HATHI TRUST
a shared digital repository

HathiTrust:

Building the Organization, Building Services

Christenson, Burton-West, Chapman, Feeman, Karle-Zenith & Wilkin

4 May 2009

Presentation structure

- Quick intro (me)

- Community-building (Heather)

- Where we go next with collaboration (me)

- Large-scale search (Tom)

- Services for users with print disabilities (Suz)

- Large-scale ingest (Jessica)

- Copyright review/CRMS (Anne)

- Q&A

Mission and Goals

- to contribute to the common good by collecting, organizing, preserving, communicating, and sharing the record of human knowledge
  - materials converted from print
  - improve access …to meet the needs of the co-owning institutions
  - reliable and accessible electronic representations
  - coordinate shared storage strategies
  - "public good" … free-riders.
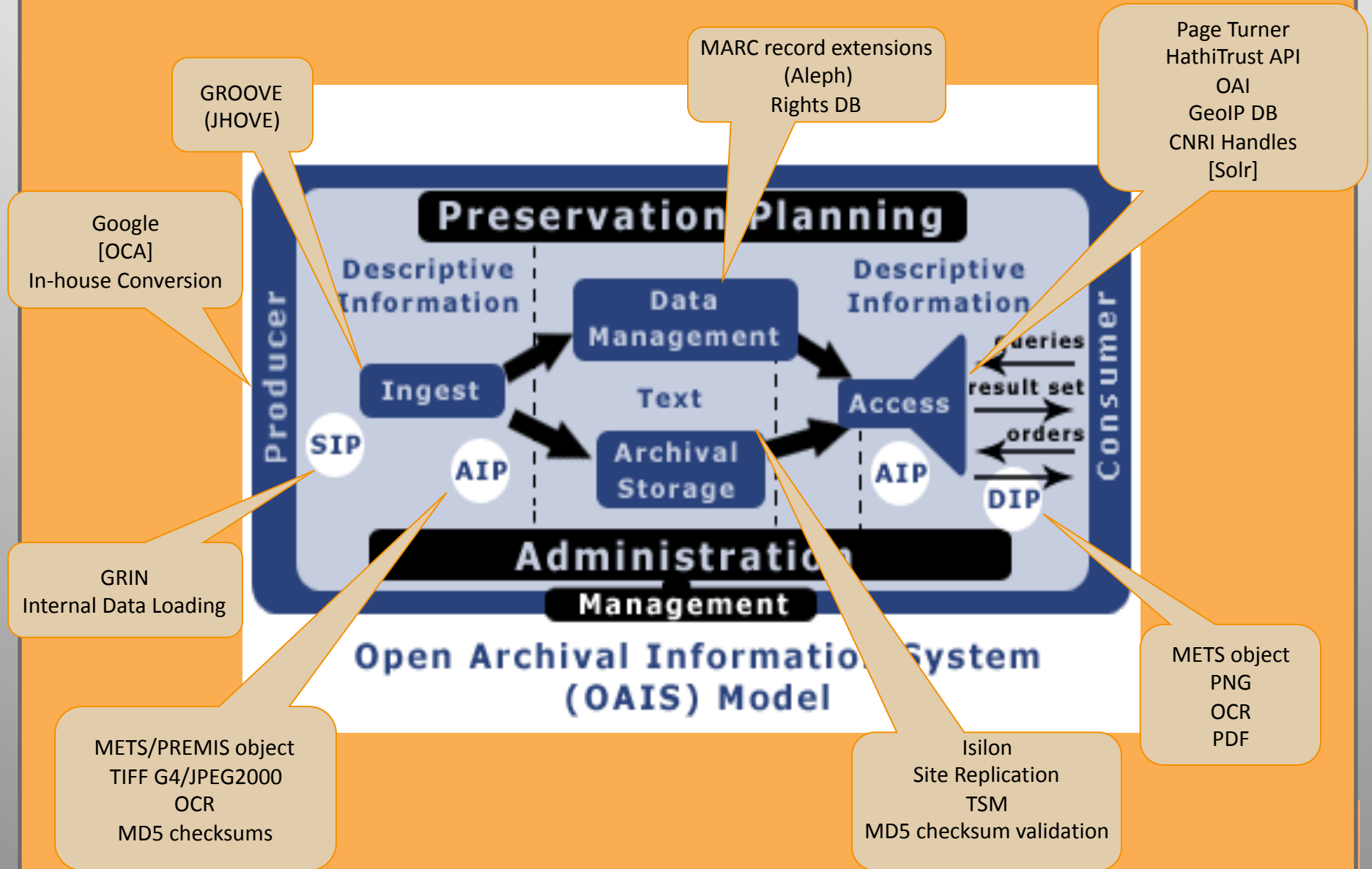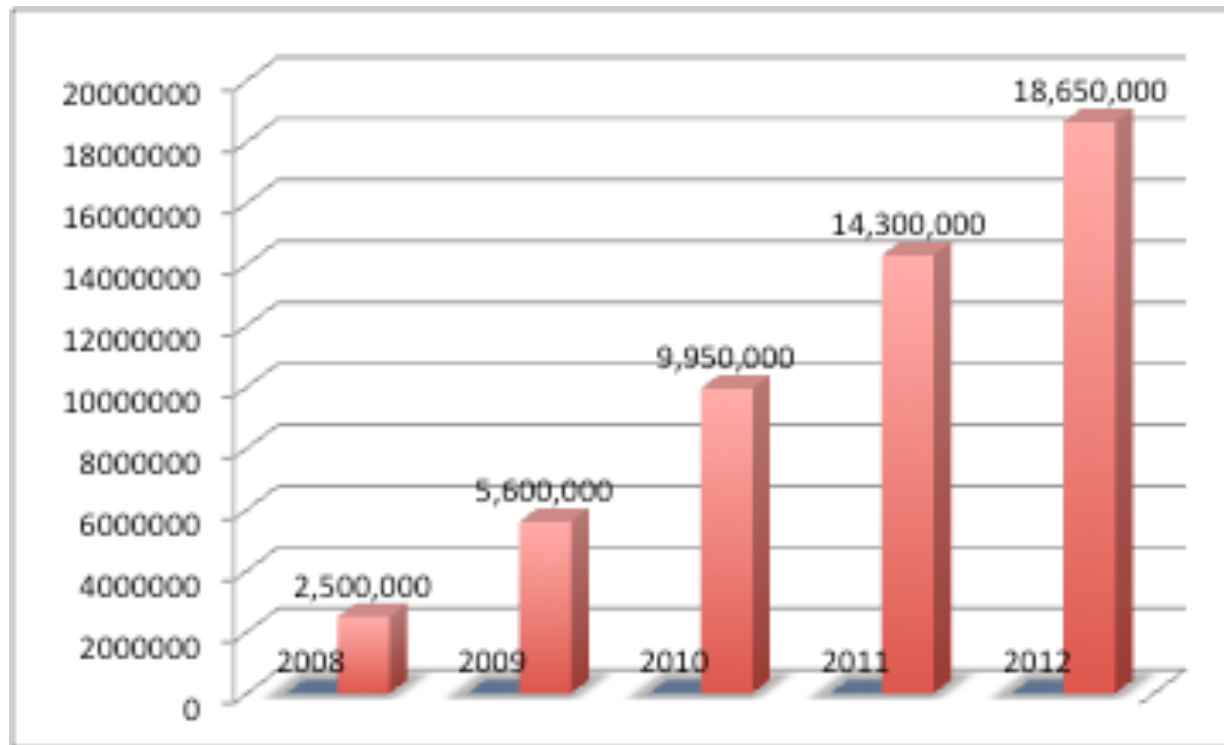 - simultaneously …centralized …open

current members

- California Digital Library
- Indiana University
- Michigan State University
- Northwestern University
- The Ohio State University
- Penn State University
- Purdue University
- UC Berkeley
- UC Davis
- UC Irvine
- UCLA
- UC Merced
- UC Riverside

- UC San Diego
- UC San Francisco
- UC Santa Barbara
- UC Santa Cruz
- The University of Chicago
- University of Illinois
- University of Illinois at Chicago
- The University of Iowa
- University of Michigan
- University of Minnesota
- University of Wisconsin-Madison
- University of Virginia

www.hathitrust.org

Preservation: OAIS Reference Model

GROOVE
(JHOVE)

MARC record extensions
(Aleph)
Rights DB

Page Turner
HathiTrust API
OAI
GeoIP DB
CNRI Handles
[Solr]

Google
[OCA]
In-house Conversion

GRIN
Internal Data Loading

METS/PREMIS object
TIFF G4/JPEG2000
OCR
MD5 checksums

Isilon
Site Replication
TSM
MD5 checksum validation

METS object
PNG
OCR
PDF

growth trajectory

accomplishments to date

1. 25 partners

2. successful ingest and millions of vols online

3. mirroring *and* backup

4. rich access

5. collection builder

6. soon, full text search and data API

www.hathitrust.org

UC's mass digitization efforts

- October 2005 to present

- 2 digitization partners, multiple sources of funding

- 10 campuses + CDL

- 7 different project locations

- CDL in coordinating role

- Over 2M volumes digitized

www.hathitrust.org

UC's mass-digitized content

- Open Content Alliance / Microsoft / Internet Archive /
  - Out-of-copyright works only
  - Primarily English language, some romance languages
- Google
  - In-copyright and out-of-copyright works
  - All languages

Why is UC participating in the HathiTrust?

- Build a sustainable model for services

- Provide an academic counterweight to Google

- Build a new organization based on trust & transparency

- Exploit full potential of aggregation

- Integrate HathiTrust offerings with existing UC services

www.hathitrust.org

UC contributions

- Deep expertise in digital preservation & discovery services

- Long experience with collaboration

- Standards, assessment, project management expertise

- Expertise & intent to pioneer ingest of Internet Archive-digitized content

www.hathitrust.org

The early days of collaboration

- Technical walkthrough: repository, development, ingest, services, policy
- Initial videoconference w/UM, follow up with smaller groups
- Ingest of UC Google content
- HathiTrust WorldCat Local team

www.hathitrust.org

UC Ingest

- File format decision TIFF G4 + JPEG2000 mix

- Requirements for ingest reporting

- UC namespace decision

- The nuances of "book packages" and bibliographic data

- Enabling the download: Google & UM & UC

The early days: challenges

- Expanding the boundaries of existing work
- Geography & travel constraints
- Communications & gaining a common understanding

The early days: what works

- Starting with what we have in common, which is a lot!

- Tapping front-line experts; minds of the best and brightest across institutions

- In the process of collaboration, participants forced to solidify their own institutional goals

www.hathitrust.org

Where next with collaboration?

- Begin sharing actual development, cf. ingest of Internet Archive content

  – Specifications

  – Validation routines?

  – Packaging?

- Collaboratively develop a collaborative framework

  – SAB and working group charges

Working groups?

- Security

- Collection management

- Non-Consumptive Research

- Digital preservation

- Discovery (bibliographic and full text)

- Externally-facing repository APIs

- Bibliographic metadata management

- Rights Management
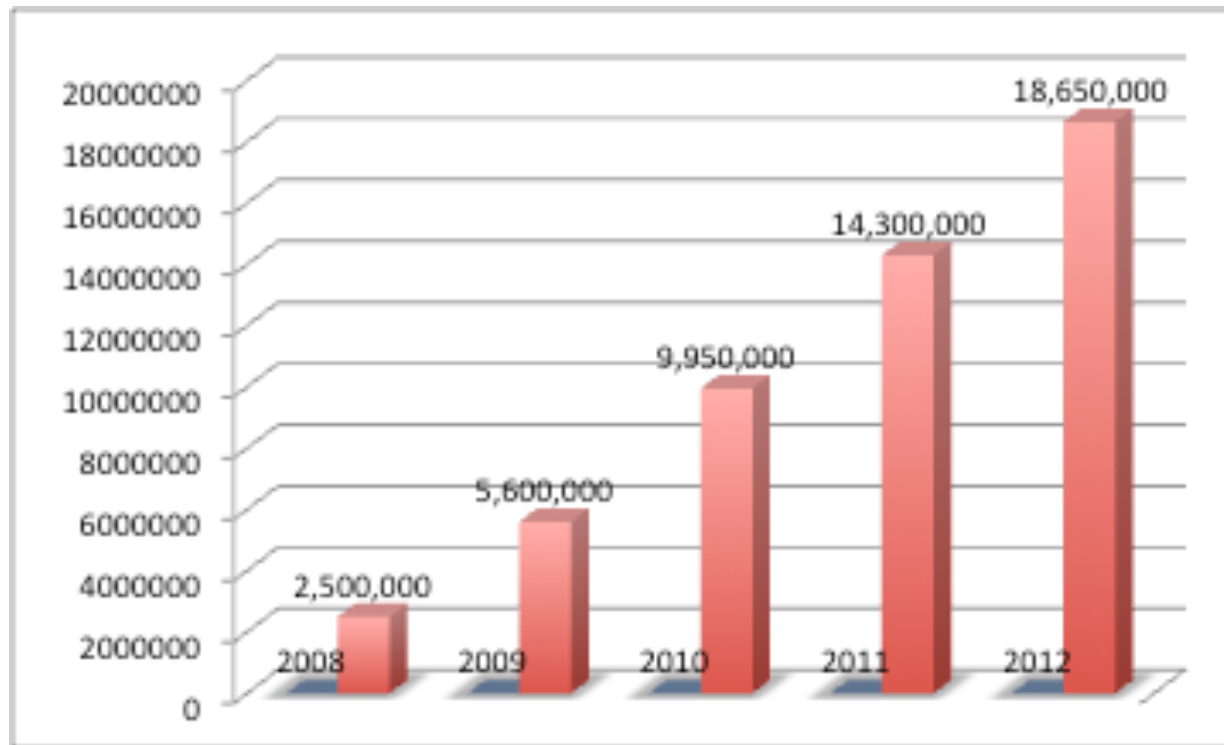
HathiTrust Large Scale Search

Tom Burton-West
Information Retrieval Programmer
May 4th 2009

Challenges

- Goal: Design a system for full-text search that will scale to 7 million -20 million volumes (at a reasonable cost.)

- Challenges:
  - Must scale to 20 million full-text volumes
  - Very long documents compared to most large-scale search applications
  - Multilingual collection
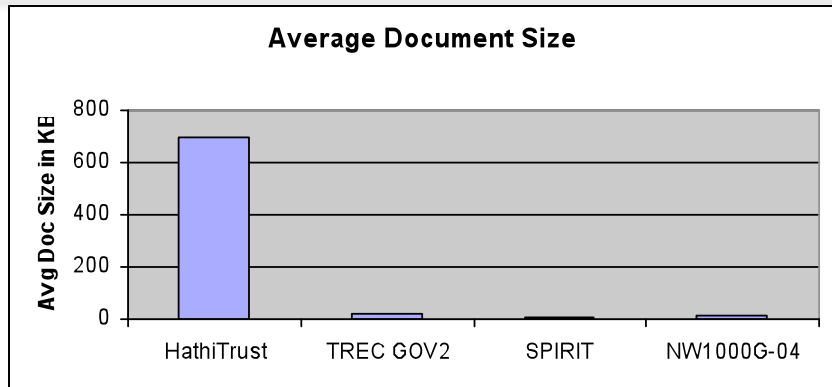 - OCR quality varies

www.hathitrust.org

growth trajectory

Long Documents

- Average HathiTrust document is 700KB containing over 100,000 words.
  - Estimated size of 7 million Document collection is 4.5TB.
- Average HathiTrust document is about 38 times larger than the average document size of 18KB used in Large Research test collections

Average Document Size

Collection	Size	Documents	Average Doc size
HathiTrust	4.5 TB (projected)	7 million	700 KB
TREC GOV2	0.456 TB	25 million	18 KB
SPIRIT	1 TB	94 million	10 KB
NW1000G-04	1.3 TB*	100 million	16 KB

www.hathitrust.org

Multilingual

- 200+ languages, 40 languages with over 1,000 volumes (some volumes in multiple languages)
- Currently all languages in one index
- Lowest common denominator tokenizing
- Some languages are challenging for Information Retrieval
  - CJK
  - Arabic
  - Vietnamese

OCR

- **The OCR varies in quality**
- **This makes the index larger and slower**
- **Example: Hebrew characters not recognized**

Testing Program

- Test scalability of Solr open source search engine with HathiTrust full-text content.
  - Solr is based on the open source Lucene library
  - Solr provides a mechanism for distributed search with sharded indexes on multiple hosts for scaling the index and with replication for scaling load.
- Tested with index sizes from 100,000 volumes up to 1 million in 100,000 volume increments

Testing Program

- Tested with memory at 4, 8, 16, and 32 GB

- Tested different configurations
  - Single index single machine
  - Split index (2 shards) single machine
  - Split index (2 shards) one on each of two machines

Testing Program: Hardware

- Servers
  - 2 PowerEdge 1950 blades
  - 2 Dual Core 3.0 GHz Processors
- NFS server
  - 4 JetStor 416S Raid Arrays RAID 6
  - 7 SATA drives 7500 RPM
- Network for NFS-Servers
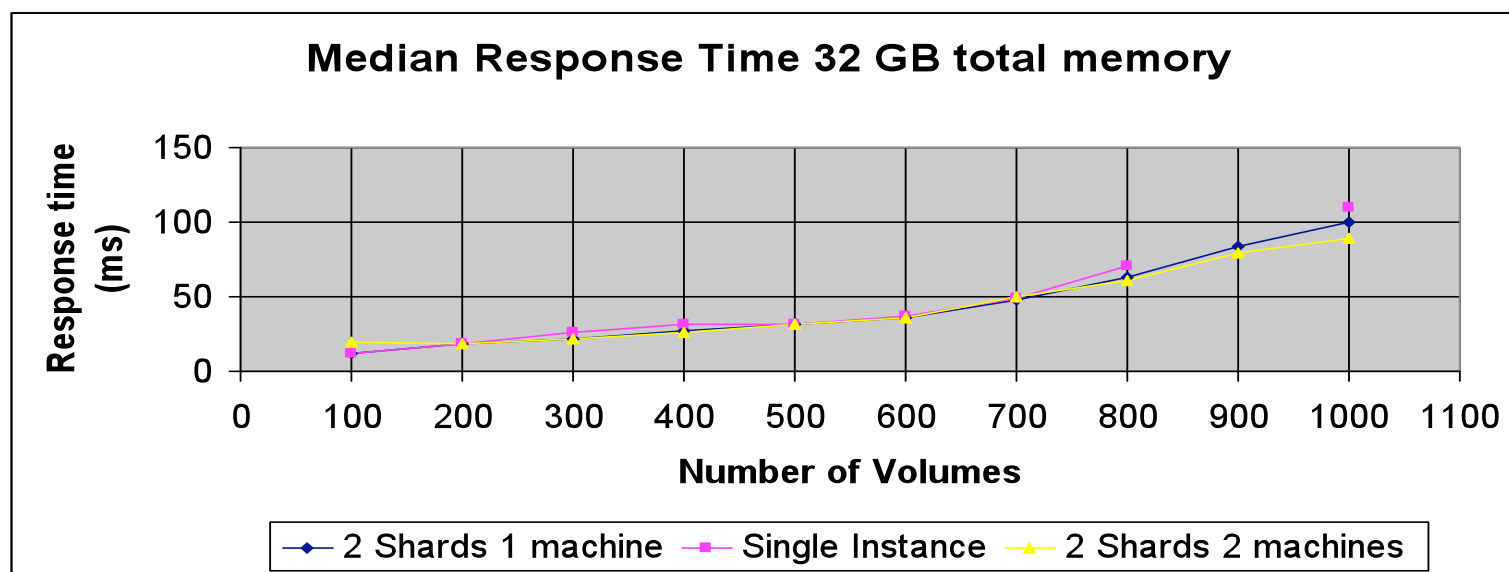 - Gigabit private/unrouted network

Testing Program: Protocol

- 10,000 warm-up queries (from OPAC logs)
- 1,000 test queries (from OPAC logs)
- More details in report:
- www.hathitrust.org/technical_reports/Large-Scale-Search.pdf

Testing Program Results
Scalability and Machine Configurations

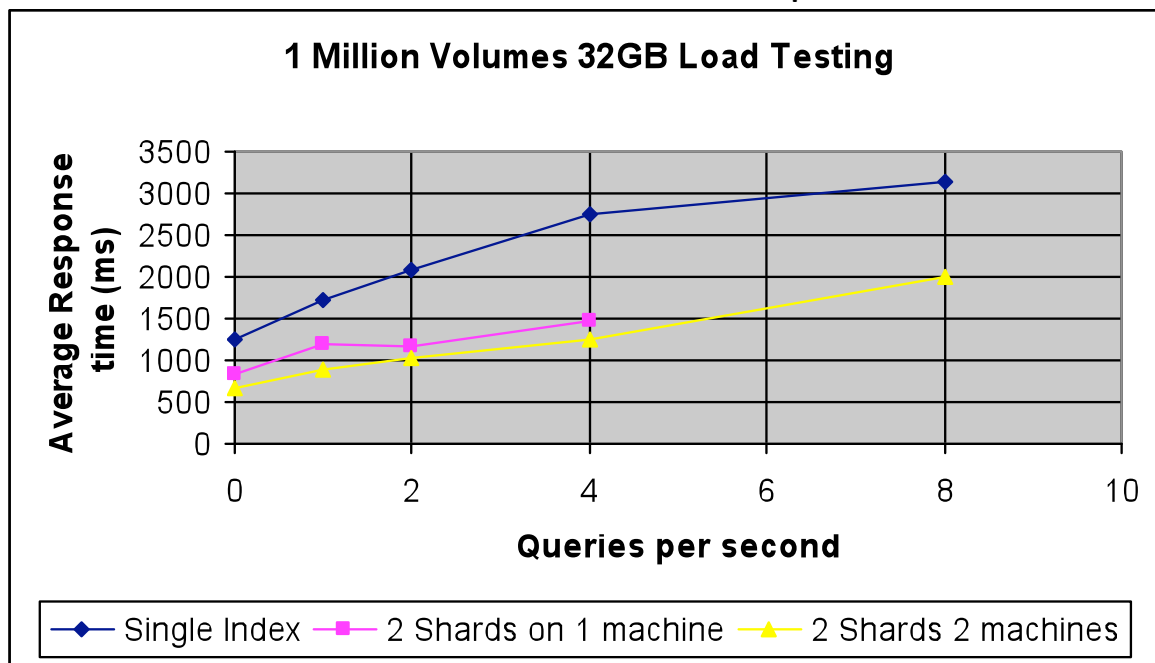- Solr scales linearly. At 32 GB total memory there is an increase in slope above 600 thousand documents.

- Best response time was for 2 shards each on its own machine

Median Response Time 32 GB total memory

Legend: 2 Shards 1 machine — Single Instance — 2 Shards 2 machines

Y-axis: Response time (ms) — 0, 50, 100, 150
X-axis: Number of Volumes — 0, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100

Testing Program Results
Load testing and Machine Configurations

- Request rates over 1 query per second increased response time.
- Rates over 4 qps were not sustainable at larger index sizes.
- 2 shards on 2 machines handled load best
- Above 600 thousand documents response time increases rapidly

1 Million Volumes 32GB Load Testing

Legend: ◆ Single Index   ■ 2 Shards on 1 machine   ▲ 2 Shards 2 machines

X-axis: Queries per second
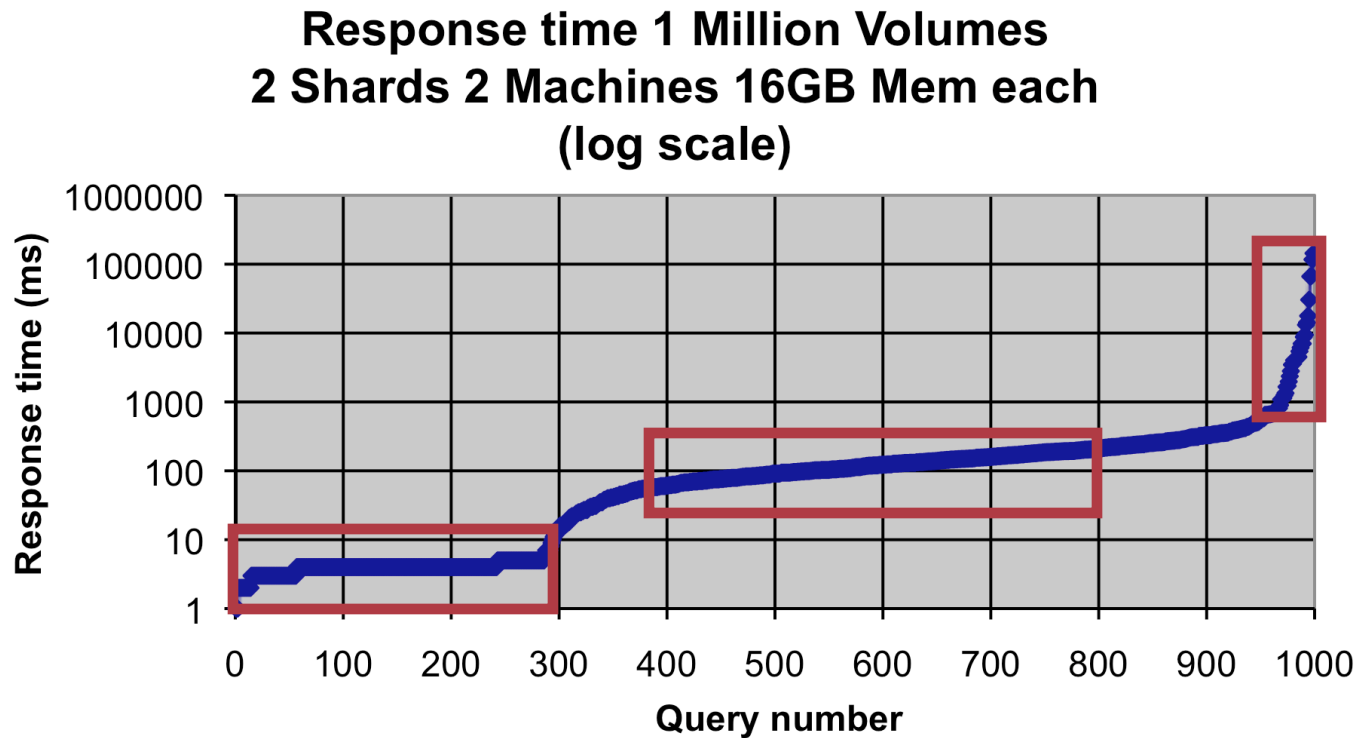Y-axis: Average Response time (ms)

Testing Program Results
Caching and Memory

- Solr uses OS memory for caching of postings

- Memory available for caching has most impact on response time

- Based on the 1 Million volume index size of 240GB, the index for 7 million documents would be 1.6TB

- Fitting entire index in memory not feasible with terabyte size index

Response time varies with query

**Response time 1 Million Volumes
2 Shards 2 Machines 16GB Mem each
(log scale)**

Y-axis: Response time (ms) — 1, 10, 100, 1000, 10000, 100000, 1000000

X-axis: Query number — 0, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000

Average:	673
Median:	91
90th:	328
99th:	7,504

www.hathitrust.org

Slowest 5% of queries

Response Time 95th percentile (seconds)

Slow Queries

- The slowest 1% of queries took between 10 seconds and 2 minutes.

- Slowest 0.5% of queries took between 30 seconds and 2 minutes

- These queries affect response time of other queries
  - Cache pollution
  - Contention for resources

- Slowest queries are phrase queries containing common words

www.hathitrust.org

Slow Queries

- Slowest test query: "the lives and literature of the beat generation" took 2.5 minutes.
- Phrase query with common terms (154,000 documents must be evaluated)

Word	Number Docs	Average tf/doc
the	800,000	10,000
of	892,000	8,453
and	769,000	4,809
literature	453,000	279
generation	414,000	211
lives	432,000	124

Query processing

- Phrase queries use position index

- position index accounts for 85% of index size

- position list for common words such as "the" can be many GB in size

- This causes lots of disk I/O and also "pollutes" the cache

www.hathitrust.org

Query Processing: Postings List

Doc ID	Content
1	The black dog chased the white cat
2	The hungry dog ate the cat's food
3	The cat sat in the sun.
4	The dog buried a bone.

Word	Doc IDs
a	4
ate	2
black	1
bone	4
buried	4
cat	1,2,3
chased	1
dog	1,2,4

Word	Doc IDs
food	2
hungry	2
in	3
sat	3
sun	3
the	1,2,3,4
white	1

Query Processing: Position List

Doc ID	Content
1	
2	The black dog chased the white cat
	1 2 3 4 5 6 7
3	
4	The dog buried a bone.

Word	(Doc ID) positions
a	(4) 4
ate	(2) 4
black	(1) 2
bone	(4) 5
buried	(4) 3
cat	(1) 7, (2) 6,(3) 2
chased	(1) 4
dog	(1) 3,(2) 3, (4) 2

Word	(Doc ID) Positionss
food	(2) 7
hungry	(2) 2
in	(3) 4
sat	(3) 3
sun	(3) 3
the	(1) 1, 5, (2) 1, 5 , (3) 1,5, (4) 1
white	(1) 6

Query Processing

- "Dog" AND "Cat"

cat	1, 2, 3
dog	1, 2 ,4

"Dog" AND "Cat"

Docs: 1,2

"The Cat"   (Phrase)

the	1, 2, 3, 4
cat	1, 2, 3

"The" AND "Cat"

Docs: 1,2,3

the	(1) 1, 5, (2) 1, 5 , (3) 1,5 (4) 1
cat	(1) 7, (2) 6, (3) 2

"The Cat"

Docs: 2,3

Stop Words

- The word "the" occurs an average of 15,000 times in about 80-90% of all documents.

- Removing "stop" words ("the", "of" etc.) not desirable

- Couldn't search for many phrases
  - "to be or not to be"
  - "the who"
  - "man in the moon" vs. "man on the moon"

Stop Words

- Stop words in one language are content words in another language

- German stopwords "war" and "die" are content words in English

- English stopwords "is" and "by" are content words ("ice" and "village") in Swedish

"CommonGrams"

- Bi-Grams triple index size

- Nutch and CDL XTF implement "CommonGrams"

- Create Bi-Grams for any two word sequence containing common terms

- "The rain in spain falls mainly" = "the-rain" "rain-in" "in-spain" "falls" "mainly"

CommonGrams Example

- Slowest query: "The lives and literature of the beat generation"
  - "the-lives" "lives-and"
  - "and-literature" "literature-of"
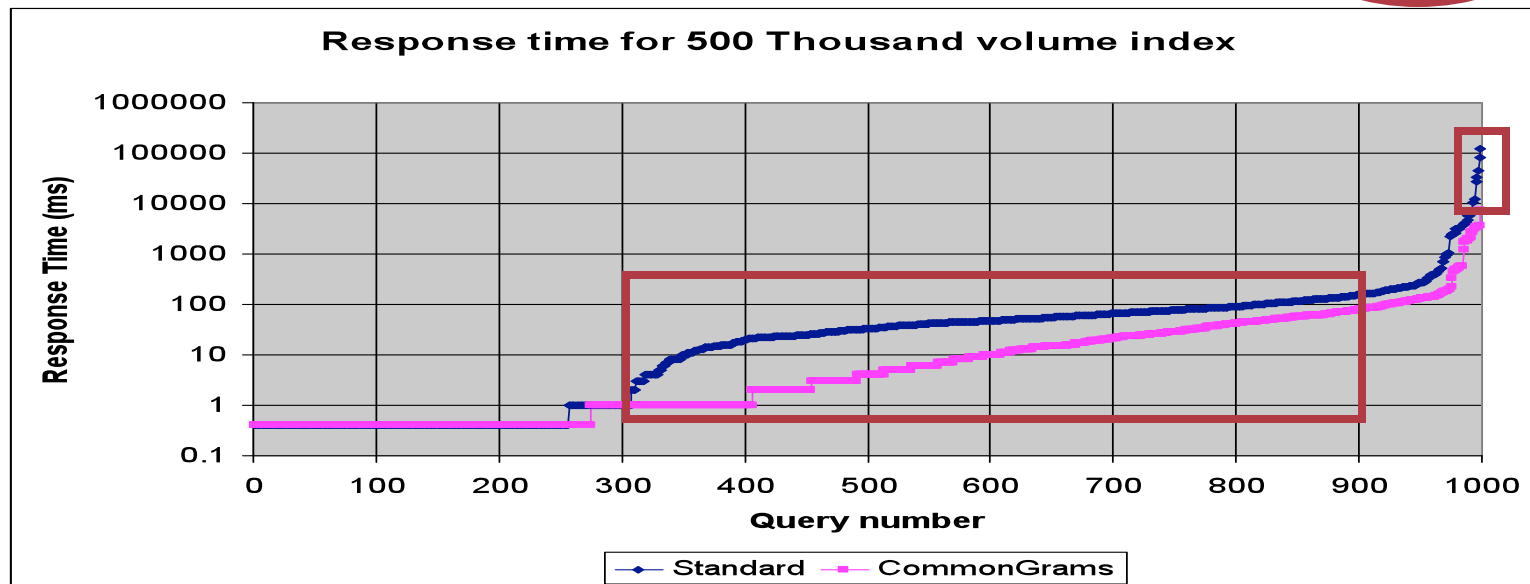 - "of-the" "the-beat" "generation"

CommonGrams

Standard Index		CommonGrams Index	
word	% of docs	word	% of docs
the	90%	the-lives	26%
lives	52%	lives-and	23%
and	86%	and-literature	15%
literature	56%	literature-of	21%
of	94%	of-the	79%
beat	34%	the-beat	5%
generation	52%	beat	34%
		generation	52%

CommonGrams
Comparison of Response time (ms)

	average	median	90th	99th	slowest query
Standard Index	459	32	146	6,784	120,595
Common Grams	68	3	71	2,226	7,800

Response time for 500 Thousand volume index

Next Steps

- Move beta search into production

- Investigate performance with 2-8 machines

- Investigate further optimizations

- Investigate facets and fielded search

- Integrate with other HathiTrust applications such as Collection Builder

Possible Future Development

- Investigate relevance ranking
  - Boost rank for words occuring in MARC subject, author, title fields.
  - Investigate document size normalization

- Improve multilingual access and retrieval

- Investigate page level indexing and integration with book indexing

Possible Future Development

- Investigate user interface/user interaction issues
  - How best to display large result sets
  - How can users best narrow their search?
  - Integration with user's tasks

"Search" URLs

- beta Search http://babel.hathitrust.org/cgi/ls

- www.hathitrust.org/technical_reports/Large-Scale-Search.pdf

- tburtonw@umich.edu

Text-only Interface & Accessibility

Suzanne Chapman
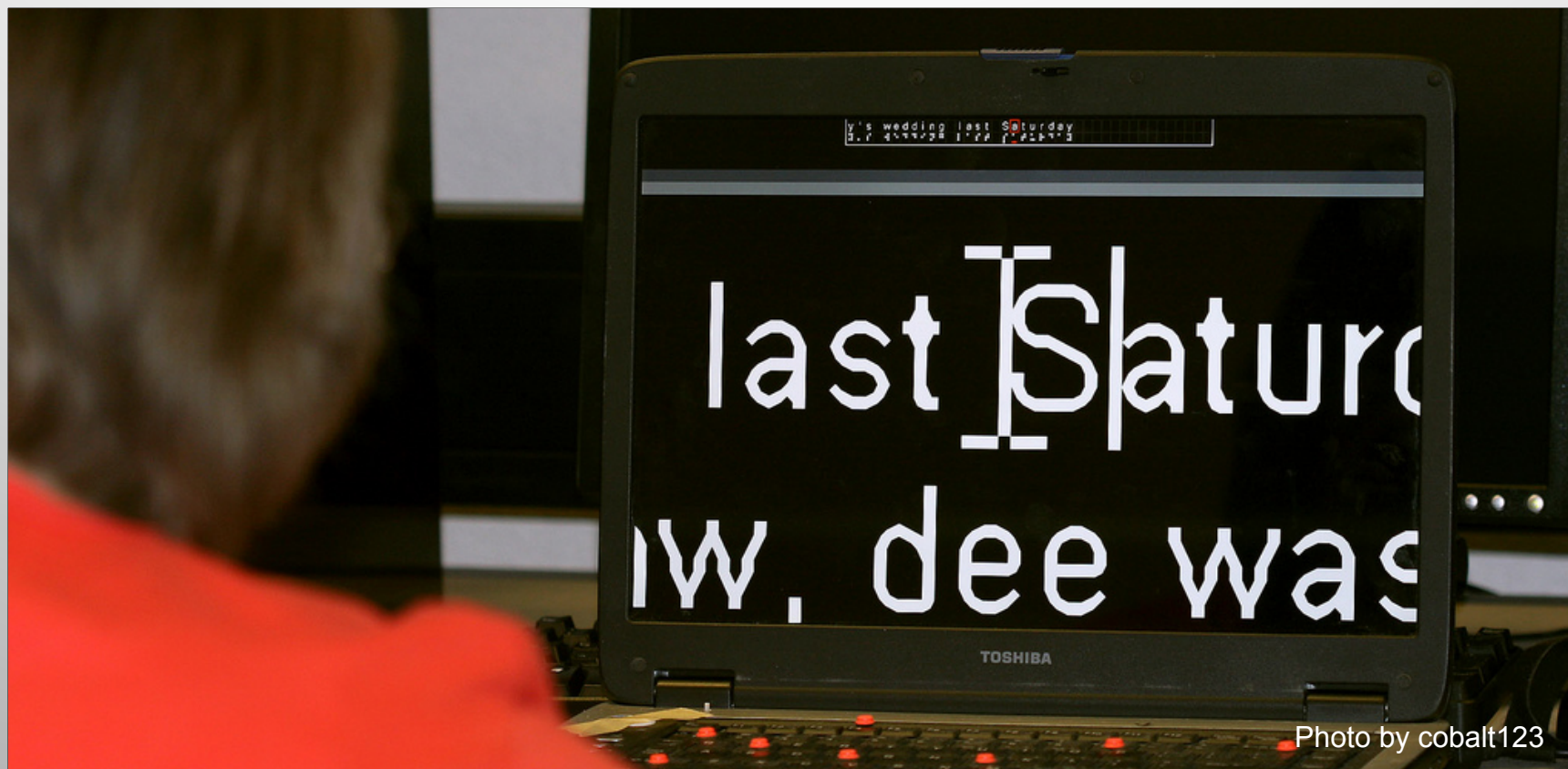
Interface & User Testing Specialist

suzchap@umich.edu

Photo by cobalt123

A system & interface designed specifically to provide access to all digitized materials (regardless of copyright) for our users with print disabilities: low-vision to no vision, learning disabilities

www.hathitrust.org

From a SXSWi Web Design Confession Booth:

I don't always care about accessibility.

Photo by allaboutgeorge

Copyright

Explicitly allowed under section 121 of U.S. Copyright law - Limitations on exclusive rights: reproduction for blind or other people with disabilities

"...it is not an infringement of copyright for an authorized entity to reproduce or to distribute copies or phonorecords of a previously published, nondramatic literary work if such copies or phonorecords are reproduced or distributed in specialized formats exclusively for use by blind or other persons with disabilities..."

http://www.copyright.gov/title17/92chap1.html#121

UM Services for Students with Disabilities (SSwD)

Precedent for offering this type of service set by SSwD and their book digitization/ OCR service to UM students with disabilities upon request.

Process

- Consulting with UM's SSwD
- Understanding assistive technology
- Coding techniques (w/ SI intern Anran Ye)
  - Evaluation & validation tools
  - Standards based code
  - Wrote set of guidelines
- Programming: Kathleen Ludewig, Phil Farber, Tim Prettyman
- Usability & Feedback (w/ SI intern Jennifer Tatomir)

www.hathitrust.org

Screen Readers

Apple Voiceover

Speaks what is happening on the computer & allows you to control the computer using just the keyboard

Navigation by:
Reading entire page
Headings
Links
Page elements

Braille Based Computers

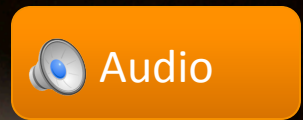Braille Sense computer with refreshable Braille display & audio voice

Audio

Photo by JRandomF

What we've done

1. Made standard interfaces more accessible

2. Created a text-only interface that is optimized for the specific needs of users with print disabilities

3. Created a system to grant additional access to the full-text of a digitized book for certain UM patrons, regardless of the book's copyright status

How it works

1. UM patron registers with the Services for Students with Disabilities office

2. Library is notified to set a status on patron's account

3. Patron checks out any book (manually or via service) that has been digitized

4. Patron receives automated email from Library catalog system containing link to text-only interface

5. Prompted to login (checks to verify patron has book checked out and is part of the program)

6. Patron has access as long as the book is checked out to their account.

THE WORLD GROWS COLD

SOMETHING was the matter with the weather.

Early man did not know what "time" meant. He kept no records of birthdays and wedding-anniversaries or days or weeks or years. He had no idea of the hour or morning he did not say "It is Light" and he used the rays of the early sun to gather food for his family. He said "It grew dark," he returned to his day's When "Behold the sun another part of the sun to sleep.

When it grew dark, he went to sleep. catch (some berries and a few birds), stuffed himself full with raw meat and a went to the and children, gave them part of the gave raw meat he kept taught him a very general way he had invariably followed.

In a Long Winter—that Spring grew into seasons. Long experience grew and the that the cold Spring when fruits ripened and the by the mild Summer were ready to be plucked the hot Summer of corn were ready to be plucked wild ears of corn.

8

THE GLACIAL PERIOD.

Example book:
Ancient Man by Hendrik Van Loon, 1922

Standard Pageturner (text view)

GUI doesn't use headings but they're useful to assistive tech users

Navigating from 1 page to another cumbersome

Standard Pageturner (with styles disabled)

Skip navigation

HathiTrust Digital Library

Go to the text-only view of this public domain book.

Special text versions of all items can be made accessible to to University of Michigan students whose disabilities do not allow them to use printed books. Go to the SSD website for more information.

Go to the web accessibility page for access keys and other accessibility related information.

HATHI TRUST Digital Library

Navigation links for login, help, feedback, etc.

- Hi suzchap!(logout)
- About
- Help
- Feedback

Try our experimental full-text search!

Navigation links for search and collections

- Catalog Search
- Public Collections
- My CollectionsJump to a collection [Jump to a Collection ▼]
- Getting Started

Bibliographic Information about this book

Title more »

Ancient man; the beginning of civilizations, by Hendrik Willem ...

Author

Van Loon, Hendrik Willem, 1882-1944.

Published

New York,Boni & Liveright, inc.[c1922]

Orig. Call No.

DS 62 .V26 1922

Description

ix p., 1 _., 208 p. incl. front., illus., plates.17 cm.

Copyright

Callout annotations:
- Skip navigation link
- Info about SSD service & link to accessibility page
- Descriptive headings added (hidden from GUI with CSS)
- Access keys for navigating pages with keyboard

Left thumbnail (smaller rendering) repeats:

Text-Only Concatenated View – Top

Entire book presented on 1 page

Skip Links

Descriptive Headings

Linked ToC with descriptive link labels

Hathi Trust Digital Library - Ancient man; the beginning of civilizations, by Hendrik Willem ...

For questions or comments, contact DLPS: dlps-help@umich.edu

Go to full bibliographic information
Go to table of contents
Skip table of contents and go to first page

Full Bibliographic Information

Title: Ancient man; the beginning of civilizations, by Hendrik Willem Van Loon.
Author Van Loon, Hendrik Willem, 1882-1944.
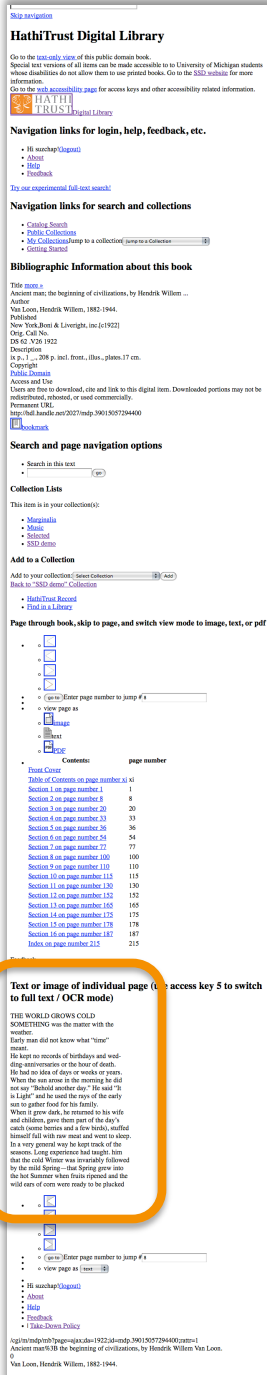**Published** New York,Boni & Liveright, inc.[c1922]
Orig. Call No. DS 62 .V26 1922
Description ix p., 1 _., 208 p. incl. front., illus., plates.17 cm.
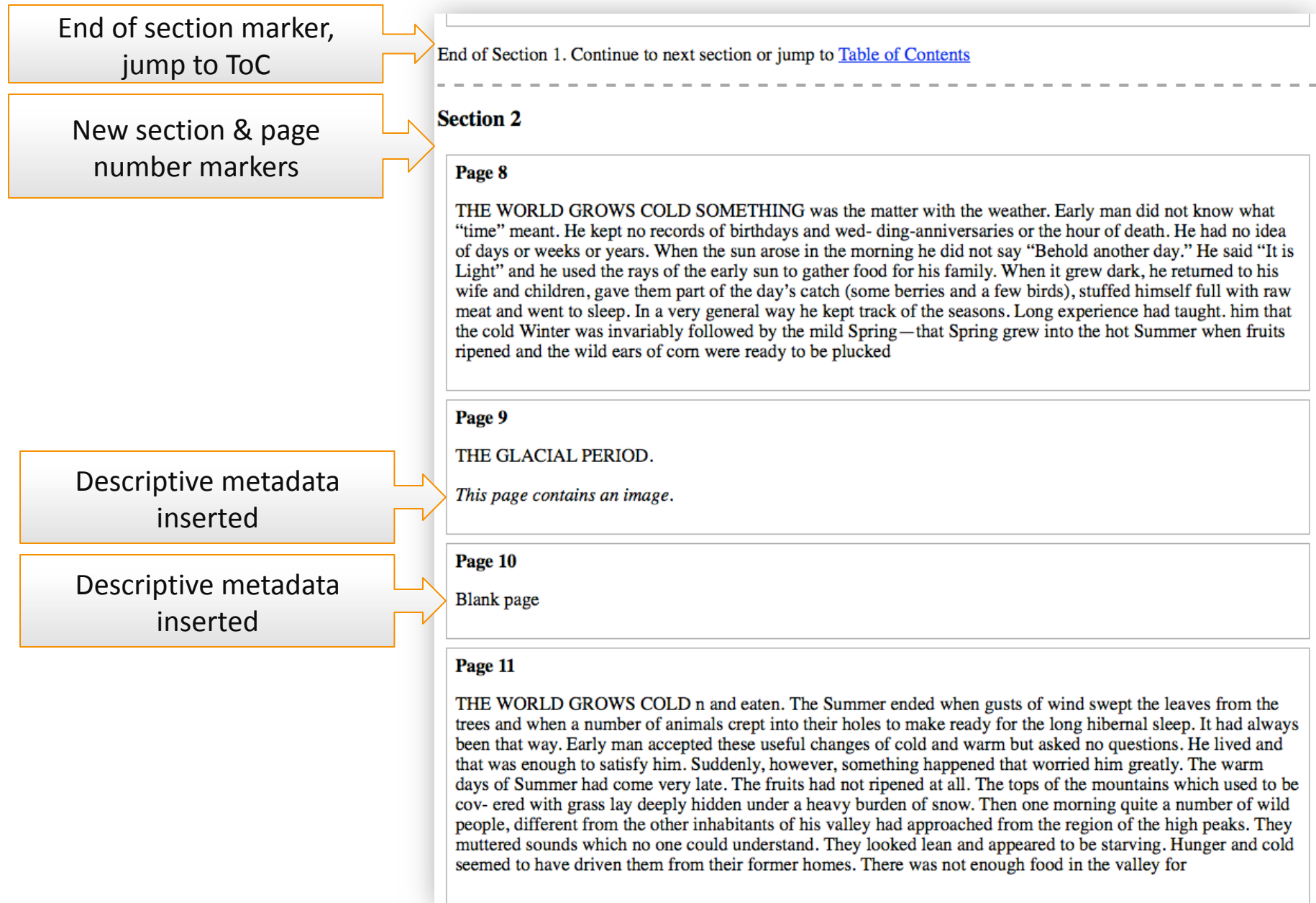**Copyright** Public Domain
Access and Use Users are free to download, cite and link to this digital item. Downloaded portions may not be redistributed, rehosted, or used for commercial purposes.
Permanent URL http://hdl.handle.net/2027/mdp.39015057294400

Table of Contents

Text-Only Concatenated View – Sections & page content

End of section marker, jump to ToC

New section & page number markers

Section 2

Page 8

THE WORLD GROWS COLD SOMETHING was the matter with the weather. Early man did not know what "time" meant. He kept no records of birthdays and wed- ding-anniversaries or the hour of death. He had no idea of days or weeks or years. When the sun arose in the morning he did not say "Behold another day." He said "It is Light" and he used the rays of the early sun to gather food for his family. When it grew dark, he returned to his wife and children, gave them part of the day's catch (some berries and a few birds), stuffed himself full with raw meat and went to sleep. In a very general way he kept track of the seasons. Long experience had taught. him that the cold Winter was invariably followed by the mild Spring—that Spring grew into the hot Summer when fruits ripened and the wild ears of corn were ready to be plucked

Page 9

THE GLACIAL PERIOD.

This page contains an image.

Descriptive metadata inserted

Page 10

Blank page

Descriptive metadata inserted

Page 11

THE WORLD GROWS COLD n and eaten. The Summer ended when gusts of wind swept the leaves from the trees and when a number of animals crept into their holes to make ready for the long hibernal sleep. It had always been that way. Early man accepted these useful changes of cold and warm but asked no questions. He lived and that was enough to satisfy him. Suddenly, however, something happened that worried him greatly. The warm days of Summer had come very late. The fruits had not ripened at all. The tops of the mountains which used to be cov- ered with grass lay deeply hidden under a heavy burden of snow. Then one morning quite a number of wild people, different from the other inhabitants of his valley had approached from the region of the high peaks. They muttered sounds which no one could understand. They looked lean and appeared to be starving. Hunger and cold seemed to have driven them from their former homes. There was not enough food in the valley for

Tools

- Web Developer Firefox Toolbar

- WAVE Firefox extension

- Accessibility Firefox extension

- Fangs Firefox extension

- FAE evaluator (UIUC)

- Cynthia Says WCAG & Section 508 evaluator

www.hathitrust.org

Outside Evaluations

- Official endorsement from National Federation of the Blind

- Favorable review from Allan Adler, Vice President for Legal and Governmental Affairs at the Association of American Publishers

- Exploring other collaborations to evaluate

www.hathitrust.org

Just the beginning

- Continue to improve accessibility & incorporate methods into development process

- We see this as a prototype & hope to expand to other HathiTrust institutions & public

- Exploring use of Shibboleth auth system

- Explore development of an OCR correction mechanism, descriptions for images

- PR to notify more potential users

Over 2,800,000 accessible volumes, and counting

"It's like we've given all of these people a library card"

-Jack Bernard, assistant general counsel and chair of the University's Council for Disability Concerns

www.hathitrust.org

HathiTrust Large-scale Ingest

Jessica Feeman
May 4th 2009

Copyright Review and the Copyright Review Management System

Anne Karle-Zenith

University of Michigan Library

Legal Context

- Copyright law is not clear-cut

- Necessitates interpretation of ambiguous info

- Liability – wrong conclusions, mistakes fall on us

- UM willing to take risks , but dictated UM Office of General Counsel

- OGC: Follow Google's rules for now

  (exception: government documents)

HATHI TRUST Digital Library

Try our experimental full-text search!

| Catalog Search | Public Collections | My Collections | Getting Started |

Hathi Trust Digital Library - Catalog Search

Catalog Search [ ] All Fields [▲▼] (Find) Search Tips

Similar Items

Hygienic fundamentals of food handling by Charles Thom and Albert C. Hunter ...
By: Thom, Charles 1872-1956
Published: (1924)

Cheese William Elsschot ; translated and with a preface by Paul Vincent
By: Elsschot, Willem 1882-1960
Published: (2002)

Book
By: Railroadians of America
Published: (1939)

The book
By: Harvey, W. H. 1851-1936
Published: (1930)

A book by Djuna Barnes
By: Barnes, Djuna
Published: (1923)

↺ Back to Search Results
← Previous record Showing record 18 of 224 Next record →

✉ Email this

The book of cheese,

Viewability:

📄 Full text

Main Author:	Thom, Charles, 1872-1956.
Other Authors:	Fisk, Walter Warner,
Language(s):	English
Published:	New York, The Macmillan company, 1918.
Subjects:	Cheese
Physical Description:	xvi p., 1 l.,392 p. front., illus. (incl. maps, plans) diagrs. 20 cm.
Original Format:	Book
Original Classification Number:	SF 271 .T45
Locate a Print Version:	Find in a library

Viewability:

📄 Full text (original from University of Michigan)

HATHI TRUST Digital Library

Try our experimental full-text search!

Catalog Search | Public Collections | My Collections | Getting Started

Hathi Trust Digital Library - Catalog Search

Catalog Search [   ] [All Fields ▼] (Find) Search Tips

Similar Items

Three novels Soft soap; The leg; Will-o'-the wisp Willem Elsschot ; [translated by A. Brotherton].
By: Elsschot, Willem 1882-1960
Published: (1965)

Kaas [deur] Willem Elsschot [pseud.] Met aantekeninge deur A.P. Grové.
By: Elsschot, Willem 1882-1960
Published: (1969)

Kaas [van] Willem Elsschot [pseud.]
By: Elsschot, Willem 1882-1960
Published: (1969)

Het dwaallicht [van] Willem Elsschot [pseud.]
By: Elsschot, Willem 1882-1960
Published: (1969)

Lijmen Het been [van] Willem Elsschot [pseud.]
By: Elsschot, Willem 1882-1960

↩ Back to Search Results
Showing record 1 of 224 Next record ⇒

✉ Email this

Cheese / William Elsschot ; translated and with a preface by Paul Vincent.

Viewability:

🔍 Search-only (no full-text)

Main Author:	Elsschot, Willem, 1882-1960.
Other Authors:	Vincent, Paul
Language(s):	English
Published:	London ; Granta Books, 2002.
Subjects:	Clerks (Retail trade) > Belgium > Antwerp > Fiction.
Physical Description:	xvi, 134 p. ; 20 cm.
Original Format:	Book
Original Classification Number:	839.38 R542ka tV77
Locate a Print Version:	Find in a library

Viewability:

🔍 Search-only (no full-text) *(original from University of Michigan)*

HathiTrust Stats

Currently Digitized (as of 4/30/09):
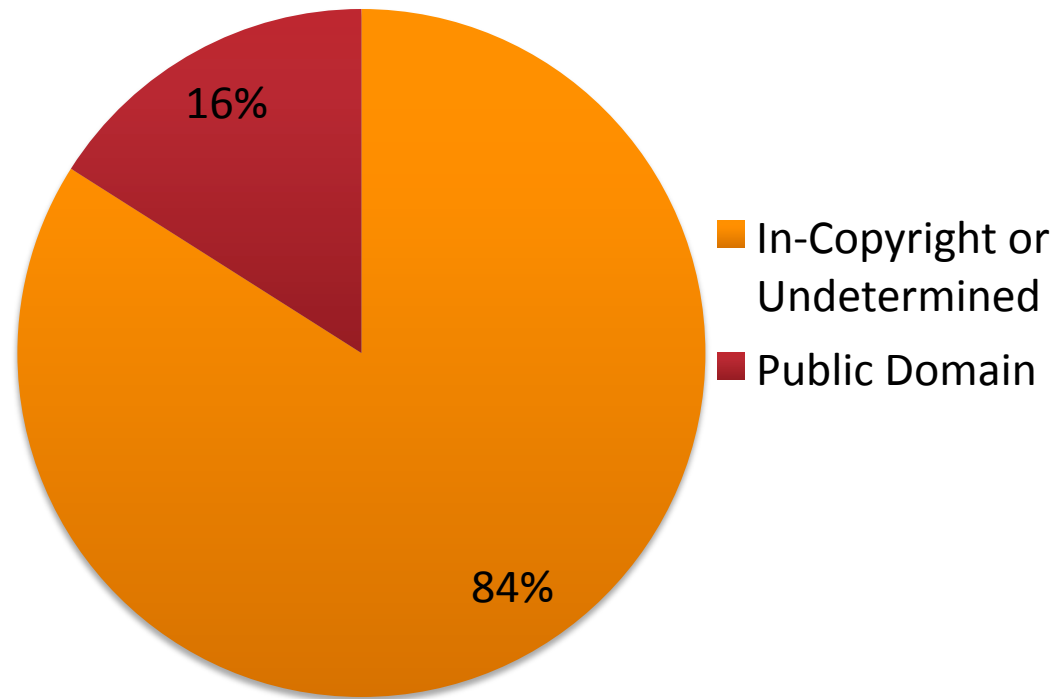
2,821,596 volumes

987,558,600 pages

105 terabytes

33 miles

2,293 tons

434,393 volumes in the public domain

Viewability Breakdown

16%

In-Copyright or Undetermined

Public Domain

84%

Initial Bibliographic Determinations

Place of publication	Publication date	Other condition	Rights extract from Aleph	Copyright status in rights db
US	Any date	US Govt. Pub*	3rd letter in location fixed field = "u"; "f" in GovPub fixed field	pd
US	1923 or later	Not US Govt. Pub	3rd letter in location fixed field = "u"; date >1922'; no "f" in GovPub fixed field	ic
US	1922 or earlier	Assuming not previously published outside US*	3rd letter in location fixed field = "u"; date<1922	pd
Any	No date information			und

Increasing Access –
Copyright Review Work

US works published 1923 through 1963 could be in the public domain if:

– they were published without a proper copyright notice on the piece (copyright/©+date)

– they were published with notice but the copyright was not renewed (after 28 years)

Increasing Access –
Copyright Review Work

- Beginning in early 2007, started copyright review work in Technical Services Division

- Work spread over 5 staff members in Electronic Access Unit

- Adds up to approximately 1.5 FTE

Increasing Access – Copyright Review Work

- To check for copyright notice:
staff have special authorization to access in-copyright works in HathiTrust via the pageturner

TROUBLEMAKERS

REBELLIOUS YOUTH IN
AN AFFLUENT SOCIETY

T. R. FYVEL

SCHOCKEN BOOKS · NEW YORK

This item is not in any of your collections

Add to your collection:
Select Collection ▼
Add

MGet It

HathiTrust Record

Find in a Library

|◄ ◄ ► ►|

go to #

size 50% ▼

rotate ↶ ↷

view page as
📄 image
📄 text
📄 PDF

Contents:

HV
7069
.F77
19~2

The present volume is published in Great Britain and Canada under the title THE INSECURE OFFENDERS

© T. R. Fyvel 1961, 1962

© T. R. Fyvel 1961, 1962

Library of Congress Catalog Card No. 62–13142

Manufactured in the United States of America

Increasing Access – Copyright Review Work

- To check for the renewal:
  staff check the Stanford Renewal Database

Copyright Renewal Database

Home

Browse

Simple Search

Advanced Search

About the Collection

Resources

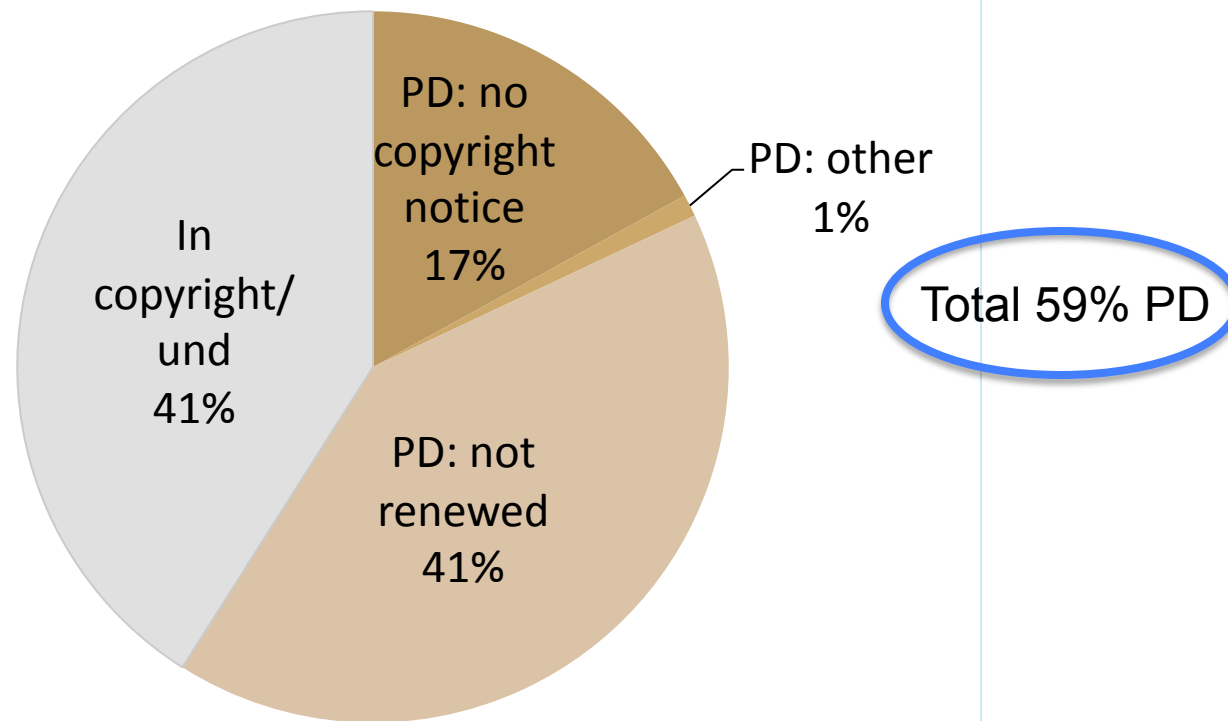## Search Results

5 results found.

great gatsby

Modify search | Search within results | Printer friendly | Download

Renewal Id	Title	Author	Registration Number	
R109367	The great Gatsby	FITZGERALD, FRANCIS SCOTT	A855444	Long record
RE448661	School edition of the Great Gatsby.	Francis Scott Key Fitzgerald, foreword and study guide: Charles Scribner's Sons, employer of author of new material in a work made for hire: Albert K. Rideout.	A527692	Long record
R283215	The great Gatsby. on introd	FITZGERALD, F. SCOTT	A75521	Long record
RE512449	The Great Gatsby; a study. Edited and with an introd. by Frederick J. Hoffman.	-	A566456	Long record
R462860	The last tycoon. Edmund Wilson, foreword. (With The great Gatsby & Selected stories) NM: foreword & The last tycoon	FITZGERALD, F. SCOTT	A158462	Long record

Copyright Review Work Results

In copyright/ und 41%

PD: no copyright notice 17%

PD: other 1%

PD: not renewed 41%

Total 59% PD

Review Stats:

- Total time spent (since 2007): 2,569 hrs
- Ave. hrs per month: 77
- Ave. number of items reviewed per month: 3300
- Ave. time spent per review: 84 secs

www.hathitrust.org

Review Stats:

As of May 1, 2009:

—over 57,000 volumes reviewed

—approx 34,000 volumes opened up/ now available as full-text

Issues w/copyright review

- Reliability of determinations
  - Stanford db (proving a negative)
  - complexities (many use cases)
  - human error
- Consistency – ditto
- Time - amount reported does not include admin time
- Data Management - spreadsheets (lots & lots!)

IMLS National Leadership Grant

Submitted proposal to build a
Copyright Review Management
  ("CRMS")

- more efficient system for managing copyright determinations work
- prioritization and workflow management
- tools to facilitate determination
- "double blind" checking to ensure reliability
- point of collaboration with other institutions

IMLS National Leadership Grant

Three years starting 12/1/08:

- Defined requirements/developed beta - 12/08-1/09

- Beta version up and running; EA Unit tests and provides feedback - 2/09

- Started Usability testing - 2/09

www.hathitrust.org

Programmer leaves to run family lumber company – 3/09

IMLS National Leadership Grant

- Interim Programmer assigned; work continues - 3/09

- Currently integrating results of usability testing; refining design and functionality before final testing and tweaking

- Plan to have CRMS version 1.0 in production by end of 5/09

- Cease manual review by 6/09

IMLS National Leadership Grant

Further out:

– Share results (e.g., OCLC Copyright Evidence Registry; OAI feed of determinations) – by fall 2009

– Copyright Office searches/compare results – Spring 2010

– Release version 2.0 by June 2010

– Train and bring on at least 3 other institutions by Fall 2010 (Indiana, Minnesota, Wisconsin)

www.hathitrust.org

For more info…

- http://www.lib.umich.edu/grants/crms/

- annekz@umich.edu

thank you!

- http://www.hathitrust.org/
  - hathitrust-info@umich.edu
  - heather.christenson@ucop.edu
  - tburtonw@umich.edu
  - suzchap@umich.edu
  - jhovater@umich.edu
  - annekz@umich.edu
  - jpwilkin@umich.edu