New Maps of the Library

Building better tools for discovery using Library of Congress Subject Headings

John Mark Ockerbloom

CNI Task Force Meeting

December 5, 2006

Take-away points

- **Subject maps enable more effective subject browsing using ontologies like LC Subject Headings**
  - An effective way to present ontologies with large number of terms and relationships
  - Navigation options not in keyword or facet-based discovery
  - May also ease care and feeding of big, messy subject ontologies
- **They can be adapted to a variety of collections (and ontologies)**
  - Both large (research library) & small (special collection) scales
- **They are feasible to implement**
  - With ordinary web browsers and servers, and without expensive proprietary software
  - Do not require a new ILS
- **They merit consideration when evaluating the future of the catalog in our institutions**

LC Subject authorities: An endangered species?

- "Opinions [of LCSH] ranged from the strongly critical to an attitude akin to quiet resignation…. There were no strong endorsements for LCSH." (p. 33)

- "[Recommendations:] ….Abandon the attempt to do comprehensive subject analysis manually with LCSH in favor of subject keywords; urge LC to dismantle LCSH" (p.18)

-- Karen Calhoun, *The Changing Nature of the Catalog and its Integration with Other Discovery Tools*, report prepared for the Library of Congress, March 2006

Rapid extinctions for controlled namespaces?

- **"The Library of Congress has determined that it will cease to provide controlled series access in the bibliographic records that its catalogers produce… The Library will implement this change on May 1, 2006."**

  -- Library of Congress announcement, April 20, 2006

  (The Library eventually waited until June, after a loud and surprised outcry from various parts of the library world)

Problems with LCSH

- **Subject headings are expensive to assign, maintain**
- **They represent a >100 year old legacy system**
  - **With all the backward compatibility issues, mismatches with user expectations, "prejudices and antipathies" that implies**
- **Our patrons aren't using subject headings-based discovery much**
  - **In large part, because Our OPACs Suck [tm, Karen Schneider]**

- **Are the costs worth the benefits?**

Benefits of LCSH

- **Enables great precision in identifying subject areas**
  - Millions of terms, 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
- **They represent a >100 year old legacy system**
  - So experts have had the chance to describe lots of resources
  - Many of those resources have no other descriptive metadata
  - Part of our descriptive infrastructure (cf. Cliff Lynch's plenary)
- **Ontology capable of supporting much richer subject browsing than most OPACs provide**
  - Users like to browse (e.g. in stacks) when they can easily do so
- **Opportunity to revive library browsing in the world of dispersed collections?**
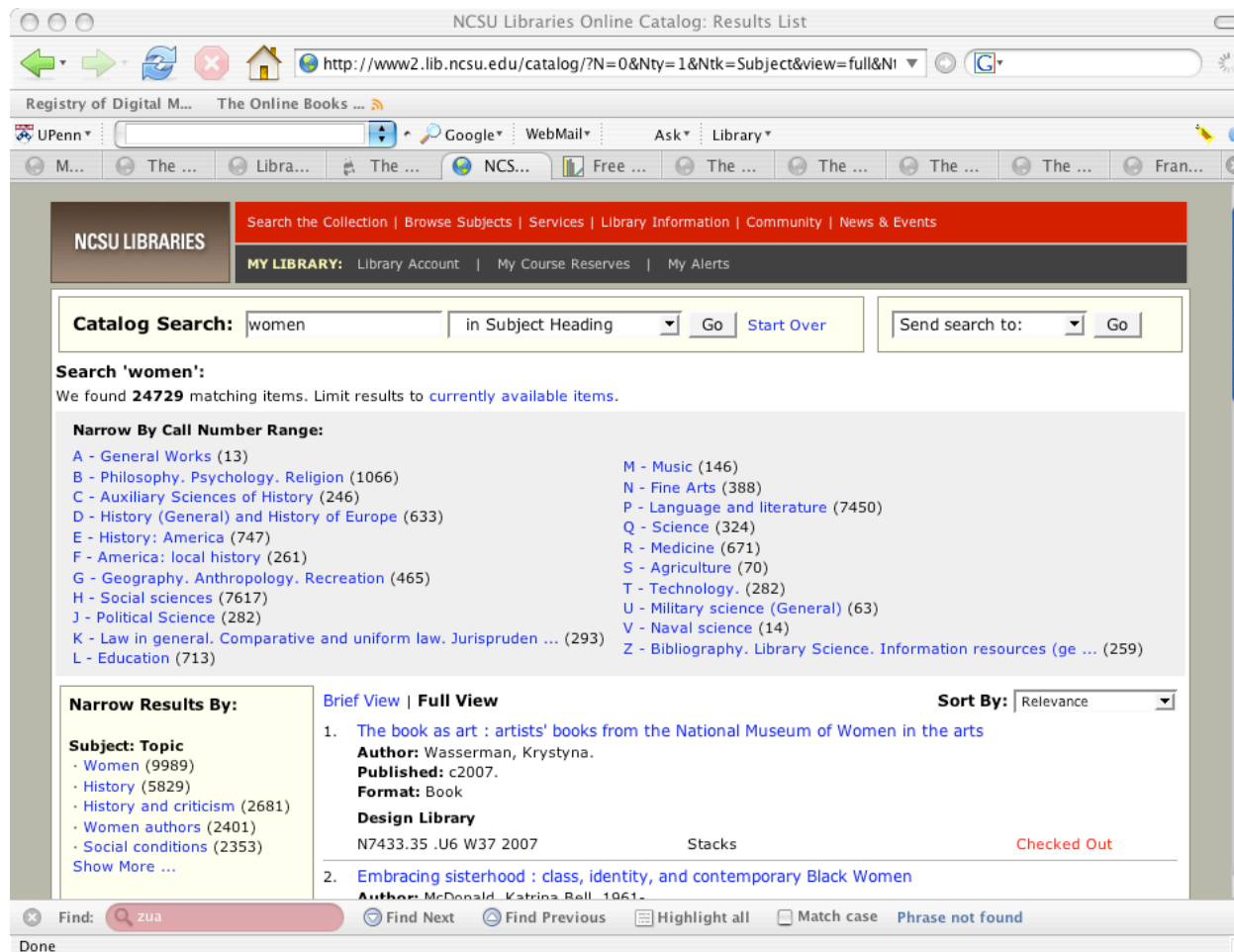
Same library, different languages

- "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

I go looking for "Cockney" in our library's catalog

- **Subject heading lookup: Nothing shown at "Cockney", nothing anywhere alphabetically close**
  - But perhaps I wanted "Cockpits of airplanes"?
- **Where Thomas Mann suggests: I get some, but miss at least as many that had been filed under a variant heading:**
  - English language -- Dialects -- London
- **Our alphabetic listing of all subjects puts**
  - English language -- Dialects -- England -- London
  - English language -- Dialects -- London

on widely separated pages
  - Other possibly relevant terms, like "English language -- Slang" are even further away

Better tools (NCSU/Endeca)

Subject maps

- **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 work 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 somewhere near the Convention Center…")
    - » Differences in concepts and names ("Wait-- THAT's the Convention Center?")

Subject maps on a small collection

- **The Online Books Page**
  - 27,000 items
  - About 13,000 subjects
  - Subject maps in production there since the summer
- **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 iBook G4)**

Building subject maps

Build them automatically:

- **First, create a collection-independent map based on authorities**

- **Second, tweak it based on local needs**
  - Log data, librarian review, user communication all helpful
  - "Tweaks" can overlay, inform centrally maintained authorities

- **Third, adapt it to a particular collection**
    - » Start by mining bibliographic data
    - » Prune "dead ends" in collection-independent version
    - » Apply rules to create additional subject relationships (see white paper for details)

Integrating subject maps with a library catalog

- **Our ILS (Voyager) gives us a few important tools**
  - We can data-mine and query authority and bibliographic metadata via SQL
  - We can rewrite catalog display pages to link out to map tools via JavaScript
  - We can link back into the catalog via constructed URLs

- **We've already used these techniques to integrate PennTags with our Franklin catalog**

- **Subject map-aware catalogs might be even nicer, but the tools above suffice to get subject maps going**

Subject maps on a large collection

- **Franklin, the Penn Library catalog**
  - **Over 2 million items**
  - **Over 1 million subject terms**
  - **Subject maps built, demonstrated, now being tested**
- **Live online demo:**
  - **http://devplw.library.upenn.edu/cgi-bin/sdi/vbook/browse?type=lcsubc&key=Women**
  - **(Now only available for limited times like this talk)**
- **Or see screenshot on the next slide**

Subject maps on Franklin

John Ockerbloom — UNIVERSITY OF PENNSYLVANIA LIBRARY — Dec. 5, 2006

Early promising signs

- **On The Online Books Page, users are making substantially heavier use of subject browsing**
  - And going deeper into the collection than before
- **Early demos with Voyager show feasibility of building, using maps with larger collections**
  - Librarian reaction enthusiastic (as has been user reaction for subject maps in the OLBP)
  - 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

What next?

- **We're evaluating interest, collaboration and application possibilities**
  - **E.g. E-Resources directory, library catalog, combination with tagging and/or facet filtering**
  - **More sophisticated map-building, display being planned**
- **See our white paper for details on how the maps are built, used**
  - **(in CNI program, and on our web site)**
- **Contact points:**
  - **Web: http://labs.library.upenn.edu/subjectmaps/**
  - **Email: ockerblo@pobox.upenn.edu**
- **What do you want your catalog to become for your users?**