Contextualizing Subject Access Across Digital Collections

The "See Also" Problem

Joseph B. Dalton
The New York Public Library
Research Libraries
Digital Library Program

DLF Fall Forum 2006 - Nov. 9, 2006
What We’ll Cover

- Overview of Problem
- Some Approaches
- Expanding Subject-Access
- Problems, Opportunities and Challenges
Background: Some Numbers

- NYPL Digital Gallery collections
  - 524,000 images
  - 318,000 bibliographic (item-level) records

- NYPL Digital Library Program uses several subject thesauri
  - LCTGM, LCSH, LCNAF, AAT, GMGPC, etc.

- Number of records containing at least 1 subject heading:
  - 260,000 (81 % of total)
The Problem

- 58,000 subject headings indexed for searching NYPL Digital Gallery
- Browsing a list this size is like trying to find a needle in a haystack
  - Lincoln:
    ‘Lincoln, Abraham’
  - Posters depicting circuses:
    ‘Posters - - Circus’ or ‘Circus Posters’?
Possible Approaches for Subject Browsing

- Parse or index list by subjects’ facets
  - 19th century - French - Posters
    - French - Posters - 19th century
    - Posters - French - 19th century

- Map subjects to selected thesauri and provide cross-references

- Build hierarchical browsing on top of some taxonomy
  - Printing & Graphics > Circus Posters
  - Culture & Society > Posters - French - 19th century
Our Approach

- Create separate index of subjects in Lucene
- Index the pointers from those subjects to their associated objects or containers
- Provide front-end context lists, like “subjects in [Collection|Title],” by filtering on the objects
- Provide free-text retrieval of subjects through some kind of “subject finder”
Indexing Field Relationships at Object-Level

- First Goal: provide object-based lists
  - Gather all subjects and some associated bibliographic identifiers: item-level, title-level, collection-level, etc.

<table>
<thead>
<tr>
<th>query</th>
<th>HITS</th>
<th>IMAGE_ID</th>
<th>PARENT_ID</th>
<th>STRUC_ID</th>
<th>SUBJECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>176</td>
<td></td>
<td>306136</td>
<td></td>
<td>Book jackets</td>
</tr>
<tr>
<td>2</td>
<td>176</td>
<td></td>
<td>569411</td>
<td></td>
<td>Quakers</td>
</tr>
<tr>
<td>3</td>
<td>176</td>
<td></td>
<td>569413</td>
<td></td>
<td>Quakers</td>
</tr>
<tr>
<td>4</td>
<td>176</td>
<td></td>
<td>604621</td>
<td></td>
<td>Abolitionists</td>
</tr>
<tr>
<td>5</td>
<td>176</td>
<td></td>
<td>580961</td>
<td></td>
<td>Quakers</td>
</tr>
<tr>
<td>6</td>
<td>176</td>
<td></td>
<td>610436</td>
<td></td>
<td>Reformers</td>
</tr>
<tr>
<td>7</td>
<td>176</td>
<td></td>
<td>138222</td>
<td></td>
<td>Baseball</td>
</tr>
<tr>
<td>8</td>
<td>176</td>
<td></td>
<td>138226</td>
<td></td>
<td>Baseball</td>
</tr>
<tr>
<td>9</td>
<td>176</td>
<td></td>
<td>138228</td>
<td></td>
<td>Baseball</td>
</tr>
<tr>
<td>10</td>
<td>176</td>
<td></td>
<td>138230</td>
<td></td>
<td>Baseball</td>
</tr>
</tbody>
</table>
Subject Display by Parent-Title Object

**Browse Source**
1 2 3

**Source:** Callin's North American Indian portfolio. Hunting scenes and amusements of the Rocky mountains and prairies of America.

<< Back to thumbnail images

<table>
<thead>
<tr>
<th>This Source</th>
<th>All Gallery Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animal behavior</td>
<td></td>
</tr>
<tr>
<td>Antelopes</td>
<td></td>
</tr>
<tr>
<td>Archery</td>
<td></td>
</tr>
<tr>
<td>Assiniboine Indians</td>
<td></td>
</tr>
<tr>
<td>Bear dance</td>
<td></td>
</tr>
<tr>
<td>Bear hunting</td>
<td></td>
</tr>
<tr>
<td>Bison</td>
<td></td>
</tr>
<tr>
<td>Bison hunting</td>
<td></td>
</tr>
<tr>
<td>Buffalo dance</td>
<td></td>
</tr>
<tr>
<td>Disguise</td>
<td></td>
</tr>
<tr>
<td>Horses</td>
<td></td>
</tr>
<tr>
<td>Indian dance -- North America</td>
<td></td>
</tr>
<tr>
<td>Indians -- Kings and rulers</td>
<td></td>
</tr>
<tr>
<td>Indians of North America -- Games</td>
<td></td>
</tr>
<tr>
<td>Indians of North America -- Hunting</td>
<td></td>
</tr>
<tr>
<td>Indians of North America -- West (U.S.)</td>
<td></td>
</tr>
<tr>
<td>Mandan Indians</td>
<td></td>
</tr>
<tr>
<td>Wolves</td>
<td></td>
</tr>
</tbody>
</table>

[clear all]
Subject Display by Group of Collections

All Collections > Explore Printing & Graphics
Advertising, book binding, dust jackets, friendship books, menus, post cards, posters, trade and greeting cards.

Search Collections

Subjects in these Collections

A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z

Aaron (Biblical priest)
Abbeys
Abraham (Biblical patriarch)
Absecon (N.J.)
Absecon Lighthouse (Atlantic City, N.J.)
Abu Simbel (Egypt)
Academic costume
Acadia -- Maps -- Early works to 1800
Acadians
Accidents
Acoma (N.M.)
Acoma Indians -- Clothing & dress
Acco
Acrobats
Acropolis (Athens, Greece)
Actaeon (Greek mythology)
Acting
Actors
Actors -- France
Actors -- Great Britain
Actresses
Adam (Biblical figure)
Adams, John, 1735-1826 -- Birthplace
Adelbert College
Adirondack Mountains (N.Y.)
Admirals -- German
Adobe buildings
Adobe buildings -- Arizona -- Tucson
Adobe buildings -- New Mexico
Adobe buildings -- New Mexico -- Albuquerque
Adobe buildings -- New Mexico -- Santa Fe
Adobe houses -- California -- Monterey
Indexing Subjects A-Z

- Early Assumption:
  1 row = 1 subject ‘field’ per object ID

- Reality:
  subjects are indexed as ‘multiple fields’

<table>
<thead>
<tr>
<th>STRUC_ID</th>
<th>SUBJECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>306136</td>
<td>Book jackets</td>
</tr>
<tr>
<td>569411</td>
<td>Quakers</td>
</tr>
<tr>
<td>569413</td>
<td>Quakers</td>
</tr>
<tr>
<td>604621</td>
<td>Abolitionists</td>
</tr>
<tr>
<td>580961</td>
<td>Quakers</td>
</tr>
<tr>
<td>610436</td>
<td>Reformers</td>
</tr>
<tr>
<td>138222</td>
<td>Baseball</td>
</tr>
</tbody>
</table>
Personal reminiscences of the anti-slavery and other reforms and reformers. By Aaron M. Powell.
New Opportunity: “Related Subjects”

As we examined this problem, some “opportunities” emerged:

- Results might be expanded by each subject’s association to its “related” (item-level) subjects

- These results resemble a reverse-mapping of NYPL Digital Gallery subjects, derived not from an applied top-level taxonomy but from the objects’ “descriptions” of themselves
Example of Basic Term-Matching Results

25 subjects for "Posters"
Your basic search for posters found these subjects:

Cine Rivoli (Buenos Aires, Argentina) -- Posters
Cine Teatro Standard (Buenos Aires, Argentina) -- Posters
Circus posters
Motion picture posters
New York (N.Y.) -- Posters
Performing arts posters
Political posters
Political posters -- Soviet Union -- History
Political posters, Russian
Posters
Posters -- 1790-1799
Posters -- 1860-1869
Posters -- 1870-1879
Posters -- French -- 19th century
Posters -- United States
Posters -- United States -- 1910-1919
Salon Garibaldi (Buenos Aires, Argentina) -- Posters

Explore Subjects (BETA)

Exact Match:

posters

Matches single word or phrase. Exact match is case sensitive.

Expanded Search:

Expands search to find more subjects.

GO

[ clear all options]

This alphabetical list of over 40,000 subjects includes people, institutions, places, and events. Library of Congress Subject Headings Thesaurus for Graphic Materials

NOTE: Many Gallery records include subject descriptions. These records may appear in the advanced searching for more results.

advanced searching for more results
“Related” Subjects Query Expands Results

781 subjects for “Posters”

Actresses
Adventure and adventurers
Advertisements
Advertisements -- New York (State) -- New York -- 1870-1879
Afforestation
African American clergy
African American farmers
African American men
African American theater
African American women
African Americans
Africans
Africans -- United States
Agricultural implements
Alcoholic beverages
Alexander, the Great, 356-323 B.C.
Alliances
Alligators

Expanded Search:

posters

Expands search to find more

GO

[ clear all options]

More about expanded search

This alphabetical list of over 781 subjects includes people, institutions, and places. Library of Congress Subject Headings (LCSH) Thesaurus for Graphic Materials

NOTE: Many Gallery records contain subject descriptions. These records are not an alternative to the names index. Try the names index advanced searching for more.
Example of “Related” Subjects Query

**Searched for:** the phrase "alligators" in Subject and "posters" in Subject

**Image Caption:** The N. Y. Sunday World

**In:** Posters : arranged by artists. > M. de Lipman (published 1893-1924)

**Published Date:** February 1895

**Library Division:** Humanities and Social Sciences Library / Art and Architecture Collection, Miriam and Ira D. Wallach Division of Art, Prints and Photographs

**Description:** 10 v., 57 cm.

**Specific Material Type:** Prints

**Subject(s):** Alligators

Night

Posters

**Collection Guide:** Turn of the Century Posters

**Digital Image ID:** 1259144

**Digital Record ID:** 611905

**NYPL Call Number:** 3-MDW+++ (Posters: arranged by artists)

[View Library Catalog Record]

[Search For More Images]
The Scrapbook Problem

- Frame-of-reference implicitly tied to single images
- In a single object (one image) a dog and cat are considered "related"
- In printer's proofs, scrapbooks of illustrations, multiple plates, etc. this notion can be problematic
- Subjects may share 1 bibliographic reference, but are they "related?"
Related Subjects: "sailboats"

Chesapeake Bay (Md. and Va.) -- 1600-1609
Children blowing bubbles
Chimneys
China
Chivalry
Christianity
Christmas cards
Cigarette cards
City & town life -- New York (State) -- New York
City & town life -- New York (State) -- New York -- 1860-1869
Cityscapes
Clogs
Clothing & dress -- Egypt -- Jizah
Clothing & dress -- Egypt -- Luxor
Clouds
Coastlines
Colossi -- Egypt -- Thebes (Extinct city)
Columns -- Egypt -- Philae
Connecticut
Costumes -- French
Croton Point (N.Y.)
Decks (Ships)
Dolphins
Dovecotes -- Egypt
Related Subjects: "sailboats"

- Children blowing bubbles
- Chimneys
- China
- Chivalry
- Christianity
- Christmas cards
- Cigarette cards
- City & town life -- New York (State) -- New York
- City & town life -- New York (State) -- New York -- 1860-1869
- Cityscapes
- Clogs
- Clothing & dress -- Egypt -- Jizah
- Clothing & dress -- Egypt -- Luxor
- Clouds
- Coastlines
- Colossi -- Egypt -- Thebes (Extinct city)
- Columns -- Egypt -- Philae
- Connecticut
- Costumes -- French
- Croton Point (N.Y.)
- Decks (Ships)
- Dolphins
- Dovecotes -- Egypt
Publisher’s Proofs
Source: Scrapbooks of colored advertising cards in English and French: volume 2.

Image Caption: [Calendars and trade cards depicting a rowing race, jesters blowing bubbles, sail boats, thread, a horseshoe, Cleopatra's Needle, fans, kimonos, a man, a woman and a dog.]

In: Publisher's proofs of the publications of L. Prang & Co., and scrapbooks of colored advertising cards, published by L. Prang & Co. and other firms. > Scrapbooks of colored advertising cards in English and French: volume 2. (published [c1876-90])

Published Date: [c1876-90]

Library Division: Humanities and Social Sciences Library / Print Collection, Miriam and Ira D. Wallach Division of Art, Prints and Photographs

Item/Page/Plate Number: 34

Medium: Chromolithographs

Specific Material Type: Prints

Subject(s): Children blowing bubbles
Fans
Sailboats
Trade cards

Notes: Publishers include: John A. Lowell & Co.; Wemple & Company.

Collection Guide: Publisher's Proofs and Related Work from L. Prang & Company

Digital Image ID: 487920

Digital Record ID: 303773
The 80/20 Problem

- How much metadata is enough metadata?
- 260,000 out of 318,000 images contain subject headings; however, 50,000 items are virtually invisible to our interface
- If user-experience proves the utility of leveraging subject headings, more staff and $ could be allocated
Expanding Subjects: The UI Problem

- Expanded ("related") subject list doesn’t fall easily into familiar categories:
  - Faceted browse display
  - “More like this” context
  - Top-down hierarchical menus

- Additional user-testing needed on GUI

- Front-end processing, though lightweight, is sometimes expensive: more can be done to optimize index and query
The Relevance Problem: Do Subjects Matter?

- Outside of specialized domains (Medline), do researchers still need subjects?
- The Big Indexers don’t care so much about subjects, they want to index all of your data: scale is where the big gains are in search now, right?
- Good subject-analysis is expensive
- Folksonomies, tags, etc. attempt to describe things the way people think of them
Flickr: "Bad Day, 1445"

Bad Day, 1445

Tags
- medieval
- illustration
- book
- men
- gibbet
- violence
- torture
- blindfold
- mutilation
- hands
- 1445
- 1440s
- cunradus
- schlapperitz
- german
- printed
- ephemera
- printedephemera
- design
NYPL: "Rubric and full-page miniature of..."
"Bad Day, 1445"

Image Caption:
Rubric and full-page miniature of men with their hands and feet being cut off.
In:
Renaissance and medieval manuscripts collection, ca. 850-ca. 1600. > [Bible History]
(created 1445)
Library Division:
Humanities and Social Sciences Library / Manuscripts and Archives Division
Description:
ff. 470, 28 x 21 cm.
Item/Page/Plate Number:
f. 372
Specific Material Type:
mss. text
Additional Name(s):
Schlaaperitz, Cunradus -- Author
Collection Guide:
Medieval and Renaissance Illuminated Manuscripts from Western Europe
Digital Image ID:
425562
Digital Record ID:
249634
NYPL Call Number:
MA 104

digitalgallery.nypl.org/nypldigital/dgkeysearchdetail.cfm...
"Bad Day, 1445"

Comments

**facecrunch pro** says:
Man, that would suck.
Posted 8 months ago. (permalink)

**facecrunch pro** says:
Do you know what language it is written in?
Posted 8 months ago. (permalink)

**pantufla pro** says:
Looks like German...see more pages by this guy at
digitalgallery.nypl.org/nypldigital/dgkeysearchresult.cfm...
Posted 8 months ago. (permalink)

**greentea flute** says:
I really like what you are doing here with your stream on flickr.... that said, this
is pretty horrific, even if fascinating.... I came across a similar book in Brittany
some few years back... I guess this is still going on, if you read the news, all
over the world... état humaine
Posted 8 months ago. (permalink)

**sadalit pro** says:
I was searching for a photo of a gibbet, came across this - your title is too
funny, but did they HAVE to? Do we still have to?
Posted 23 hours ago. (permalink)

Would you like to comment?

Sign up for a free account, or sign in (if you're already a member).
Search: violence or torture

Stringed instruments
Stroud, William
Suffolk, William de la Pole, Duke of, 1396-1450
Suriname
Swindlers & swindling
Synagogues -- Netherlands -- 1700-1799
Süss-Oppenheimer, Joseph, 1698- or 9-1738
Tacks
Tarring & feathering -- Massachusetts -- 1860-1869
Tarring & feathering -- Massachusetts -- Boston -- 1770-1779
Tarring & feathering -- Pennsylvania -- 1790-1799
Tarring & feathering -- United States
Tarring & feathering -- United States -- 1760-1769
Tax collection -- United States -- 1760-1769
Thomas, Clément
Throwing
Torture
Torture of self -- Great Plains -- 1800-1899
Torture victims
Tourniquets
Towers -- England
Tramps
Treadwheel
Trees
Search: "mutilation"

Your expanded search for mutilation found these related subjects:

Books -- Mutilation, defacement, etc.

New York Public Library

2 subjects found for this search

View all images with mutilation in subject
The New York Public Library ~ Digital Library Program

November 9, 2006

Drilled books, Mar. 27, 1913

In: New York Public Library Visual Materials > Lantern Slides > Office of Special Investigation
The Answer: [Maybe|No|Yes]?

- Images, at least, need description, and likely will for the foreseeable future(?)
- Subjects and other controlled vocabularies are good at describing while minimizing ‘noise’
- Tags are subjects, just loosely typed?
Some Further References & Acknowledgments


- Thanks to: Lee Horowitz (NYPL DGTL Oracle database consulting), Janet Murray (NYPL DGTL Metadata Coordinator), Tom Robertson (Lucene consulting, Stanford), Barbara Taranto (NYPL DGTL Director)

- Contact: jdalton@nypl.org