

# Mapping the library future

Subject navigation for today's and tomorrow's  
library catalogs

**John Mark Ockerbloom**

**ALA Catalog Form and Function Group**

**January 12, 2008**

# Take-away points

- **LC Subject Headings form a rich basis for discovery, crippled by poor discovery tools, slow evolution**
  - A deep ontology with many detailed terms and relationships
  - Network of concepts not captured by keywords or facets alone
- **Subject maps enable better subject browsing with ontologies like LCSH**
  - Multi-dimensional browsing, highlighting relevant connections
  - May also make it easier to maintain, evolve big, messy subject ontologies
- **They are practical for present and next-generation catalogs**
  - Build on legacy data, but accommodate new collecting practices
  - Both large (research library) & small (special collection) scales
  - With ordinary web browsers, servers, and ILS, and without expensive proprietary software
  - Can work well with facet browsing, tagging
- **We can build on library strengths and Web innovations, to better connect our users with the resources they need**

# The appeal of browsing

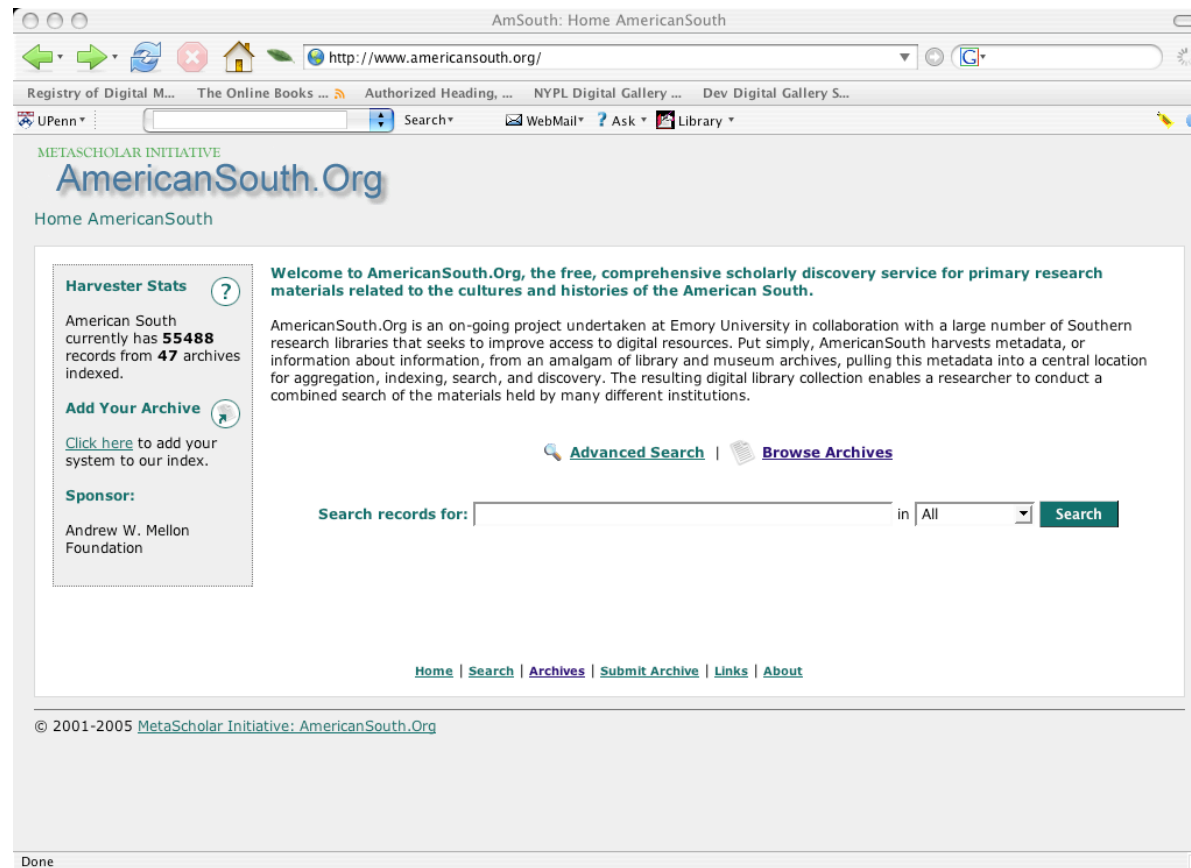


Green Apple Books (San Francisco), March 2007  
Photo by Dolan Halbrook (Copyright 2007. Creative Commons Licensed)

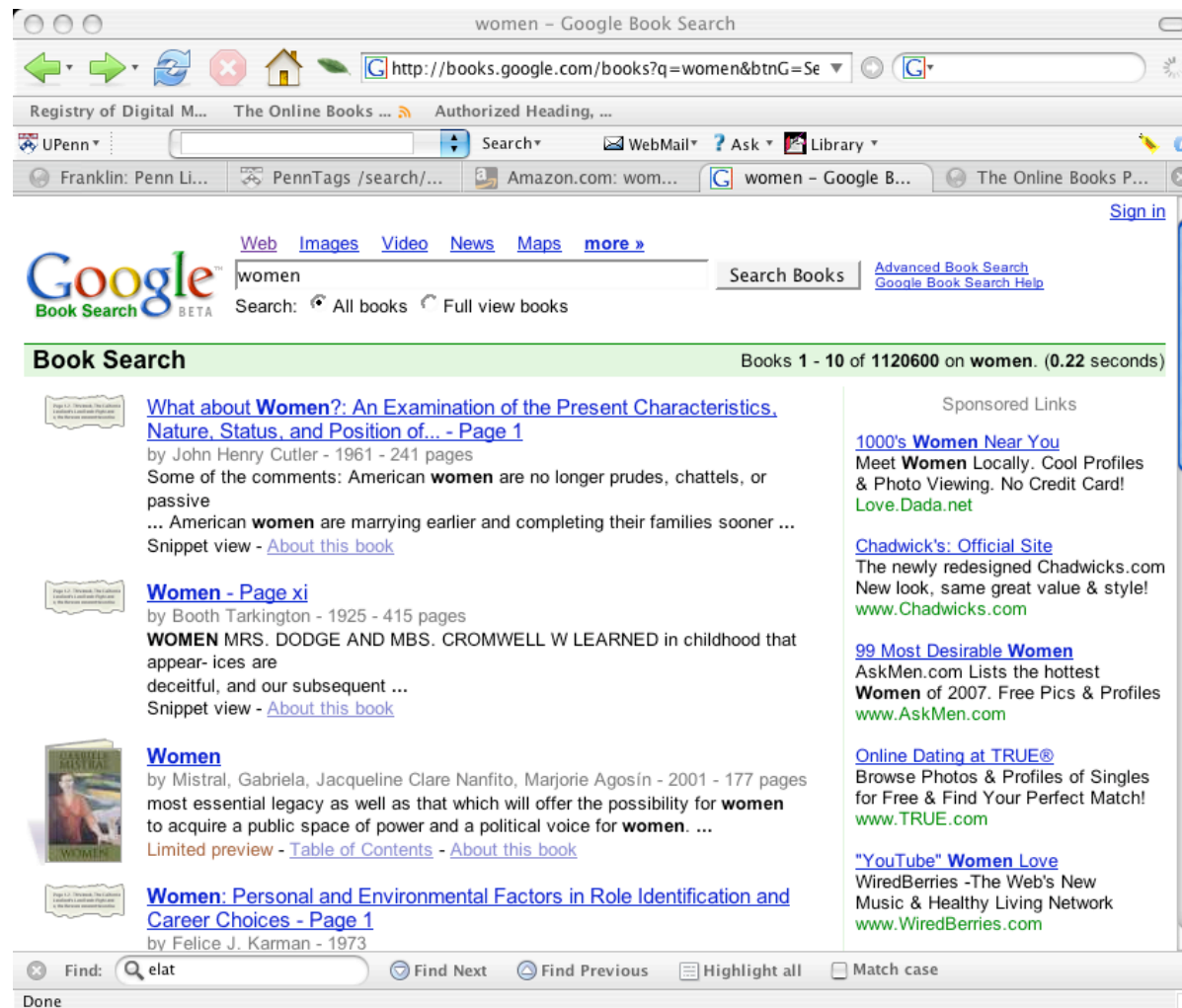
# More content not on shelves...



# ...and hard to browse

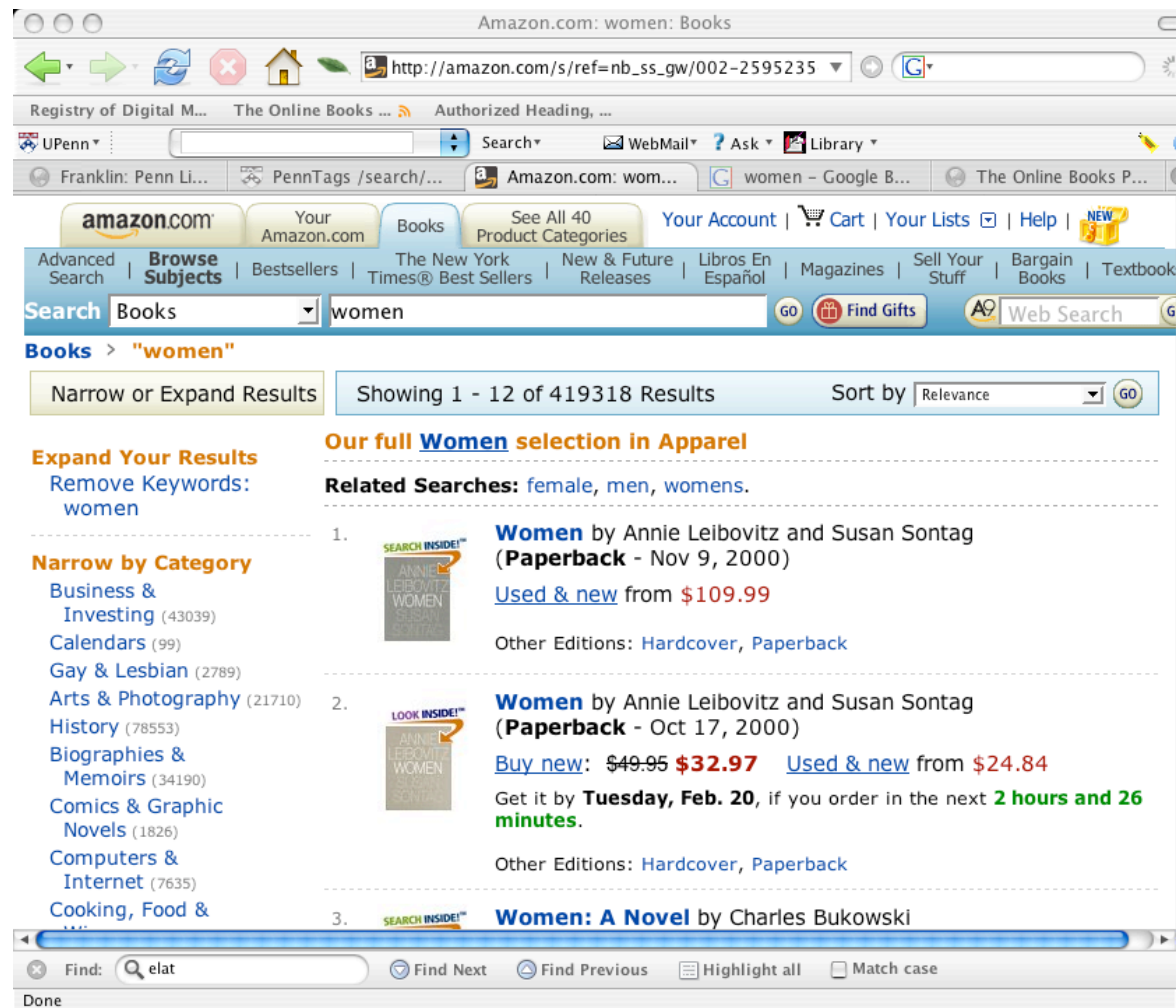


# Women in Google Books

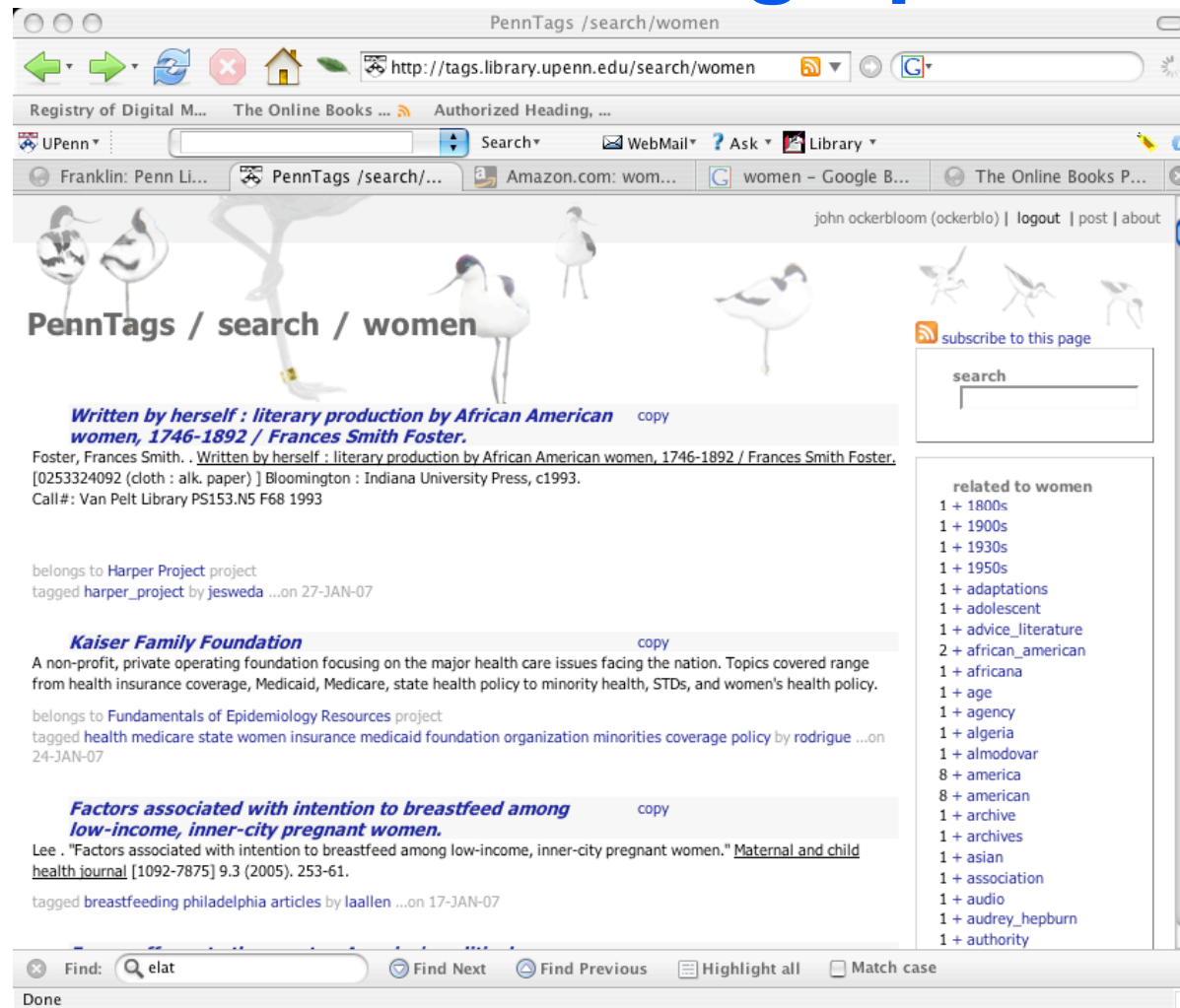




# In Amazon



# In academic tag-space





# In our catalog

Franklin: Penn Library Catalog

http://www.franklin.library.upenn.edu/cgi-bin/Pweb1

Registry of Digital M... The Online Books ... Authorized Heading, ...

UPenn Search WebMail Ask Library

Franklin: Penn Li... PennTags /search/... Amazon.com: wom... women - Google B... The Online Books P...

Other Catalogs Find Articles BorrowDirect+ Ask Us/Get Help

**Penn Library** *Franklin: Penn Library Catalog*

new search search history saved searches my account place request bookbag user guide

Search Request: Subject = women  
Search Results: Displaying 1 through 50 of 50 entries.

← previous next →

#	Titles	Headings	Headings Type
1	121	Women.	Moving image genre-form guide
2	377	Women.	Library of Congress
3	161	Women.	MESH
4	2	Women--19th century--Biography.	Library of Congress
5	1	Women--20th century--Biography.	Library of Congress
6	1	Women--20th century--Fiction.	Library of Congress
7	1	Women--20th century--Pakistan.	Library of Congress
8	1	Women abolitionists--Fiction.	Library of Congress
9	2	Women abolitionists--Great Britain.	Library of Congress
10	2	Women abolitionists--Great Britain--History.	Library of Congress
11	1	Women abolitionists--Massachusetts.	Library of Congress
12	1	Women abolitionists--Massachusetts--Boston--History--19th century.	Library of Congress
13	1	Women abolitionists--Pennsylvania--Biography.	Library of Congress

Find: elat Find Next Find Previous Highlight all Match case

Done

# The strength and the weakness of the subject catalog

- “One researcher... was interested in linguistic studies of the Cockney dialect. He simply typed ‘Cockney’ as a keyword into our catalog... [and] missed most of the linguistic studies..... The proper LC subject heading ‘English language -- Dialects -- England -- London’ rounds up in one categorical grouping all such works scattered by variant keywords-- and variant languages”
  - Thomas Mann, “Research at Risk”, in *Library Journal* online site, article dated June 15, 2005

# How do we get there?

Franklin: Penn Library Catalog

http://www.franklin.library.upenn.edu/cgi-bin/Pweb1

Registry of Digital M... The Online Books ... Authorized Heading, ...

UPenn Search WebMail Ask Library

Franklin: Penn Li... Franklin: Penn Libr... Amazon.com: wom... women - Google B... The Online Books P...

Other Catalogs Find Articles BorrowDirect\* Ask Us/Get Help

**Penn Library** *Franklin: Penn Library Catalog*

new search search history saved searches my account place request bookbag user guide

Search Request: Subject = cockney  
Search Results: Displaying 1 through 50 of 50 entries.

← previous next →

#	Titles	Headings	Headings Type
<a href="#">more info</a> 1	0	Cockpits of airplanes	Library of Congress
<a href="#">more info</a> 2	0	Cockran, Bourke, 1854-1923	Library of Congress
<a href="#">more info</a> 3	0	Cockran, W. Bourke (William Bourke), 1854-1923	Library of Congress
4	2	Cockran, William Bourke, 1854-1923	Library of Congress
5	1	Cockrell, Alan, 1949-	Library of Congress
<a href="#">more info</a> 6	0	Cockrell, F. M. (Francis Marion), 1834-1915	Library of Congress
7	2	Cockrell, Francis Marion, 1834-1915	Library of Congress
<a href="#">more info</a> 8	0	Cockrell, Mr. (Francis Marion), 1834-1915	Library of Congress
<a href="#">more info</a> 9	0	Cockrell, Senator (Francis Marion), 1834-1915	Library of Congress
10	1	Cockrill, Maurice, 1936---Criticism and interpretation.	Library of Congress
<a href="#">more info</a> 11	6	Cockroaches	Library of Congress
12	3	Cockroaches.	MESH

Find: elat Find Next Find Previous Highlight all Match case

Done

# Cataloging is messy

Franklin: Penn Library Catalog

Registry of Digital M... The Online Books ... Authorized Heading, ...

UPenn ... Search ... WebMail ... Ask ... Library ...

Franklin: Penn Li... Franklin: Penn Libr... Amazon.com: wom... women - Google B... The Online Books P...

**Penn Library** *Franklin: Penn Library Catalog*

[new search](#) [search history](#) [saved searches](#) [my account](#) [place request](#) [bookbag](#) [user guide](#)

Search Request: Subject = english language dialects england I  
Search Results: Displaying 1 through 50 of 50 entries.

[← previous](#) [next →](#)

#	Titles	Headings	Headings Type
1	1	English language--Dialects--England--Lake District	Library of Congress
2	2	English language--Dialects--England--Lancashire	Library of Congress
3	1	English language--Dialects--England--Lancashire--Texts.	Library of Congress
4	1	English language--Dialects--England--Leeds.	Library of Congress
5	1	English language--Dialects--England--Leicestershire.	Library of Congress
6	4	English language--Dialects--England--Lincolnshire.	Library of Congress
7	1	English language--Dialects--England--Lindsey.	Library of Congress
8	22	English language--Dialects--England--London	Library of Congress
9	1	English language--Dialects--England--London--Texts.	Library of Congress
10	1	English language--Dialects--England--Lorton (Cumberland)	Library of Congress
11	5	English language--Dialects--England--Maps.	Library of Congress
12	1	English language--Dialects--England--Newcastle upon Tyne (Tyne and Wear)--Social aspects.	Library of Congress
13	3	English language--Dialects--England--Norfolk	Library of Congress
14	1	English language--Dialects--England, North--Maps.	Library of Congress
15	1	English language--Dialects--England--North--Texts.	Library of Congress

Find:  [Find Next](#) [Find Previous](#) [Highlight all](#) [Match case](#)

Done

# Better browsing with facets

The screenshot shows a web browser window titled "NCSU Libraries Online Catalog: Results List". The address bar shows the URL: <http://www2.lib.ncsu.edu/catalog/?N=0&Nty=1&Ntk=Subject&view=full&Nl>. The browser has several tabs open, including "Registry of Digital M...", "The Online Books ...", "UPenn", "Google", "WebMail", "Ask", and "Library".

The main content area displays the search results for the query "women". At the top, there is a navigation bar with links: "Search the Collection", "Browse Subjects", "Services", "Library Information", "Community", and "News & Events". Below this is a "MY LIBRARY" section with links: "Library Account", "My Course Reserves", and "My Alerts".

The search results section shows "Catalog Search: women in Subject Heading" with a "Go" button and a "Start Over" link. Below this, it states "Search 'women': We found 24729 matching items. Limit results to currently available items.".

There are two main sections for narrowing results:

- Narrow By Call Number Range:** This section lists various call number ranges and their corresponding item counts, such as "A - General Works (13)", "B - Philosophy, Psychology, Religion (1066)", "C - Auxiliary Sciences of History (246)", "D - History (General) and History of Europe (633)", "E - History: America (747)", "F - America: local history (261)", "G - Geography, Anthropology, Recreation (465)", "H - Social sciences (7617)", "J - Political Science (282)", "K - Law in general. Comparative and uniform law. Jurispruden ... (293)", "L - Education (713)", "M - Music (146)", "N - Fine Arts (388)", "P - Language and literature (7450)", "Q - Science (324)", "R - Medicine (671)", "S - Agriculture (70)", "T - Technology, (282)", "U - Military science (General) (63)", "V - Naval science (14)", and "Z - Bibliography. Library Science. Information resources (ge ... (259)".
- Narrow Results By:** This section lists various subject topics and their corresponding item counts, such as "Subject: Topic", "Women (9989)", "History (5829)", "History and criticism (2681)", "Women authors (2401)", and "Social conditions (2353)". There is a "Show More ..." link.

The search results are displayed in a list format. The first result is "The book as art : artists' books from the National Museum of Women in the arts" by Wasserman, Krystyna. It is published in 2007 and is a book. The second result is "Embracing sisterhood : class, identity, and contemporary Black Women" by McDonald, Katrina Bell. It is published in 1961.

At the bottom of the page, there is a search bar with the text "Find: zua" and several buttons: "Find Next", "Find Previous", "Highlight all", "Match case", and "Phrase not found".

# Better browsing with facets

## But what are you missing?

The screenshot shows a web browser window displaying the NCSU Libraries Online Catalog search results for the query 'women'. The browser's address bar shows the URL: <http://www2.lib.ncsu.edu/catalog/?N=0&Nty=1&Ntk=Subject&view=full&Nl>. The page features a navigation bar with links to 'Search the Collection', 'Browse Subjects', 'Services', 'Library Information', 'Community', and 'News & Events'. Below this, a 'Catalog Search' section shows the search term 'women' and the selected filter 'in Subject Heading'. The results indicate that 24729 items were found. A 'Narrow By Call Number Range' section lists various subject categories and their corresponding item counts, such as 'A - General Works (13)', 'B - Philosophy, Psychology, Religion (1066)', 'C - Auxiliary Sciences of History (246)', 'D - History (General) and History of Europe (633)', 'E - History: America (747)', 'F - America: local history (261)', 'G - Geography, Anthropology, Recreation (465)', 'H - Social sciences (7617)', 'J - Political Science (282)', 'K - Law in general, Comparative and uniform law, Jurisprudence (293)', 'L - Education (713)', 'M - Music (146)', 'N - Fine Arts (388)', 'P - Language and literature (7450)', 'Q - Science (324)', 'R - Medicine (671)', 'S - Agriculture (70)', 'T - Technology, (282)', 'U - Military science (General) (63)', 'V - Naval science (14)', and 'Z - Bibliography, Library Science, Information resources (general) (259)'. A 'Narrow Results By' section on the left lists facets for 'Subject: Topic', including 'Women (9989)', 'History (5829)', 'History and criticism (2681)', 'Women authors (2401)', and 'Social conditions (2353)'. The main results list shows two items: '1. The book as art : artists' books from the National Museum of Women in the arts' by Wasserman, Krystyna, published in 2007, and '2. Embracing sisterhood : class, identity, and contemporary Black Women' by McDonald, Katrina Bell, published in 1961. The first item is marked as 'Checked Out'.

NCSU LIBRARIES

Search the Collection | Browse Subjects | Services | Library Information | Community | News & Events

MY LIBRARY: Library Account | My Course Reserves | My Alerts

Catalog Search: women in Subject Heading Go Start Over

Send search to: Go

Search 'women':

We found 24729 matching items. Limit results to currently available items.

Narrow By Call Number Range:

- A - General Works (13)
- B - Philosophy, Psychology, Religion (1066)
- C - Auxiliary Sciences of History (246)
- D - History (General) and History of Europe (633)
- E - History: America (747)
- F - America: local history (261)
- G - Geography, Anthropology, Recreation (465)
- H - Social sciences (7617)
- J - Political Science (282)
- K - Law in general, Comparative and uniform law, Jurisprudence (293)
- L - Education (713)
- M - Music (146)
- N - Fine Arts (388)
- P - Language and literature (7450)
- Q - Science (324)
- R - Medicine (671)
- S - Agriculture (70)
- T - Technology, (282)
- U - Military science (General) (63)
- V - Naval science (14)
- Z - Bibliography, Library Science, Information resources (general) (259)

Narrow Results By:

Subject: Topic

- Women (9989)
- History (5829)
- History and criticism (2681)
- Women authors (2401)
- Social conditions (2353)
- Show More ...

Brief View | Full View

Sort By: Relevance

- The book as art : artists' books from the National Museum of Women in the arts  
Author: Wasserman, Krystyna.  
Published: c2007.  
Format: Book  
Design Library  
N7433.35 .U6 W37 2007 Stacks Checked Out
- Embracing sisterhood : class, identity, and contemporary Black Women  
Author: McDonald, Katrina Bell, 1961-

Find: zua Find Next Find Previous Highlight all Match case Phrase not found

Done



# Problems with LCSH

- **Subject headings are expensive and difficult to assign, maintain**
  - And are inconsistently assigned across time and collections
- **They are difficult to use**
  - a >100 year old legacy system, with all the backward compatibility issues, mismatches with expectations that implies
  - They tend not to keep up with ongoing terminology well
- **Our patrons aren't using subject headings-based discovery much**
  - Hard to find all the terms that are relevant, or even the first one
  - With the tools we give them, can you blame them? [see critiques by Karen Schneider and others]
- **Are the costs worth the benefits?**

# Some benefits of LCSH

- **It's the metadata and ontology we have at hand**
  - That's the flip side of it being a >100 year old legacy system
  - Many resources of interest are only described by it
- **Enables great precision in identifying subject areas**
  - Millions of terms (including 200,000+ authorized headings), careful definitions, lots of explicit and implicit relationships
  - Can pinpoint items with subject as main topic more precisely than one can with keyword search
  - Much larger array of effectively controlled areas than tags or keywords
- **Ontology can support much richer subject browsing, with good tools**
  - Users like to browse (e.g. in stacks) when they can easily do so
  - Allows level of specificity not easily found with facets (e.g. general works on a topic)
  - May also allow useful browsing, adaptation when assigning headings
- **Can we revive focused topical browsing with dispersed collections?**

# Subject maps: It's the network

- **Organized networks of well-defined subject terms *and relationships between them***
  - Applied and customized to particular collections of items
- **Clustered, side-by-side displays of subjects, items, and relationships in ordinary text-based Web browsers**
  - Other display options also possible
- **Designed to function like geographic maps**
  - Get users to a [subject] area they're interested in
  - Let them see what's there, and in nearby areas, at a glance
  - Show them routes to get to nearby areas, so they can home in on what's most useful to them
  - Detail and layout helps compensate for
    - » Imprecise cataloging ("It's near the underground station")
    - » Differences in concepts and names ("Subway? Trolley? Regional rail?")
- **Work well with detailed ontologies like LCSH**

# Subject maps on a small collection

- **The Online Books Page**
  - 30,000+ items
  - 25,000+ subjects
  - Subject maps in production there since 2006
- **Live online demo:**
  - <http://onlinebooks.library.upenn.edu/subjects.html>
- **Live laptop demo:**
  - <http://127.0.0.1/books/subjects.html>
  - (it's just as fast, on this MacBook)

# Building subject maps

## Build them automatically:

- **First, create a collection-independent topic map based on authorities**
- **Second, tweak it based on local needs and data**
  - “Tweaks” can overlay, inform centrally maintained authorities
  - Sources include usage data, domain-specific knowledge bases
- **Third, adapt it to a particular [set of] collection[s]**
  - Start by mining bibliographic data
  - Add terms not already in collection-independent version
  - Prune “dead ends” from collection-independent version
  - Apply rules to create additional subject relationships

# Where do terms and relationships come from?

- **From bibliographic data**
  - Terms used in a catalog generally assumed to be valid
  - If terms not present, normalization may be possible
  - For relationships, consider assignment patterns: e.g. name as first subject followed by “<topic> -- Biography”
- **From authority files and other “canonical” data**
  - “BT”, “NT”, “UF”; lists of state names and abbreviations
- **From user data**
  - Look at search logs, tag usage and correlations, etc.
- **From facet analysis**
  - Which facets can be added or dropped? Which can have broader or narrower terms substituted? Permutations?
- **From lexical and domain analysis**
  - If “X and Y” relates to Y, it probably also relates to X
  - Recognizing geographic terms and abbreviations
- **Inclusion rules can be added or removed as required**



# Subject maps on a large collection

- **Franklin, the Penn Library catalog**
  - Over 3 million items
  - Over 1 million subject terms
  - Subject maps built, demonstrated, now being tested
    - » Can be newly generated from harvested data in ~40 minutes on up-to-date hardware with 8GB memory
- **Live online demo:**
  - <http://devplw.library.upenn.edu/cgi-bin/sdi/vbook/browse?type=lcsbc&key=Women>
  - (Now only available for limited times like this talk)
- **Or see screenshot on the next slide**

# Subject maps on Franklin

Franklin Subject Browse Demo

http://devplw.library.upenn.edu/cgi-bin/sdi/vbook/browse?type=lcsbck&k

Registry of Digital M... The Online Books ...

UPenn \* Google \* WebMail \* Ask \* Library \*

M... The ... Libra... The ... Load... Free ... The ... The ... The ... The ... Fran...

Penn Library Franklin Discovery - from the Penn Library Catalog

Other Catalogs Find Articles BorrowDirect+ Ask Us/Get Help

BETA

Browsing subject area: **Women** ([About this browser](#))

You can also [browse an alphabetical list](#) from this subject or from:

**Women**

Here are entered works on the human female. Works on female organisms in general are entered under [Females](#).

**Broader terms:**

- [Anthropology](#)
- [Females](#)
- [Man](#)
- [Sociology](#)

**Related term:**

- [Misogyny](#)

**Narrower terms:**

- [Women -- 19th century](#)
- [Women -- 20th century](#)
- [Women -- Abstracts](#)
- [Women -- Abuse of](#)

**Filed under: [Women](#)**

Full Title	Author	Date
<a href="#">Solas [videorecording] = Alone / una producci[on] de Maestranza Films (Sevilla) con la participaci[on] de Vil[A] Digital/Canal Sur TV ; producida por Antonio P. Pel[A]rez ; escrita y dirigida por Benito Zambrano.</a>		200u
<a href="#">Devaki [videorecording] / [presented by] K.D. Singh ; produced by Alchemist Ltd. ; directed by Bappaditya Bandopadhyay.</a>		2006
<a href="#">Novia que te vea [videorecording] = Like a bride / produced by Instituto Mexicano de Cinematograf[A]/Fondo de Fomento a la Calidad Cinematograf[A]lica/Producciones Arte Nuevo ; screenplay by Hugo Hiriart ; Guita Schyfter film.</a>		2006
<a href="#">Female brain / Louann Brizendine.</a>	<a href="#">Brizendine, Louann.</a>	2006
<a href="#">What Lacan said about women : a psychoanalytic study / Colette Soler ; translated by John Holland.</a>	<a href="#">Soler, Colette.</a>	2006
<a href="#">Androgens and reproductive aging / [edited by] ...</a>		

Find:  zua Find Next Find Previous Highlight all Match case Phrase not found

Done

# Integrating subject maps with a library catalog

- Our ILS (Voyager) gives us a few important tools
  - **SQL**: Data-mine and query authority and bibliographic metadata
  - **JavaScript**: Rewrite catalog web pages to link out to map tools
  - **URLs with fielded queries**: Link back to the catalog from maps
- We've already used these techniques to integrate PennTags with our existing Franklin catalog
- Better: Subject map-aware catalogs
- Better still: ILS's with well-defined APIs providing data and services to subject map and other displays
  - That's the subject of an ongoing Digital Library Federation initiative

# A larger repertoire of discovery tools

**Subject maps**

**Facets**

**Full text search**

**Social tagging**

# A larger repertoire of discovery tools

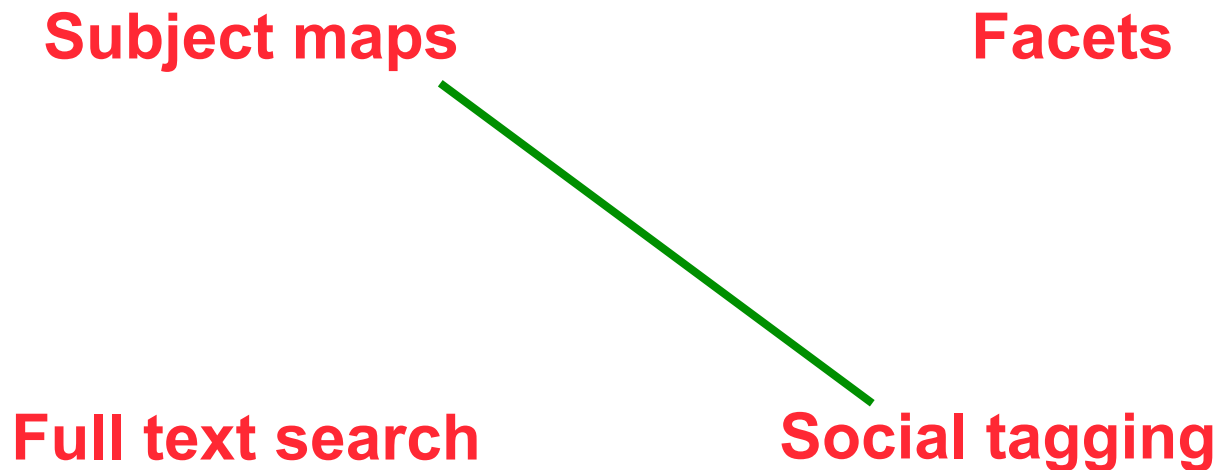
**Subject maps** ————— **Facets**

**Full text search**

**Social tagging**

- \* Subject maps, non-subject facets (e.g. media type, language, availability), can be used simultaneously
- \* Subject maps can be annotated when built so only terms that apply under arbitrary facet limitations get shown

# A larger repertoire of discovery tools



- Correlated with formal ontology terms (perhaps automatically?)
- Useful both as leadin to catalog, leadout to user-tagged items



# A larger repertoire of discovery tools

**Subject maps**

**Facets**

**Full text search**

**Social tagging**

- Search of full text (or metadata) can be used to identify frequently occurring subjects to browse
- Or, full text search limit results to items falling within a particular subject cluster

# Some early promising signs and speculations

- **Since maps introduced to The Online Books Page, users making substantially heavier use of subject browsing**
  - And going deeper into the collection than before
  - Adding basic geographic smarts was easy
- **Demos with Franklin, multiple sources show feasibility of building, using maps with larger collections**
  - Noise filtering / normalization an issue with any large collection
  - Scaling issues seem to be increasingly tractable
    - » As memory gets bigger and cheaper, map building for larger collections can be done on cheaper hardware
    - » Map navigation is computationally cheap already (since it's only looks at local portions of map)
    - » We're looking at more advanced techniques for presenting heavily populated map portions more effectively to users
- **Could we scale this up to the global level?**
  - Why not include what your users can get via ILL, other means?

# What would we like to do next?

- **Put full Penn catalog maps in front of users**
  - Will require some refinement of maps, tweaking of Voyager, possibly normalization of obsolete/improper terms
  - User studies could be useful as well, not sure we'll do this
- **Try combining it with ongoing facet work**
  - Test idea of subject maps that adapt with facet restrictions
  - See if we can combine it with Lucene/Solr views
- **Develop smarter map building and display techniques**
- **Create a subject map builders toolkit**
  - Configurable for different authorities, collections, rulesets
    - » Authorities web service could be useful here
  - Make it easier to fix headings, add tweaks, correlate with tags...
- **Think about how our users will discover the best resources (ours or others') in a networked, highly digital world**

# Conclusions

- Subject browsing is an important mechanism for resource discovery
- Detailed ontologies like Library of Congress Subject Headings can form the basis for useful browsing, if used with appropriate tools
- Subject maps are useful and practical tools for exploring collections with complex ontologies, maintaining those ontologies

**Subject maps build on library strengths to better connect our users with the resources they need in a distributed, digitized information environment**

- **For more information (demos, whitepapers...):**
  - Web: <http://labs.library.upenn.edu/subjectmaps/>
  - Email: [ockerblo@pobox.upenn.edu](mailto:ockerblo@pobox.upenn.edu)
- **Thanks!**