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

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**
- **日本文献の例**
  - **課題**
  - **将来への展望**

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

- Community:　協同事業　ー　資料保存の共有
- Scale:　大規模の協同作業、資料による　universal library
- Access and Preservation: 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構築

2008 年に開始　Depository ー　ミシガン大学・インディアナ大学

法的財務管理　ー　ミシガン大学にオフィスがあるが参加校による協同事業であり、NPOではない。

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

⇨共有で管理・保存し、利用する

- 10.7 million volumes
- 5.6 million book titles
- 276,000 serial titles
- 3.4 million public domain volumes (~32%)

- 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 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)

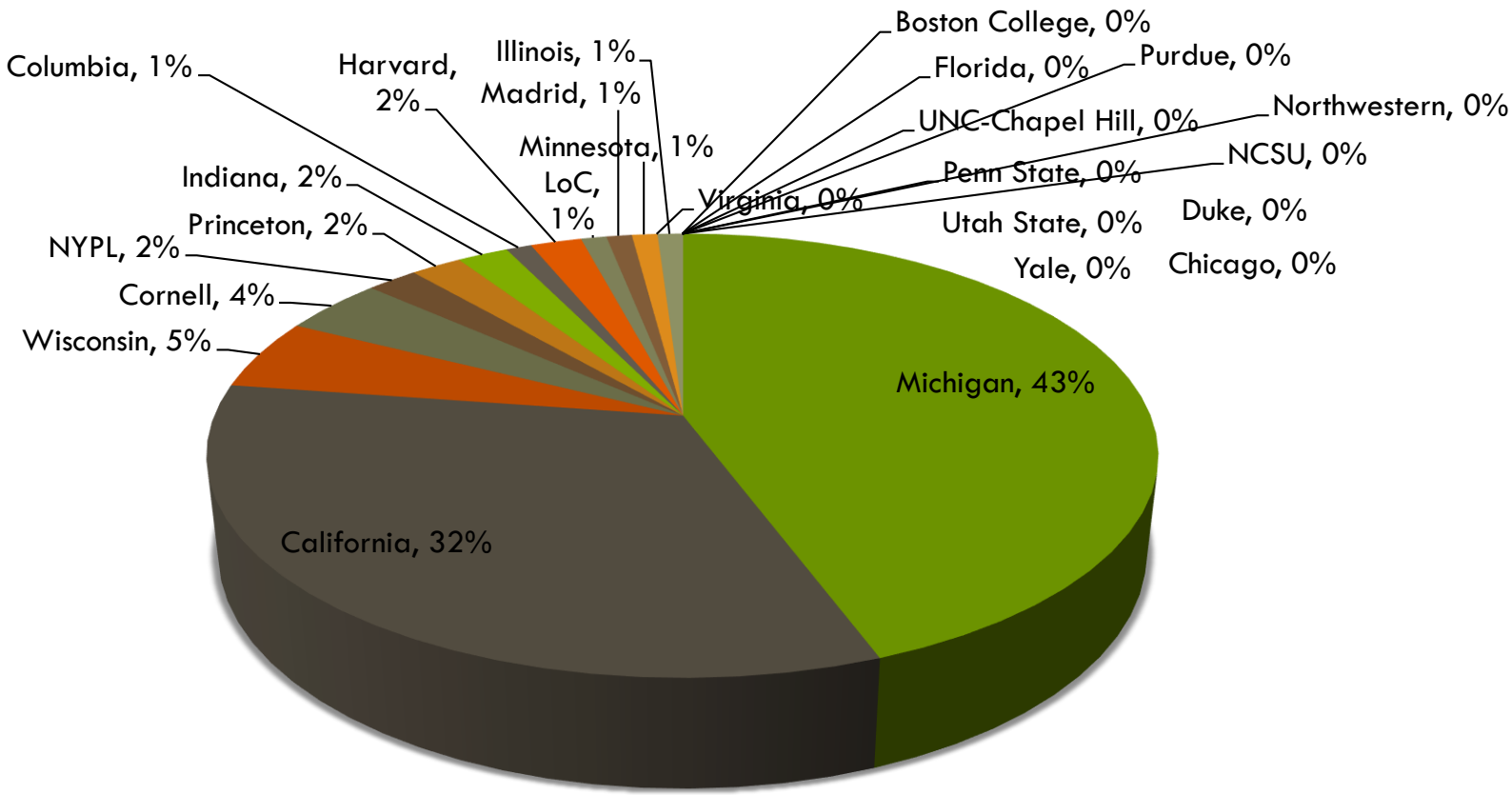
70 以上のパートナー校

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

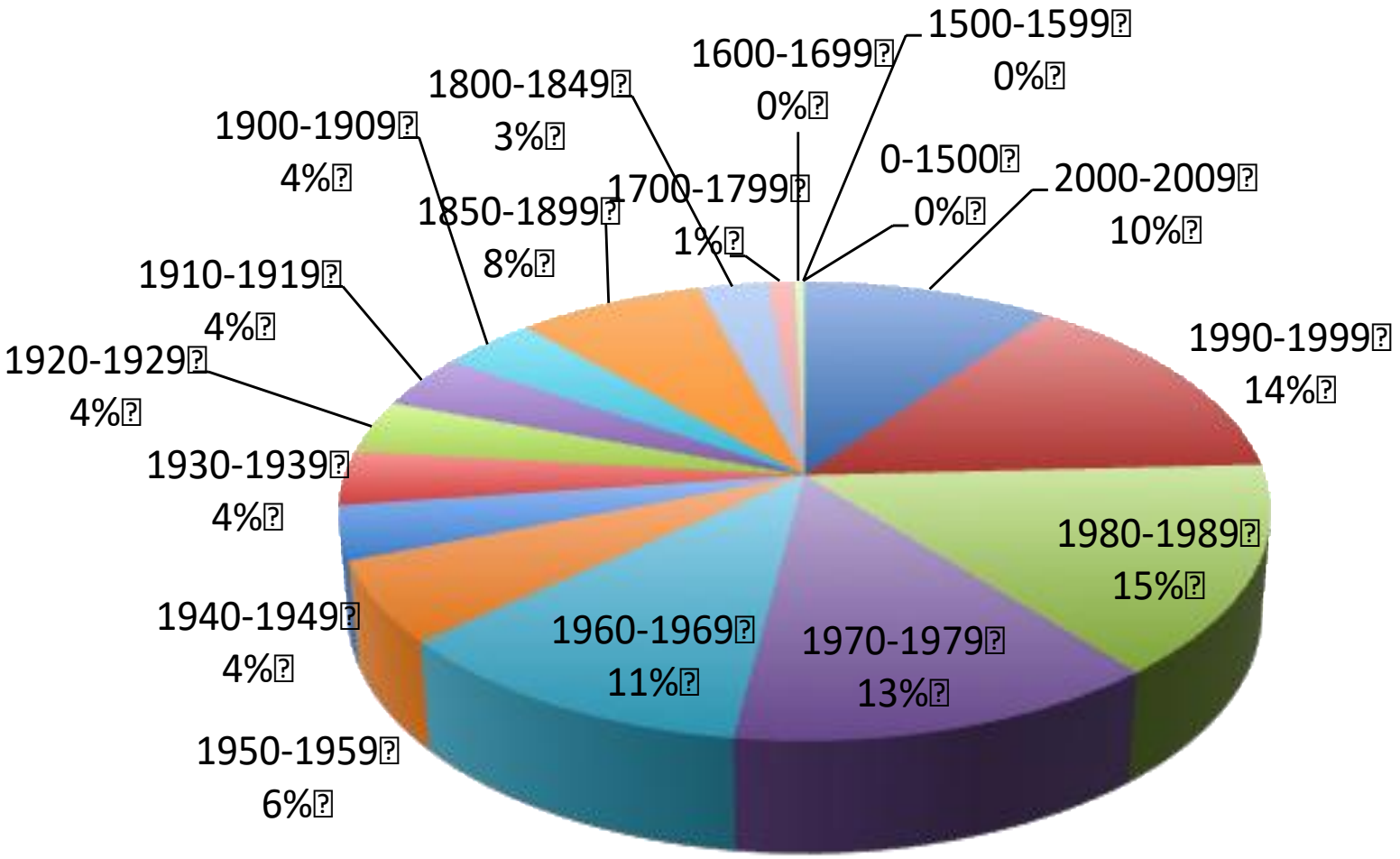
各図書館提供の資料分布

千百万冊 のうちの4万冊は0%として表示



年代別

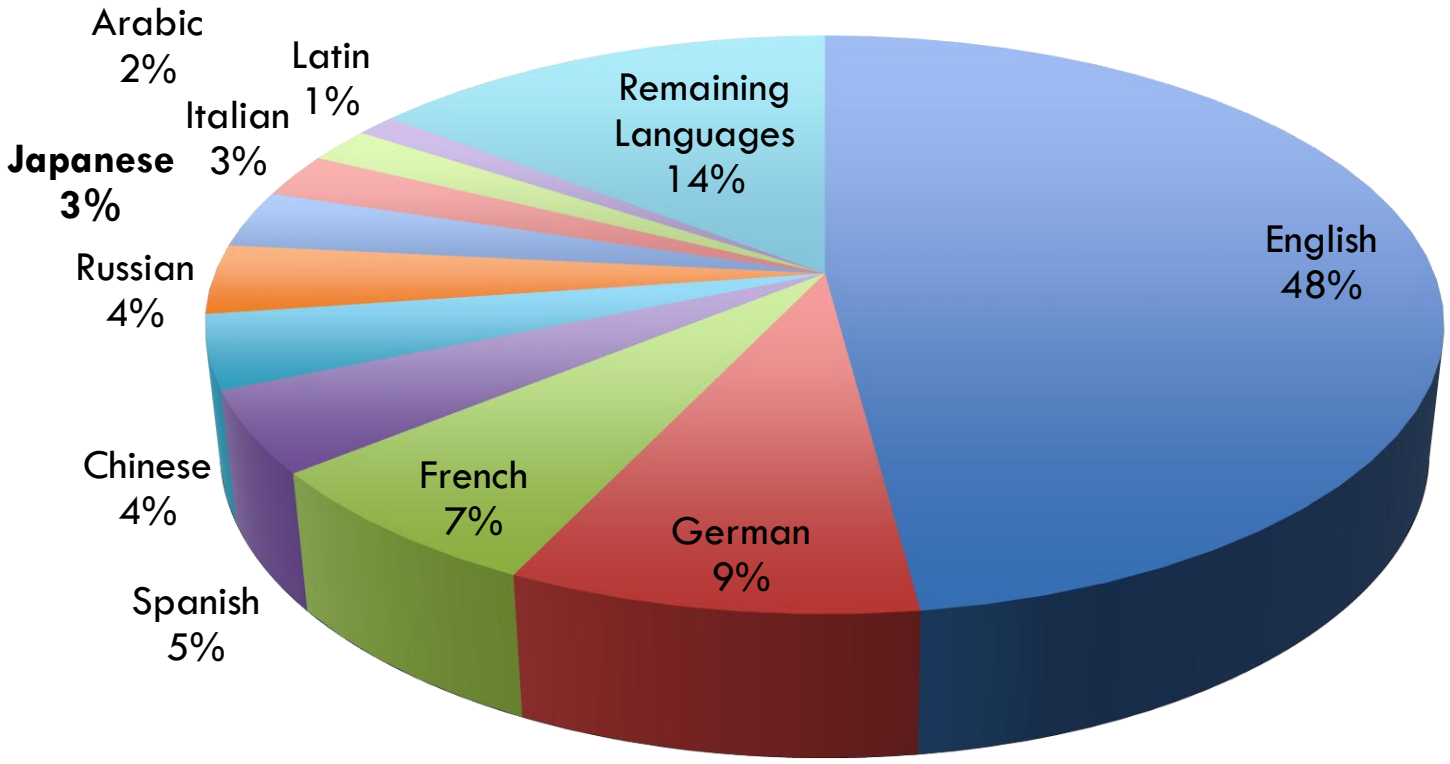
1923年 以前の出版物（Public domain) ~ 全体の 20%



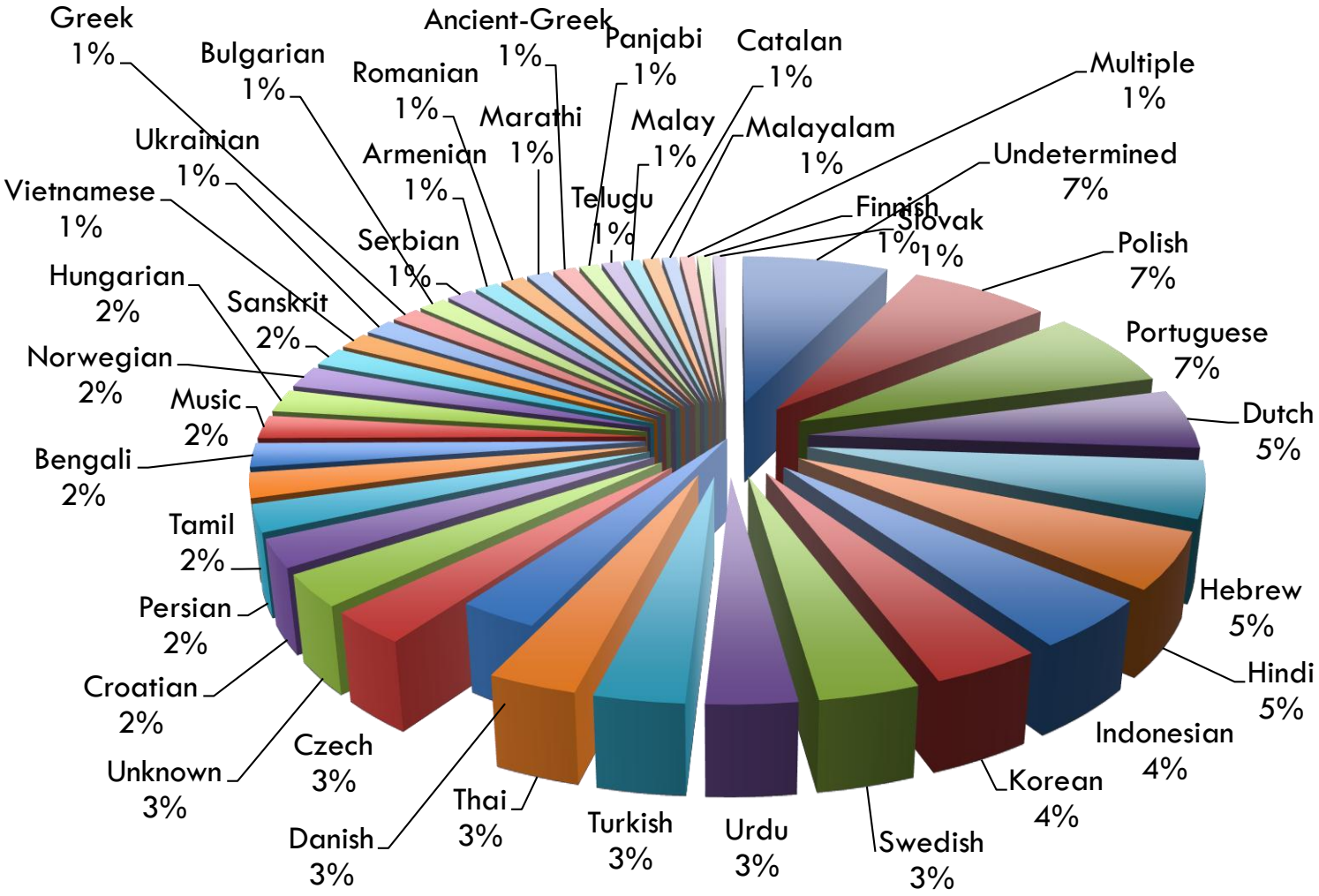
言語別

日本語資料は全体の3%　187,672冊

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



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

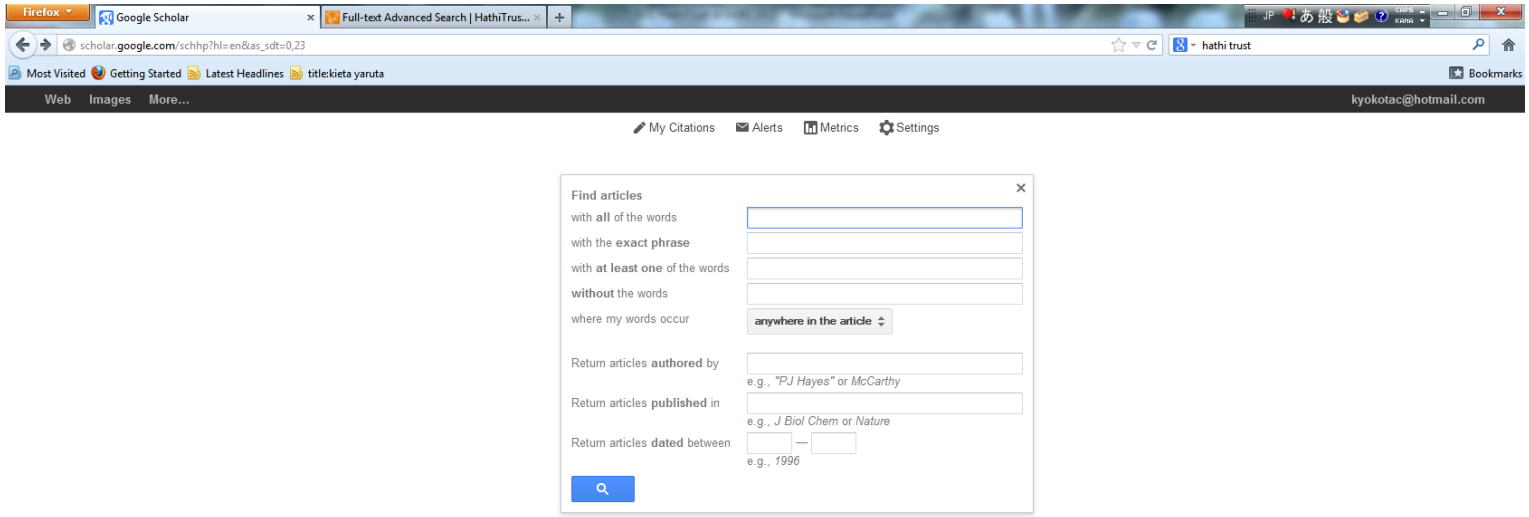
⬇

Google **よりも50万点以上のコンテンツ**

Google とHathiTrust の違い

書誌データの質・検索機能

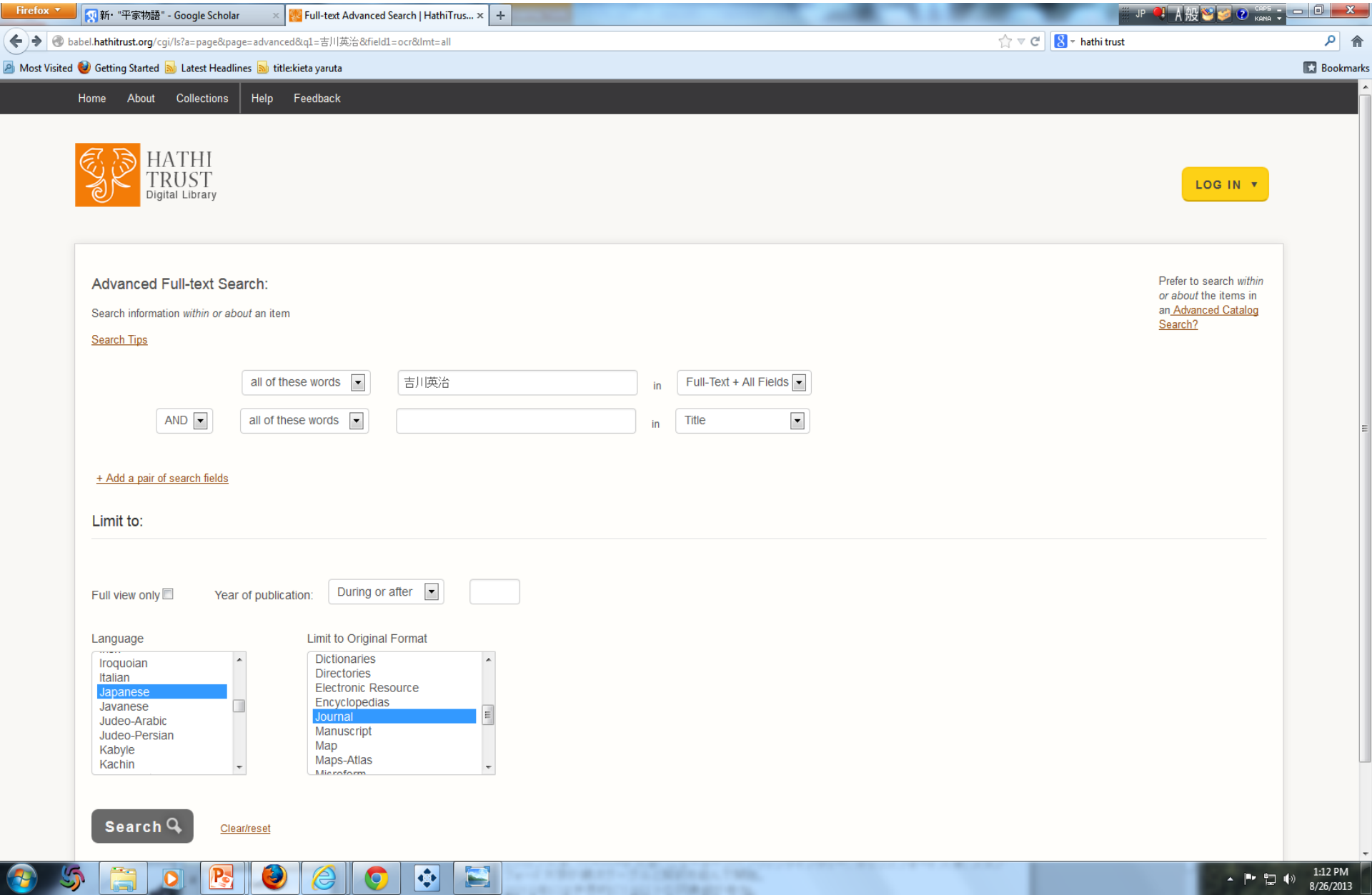
Google scholar



Google「東照宮」で検索

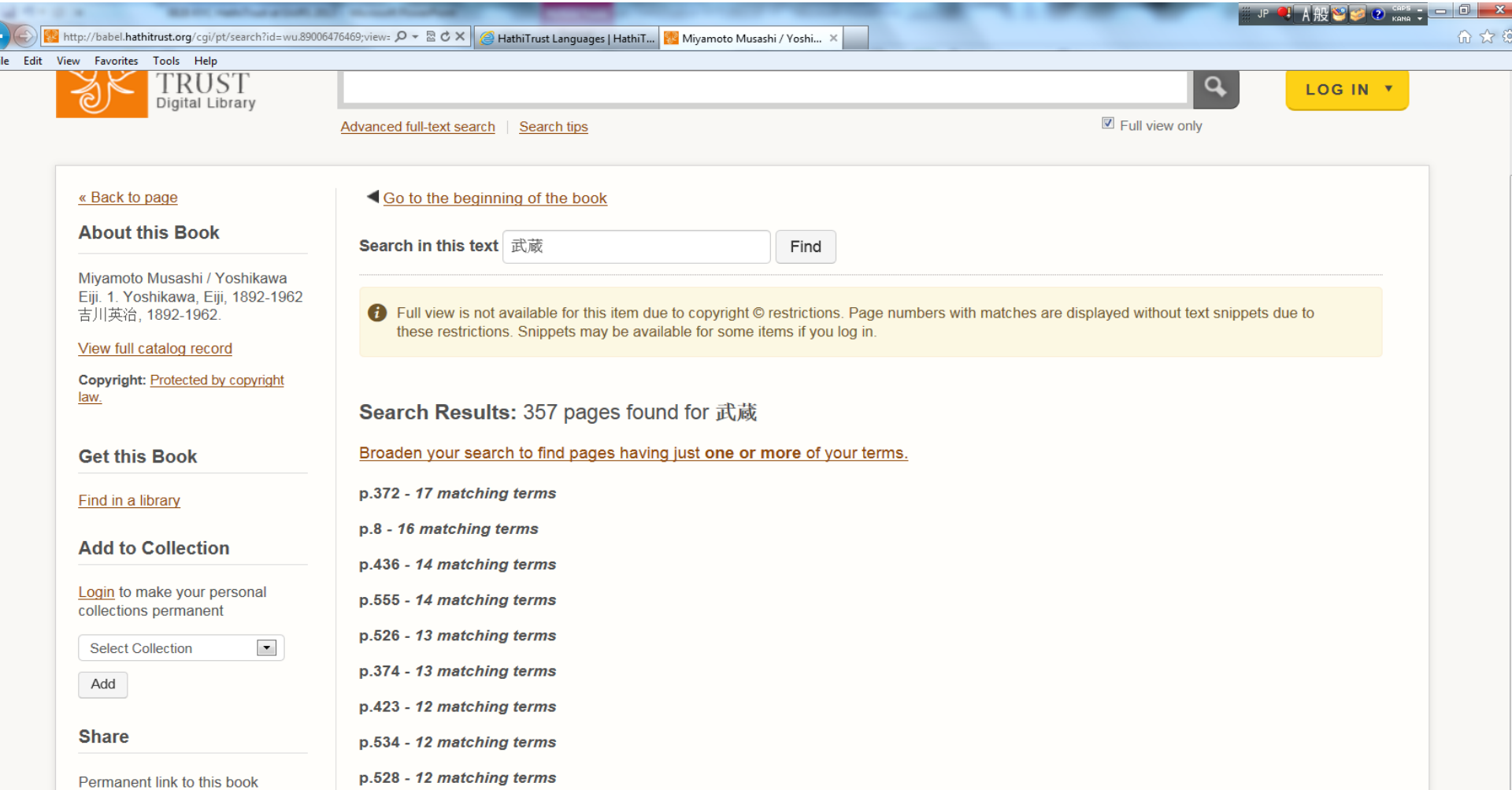


HathiTrust 図書館機能を提供



HathiTrust : Limited Search でも全文検索する

「宮本武蔵」(吉川栄治著)「武蔵」で検索



アクセスと利用　Access use policies

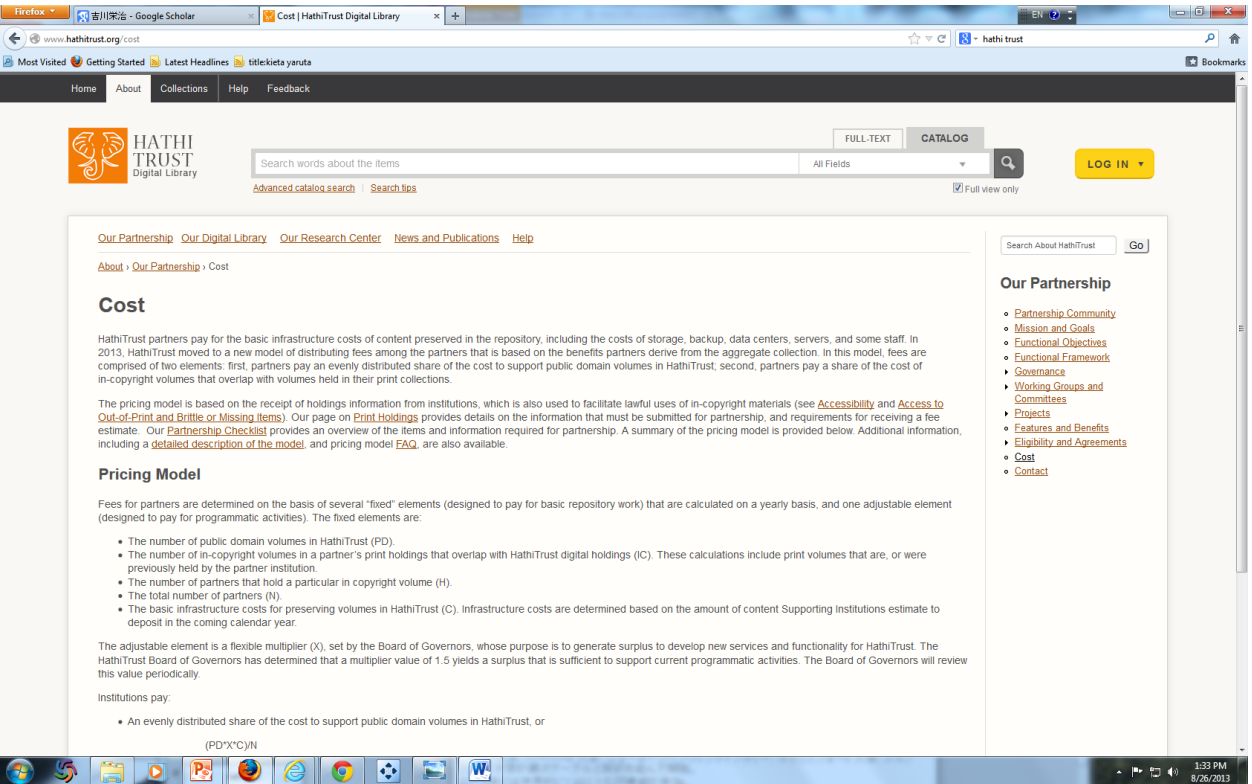
http://www.hathitrust.org/access_use#

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数、全体数等々により機関ごとに多様

Hathi Trust と相談要



Application Programming Interface (API)

- 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

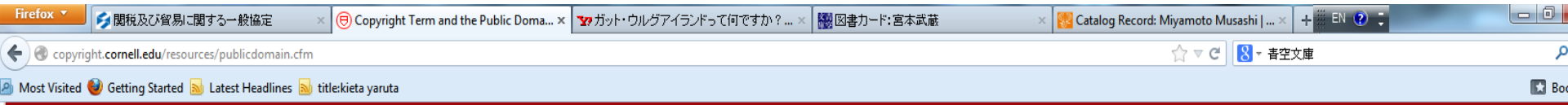
　15位　5,882　0. 34%

- 日本語理解のないスタッフ/ 大量プロセスによる電子化
- OCR (Optical Character Recognition) 改良は進んでいる
- 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



Works First Published Outside the U.S. by Foreign Nationals or U.S. Citizens Living Abroad[9]		
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 1978[10]		
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：Hathi Trust に含まれている

吉川栄治　1892-1962「宮本武蔵」
東京 : 講談社, 昭和 46 [1971]

- 日本国内 「青空文庫」にてフルテキスト公開
  著作権　著者の没後50年間

- 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
  GATT（関税および貿易に関する一般協定）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）

⇩

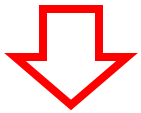
HathiTrust　Limited (search only)1

将来の課題

- 慶応大学がHathi Trust に参加

  1923年以前の資料　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はフルテキストで閲覧できるようにする。



新しいスタッフが必要・日本からの協力？

How to find out more

- 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