

Hathi Trust 101

日本語資料との関連性において

Keiko Yokota-Carter University of Michigan

- **Hathi Trust Digital Library 101 基本的情報**
- **Google project との違い**
- **日本語文献の例**
 - 課題
 - 将来への問題提起

2013 European Association of Japanese Resource Specialists annual meeting,
Paris

Hathi Trust 101 日本語資料との関連性において

2013 European Association of Japanese Resource Specialists annual meeting, Paris



Hathi (hah-tee)

象 ヒンディー語
 巨大
 永遠の記憶
 知恵
 安全
 信頼できる

Mission 使命



To contribute to the common good by collecting, organizing, preserving, communicating, and sharing the record of human knowledge.

「人類の知識の記録の収集、組織化、保存、伝達、共有により、公益に貢献すること」

(田中敏訳 カレントアウェアネス-ENo.230 2013.01.24)

基本概念

4

2008 年に開始

- Depository — ミシガン大学・インディアナ大学
- 法務財務管理 — ミシガン大学にオフィスがある

参加校による協同事業

- Community: 協同事業 — 資料保存の共有と利用
- Scale: 大規模な協同作業 —> universal library
- Preservation and Access: 電子と紙の資料
- Openness: 法的な資料の利用（閲覧、印刷、ダウンロード）
協同作業による保管
オープンソースを可能にする技術の協同構築

図書館のプロジェクト （出版社による電子化資料は含まない）

70校以上 （国際的に広がりつつある）

各校が独自で、あるいはグーグルプロジェクトで電子化した資料を



共同で管理・保存し、利用する

- 10.7 million volumes 1070万冊 (2013 6月)
- 5.6 million book titles 560万タイトル
- 276,000 serial titles 27万6千タイトル
- 3.4 million public domain volumes (~32%) 340万冊

- すべての資料の書誌データ、フルテキストの検索が可能

Mark format 書誌データのない資料はHathi Trust には含まない

Mark format 書誌データが提供できることがHathi Trust への入会規則

Q : 米国以外の図書館の電子資料は？

A : 基本的書誌情報があれば可能 **HathiTrust事務局 に相談要**

- Discovery Tool (ProQuest, EBSCO, OCLC, Ex Libris) 内で検索可能
U of Michigan ProQuest – Summon で検索可能

HathiTrustのコンテンツを含むかどうかは各参加校がDiscovery Tool会社と検討すること

- 著作権法を遵守した利用 (検索、閲覧、ダウンロード・印刷)

70 以上のパートナー校

Arizona State University
Baylor University
Boston College
Boston University
California Digital Library
Columbia University
Cornell University
Dartmouth College

Duke University

Emory University
Florida State University
Getty Research Institute
Harvard University Library
Indiana University
Johns Hopkins University
Lafayette College
Library of Congress
MIT

McGill University`
Michigan State University
New York Public Library
New York University
North Carolina Central

North Carolina State
Northwestern University
The Ohio State University
The Pennsylvania State
Princeton University
Purdue University
Stanford University
Texas A&M University
Universidad Complutense de
Madrid

University of Arizona
University of Calgary
University of California

Berkeley
Davis
Irvine
Los Angeles
Merced
Riverside
San Diego
San Francisco
Santa Barbara
Santa Cruz

University of Chicago

University of Connecticut
University of Florida
University of Illinois
University of Illinois at Chicago
The University of Iowa
University of Maryland
University of Miami
University of Michigan

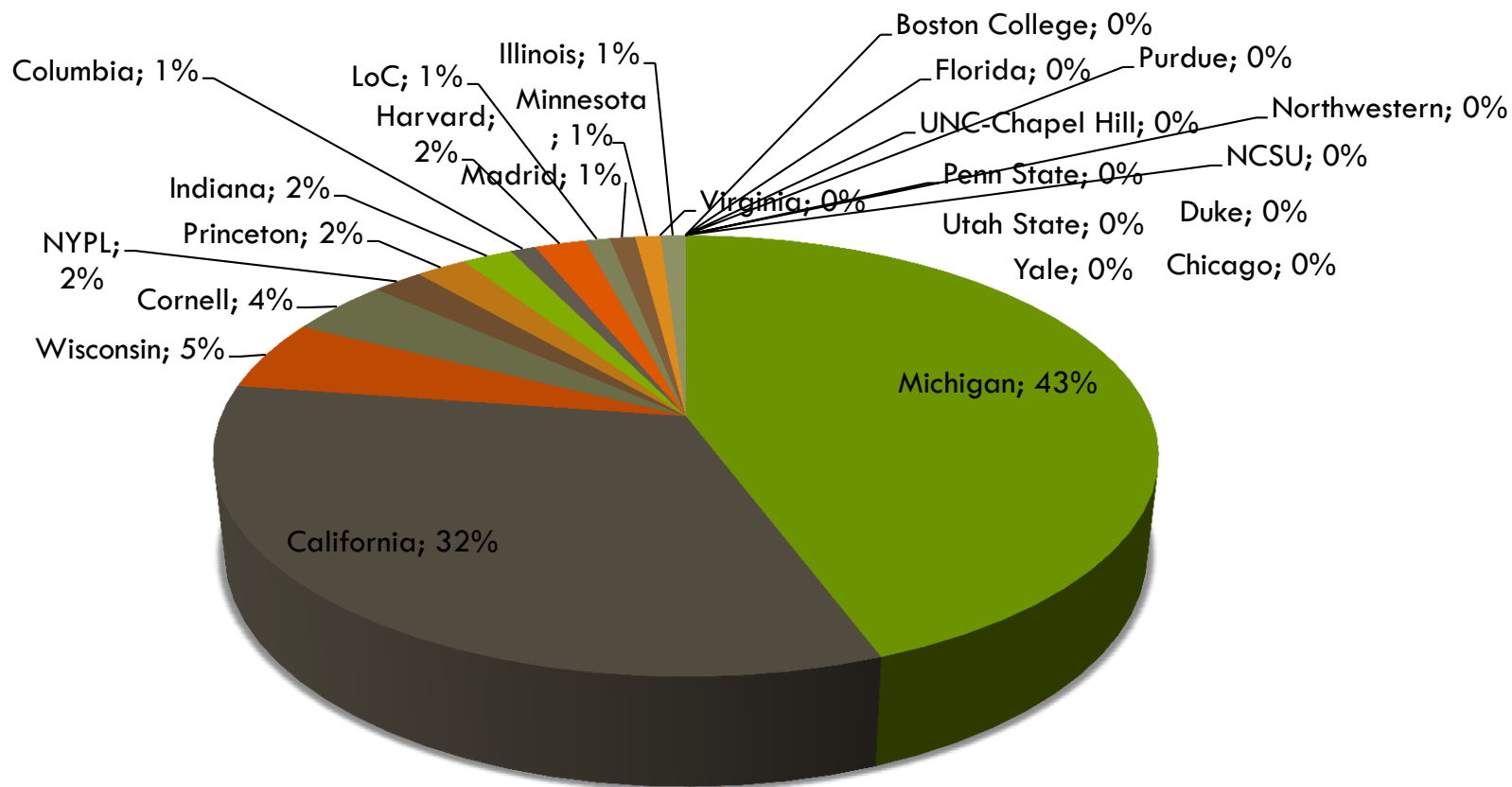
University of Minnesota

University of Missouri
University of Nebraska-Lincoln
University of North Carolina at Chapel Hill
University of Notre Dame
University of Pennsylvania
University of Pittsburgh
University of Utah
University of Virginia
University of Washington
University of Wisconsin-Madison
Utah State University
Washington University
Yale University Library

各図書館提供の資料分布

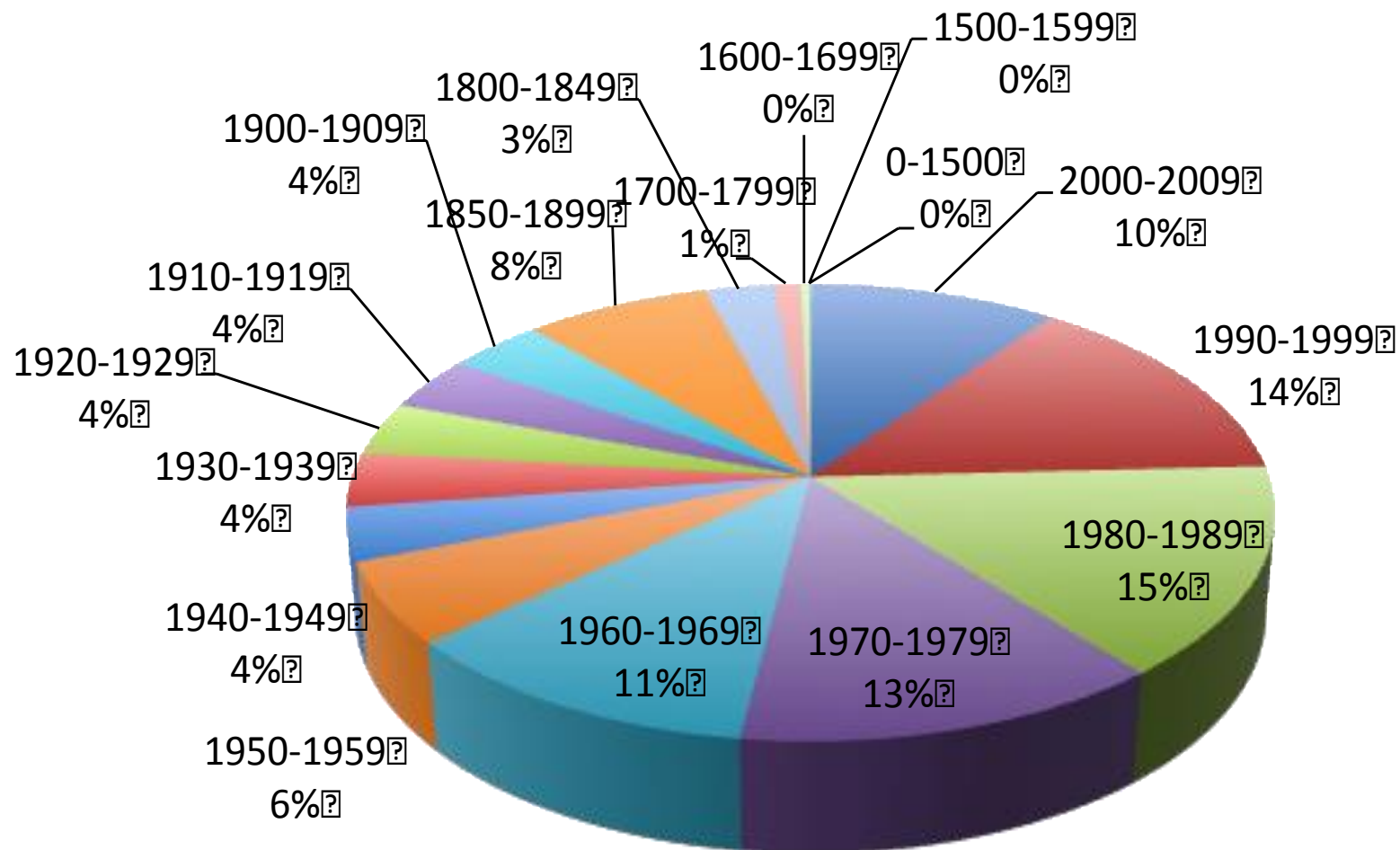
4万冊レベルは0%として表示

8



年代別

1923年 以前の出版物 ~ 全体の 20%

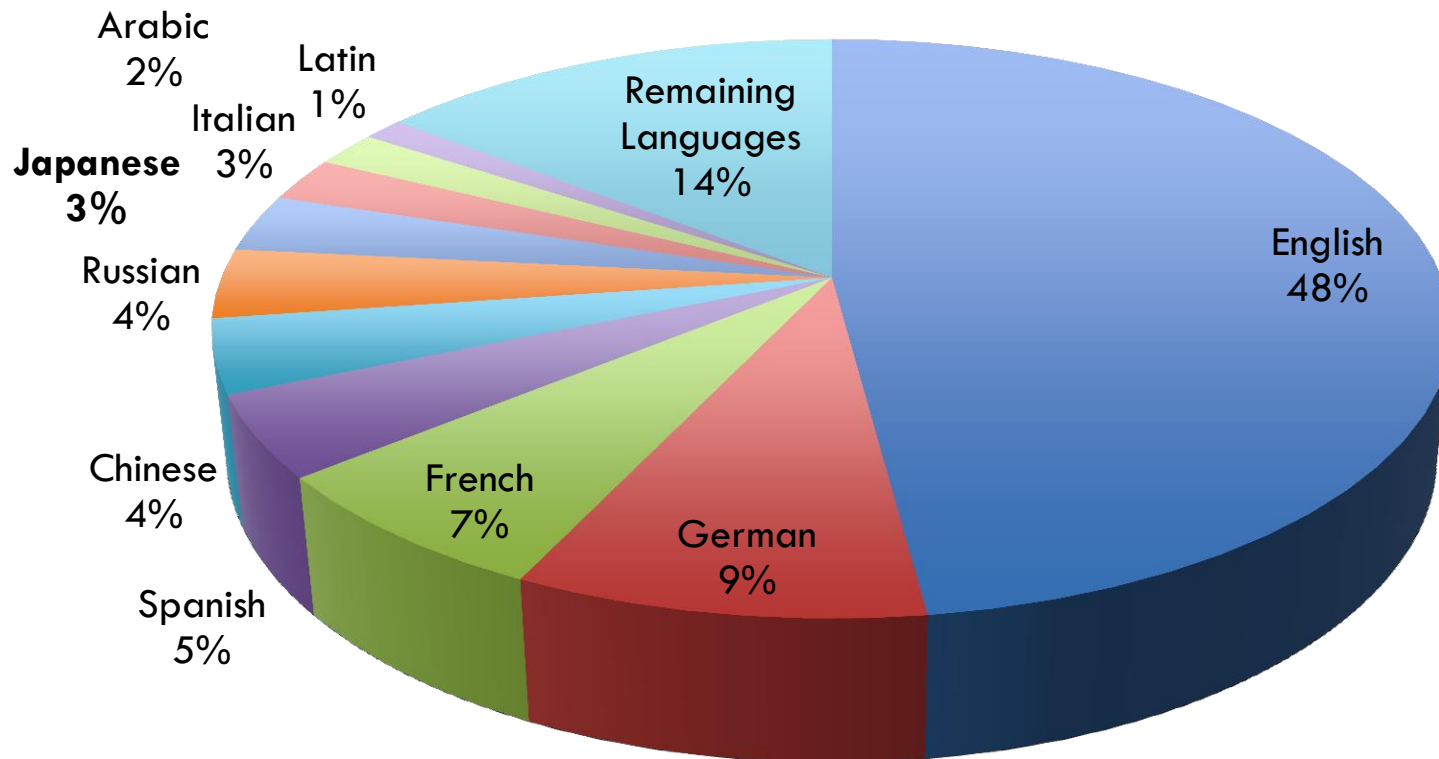


言語別

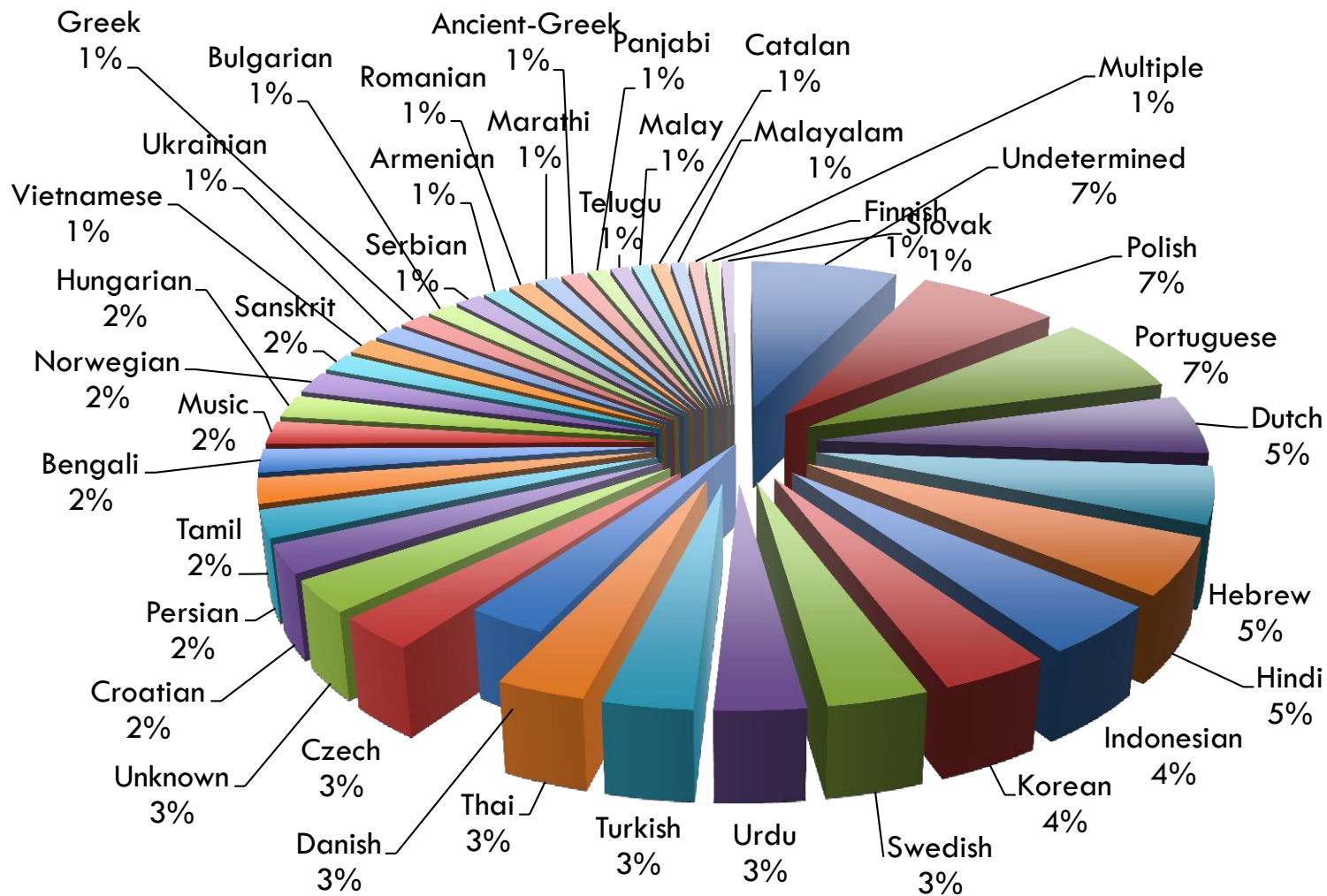
日本語資料は全体の 3 % 187,672冊

The top 10 languages make up ~86% of all content.

10



次の 40 言語は 全体の13%



Google と HathiTrust の違い -- 組織とMission 使命

□ **Google = Corporation** **Access**

“To organize the world's information and make it universally accessible and useful. Facilitating library access to scholarly texts brings us one step closer to this goal”.

□ **HathiTrust = Libraries** **Preserving and Access**

“To contribute to the common good by collecting, organizing, preserving, communicating, and sharing the record of human knowledge”.

Google と HathiTrust の違い -- コンテンツ

Goolge – Goolge book projecct

Hathi – 参加校が提供するすべてのコンテンツ

Goolge book projecct によるコンテンツ

+

参加校独自のコンテンツ

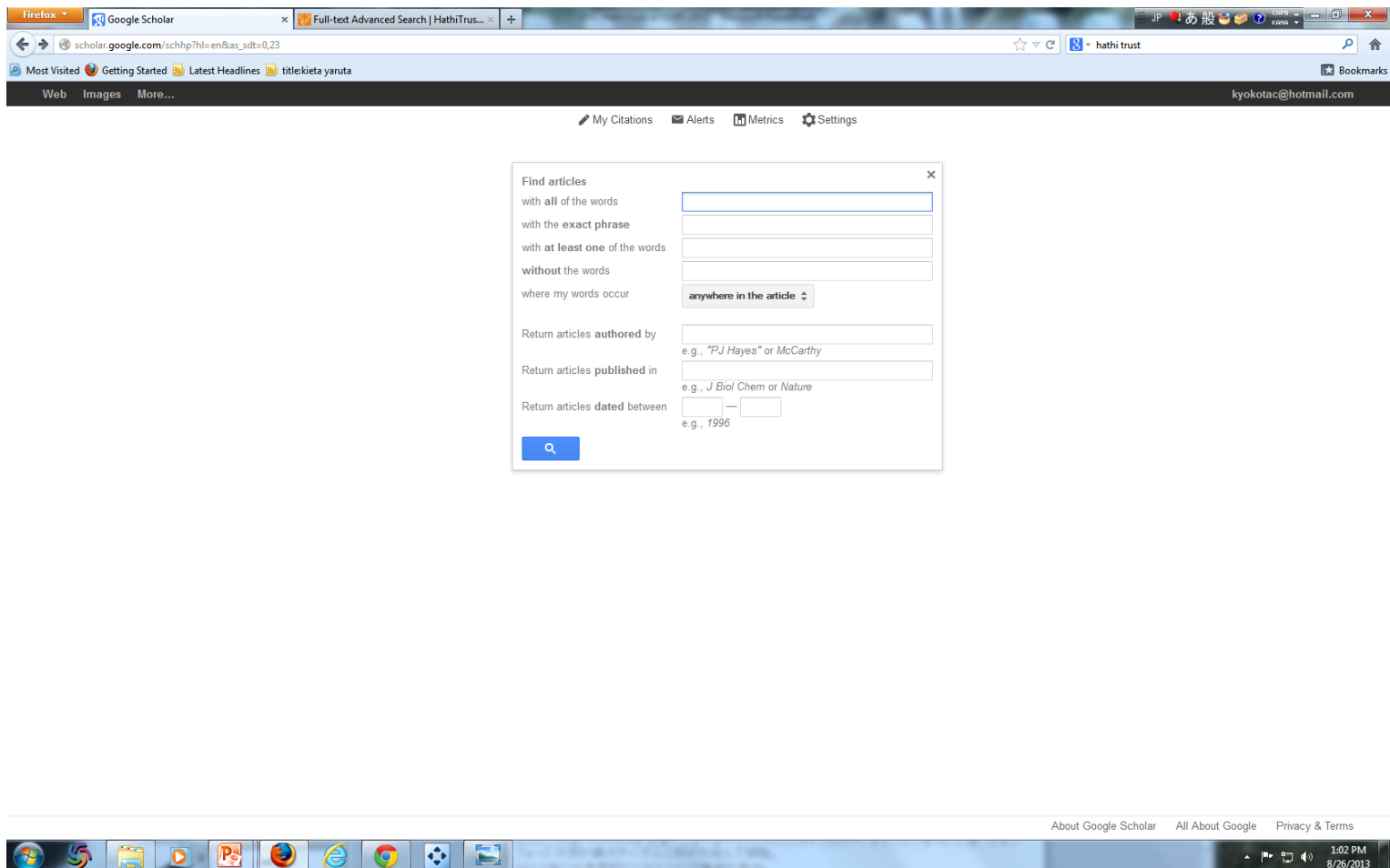
↓

Google よりも50万点以上多いコンテンツ

Google と HathiTrust の違い

書誌データの質・検索機能

Google scholar



Google 「東照宮」 で検索

Firefox | Papers and Presentations | HathiTrust... | Kojiki. Translated with an introd. and ... | 東照宮 - Google Scholar

scholar.google.com/scholar?hl=en&q=東照宮&btnG=&as_sdt=1%2C23&as_sdt=...

Most Visited | Getting Started | Latest Headlines | title:kieta yaruta

Web | Images | More...

kyokotac@hotmail.com

Google

東照宮

Scholar

About 2,380 results (0.05 sec)

My Citations

Articles

Legal documents

Any time

Since 2013

Since 2012

Since 2009

Custom range...

Sort by relevance

Sort by date

☒ include patents

☒ include citations

Create alert

Tip: Search for English results only. You can specify your search language in Scholar Settings.

いる可能性も考えられる。その他調査した市内 5カ所は全て生息を確認できなかった。
...、青葉山公園、青葉区東照宮、東照宮... - ヒメグモのオス間 ... - kakureobi.sakura.ne.jp
KISHIDAIA, No. 97, Mar. 2010
—— DRAGLIN ES—— 38 いる可能性も考えられる。その他調査した市内 5カ所は全て生息を確認できなかった。表宮城県仙台市におけるキシノウエタテグモにおける ...
All 2 versions Cite More▼

走近「日光東照宮」
韓瑞燦 - 日語知識, 2007 - 万方数据资源系统
摘要 [東照宮] は供奉(徳川家康)の神社、遍布在日本各地。其中最有名な要数[日光東照宮]。[日光東照宮] 于1617 年创建。由于是徳川家の庙堂。所以只规划了朴实的社殿。但由于第三代将军[徳川家光] 非常崇拜祖父[徳川家康]。于是在一片原始自然的杉林间。用了 ...
Cite More▼

[CITATION] 御番所日記
日光東照宮 社務所 - 1931 - 日光東照宮社務所
Cite More▼

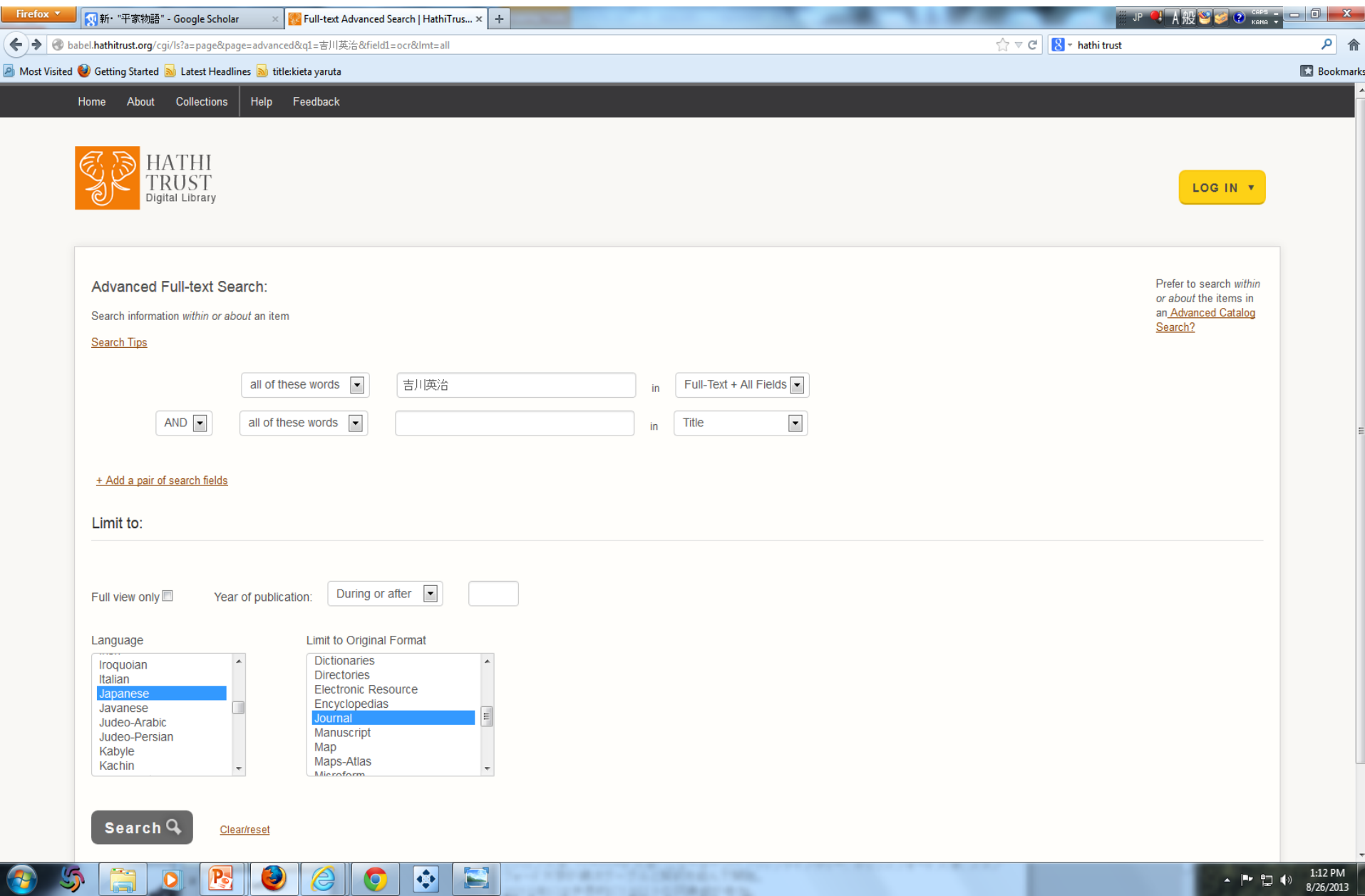
[CITATION] 日光東照宮百話
東照宮 (日光市) - 1925 - 日光東照宮社務所
Cite More▼

[CITATION] 近世仏教治国論の史料と研究--平松開運録/東照宮御遺訓
大桑資 - 2011 - buddhism.lib.ntu.edu.tw
序號: 378346. 題名: 近世仏教治国論の史料と研究--平松開運録/東照宮御遺訓。作者: 大桑資=Okuwa, Hitoshi; 平野寿則=Hirano, Toshinori. 出處題名: 卷期: 日期: 2011.01.01. 頁次: 752. 叢書名: 叢書號: 出版者: 清文堂出版。出版地: 大阪, 日本[Osaka, Japan] (http://www.seibundo- ...
Cite More▼

[CITATION] 日光東照宮の奥院寶塔中神道秘式
S Sugawara - 天台學報, 1992 - buddhism.lib.ntu.edu.tw
序號: 321303. 題名: 日光東照宮の奥院寶塔中神道秘式=A secret ceremony of Shinto in a stupa of a the depths temple in Nikko-Toshou. 作者: 菅原信海=Sugawara, Shinkai. 出處題名: ...

基本は図書館

協同で事業全体を向上させる



Gunsho ruijū / [Hanawa Hokinoichi shū] v.01.

[View full catalog record](#)

Copyright: [Public Domain in the United States; Google-digitized.](#)

Get this Book

[Find in a library](#)

[Download this page \(PDF\)](#)

[Download whole book \(PDF\)](#)

Add to Collection

[Login](#) to make your personal collections permanent

Select Collection

Add

Share

Permanent link to this book

<http://hdl.handle.net/2027/uc1.100>

Link to this page

<http://hdl.handle.net/2027/uc1.100>

Version: 2010-05-26 04:23 UTC [?](#)



第一輯 目次

群書類從第一輯目次

神祇部 凡七十種

卷第一

皇太神宮儀式帳

卷第二

止由氣宮儀式帳

卷第三

太神宮諸雜事記

卷第四

神宮雜例集

卷第五

二所太神宮例文

卷第六

內宮長曆送官符

外宮嘉祿三年山口祭記

卷第七

貞和御飭記

內宮臨時假殿遷宮記

卷第八

治承元年公卿勅使記

正應六年七月十六日公卿勅使御

第

卷第九

神風抄

卷第十

古老口實傳

詔刀師沙汰文

卷第十一

元亨元年十一月廿二日高宮假殿

附高宮御裝束奉飭日記

卷第十二

小朝熊社神鏡沙汰文

卷第十三

八幡愚童訓

HathiTrust : Limited Search – 全文閲覧不可でも全文検索する

「宮本武蔵」 （吉川栄治著） 「武蔵」 で検索

The screenshot shows a web browser window with the HathiTrust search results page. The address bar shows the URL: <http://babel.hathitrust.org/cgi/pt/search?id=wu.89006476469;view=...>. The browser tabs include 'HathiTrust Languages | HathiT...' and 'Miyamoto Musashi / Yoshi...'. The page header features the HathiTrust Digital Library logo and a search bar. Below the search bar, there are links for 'Advanced full-text search' and 'Search tips', and a 'Full view only' checkbox. The main content area is divided into two columns. The left column contains information about the book, including the author 'Miyamoto Musashi / Yoshikawa Eiji', the title '1. Yoshikawa, Eiji, 1892-1962 吉川英治, 1892-1962.', and a link to 'View full catalog record'. It also includes a 'Copyright' notice: 'Protected by copyright law.' The right column shows the search results for the term '武蔵'. It includes a message stating: 'Full view is not available for this item due to copyright © restrictions. Page numbers with matches are displayed without text snippets due to these restrictions. Snippets may be available for some items if you log in.' Below this, the search results are listed: 'Search Results: 357 pages found for 武蔵'. A link is provided to 'Broaden your search to find pages having just one or more of your terms.' The results are listed in a table format with page ranges and the number of matching terms.

« [Back to page](#)

About this Book

Miyamoto Musashi / Yoshikawa Eiji. 1. Yoshikawa, Eiji, 1892-1962 吉川英治, 1892-1962.

[View full catalog record](#)

Copyright: [Protected by copyright law.](#)

Get this Book

[Find in a library](#)

Add to Collection

[Login](#) to make your personal collections permanent

Select Collection

Share

Permanent link to this book

◀ [Go to the beginning of the book](#)

Search in this text

Full view is not available for this item due to copyright © restrictions. Page numbers with matches are displayed without text snippets due to these restrictions. Snippets may be available for some items if you log in.

Search Results: 357 pages found for 武蔵

[Broaden your search to find pages having just one or more of your terms.](#)

p.372 - 17 matching terms
p.8 - 16 matching terms
p.436 - 14 matching terms
p.555 - 14 matching terms
p.526 - 13 matching terms
p.374 - 13 matching terms
p.423 - 12 matching terms
p.534 - 12 matching terms
p.528 - 12 matching terms

アクセスと利用

Access use policies

http://www.hathitrust.org/access_use#

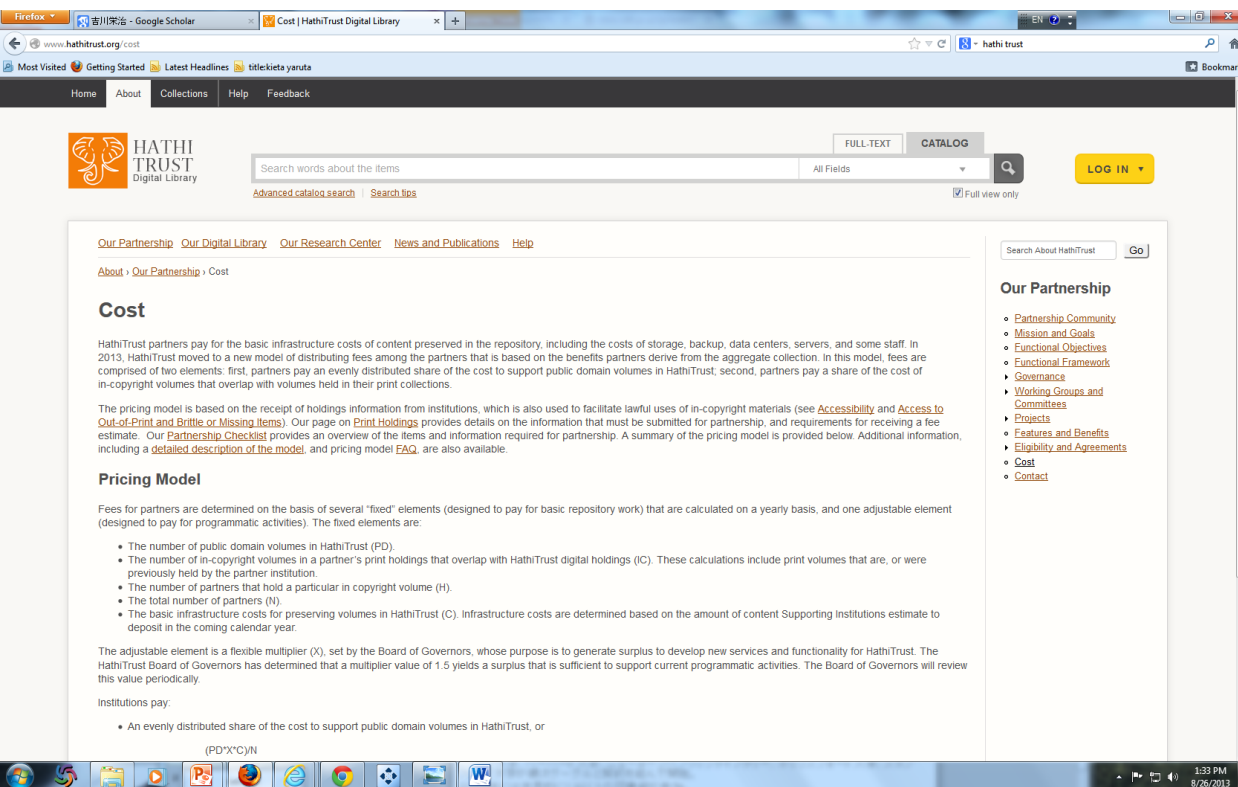
19

Type of work	Searchable (bibliographic and full-text)	Viewable*	Full-PDF download	Print on Demand	Print disabilities*	Preservation uses (Section 108)*
Public domain worldwide	Worldwide	Worldwide	Partners only if scanned by Google, if not, worldwide.	Worldwide	Partners worldwide	N/A
Public domain (US) – Non-US works published between 1872 and 1923.	Worldwide	When accessed from with the United States	Partners in the US if scanned by Google, if not, anyone US	Available within the United States	Partners in the US; partners worldwide where similar laws in effect	N/A
Works that rights holders have opened access to in HathiTrust	Worldwide	Worldwide	Worldwide (if digitized by Google, full-PDF only available if opened with CC license)	Worldwide with permission	Partners worldwide	N/A
Works that are in-copyright or of undetermined status	Worldwide	Not available	Not available	Not available	Partners in the US; partners worldwide where similar laws in effect	Partners in the US; partner worldwide where similar laws in effect

パートナーシップ コスト

各機関の所蔵数、所蔵文献内の著作権許諾数、全体数との割合等々により機関ごとに算出一多様

Hathi Trust と相談要



The screenshot shows the HathiTrust Digital Library website in a Firefox browser. The page is titled 'Cost' and is part of the 'Our Partnership' section. It details the pricing model for partners, including infrastructure costs and a flexible multiplier (X) set by the Board of Governors. The page also includes a search bar, navigation links, and a sidebar with 'Our Partnership' links.

Cost

HathiTrust partners pay for the basic infrastructure costs of content preserved in the repository, including the costs of storage, backup, data centers, servers, and some staff. In 2013, HathiTrust moved to a new model of distributing fees among the partners that is based on the benefits partners derive from the aggregate collection. In this model, fees are comprised of two elements: first, partners pay an evenly distributed share of the cost to support public domain volumes in HathiTrust; second, partners pay a share of the cost of in-copyright volumes that overlap with volumes held in their print collections.

The pricing model is based on the receipt of holdings information from institutions, which is also used to facilitate lawful uses of in-copyright materials (see [Accessibility](#) and [Access to Out-of-Print and Braille or Missing Items](#)). Our page on [Print Holdings](#) provides details on the information that must be submitted for partnership, and requirements for receiving a fee estimate. Our [Partnership Checklist](#) provides an overview of the items and information required for partnership. A summary of the pricing model is provided below. Additional information, including a [detailed description of the model](#), and pricing model [FAQ](#), are also available.

Pricing Model

Fees for partners are determined on the basis of several "fixed" elements (designed to pay for basic repository work) that are calculated on a yearly basis, and one adjustable element (designed to pay for programmatic activities). The fixed elements are:

- The number of public domain volumes in HathiTrust (PD).
- The number of in-copyright volumes in a partner's print holdings that overlap with HathiTrust digital holdings (IC). These calculations include print volumes that are, or were previously held by the partner institution.
- The number of partners that hold a particular in copyright volume (H).
- The total number of partners (N).
- The basic infrastructure costs for preserving volumes in HathiTrust (C). Infrastructure costs are determined based on the amount of content Supporting Institutions estimate to deposit in the coming calendar year.

The adjustable element is a flexible multiplier (X), set by the Board of Governors, whose purpose is to generate surplus to develop new services and functionality for HathiTrust. The HathiTrust Board of Governors has determined that a multiplier value of 1.5 yields a surplus that is sufficient to support current programmatic activities. The Board of Governors will review this value periodically.

Institutions pay:

- An evenly distributed share of the cost to support public domain volumes in HathiTrust, or

(PD*X)/N

Application Programming Interface (API)

参加校による技術構築協同作業

21

- Data API
 - ▣ Volume and rights information
 - ▣ Page images
 - ▣ OCR
 - ▣ http://www.hathitrust.org/data_api
- Bibliographic API
 - ▣ Volume and rights information
 - ▣ MARC records
 - ▣ http://www.hathitrust.org/bib_api
- OAI
 - ▣ <http://www.hathitrust.org/data>
- “Hathifiles”
 - ▣ <http://www.hathitrust.org/hathifiles>

日本語資料の課題

日本語資料数 7 位、187,672 3 %

そのうちパブリックドメインにある日本語資料

15 位 5,882 0.34 %

- 資料の質 :

- 日本語理解のないスタッフ

- 大量プロセスによる電子化

- オリジナルの画像の質が悪く検索が不十分な資料もある

- 例 : 「朝日新聞外地版」

- OCR (Optical Character Recognition) 改良は進んでいる (約 80 % は検索可能)

- 複雑な著作権

HathiTrust（米国における）資料の著作権 チェックポイント

出版年 / 出版された国 / 著者の没年

<http://copyright.cornell.edu/resources/publicdomain.cfm>

Firefox

開税及び貿易に関する一般協定

Copyright Term and the Public Doma...

ガット・ウルグアイランドって何ですか？ ...

図書カード: 宮本武蔵

Catalog Record: Miyamoto Musashi | ...

EN

青空文庫

copyright.cornell.edu/resources/publicdomain.cfm

Most Visited

Getting Started

Latest Headlines

title:kieta yaruta

Boo

Works First Published Outside the U.S. by Foreign Nationals or U.S. Citizens Living Abroad⁹

Date of Publication	Conditions	Copyright Term in the United States
Before 1923	None	In the public domain (But see first special case below)

Works Published Abroad Before 1978¹⁰

1923 through 1977	Published without compliance with US formalities, and in the public domain in its source country as of 1 January 1996 (but see special cases) ²⁰	In the public domain
1923 through 1977	Published in compliance with all US formalities (i.e., notice, renewal) ¹¹	95 years after publication date
1923 through 1977	Solely published abroad, without compliance with US formalities or republication in the US, and not in the public domain in its home country as of 1 January 1996 (but see special cases)	95 years after publication date
1923 through 1977	Published in the US less than 30 days after publication abroad	Use the US publication chart to determine duration
1923 through 1977	Published in the US more than 30 days after publication abroad, without compliance with US formalities, and not in the public domain in its home country as of 1 January 1996 (but see special cases)	95 years after publication date

Works Published Abroad After 1 January 1978

1 January 1978 - 1 March 1989	Published without copyright notice, and in the public domain in its source country as of 1 January 1996 (but see special cases) ²⁰	In the public domain
1 January 1978 - 1 March 1989	Published without copyright notice in a country that is a signatory to the Berne Convention and is not in the public domain in its source country as of 1 January 1996 (but see special cases) ²⁰	70 years after the death of author, or if work of corporate authorship, 95 years from publication
1 January 1978 - 1 March 1989	Published with copyright notice by a non-US citizen in a country that was party to the Universal Copyright Convention (UCC)	70 years after the death of author, or if work of corporate authorship, 95 years from publication
After 1 March 1989	Published in a country that is a signatory to the Berne Convention	70 years after the death of author, or if work of corporate authorship, 95 years from publication
After 1 March 1989	Published in a country with which the United States does not have copyright relations under a treaty	In the public domain

Special Cases

例：吉川栄治 1892-1962 「宮本武蔵」

東京：講談社, 昭和 46 [1971] HathiTrust [Limited \(search only\)1](#)

□ 日本国内 著作権 著者の没後50年間

吉川栄治没後 1 1 年目



「青空文庫」にてフルテキスト公開

□ 米国内 外国で出版された資料

1923 年以前 パブリックドメイン（著者の規制以外）

1923年—1977年内にGATT加盟国で出版された著作には

「知的所有権の貿易関連の側面に関する協定（TRIPS協定：

Trade-Related Aspects of Intellectual Property Rights）が適用

出版された年より95年間著作権有効（著者承諾以外）



HathiTrust [Limited \(search only\)1](#)

将来への問題提起

- 慶応義塾大学がHathi Trust に参加
12万冊
- 著作権：
出版された国でパブリックドメインで公開されている資料はその国内でフルテキストで閲覧できるようにする。
(UK, Canada, Australia で進行中)

例： Hathi Trust内にある吉川栄治の著作（日本ではパブリックドメイン）

日本国内での公開 ——— 検討可能

現在：著作権関係スタッフ40名・弁護士1名

新しいスタッフが必要・日本からの協力？

- 新機能：オリジナル文献の所蔵機関情報を加える

グローバルな日本研究資料所蔵情報の把握？

コレクション構築協同作業のプラットフォーム？

資料

26

- About: <http://www.hathitrust.org/about>
- Twitter: <http://twitter.com/hathitrust>
- Facebook: <http://www.facebook.com/hathitrust>
- Monthly newsletter:
 - <http://www.hathitrust.org/updates>
 - RSS http://www.hathitrust.org/updates_rss
- Contact us: feedback@issues.hathitrust.org
- Blogs: <http://www.hathitrust.org/blogs>
 - Large-scale Search
 - Perspectives from HathiTrust