INTEGRATING FOLKSONOMIES INTO CULTURAL HERITAGE DIGITAL COLLECTIONS: THE CHALLENGES AND OPPORTUNITIES OF WEB 2.0

Daniel Gelaw Alemneh
University of North Texas,
and
Dr. Samantha Kelly Hastings
University of South Carolina

ICKM 2008
INTERNATIONAL CONFERENCE ON KNOWLEDGE
Currently, the information retrieval support is limited.
Do you recognize these icons?
“The new Web is a very different thing. It is a tool for bringing together the small contributions of millions of people and making them matter.”

“…It’s a story about community and collaboration on a scale never seen before. It is about cosmic compendium of knowledge Wikipedia and the million-channel people’s network YouTube and the online metropolis MySpace. It’s about the many wresting power from the few and helping one another for nothing and how that will not only change the world, but also change the way the world changes…”

[OCLC’s report: Sharing, privacy and Trust in our Networked World, Oct. 2007:]

Graph showing the growth of total sites and active sites over time from Aug. 1995 to Sept. 2008.
Trends

- Huge multimedia digital libraries instead of documents
- Complex retrieval systems instead of matching queries and document representations
- Visualization of the information space instead of a ranked list of search results
- Human information behavior instead of information need
Trends...

User as creator, annotator, indexer, searcher, and eventual user of his/her content instead of authors and professional indexers

User’s language and vocabulary instead of controlled vocabulary

Tags and folksonomies instead of subject headings, taxonomies and classification system
Folksonomy

- Folksonomy is a user-generated system that allows users to tag their favorite digital resources with their chosen natural-language words or phrases.

- Thomas Vander Wal is credited with first coining of the term in 2005, when he mashed up the words “taxonomy” and “folk”
  - “result of personal free tagging of information and objects (anything with a URL) for one's own retrieval”
Why create tags?

- To organize information
- To support search
- To find them again later
- To discover website and share them with others
- To organize a large collection into categories in tune with the user’s own idiosyncratic mental model
- To get exposure and traffic
- To take advantage of functionality
- As a way of voicing their opinions
- To play a game…
User-Based Tagging

- **del.icio.us** is a social bookmarking site.
  - Instead of saving a Webpage link in your “Favorites” (IE) or “Bookmarks” (Firefox) folder, you save it to your del.icio.us page.

- **43Things** is like a giant, global to-do list.
  - You can add all of those things that you have been meaning to do,

- **Flickr** is a digital image storage/management site.
  - It is a place for you to organize all of your photos into albums, tag them with descriptive keywords, and view others’ images

- **Technorati** allows you to perform searches on blog content.
Advantages

Folksonomies are:

- Inclusive, democratic and self-moderating
- Current
- Lightweight
- Predisposed to discovering unknown/unexpected resources
- Folksonomies engender community
- Folksonomies offer a low cost alternative
- Folksonomies offer usability
- Expression of the direct information needs and desires of the user
  - offer insight into user behavior
Weaknesses

- There has been considerable debate concerning folksonomies flaws:
  - Folksonomies have no synonym control.
  - Folksonomies have a lack of precision.
  - Folksonomies lack hierarchy.
  - Folksonomies have a lack of recall.
  - Folksonomies are susceptible to malicious tag
Folksonomies in Cultural Heritage Institutions

- Many institutions are using or exploring social software to supplement their existing systems:
  - **PennTags**, The University of Pennsylvania library tagging system
    - allows users to bookmark and tag websites as well as library cataloging records.
  - **MBooks Collection Builder**, University of Michigan’s interface.
    - allows users to create their own collections and view public collections created by others.
    - Enhancements planned to include **MTagger**
Folksonomies in ...

- **Bibliocommons**, a social discovery system for libraries
  - transforming online library catalogues from searchable inventory systems into engaging social discovery environments.

- **CiteULike**, a social bookmarking site for academic citations
  - Allows user to use tags to store, organize, and share the scholarly papers they are reading.
  - allows users to export their libraries to BibTex or Endnote
Folksonomies in …

- **Conntoea**, references management tool
  - Allows users to find, save, customize, export or import their references

- **BibSonomy**, a system for sharing bookmarks and lists of literature.
  - allows users to import from delicious their libraries to BibTex or Endnote
Challenges and Opportunities

Digital Rights Management

“We dream of a world with free access to content. In the meantime, there’s DRM.”

Metadata Driven access

Example
Personal Digital Assistant

Synonyms
- Handheld Computer

"Alternate" Spellings
- Persenal Digitel Asistent

Abbreviations / Acronyms
- PDA

Broader Terms
- Wireless, Computers

Narrower Terms
- PalmPilot, PocketPC

Related Terms
- WindowsCE, Cell Phones
Challenges

Nature of information creation, organization, retrieval, use, and preservation is becoming more complex

- Aspects of data stewardship throughout the data lifecycle.

Changing users’ roles

- Everyone is a data provider, a search provider, and a metadata harvester.
- There is no single model that explains the behavior of users who create, index, search, & use their own content.

Free & uncontrolled users’ language and vocabulary

- Users may lack domain knowledge and/or knowledge about system.
Opportunities

Users are willing to provide descriptions of their own and others’ contents

- “Every user his/her overview”
  (Abebe’s version of Ranganathan’s 2nd law)

Rich data to study their tagging behavior

- Design of browsing interfaces could be informed by research on social tags

Designers of indexing tools & systems have a real opportunity to implement user-centered indexing

- Browsing facilitates searching by users with a vague idea about their need
Emerging concepts & agendas

Next generation catalogs

- RDA; FRBR; cataloging in hybrid & digital libraries; changes in basic cataloging work and catalogers' responsibilities; emerging perceptions of cataloging quality; how end users' expectations and behaviors affect cataloging norms; metadata records and elements in different contexts; etc.

Repositories are being deployed in a variety of settings and across a range of scales (subject, national, regional, institutional, project, lab, personal).

- The technical, managerial, practical and theoretical issues that arise from diverse applications of repositories in the increasingly pervasive information environment.
A number of professional groups are evaluating Web 2.0 tools and assess impacts and metrics relating to using the Web as a library’s core service space.

The theme of the IS related conferences, such as ASIS&T, try to address the entire phenomenon associated with Web 2.0:

- When users become active producers and contributors in the information sphere, what are the implications for LIS?
- How are social computing and Web 2.0 trends affecting the work of information professionals?
- What current research and applications are shaping future directions? …
Implications...

A growing number of non LIS Conferences and Journals reflected upon the Web 2.0 issues:

- What is the future of "top-down", cultural institutions in the age of "bottom-up", access to knowledge and cultural artifacts through Web 2.0 technologies?
- Will such institutions respond to this threat to their cultural hegemony by resistance or adaptation?
- How can a cultural heritage institution appeal to an audience which has unprecedented access to cultural resources?
- How can institutions predicated on a cultural economy of scarcity compete in an emerging state of cultural abundance?
Some things are timeless

Ranganathan’s 1931 five laws of LS

“Application of Ranganathan's Laws to the Web” (Noruzi’s 2004 versions):

1. Web resources are for use.
2. Every user his or her web resource.
3. Every web resource its user.
4. Save the time of the user.
5. The Web is a growing organism.
Summary

The ability to tag onto an OPAC or other 1.0 technologies may not fix the underlying issues of IR.

Let’s keep on tagging, anyway.

When social networking moves out of the friend’s circle and into the wider information sphere, the impact would be...
References & Web Sites Consulted

- ASIS&T: http://www.asis.org/
- CiteULike: http://www.citeulike.org/, Connotea: www.connotea.org
- del.icio.us: http://del.icio.us/, Flickr: http://www.flickr.com/
- 43Things: http://www.43things.com/
References...


- Technorati: http://www.technorati.com/


Thank you!