Subject Cataloging and Social Tagging in Library Systems

Dongyun Ni
Catalog Librarian,
University of Hawaii at Manoa Library, U.S.A.
Email: dni@hawaii.edu

Keywords (關鍵詞): Social Tagging (社會性標記); Subject Cataloging (主題編目); Tags (標記); Subject Headings (主題標目); Information Retrieval (資訊檢索); Library Data Management (圖書館資料管理)

【Abstract】
In the library world, controlled vocabularies have long been created and maintained by professional catalogers for ease of retrieving relevant library materials. Recently, social tagging has been widely applied in many social network sites such as Flickr and Delicious. This phenomenon reflects the needs of online users for information organization and retrieval in online communities. This report focuses on current development in social tagging applications in libraries and uses of subject headings data beyond traditional library catalog model. Some good practices and library system developments are introduced to demonstrate improvements information professionals have made to provide more personalized and friendly bibliographical services and assist users to explore library collections more efficiently. It also attempts to point out possible future development directions in social tagging and subject cataloging in library information systems.

【Abstract】
在圖書館界，控制詞彙早已被專業編目人員為方便使用者檢索圖書館資料所使用並加以維護。最近，社會性標記已被廣泛應用於許多社交網站，如Flickr和Delicious。這種現象反映了網路使用者對於網路社群中資訊組織和檢索的需求。本報告著重闡述當前社會性標記在圖書館的應用和主題標目資料在傳統圖書目錄模式之外的使用。同時介紹一些圖書館系統的發展和值得推薦的具體實踐以體現資訊領域專業人士為使用者提供更個性化的友善的書目服務所做的努力。本文還試圖探討主題編目和社會性標記在圖書館資訊系統未來可能的發展方向。

Introduction
Library catalogs have long provided well organized and well managed information for materials acquired by libraries. Nevertheless, this does not mean that people are willing and motivated to use integrated library
systems for retrieving their needed information. In fact, several recent studies (Kibirge & DePalo, 2000; Haglund & Olsson, 2008) have shown that a majority of library users, including faculty and professional researchers, prefer getting information online directly without ever setting foot in a library, which is actually very much in line with "the principle of least effort" (Mann, 1993). Faced with this challenging situation, librarians and information professionals have taken measures to improve integrated library systems vigorously over the past couple of decades so that they can be more responsive to users' online searching experience without having to sacrifice the quality of access to valuable library resources. Such actions include allowing users to add data to library catalogs and integrating user-supplied data into online catalogs (Library of Congress Working Group on the Future of Bibliographic Control, 2008).

Besides self reference and task organization (Golder & Huberman, 2006), online tagging adds a different layer of indexing from traditional well-controlled subject access. Rollar’s study (2009) demonstrated that user tags can enhance subject access to library collections, but cannot replace functions provided by controlled vocabularies. This report focuses on two sets of metadata that have been employed in library systems, namely, subject headings and tags. It discusses how social tagging and subject cataloging can inform each other and make library systems more usable. The methodological approach of this report is to review relevant research literature and introduce current applications and ongoing projects in this particular area of professional practice.

**Background**

**A Brief History of LCSH**

With the exploding growth of human knowledge, library collections increase rapidly in size, depth and breadth. Consequently, more logical and specific content-driven arrangement and organization of library materials is required. This is how modern classification schemes and later subject heading lists came into being around the beginning of the twentieth century (Metcalfé, 1976).

Compared to descriptive cataloging, there is considerably more intellectual effort and judgment involved in subject cataloging. This effort goes beyond mastering the cataloging rules and applying them. Subject cataloging includes two processes: subject analysis and authority control. The purpose of subject analysis and subject heading assignment is to facilitate user access to items in the catalog on a particular topic. It is especially important when the title words reveal little about the item’s content. The final step of subject analysis is to assign an appropriate call number from whatever classification system a library uses. This ensures that the item cataloged will be shelved in the library collection with other materials on the same topic and/or near closely related topics.

Authority control is the practice of creating and maintaining unique forms of name and subject headings to be used as access points in bibliographic records, and ensuring that all of the bibliographic records in the library catalog contain these authorized forms. Variant forms by which a name or subject might be searched are included as “see from” references in authority records, and these references serve to direct the searcher from the term she searched under to the authorized term.

In the United States, two major subject heading lists are used – the Library of Congress Subject Headings[1] (LCSH) and the Sears List of Subject Headings[2]. Both are based on the same principles: the reader as focus, unity, usage, and specificity (Taylor, 2004, p. 345). Because LCSH aims to cover all of the subject areas about which books in its large collections and other large North American library collections have been written, and is used as a model for many other controlled vocabulary lists worldwide, in the rest of this article, the policies and procedures of LCSH are cited to demonstrate how subject headings are created and structured.
As originally conceived, LCSH was designed to cover all of the subjects written about in the books collected by the Library of Congress. Later, a program called the Subject Authority Cooperative Program (SACO) was established to provide a means for other libraries in the United States to submit new subject headings and classification numbers to the Library of Congress for inclusion in LCSH. The cooperative effort of SACO enables the establishment of headings and classification numbers needed for items in areas in which the Library of Congress (LC) does not collect. Once those terms and classification numbers are accepted into the LC master subject authority file and classification schedule, they are available for use in library catalogs that employ LCSH.

Overview of Social Tagging

For the purpose of identification and classification, tags can be defined as "a freely chosen set of textual keywords" (Guy & Tonkin, 2006). Most keyword search devotees will find tags helpful in many cases and welcome them. Because of the huge successes of web 2.0 sites in the last few years, these sites have obtained mass recognition and attracted the mainstream media's attention. The rapid accumulation of online social activities and information sharing on these sites demands a powerful yet low-cost tool to aid in the organization and retrieval of user posts. Because "tags enable a huge amount of user-produced organizational value, at vanishingly small cost" (Shirky, 2005), many sites have adopted “tag” feature to facilitate searches and explore related tags.

Tagging in online social environment grants people a more personalized and more social user experience. It is a relatively new indexing tool developed to satisfy the needs of websites such as social network sites, wikis, and blogs, in which online information sharing and communication are supported and encouraged. However, not all online tagging practice is considered to be collaborative tagging. Sites such as MetaFilter and YouTube that only allow contributors to add tags to their own posts do not exactly fall into the category of collaborative tagging even though tag usage data is used to illustrate popular topics/interests and monitor changes.

Outside of a particular online communication environment, an individual's tagging may be trivial and useless to others; however, collaboration among large numbers of users makes the magic happen. Collaborative tagging emphasizes utilizing collective effort to classify and index resources. In stark contrast to professional cataloging and indexing, this bottom up approach relies on “the wisdom of the crowd” – the collective opinion of contributors. Collaborative tagging systems have been widely adopted to "organize, browse and publicly share personal collections of resources on the World Wide Web ..." (Quintarelli, Resmini, & Rosati, 2006).

In collaborative tagging, each user contributes tags based on his/her personal and social needs. It is possible that some users may tag a resource with biased or improper terms. However, an aggregated tag frequency graph can show a drop off (either steeper or more gradual) among these tags. Therefore, high frequency tags can indicate that users in a particular online community have reached a consensus on what a given resource is about (Kipp & Campell, 2006), and those low frequency tags will consequently be weeded out. From a statistical point of view, the larger the volume of tags is, the better a resource will be described. This is the significance of collaborative tagging.

However, a study on Delicious, Flickr, and YouTube (Ding et al., 2009) found that the tag frequency analysis cannot be used as a reliable indicator reflecting changing trends in user interests because "the focus of tagging activities ... is not on the intellectual content of resources but on more superficial features..." This finding also supports Guy and Tonkin's view on the real problem of social tagging, which is that tags are "trying to serve two masters at once; the personal collection, and the collective collection" (2006). Therefore, tag analysis in online catalog need to be carefully evaluated and justified as well.
The most important reason for adopting social tagging in the library world is to achieve "the benefits of evolution and growth" (Steele, 2009). Evolving technologies, like Internet and social networking applications, foster new information seeking patterns among young users. Attwood (2009) reported that a “want it now” culture has formed among students. Facing this kind of challenge, libraries play a more vital role in demonstrating explicitly how to gather and evaluate information and leading the young generation to understand what true learning is. In fact, it has been reported that the need for information literacy is growing (New Media Consortium & EDUCAUSE, 2009).

Meanwhile, libraries always strive to provide excellent collections and services to users and are motivated to make changes to accommodate users’ needs. Inviting tagging is a means of making library services more social and personalized (Dempsey, 2009). In addition, all library users will find tags helpful when library materials on new concepts, trends, and events are available while new corresponding subject headings have not been established. Beyond that, collaborative tagging functionality could be added to the integrated library system and utilized as another set of metadata to enhance information organization and retrieval.

Current Development

Most librarians and information specialists agree that the use of social discovery technologies within the existing online library catalog is the most desirable and realistic approach for library information system development (Mathes, 2004; Hammond, etc., 2005; Peterson, 2008). Natural language metadata created through social tagging should be supplemental to the bibliographic data already contained in library catalogs.

Tagging Applications in Libraries

Integrating Tagging to Library Online Catalogs

Some libraries use federated search modules, which search across multiple databases (including the library catalog) and present results in an integrated access interface. Because large academic libraries have subscriptions to many large databases and also develop their own digital collections (including institutional repositories), products such as Primo® (an add-on component to the integrated library systems produced by Ex Libris) and AquaBrowser® (a SerialsSolutions product that can be used with a variety of different ILS products) provide the system to search across all the different type of resources at one time and return one integrated list of results. This one-stop search mode is becoming popular and is seen by some as the next generation library discovery tool.

In the new generation of discovery tools, social tagging has been applied to help individual library users remember and organize their resources and to support group activities and share resources. Usually, tags assigned by a community member can be viewed by others and all tags are searchable and arranged by popularity or chronology on a tags page. At the resource level, tags assigned to a particular resource are listed in a tagging panel. Tags can be edited and deleted by users and the users need to follow guidelines to do so. Tag usage and the total number of tags being used in new library discovery tools vary greatly by institution, but in no library are they as large as on some other Web-enabled sites carrying bibliographic information, such as Amazon and Google Books. A long term observation and study is needed to follow developments in social tagging in library information systems.

Acquiring External Tagging Information to Promote Library Collections

Instead of having a built-in tagging feature in the integrated library system, some public libraries choose to import tags from other social network sites to the existing public online catalog to enrich its information. A typical example is Seattle Public Library Catalog. At the end of the description of each title, there is an item called “Tags from LibraryThing.com,” which contains the tags that have been added to the same title in the
LibraryThing. The weakness of this combination is that once the user accesses the LibraryThing tab, s/he navigates away from the library catalog interface and into LibraryThing interface. For the same reason, library users cannot add tags through the tag browser.

**New Platform Supporting Social Interactions**

Another approach to encourage social interactions is to integrate a content management system supporting various user networking features into the existing library online catalog.

BiblioCommons can be considered as a product of this type. It claims to “enable rich connections around library collections — connections between our users and the content, conversations, and communities they’re most interested in” (2009). It has been adopted by Oakville Public Library. In figure 1, we can see that tags are in facet display to assist exploration and discovery. Library users can contribute tags and reviews, and communicate with others who read the same book. Users also can mark other users’ collections as trusted resources.

**Different Views of Subject Headings**

Although user-supplied tags can be served as an alternative subject search method, controlled vocabulary, such as subject headings, is the core concept in information organization and retrieval. Subject headings provide unique and consistent access points to assist users in finding items of interest to them in a fast and
accurate manner, especially for unknown titles and unfamiliar topics. This advantage should be appreciated and utilized to enhance the searchability and usability of library resources.

From the constructs of the controlled vocabulary perspective, subject headings are well defined in terms of data fields and relationship establishment. In addition, valid headings are assigned and linked based on cataloging policy and procedures to ensure content-driven organization of library materials. Therefore, most current online catalogs have perpetuated the hierarchical arrangement of subject headings which has been used since card catalog times to guide users from the general topic to the specific.

Meanwhile, international cooperation among cataloging professionals and FRBR[4] implementation in library systems will make it possible to provide easier-to-understand subject information and a better structure to assist users’ information seeking process.

**Faceted Application and Subject Cloud**

Inspired by the tag cloud concept, OCLC[5] has been experimenting with a “subject cloud” display in some of its projects. FictionFinder (http://fictionfinder.oclc.org/) is one of them. Because it is FRBR based, it integrates different types of cataloging records created in different languages by different cataloging agencies into a shared platform. Users can choose subjects in their preferred language (Fig. 2).

The controlled vocabulary supported by this site is Faceted Application of Subject Terminology (FAST), an LCSH vocabulary with simpler syntax, which makes automatic indexing possible. This “tiered” approach does not pursue precise representation as in subject strings, but instead delivers different levels of subject representation. Here’s an example of subjects display at the manifestation level (Fig. 3). In addition, FRBR enables users to limit results by edition (language or format), genre, and characters in the same place.

![Subject Clouds](image-url)
One issue with FAST is how to handle precoordi-
nated LCSH phrases (“Women in literature,” for in-
stance) so as to maintain their original meanings. Some
attempts, such as indexing their constituents separately
(Anderson and Hofmann, 2006), have been made but
have not proven satisfactory (“Women” + “Literature”
is much broader than “Women in literature” in scope).

**Useful Tools to Assist Bibliometric
Research**

Rich subject information in online catalogs can also
be used as a tool to assist research in other fields. For
example, WorldCat Identities (http://orlabs.oclc.org/
Identities/) is a useful tool for bibliometric research on
people and corporate bodies. This large data mining
project is based on WorldCat’s huge and fast-growing
numbers of bibliographic records and authority records.
It attempts to give an exhaustive collective view of
works about and/or by a particular person or corporate
body and shows publication patterns and related stud-
ies. On the home page, there is a subject cloud view to
show the top 100 identities in WorldCat. For each
identity record, authority information and bibliog-
ographic information are well organized and provide a
thorough bibliometric overview of the person. For
some famous people, a Wikipedia link is provided.
Making Library Collections More Visible to the Public

Libraries have been working hard to promote libraries’ rich collections, especially their “hidden treasures” to the public. Many libraries have published their own pages on popular social network sites to attract more people.

At the same time, continued efforts have been made to exchange bibliographic data among libraries and other types of organizations. One example of such efforts is Flickr: the Commons (http://www.flickr.com/commons), a collaborative pilot project launched in early 2008 by the Library of Congress and Flickr. An early report shows that it has increased awareness of the hidden treasures in the LC collection and “also given Library staff experience in social tagging and Web 2.0 community input and cast the Library in a leadership role for other cultural and government communities exploring Web 2.0 possibilities” (Raymond, 2008).

Tags created for the images have been helping others identify the context with more detailed descriptions. LC users are directed to the Flickr site if they want to create a tag for an image. More libraries and museums have joined the Commons. And because of the unexpected positive feedback from users in the Commons project, LC has started to contribute their historical videos to its YouTube channel (http://www.youtube.com/user/LibraryOfCongress).

Issues and Future Development

In a newly published OCLC report (2009) on a recent WorldCat end-user survey, “more subject information” is listed as #2 (32%) among the most helpful enhancements OCLC could make to assist end users in identifying the item they need. Addressing this user need will require the combination of headings and tagging to play a more important role in library arena.

Subject headings provide not only an intellectual overview of the resources in a collection, but also more search options and strategies to users. To increase subject access functionalities, Markey in his paper on future library online catalog (2007) suggested testing ranking algorithms and Relevance feedback algorithms that give much higher weights to titles and subject headings than to the words buried deep in the text. This approach takes the advantage of existing subject cataloging data to enhance the quality of searching and retrieval in online catalog.

Although there have been some efforts to make subject headings more useful and visible to library users, there is still a long way to go before the goal of searching and browsing seamlessly in both the authority file and the bibliographic catalog is achieved from the end user’s view. More importantly, information professionals need to employ more creative and flexible designs that allow users to select and form valid heading strings that meet their needs more intuitively and guide users through this process. Facet-based searching and navigation with LCSH, such as that demonstrated by OCLC’s FAST project, [6] is a promising attempt, notwithstanding the fact that it has encountered barriers handling LCSH’s complicated syntax and structure (McGrath, 2007). Equally challenging in innovative projects such as FAST is maintaining the rich meaning a heading string might convey in the process of simplifying heading structure. On the other hand, how to handle records without any subject headings assigned or more than one subject schemes (LCSH and MeSH, for example) applied in existing online catalogs should be well considered in implementing the faceted subject approach (Olson, 2007).

In the meantime, ongoing development and applications of new social discovery systems for libraries, including ongoing open source applications that support more social information discovery features, such as SOPAC, have been developed. They support more social discovery options for library users. This type of
software is built upon the existing library online catalog. How to utilize subject cataloging data to facilitate social discovery features with this new platform should be taken into consideration.

It seems that everyone agrees that the library catalog contains high quality bibliographic data. The persistent argument against authority control and subject analysis is the cost of these operations. To solve this problem, some organizations have been experimenting and exploring the possibilities of automatic subject cataloging and classification.

Compared to LCSH’s complex rules and syntax that ensure quality in subject headings, tag management is at the beginning stage and far from perfect. Common methods employed in tag management include folksonomies, taxonomies, and simple vocabulary control (combine or separate tags by their meanings) (Smith, 2008). However, there is no enforcement of such controls, which affects the accuracy and consistency of tags. The current shortcomings in tagging could discourage participation in the long run (Tedjamulia, et al., 2005).

How to enhance the quality of social tagging is an imminent task and no viable solution has yet emerged, even though some theories on structural enhancement have been proposed (Peters & Weller, 2008). Existing development models in other areas, such as “laws of quality,” which have proved successful in open-source software development, do not necessarily succeed when employed in peer production projects (Duguid, 2006). To enhance the quality of tag management, good strategies that have been applied and proved beneficial on other types of social networking sites, such as voting and badges for contributors, might be taken into consideration and examined in the context of tagging systems as well.

On the other hand, in the process of enhancing quality of tagging, the freedom of using any words as tags is restricted and it will discourage new user participation since the tagging management system will recommend conventional or popular tags based on earlier user population (Guy & Tonkin, 2006). How to solve this dilemma is a big challenge we have to face. These shortcomings and development issues need to be worked out, but collaborative tagging is surely a good strategy to balance effort and benefit in improving subject access to information.

In the online library catalog, collaborative tagging can empower library users not only to add their personal touches to the system, but also to create a community atmosphere to share good results and information through aggregation of library users’ efforts. Since online communities are established based on similar passions or interests and require significant numbers of active participants for the aggregated information to be useful, social tagging will have more impact in the online library catalog when it is used on general or popular topics. Because public libraries collect popular materials (movies, fiction, etc.) and provide information to the general public, social tagging applications might be more helpful to public library users.

Librarians dealing with non-textual materials (music, for instance) in specialized libraries have found that social tagging helps users conduct more flexible, less precise, more exploratory music information searches (Evans & Cleveland, 2008). Since school library collections are often limited and specific in their size and scope, it would be interesting to see how students create tags for books after reading them. This valid and real-time participation can be part of classroom activities to encourage students’ critical thinking.

Beyond library online catalogs, discovery of other valuable library resources with collaborative tagging seems promising. It is more so at a time when library budgets are contracting while the supply of and demand for information is growing. It is incumbent upon libraries to consider the selective adoption of Web 2.0 technologies. For example, all of the digital documents in an institutional repository should be fully cataloged. However, due to personnel and financial constraints, most materials in institutional repositories are not fully cataloged. In this case, the crowd sourcing power of social
tagging could be utilized as a complement to preexisting metadata to improve access and retrieval. In addition, social tagging is valuable in image reorganization; it can enrich bibliographic data for a library’s unique digital collections and archives, especially for electronic reproductions of rare images and manuscripts.

From the above analysis, we can see that people with different kinds of expertise are needed to develop new subject access technologies. From a system designers’ perspective, all related or interested parties should get involved and become well informed about the process of design and implementation. Collaboration is the key to realizing and maximizing the power of subject cataloging and collaborative tagging in library systems, thus ensuring the success of any future development. With a mixture of tags and traditional high quality headings in library catalogs, libraries could become the information portal of choice and thereby encourage learning and knowledge acquisition.

Notes

[1] Library of Congress Subject Headings: the current print edition (LCSH 30) was published by Library of Congress in 2007 containing headings established through December 2006 and the approximately 299,000 authority records in the file then. The up-to-date LCSH (both established and proposed) can be accessed at Library of Congress Authorities (http://authorities.loc.gov/). The guidelines for proposing new subject headings are stated in the Library of Congress publications Subject Headings Manual and Free-floating Subdivisions: An Alphabetical Index.

[2] Sears List of Subject Headings has served the needs of small and medium-sized libraries, delivering a basic list of essential headings, together with patterns and examples to guide the cataloger in creating further headings as needed. Practical features include a thesaurus-like format, an accompanying list of cancelled and replacement headings, and legends within the list that identify earlier forms of headings.

[3] Web 2.0 refers to “a second generation of web development and design which facilitates communication secure information sharing, interoperability, and collaboration on the World Wide Web. Web 2.0 concepts have led to the development and evolution of web-based communities, hosted services, and applications.”—Wikipedia.

[4] FRBR stands for Functional Requirements for Bibliographic Records. It is developed by the International Federation of Library Associations (IFLA).


References


