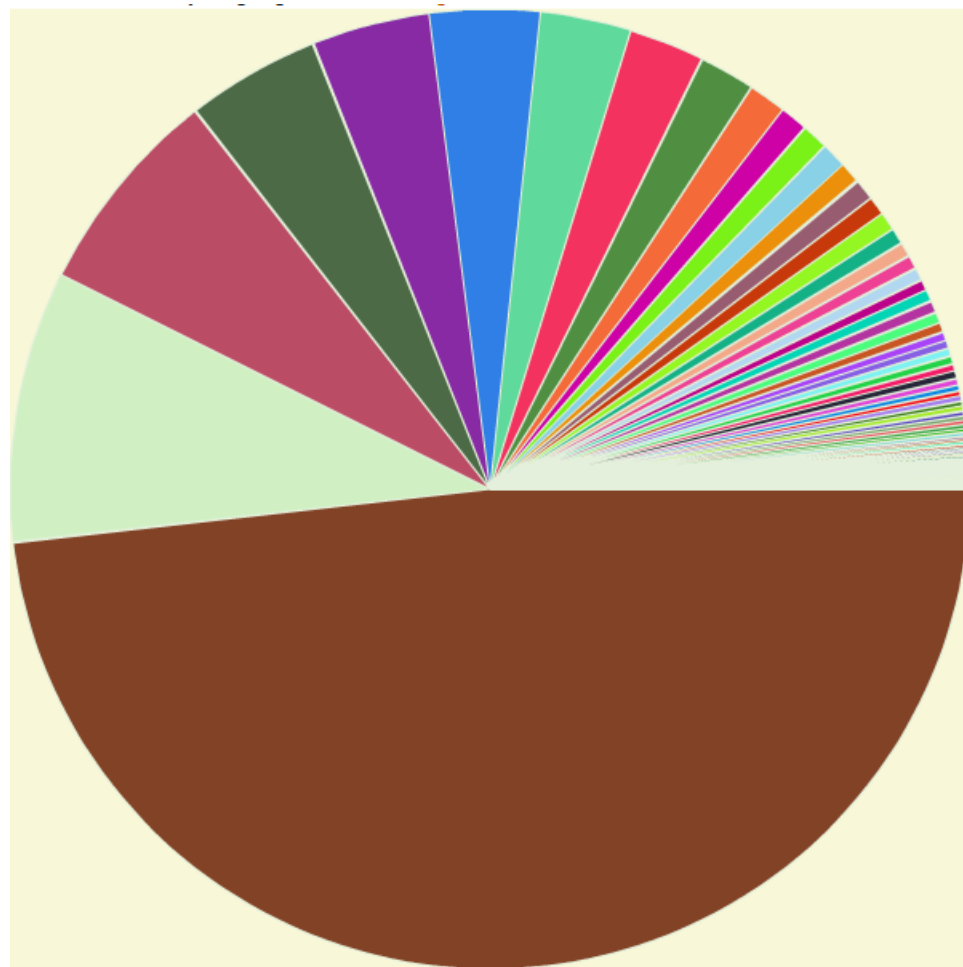
10,497,399 total volumes
5,533,530 book titles
273,684 serial titles
3,674,089,650 pages
471 terabytes
124 miles
8,529 tons
3,189,660 volumes (~30% of total) in the public domain
[View visualizations of HathiTrust call numbers, languages, and dates](#)
[statistics information >>](#)

Featured Collection

About Help Feedback Take-Down Policy Privacy Contact

内容（按语言分）

- 英文：48.22%
- 德文：9.17%
- 法文：7.09%
- 西班牙文：4.48%
- 中文：3.96%
- 俄文：3.76%
- 日文：3.08%
- 意大利文：2.53%
- 阿拉伯文：1.88%
- 拉丁文：1.31%
- 未确定：0.93%
- 波兰文：0.89%
-



中文内容

- HT中的中文书刊目前大概有40万册。其中大多来自密歇根大学和加州大学。绝大多数部分由谷歌扫描制作，并经过OCR识别。
- OCR识别完全依靠机器，没有经过人工校对，准确率极低。尽管近几年有改善趋势，但是还是远远达不到实用的水平。
- 很多直排本书籍出现顺序和页面颠倒现象。因此尽管有不少进入公有领域的中文古籍资源，基本上无法使用。

版权管理

- HT的法律基础：美国版权法(Copyright Act)的第107条(合理使用)和108条(图书馆和档案馆的复制权)。
- 公有领域(public domain)的定义：1923年以前出版的美国书籍以及1872年以前出版的国外书籍(在美国境内可以全文阅读1923年以前的外文书籍)，以及由作者授权公开的书籍(比如加州大学出版社的许多出版物)。
- 出版年份由图书编目记录决定，存在不准确的情况。

版权争议

- “谷歌图书”项目的版权争议：美国作家公会 (Authors Guild) 与谷歌的官司纠纷。
- 2011年9月美国作家公会、澳大利亚作家协会、魁北克作家公会和八位作家联合起诉HT，目前官司正在进行中。
- HT曾于前年启动“孤儿图书计划” (Orphan Book Project)，但是在遭到起诉后停止。

前景

- 尽管受到法律上的挑战，HT作为一项受到北美大多数主流高校支持的数字图书馆项目，应该不会终止，将继续存在、发展并吸收更多成员馆。
- 与美国公共数字图书馆(Digital Public Library of America)的合作(以及竞争)。
- HT对大中华地区高校图书馆的意义。