

SOCIAL CLASSIFICATION: PANACEA OR PANDORA?
17th Annual ASIS&T SIG/CR Classification Research Workshop
Saturday, November 4, 2006 -- Austin, TX

Workshop Co-Chairs:
Jonathan Furner (University of California, Los Angeles, USA)
Joseph T. Tennis (University of British Columbia, Canada)

ABSTRACTS OF POSTERS

October 1, 2006

<i>Social bookmarking in the enterprise</i> Michael D. Braly and Geoffrey B. Froh (University of Washington, USA)	2
<i>Cognitive operations behind tagging for one's self and tagging for others</i> Judd Butler (Florida State University, USA)	5
<i>Ranking patterns: A Flickr tagging system pilot study</i> Janet Capps (Florida State University, USA)	7
<i>Folksonomies vs. bag-of-words: The evaluation and comparison of different types of document representations</i> Anatoliy Gruzd (University of Illinois at Urbana-Champaign, USA)	9
<i>Social classification and online job banks: Finding the right words to find the right job</i> Kevin Harrington (Florida State University, USA)	12
<i>Tag distribution analysis using the power law to evaluate social tagging systems: A case study in the Flickr database</i> Hong Huang (Florida State University, USA)	14
<i>@toread and cool: Tagging for time, task, and emotion</i> Margaret E. I. Kipp (University of Western Ontario, Canada)	16
<i>Ne'er-do-wells in Neverland: Mediation and conflict resolution in social classification environments</i> Chris Landbeck (Florida State University, USA)	18
<i>Exploratory study of classification tags in terms of cultural influences and implications for social classification</i> Kyoungsik Na and Changwoo Yang (Florida State University, USA)	20
<i>Folksonomies or fauxsonomies: How social is social bookmarking?</i> Marina Pluzhenskaia (University of Illinois at Urbana-Champaign, USA)	23
<i>Shared, persistent user search paths: Social navigation as social classification</i> Robert J. Sandusky (University of Tennessee, Knoxville, USA)	25
<i>The use of collaborative tagging in public library catalogues</i> Louise Spiteri (Dalhousie University, Canada)	29
<i>Using social bookmarks in an academic setting: PennTags</i> Jennifer Erica Sweda (University of Pennsylvania, USA)	31

SOCIAL BOOKMARKING IN THE ENTERPRISE

Michael Braly <mbraly@u.washington.edu>

University of Washington, Seattle, WA, USA

Geoffrey Froh <geofff@u.washington.edu>

University of Washington, Seattle, WA, USA

Abstract

In this practitioner-oriented overview of a pilot project at a medium-sized software company, we outline the early phases of an effort to implement a *Social Bookmarking System* (SBS) within an enterprise. In particular, we discuss some of the unexpected challenges encountered with regards to potential user adoption, and the design strategy we used to address those challenges.

1. Introduction

Findability in the enterprise intranet has become an increasing critical issue with the growth in size and complexity of corporate information environments. To date, much of the solution space has focused on approaches such as the construction of rich, domain specific taxonomies and the development of sophisticated full-text search algorithms (Barrows & Traverso, 2006). These methods can be extremely expensive and require careful ongoing maintenance to succeed. While they have proved valuable, some organizations, are beginning to seek out new innovations (Damianos et al., 2006).

Social Bookmarking Systems (SBS) are a class of collaborative applications that allow users to save, access, share and describe shortcuts to web resources. Initially conceived as personal information management tools, they were designed to function as centralized storage repositories to simplify the collection of bookmarks for users who browse the Internet with more than one machine in different locations. Later, systems such as the now archetypical del.icio.us added two key features: 1) description of bookmarks with arbitrary free keywords (“tagging”), and 2) sharing of bookmarks and tags across users.

We decided to undertake a small pilot project within our own enterprise to determine whether an SBS might aid in re-findability, term extraction, and identification of communities of practice. Recent technology experiments such as IBM's Dogear (Millen et al., 2006) have suggested some promise for del.icio.us-style systems inside the corporate firewall.

2. Assessing User Readiness

One of the attractive features of social software is that they tend to be inexpensive to implement from a technical standpoint. However, because their success relies entirely on user participation, the organizational cost can be quite high. Therefore, instead of moving directly into implementation, we first conducted a user survey and series of interviews to both validate the deficiency in existing information retrieval mechanisms and gauge the receptivity to bookmarking as a possible solution. Rather than definitive data about user attitudes towards tagging, we found it difficult to elicit constructive feedback because most users – even those familiar with existing systems such as del.icio.us – did not fundamentally understand core social bookmarking concepts.

3. Communicating Concepts to Users

Based on our initial findings, we modified our project plan to focus efforts on user education. We employed a non-traditional design approach in which we identified the central features of an SBS, mapped those features to user activities, and then translated the activity scenarios into graphical comics. In architecting complex systems, comics can more effectively communicate concepts by abstracting away technical details such as the user interface (Cheng & Jao, 2006).

4. Future Work and Implications

This education strategy is incorporated into the roadmap for the future phases of the project that also includes milestones related to technical extensibility, data collection, and internal marketing to drive usage.

We believe that the most critical aspect of implementing social classification within an enterprise context may be preparing users to both understand and embrace tagging as a conceptual framework.

References

Barrows, R., and Traverso, J. (2006). Search considered integral. *ACM Queue* 4(4): 30-36.

Cheng, K., and Jao, J. (2006, March 9). Communicating concepts through comics. Retrieved April 9, 2006, from <http://kevnnull.com/presentations/iasummit2006/Communicating%20Concepts%20Through%20Comics.pdf>.

Damianos, L., Griffith, J., and Cuomo, D. (2006). Social bookmarking on a corporate intranet. *WWW2006 Tagging Workshop*, Retrieved May 3, 2006, from <http://www.rawsugar.com/www2006/28.pdf>.

Millen, D. R., Feinberg, J., and Kerr, B. (2006). Dogear: Social bookmarking in the enterprise. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Montréal, Québec, Canada, April 22-27, 2006)*: 111-120.

COGNITIVE OPERATIONS BEHIND TAGGING FOR ONE'S SELF AND TAGGING FOR OTHERS

Judd Butler <jbutler01@comcast.net>

Florida State University, Tallahassee, FL, USA

An increasingly common practice in current Web life and work is the social classification of music, images, Web sites, and other digital objects. It is through this distributed tagging that folksonomic classification and retrieval systems are formed. Cognitive psychology, through its study of perception, information processing, and memory, has helped shape human/computer interfaces and interactions. Cognitive psychology, again, has contributed to a general understanding of social classification practices through its research in recall, recognition, and categorization. Still, little if any research has isolated the cognitive operations that underlie social tagging and measured their results. Many of the questions raised about tagging may be answered through the methods of cognitive psychology and verbal protocol analysis specifically.

Verbal protocol analysis has proved to be an effective approach to exploring the mental strategies employed to accomplish tasks. A think-aloud activity requires subjects to verbalize their thoughts while performing a clearly defined task, which gives researchers access to the cognitive processes responsible for behavior (Karsenty, 2001). The identified cognitive operations vary by domain and task, but examples include recognition, labeling, hypothesis forming, confirmation and disconfirmation, conditional planning, goal referencing, summarizing, and evaluation of alternatives. The method has yielded data on decision making and problem solving in a number of domains (management, medicine, computer programming, sports, law enforcement). Verbal protocol analysis has also been effective in the study of information-seeking behavior (Chatman, 1992; Hirsh, 1999; Sullivan & Seiden, 1995; Xie & Cool, 1998; Yang, 1997).

This project examines and compares the cognitive strategies used to tag digital objects for personal retrieval (Treatment 1) and for retrieval by others (Treatment 2). Subjects will tag photographic images of objects from the domain of landscape plants (e.g. trees, vines, flowers, shrubs). The images will be selected from online collections and pilot tested for their efficacy in the study to assure they show separation between experts and novices in tag number and tag specificity. A group of 10 domain novices and a group of 10 domain experts will be selected. Each group will be assigned three tasks: tag a set of 20 photographic images only for their own

personal retrieval, tag 20 images only for others to retrieve, and tag 20 images for both their own personal retrieval and for others to retrieve.

This research will involve three data collection activities: recorded times required by each subject to tag each set of images, tag patterns that reveal behaviors of experts and novices tagging for themselves and for others, and cognitive operations at work when tagging for one's self and when tagging for others. Data analysis will demonstrate and perhaps lead to explanation of the dimensions of tagging behavior. Findings may lead to recommendations for software features that facilitate effective tagging and enhance retrieval. Successful completion of this study will inform the design of follow-up studies on cognitive operations underlying tagging behavior.

References

Chatman, E. (1992). *The information world of retired women*. New York: Greenwood Press.

Hirsh, S. (1999). Children's relevance criteria and information seeking on electronic resources. *Journal of the American Society for Information Science* 50(14): 1265-1283.

Karsenty, L. (2001). Adapting verbal protocol analysis methods to investigate speech systems use. *Applied Ergonomics* 32(1): 15-22.

Sullivan, P., and Seiden, P. (1995). Educating online catalog users: The protocol assessment of needs. *Library Hi Tech* 3(2): 11-19.

Xie, H., and Cool, C. (1998). The importance of teaching "interaction" in library and information science education. *Journal of Education for Library and Information Science* 39(4): 323-331.

Yang, S. (1997). Information seeking as problem solving using a quantitative approach to uncover the novice learners' information-seeking processed in a Perseus hypertext system. *Library and Information Science Research* 19(1): 71-92.

RANKING PATTERNS: A FLICKR TAGGING SYSTEM PILOT STUDY

Janet Capps <capps_janet@yahoo.com>

Florida State University, Tallahassee, FL, USA

Drawing upon literature from cultural linguistics, cognitive anthropology, cultural anthropology, and information science, this dual-phase content analysis pilot study of Flickr's digital social classification tagging system is designed to investigate sub-culture within and between differences in image tag annotations. Based on research of cultural influences on cognition and perception, it would be natural to expect differences in tagging behaviors from Yahoo's free image sharing website's sub-cultures to emerge reflecting both high/low-context cultural schemas.

While tagging is not a requirement, Flickr's non-hierarchical system with no previous keyword agreement allows the members to create and control the tag terms. The members may opt to join groups on a self-selected basis thus creating themed sub-cultures. These dynamically interacting online groups populated by users of varying ages, genders, ethnicity, and cognitive schemas are creating information via the descriptor tagging system.

Work by anthropologist Edward T. Hall (1997) identifies cultures as falling within a high-context to a low-context framework. This model predicts that cultures vary with the degree of explicit and implicit information in communication and cultural characteristics. For this study, *culture* embraces socially constructed situated collectivities that are united through shared schemas, goals, and symbolic communication. Sub-culture groups may share numerous cultural ideas with a larger population but are unified through ideas and interests.

The folk taxonomy (folksonomy) five typical ranking levels (unique beginner, life-form, generic, specific, and ultra specific) combined with the semantic structure of Jørgensen et al.'s (2001) *Pyramid* model (e.g., generic, specific, and abstract objects and scenes) will backbone the coding analysis process. The usefulness of the ranking patterns would have implications that a wider-range of shared cues may be communal across larger populations to facilitate personal categorization of objects. Perceived to contain the most efficient or useful information, the *basic level terms* (e.g., woman, table, oak) described by Rosch's prototype research is associated with the folksonomy generic level (1978). Measuring for distribution patterns of information delivered

through inter and intra sub-culture tagging behaviors would assist in identifying group differences and would flag future research needs.

For the inter-group analysis, Language and Culture & Society were selected for data collection from Flickr's Groups_Browse_Groups for a high probability of reflecting cultural attribute differences within the tagging conventions. From these 200 groups, memberships numbering 50-100 with 100-200 pictures were isolated for intra-group rank coding of the most commonly used tags at the individual member level.

The guiding questions are:

- What inter-group ranking patterns emerge within the tagging systems of Flickr self-selected subcultures?
- What intra-group ranking patterns emerge within the tagging systems of Flickr self-selected subcultures?
- In what ways do these inter/intra emerging patterns differ from each other?

As the debate continues between top-down driven systems versus bottom-up constructed social classification systems the presence of cultural markers within Flickr's folksonomy would fuel additional research with implications to classification systems and retrieval designs. Emerging patterns may also contribute to the evolving international cognitive models associated with 21st century information image seeking needs and will provide additional data for expanding understanding of internet tagging behaviors.

References

Hall, E. T. (1989). *Beyond culture*. Garden City, NY: Anchor Books.

Jørgensen, C., Jaimes, A., Benitez, A. B., and Chang, C.-F. (2001). A conceptual framework and empirical research for classifying visual descriptors. *Journal of the American Society for Information Science and Technology* 51(11): 938-947.

Rosch, E. (1978). Principles in categorization. In E. Rosch & B. B. Lloyd (eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.

FOLKSONOMIES VS. BAG-OF-WORDS: THE EVALUATION & COMPARISON OF DIFFERENT TYPES OF DOCUMENT REPRESENTATIONS

Anatoliy Gruzd <agrudz2@uiuc.edu>

University of Illinois at Urbana-Champaign, Champaign, IL, USA

1. Purpose

Among the factors that influence the *effectiveness* of retrieval systems, the most influential is the quality of *document representation (docrep)* (Lancaster, 1998). Most Internet search engines rely on *docreps* automatically extracted from web pages (commonly called *Bag-of-Words*). Unfortunately, this automatic approach often introduces *noise* (items unrelated to the page's core topic) to *docreps*. One way to reduce noise is to utilize user-created *docreps* which are less susceptible to it. Until recently, it was impractical to rely on user-created *docreps* on Internet-size collections. This all changed when online bookmarking web-services such as *citeulike.org* and *del.icio.us* started to appear. These bookmarking web-services made it easier for the vast Internet communities to collaborate and produce community-generated descriptors (known as *folksonomies*). Due to their multi-representational nature (from various community members), *folksonomies* provide retrieval systems with *docreps* that tend to be more user-oriented. With this observation in mind, I am investigating whether *folksonomies*-based retrieval systems would yield more relevant results than conventional systems.

2. Approach

To formally answer this question, I followed White & Griffith's (1987) methodology to determine how well descriptors from *docreps* discriminate among related (*exhaustivity level*) and unrelated (*specificity level*) documents. First, I identified clusters of documents that are related to each other by their content. Second, I counted how many descriptors on average span more than one document in the cluster (referred to here as **Count1**). Finally, I counted how many documents on average outside the cluster are spanned by the descriptors (referred to here as **Count2**). Using the average values of **Count1** and **Count2**, I compared the two different *docrep* types. To visualize and interpret results, I used Paijmans' (1993) approach to plot **Count1** and **Count2** in a manner similar to that of a precision/recall graph.

3. Data collection

For my pilot study, I randomly selected a relatively small collection consisting of 190 web pages out of more than 42,000 web pages tagged as “*article*” in *del.icio.us*. These “*articles*” were selected because they contain substantial amount of textual information and usually focus on a single topic. All 190 web pages were then group into seven topical clusters (academics, economy, science, etc. ...). Due to the small size of this pilot sample, I was able to group them manually. However, for larger collections, we will need other criteria (independent from both *foksonomies* and *Bag-of-Words*) that can be used to automatically group related web pages. Some possible candidates may include metadata generated by web pages’ creators, manually created Internet subject directories (e.g. *Yahoo! Directory*), or hyperlinks found on web pages.

4. Findings

Interestingly, my results demonstrated that *foksonomies*-based and *Bag-of-Words*-based *docreps* yielded a similar level of *exhaustivity*. On average, the number of descriptors that span three or more documents in each cluster are higher by only 1% for *foksonomies* vs. *Bag-of-Words*. However, *foksonomies*-based *docreps* have a higher *specificity* level than *Bag-of-Words*-based *docreps*. On average, for *foksonomies*-based *docreps*, the number of documents outside the cluster that are spanned by descriptors are about 10.43% less than for *Bag-of-Words*-based *docreps*. The preliminary results from this limited study demonstrated the potential advantages of *foksonomies* vs. *Bag-of-Words*. The difference probably comes from the fact that *Bag-of-Words* tends to include more common words; however, a larger scale study is needed to make more conclusive decisions.

In sum, the tools and techniques developed in this study, the implementation of White and Griffith’s methodology and Paijmans’ visualization proved to be an effective toolkit to evaluate and compare *foksonomies* vs. *Bag-of-Words*.

References

Lancaster, F. W. (1998). *Indexing and abstracting in theory and practice* (2nd ed.). Champaign, IL: GSLIS, University of Illinois at Urbana-Champaign.

Paijmans, H. (1993). Comparing the document representations of two IR-systems: CLARIT and TOPIC. *Journal of the American Society for Information Science* 44(7): 383-392.

White, H. D., and Griffith, B. C. (1987). Quality of indexing in online data bases. *Information Processing & Management* 23(3): 211-224.

SOCIAL CLASSIFICATION IN ONLINE JOB BANKS: FINDING THE RIGHT WORDS TO FIND THE RIGHT JOB

Kevin Harrington <krh04@garnet.acns.fsu.edu>

Florida State University, Tallahassee, FL

Employers seeking new employees use several methods to disseminate information about opportunities in their organizations, including the use of traditional print methods (e.g., newspaper want ads, trade journals) and newer technologies such as online databases (e.g., Monster.com and Careerbuilders.com) available on the Internet. In each of these systems, the taxonomy of classification protocols is standardized by “experts” retained by the publishers. The employers are therefore required to post their want ads in categories designed by the publishers.

Print ads are published in the classified sections of newspapers and organized under the header “Help Wanted.” The job titles are placed in hierarchical job categories such as “Medical: nursing” or “Professional: education.” In this format, the taxonomy of classification protocols tends to follow a common practice across different publishers. An employer submits an ad to a newspaper and the paper sets the category for the ad.

Today’s employers have access to online job databases to disseminate employment opportunities. Again, however, the employer is required to choose from among job headers created by the publishers of the databases. While an employer may be allowed to post additional job information in an open dialog text box, it is still constrained by the categories designed by the database publishers.

This system may lead to confusion in the job market. An employer’s human resources personnel are often required to navigate a particular job publisher’s taxonomy to determine where to place a particular job advertisement. If the publisher’s taxonomy lacks sufficient depth to include Information Technology (IT) jobs in a separate category, which is then subcategorized by area of specialization, the job advertisement will likely be placed in any one of a number of far less specific categories (e.g., Category - Manufacturing: Job Title - Lotus Notes Developer).

It is believed that deficiencies in present folksonomic online job database taxonomies result in job advertisements in the Information Technology field being placed in a wide variety of non-intuitive categories: the categorical hierarchy used by the major online job database providers is therefore inadequate to allow Information Technology workers to efficiently find available jobs. The purpose of this research is to determine whether the use of professionally-designed classification systems (e.g., those used by the Department of Labor in its Standard Industrial Classification (SIC) descriptions or those used by professional organizations such as the Association for Computing Machinery (ACM) and American Society for Information Science and Technology (ASIS&T)) on their job boards would provide a more efficient match between job seekers and employers. For example, the SIC for organizes IT jobs using the following system: Major Group 73: Business Services → Industry Group 737: Computer Programming, Data Processing.

This exploratory research project will examine the differences between folksonomic job classifications designed by the online publishers and those designed by professionals in the labor and IT fields. A qualitative analysis research method, content analysis, will be used to examine whether online database publishers' classification systems and selected professional classifications with the actual job descriptions posted by the employers in order to determine which classification system best serves the needs of employers. It is hoped that this research will lead to the design of better IT job classification systems by database publishers and will thus make the process of matching employers with IT experts far more efficient.

TAG DISTRIBUTION ANALYSIS USING THE POWER LAW TO EVALUATE SOCIAL TAGGING SYSTEMS: A CASE STUDY IN THE FLICKR DATABASE

Hong Huang <hhh3625@garnet.acns.fsu.edu>

Florida State University, Tallahassee, FL, USA

With the availability of shared systems through the Internet, social tagging systems have grown rapidly in recent years. The Flickr database is a popular open tagging photo-sharing system. Users can use self-defined tags to label their uploaded pictures and categorize their photos under different kinds of groups and communities. The rich tag clouds within Flickr serve as an important navigational tool for finding people or resources with similar social concepts. The archive of the tags, user groups' classification and abundance of other tools within Flickr enable it to become a key photo management platform used by many communities.

Power laws are observed in many fields, including physics, biology, geography, sociology, economics, linguistics, war and terrorism. The power law is among one of the most frequent scaling laws that describe the scale invariance found in many natural phenomena. Zipf's law is an applicable theory of Power law to describe the linguistic terms distribution in the literature.

The objective for our studies here is to explore the influence of images themselves to see whether large image collection mimics large textual collection and understand potential impact of cultural influence of language in social classification. The tags stored in Flickr were collected and investigated for the distribution and compared with Zipf's law distribution. Even though the Flickr or other open tagging systems like del.icio.us and Technorati may seem to be different because of the different content types they store, the social abstraction or logical data model is remarkably similar at the core which complies with the power law at some levels since they are all scale free topology networks (Barabási & Albert, 1999). However, the huge difference in social relationships between different users, groups or communities might be a cause in variations for tag usage. The tag distribution analysis using the power law will reveal the nature of the language distribution within the Flickr system's taxonomy and see whether it conforms to the power law or if the nature of the social classification process has an impact in the ecology of the language.

We randomly selected a set of 2500 users by using Flickr API and are analyzing their tag distribution when compared with the Power law. We express the tag clouds as an ordered list of tags, and a number, with the number representing the steepness of the curve, and the list representing the tags, ordered from the most popular to the least popular (Anderson, 2004). Such ordering/ranking of terms might improve the automating processing and bootstrapping in social tagging systems. The poster will present the results of this and other analysis.

References

Anderson, C. (2004, October). The long tail. *Wired* 12(10).

Barabási, A.-L., and Albert, R. (1999). Emergence of scaling in random networks. *Science* 286: 509-512.

@TOREAD AND COOL: TAGGING FOR TIME, TASK AND EMOTION

Margaret E. I. Kipp <mkipp@uwo.ca>

University of Western Ontario, London, ON, Canada

1. Purpose

Social Classification or tagging is increasingly a subject of interest in library and information science (and related fields) as social bookmarking tools such as del.icio.us have become increasingly popular. Simple visualisations such as sorting tags by frequency or displaying tag clouds in which tag size denotes popularity suggest that tagging systems form interesting new taxonomies or folksonomies of related terms. This study examines these tagging systems seeking elements of convergence and divergence with traditional cataloguing and classification theory and practice.

2. Scope

This study examines the use of unusual tags which do not fit the mould of traditional cataloguing and classification, namely, that they are not subject related. These tags include two major categories:

- affective (emotional) tags
- time and task related tags.

Examples of affective tags include interesting, fun and cool. Examples of time and task related tags include @toread, todo, and tobuy.

3. Approach

Data has been collected from del.icio.us, citeulike and connotea via python scripts designed to gather information on specific tags or URLs. The data used for this study is part of a larger study by Kipp and Campbell (2006) examining patterns in tagging. Analysis of this data showed approximately 16% of tags were time and task related. Time and task or affective tags were located in multidimensional scaling graphs of cotag (coword) data (Kipp & Campbell, 2006).

4. Findings

Many users of del.icio.us, citeulike and connotea appear to want to store more than just the subject of the documents they are bookmarking. Tags such as @toread, tobuy, todo, fun and cool suggest that users see their relationship to these documents in different ways. While the latter tags express an emotional connection to the document, the former show evidence of a desire to attach personal information management information to documents. This desire to combine personal information management and document classification echoes findings in document use research at Xerox in which users categorised items in order to better understand their relationship to other items and to tasks the users wished to perform (Sellen & Harper, 2002).

5. Implications

A large part of library science research is involved in the examination of how users seek and use information. Another important aspect of this is how they relate to information (Bates, 1998, p. 1048). Findings from this study suggest that users relate information to time related tasks, activities and their own emotional reactions.

References

- Bates, M. J. (1998). Indexing and access for digital libraries and the Internet: Human, database, and domain factors. *Journal of the American Society for Information Science* 49(13): 1185-1205.
- Kipp, M. E. I., and Campbell, D. G. (2006). Patterns and inconsistencies in collaborative tagging practices: An examination of tagging practices. To be presented at the Annual Meeting of the American Society for Information Science and Technology, Austin, TX, November 3-8, 2006.
- Sellen, A. J., and Harper, R. H. R. (2002). *The myth of the paperless office*. Cambridge, MA: MIT Press.

METHODS OF FIGHTING MADNESS: CONFLICT RESOLUTION IN SOCIAL CLASSIFICATION ENVIRONMENTS

Chris Landbeck <clandbeck@comcast.net>

Florida State University, Tallahassee, FL, USA

The advent of personal computers and the Internet has also brought about social indexing groups on a scale never before seen. These “folksonomies” organize information with little regard for the methods used in library or information sciences. The collaborative nature of folksonomies almost guarantees some sort of conflict will arise among users over the mechanisms by which subjects of interest to the group can be discussed and over how to organize and describe information of interest (Mote, 2006). If we assume that the members of the groups wish for their communities to persist, it is clear that they must protect themselves from the potential negative outcomes of either of these scenarios with some method for resolving conflicts.

What conflict resolution methods exist? More importantly, what is the relationship between these and the types of data collected, methods used to collect them, and the community built to continue the work?

This exploratory study seeks to answer these questions through content analysis. For this study, flickr.com (a collection of personal images with descriptive tags), wikipedia.org (an attempt at creating an encyclopedia of general knowledge), and espgame.org (an effort to tag all the images of the Internet) are the units of analysis; the administrative pages of each community’s websites are the units of observation. These sites were chosen because they are large, ongoing representations of the work of noteworthy and noted social classification organizations (Bearman and Trant, 2005). Note that the content of the various collections is not being studied as the data therein are, for the purposes of this work, irrelevant; pages speaking to the organization’s efforts to manage conflict among its users were sought for analysis. All pages where terms such as “conflict”, “argument”, “disagreement”, and “dispute” were used in the context of the research questions of this paper were sought; other terms were pursued if judged to be pertinent. The number of such pages was counted to determine the level of development of the conflict resolution system, and the content of the pages was analyzed to determine their nature. The results show that Wikipedia invites discussion and argument, that anyone can initiate a disagreement on any subject, at which time a resolution hierarchy should (but does not always)

come into play; it is worth noting that Wikipedia also has a highly developed vocabulary for those who “game the system” (Kroski, 2005). In Flickr, the person who posts an image is the owner of it, which gives that person the right to veto others’ descriptions of their images. At espgame.org, no conflict is possible (regardless of any disagreements), as the two people randomly paired to blindly assign mutual tags to images are both anonymous and isolated from one another; here, the system precludes conflict by eliminating the means for it to occur.

While the methods of conflict resolution (or lack thereof) are evident, the effect on both the things being classified and on the activity of classification is more elusive. Note that each of these organizations collect data differently: Wikipedia seems to be seeking what might be termed “truth” or “reality,” recognizing only the authority of consensus; Flickr seems to equally value social indexing and community building; ESPGame simply exchanges entertainment for image tags. While there is a correlation between the type of information being classified and the state of development in their conflict resolution systems in these communities, one cannot be said to cause the other; this is a topic for future research.

References

Bearman, D. & Trant, J. (2005, September). Social terminology enhancement through vernacular engagement. *D-Lib Magazine* 11(9). Retrieved July 10, 2006, from <http://www.dlib.org/dlib/september05/bearman/09bearman.html>.

Kroski, E. (2005). Tag! You’re it!. Retrieved July 10, 2006, from http://blocklevel.com/weblog/information_architecture/tag_youre_it/.

Mote, N. (2006). What we can learn from folksonomy. Retrieved July 8, 2006, from <http://fairuz.isi.edu/blog/index.php/2006/06/24/what-we-can-learn-from-folksonomy-and-delicious/>.

EXPLORATORY STUDY OF CLASSIFICATION TAGS IN TERMS OF CULTURAL INFLUENCES AND IMPLICATIONS FOR SOCIAL CLASSIFICATION

Kyoungsik Na <kn05d@fsu.edu>

Florida State University, Tallahassee, FL, USA

Changwoo Yang <cy3771@fsu.edu>

Florida State University, Tallahassee, FL, USA

These days, due to the advent of computers and the Internet, people can freely classify images, photos, personal information, and can bookmark web pages. Unlike formal classification systems (e.g., Dewey Decimal Classification, Library of Congress Classification) that have been developed and used by librarians and catalogers, indexers, or information professionals, this personal indexing phenomenon is gaining a number of casual and paraprofessional users and is growing fairly quickly by virtue of the ease and simplicity of using user-generated tags for objects on the Web.

Even though such social classification systems can facilitate information retrieval on the Web, there might be problems of precision with user-generated tagging in the process and variations of user-generated tagging because of different environments such as class, occupational, or cultural influences.

According to Nisbett et al. (Nisbett & Mascuda, 2003; Nisbett & Norenzayan, 2003) cultural cognitive theory (CCT), people from different culture have differing cognitive process. For example, East Asian cognition tends to interpret the field of view as a whole, while western cognition focuses more on separate objects oriented within the field (Nisbett, Peng, Choi, & Norenzayan, 2001; Witkin, Moore, Goodenough, & Cox, 1977). Would this tendency also be reflected in tagging and thus in social classification schemas? Culture may facilitate or impede various cognitive processes such as tagging, but little is known regarding the influences such as different interpretations of information, cognitive learning processes, or cognitive load.

This poster outlines research exploring the similarities and differences in tagging behavior (and by implication, social classification schemas) between two different cultures, Korean and American. In order to do so, cultural cognitive theory (CCT) was employed as a theoretical

framework. CCT will be used to examine how much facilitation or impedance related to cultural factors is involved in cognitive processes such as understanding the representation of images.

This pilot study employs an experimental test in which six images will be shown to ten subjects for each culture. Jørgensen's research (1997) asked subjects to perform three tasks: image describing task, image tagging task, and image descriptive memory task, and demonstrated that six images can generate enough data when individuals were writing descriptions of these images to study a variety of attributes associated with these descriptions. During the test the subjects will perform these three similar tasks. First, the subjects will describe six projected images in a few sentences. Secondly, they will tag the image attributes they observe on the screen one by one. Lastly, after finishing the two previous tasks, they will complete descriptive memory task. The last task will be performed separately after two previous tasks are done. The representational images (which are not related to each other, and are culturally neutral) will consist of several objects that combine foreground and background objects. The subjects will start off with an image that will not be difficult to tag and the complexity of the images will increase.

This study examines the differences and the similarities, if any, in tagging between two different cultural groups. It also discusses why cultural groups interpret images differently or similarly when tagging a series of images. Finally, it will discuss the results of the research for social classification.

References

- Jørgensen, C. (1997). Attributes of images in describing tasks. *Information Processing & Management* 34(2/3): 161-174.
- Nisbett, R. E., and Masuda, T. (2003). Culture and point of view. *PNAS* 100(19): 1163-1170.
- Nisbett, R. E., and Norenzayan, A. (2002). Culture and cognition. In H. Pashler & D. Medin (eds.), *Stevens' handbook of experimental psychology: Cognition* (3rd ed., Vol. 2, pp. 561-597). New York: John Wiley & Sons.
- Nisbett, R. E., Peng, K., Choi, I., and Norenzayan, A. (2001). Culture and systems of thought: Holistic vs. analytic cognition. *Psychology Review* 108: 291-310.

Witkin, H. A., Moore, C. A., Goodenough, D. R., and Cox, P. W. (1977). Field-dependent and field-independent cognitive style and their educational implication. *Review of Educational Research* 47: 1-64.

FOLKSONOMIES OR FAUXSONOMIES: HOW SOCIAL IS SOCIAL BOOKMARKING?

Marina Pluzhenskaia <pluzhens@uiuc.edu>

University of Illinois at Urbana-Champaign, Champaign, IL, USA

The idea of social bookmarking might be considered quite invigorating for LIS practitioners and researchers after the customary complaints about standard classifications being unable to keep up with rapid social, political, and technological changes.

Much has been written on folksonomy's potentials and its drawbacks (Mathes, 2004). It empowers end users to index web-based information according to their own cognitive maps and can be beneficial for communities through sharing tags. But there are some shortcomings as well. One of the most commonly admitted weakness of social tagging is its ability to work for individual collections more efficiently than for communities (Guy & Tonkin, 2006). In this sense, social bookmarking does not seem always completely social. Many people use tags that cannot be of any help to others because they make sense only in the context of their authors' interests. As a classification system per se, folksonomy, thus, exists only for its creator who knows the whole set of his/her tags and the logical connections between them, which makes it more individualistic rather than social.

This study focuses on three groups of social tags with 'different degrees of sociality'. The first group consists of tags that reflect web sites' content ("web2.0", "folksonomy", etc.). Such tags are context-independent and can be easily shared by all users. The second group is constituted by format-based tags ("article", "paper", "tutorial" and the like) and can be useful only within individual collections either for their authors or for those whose interests lie in the same knowledge domain. The third group contains user-specific tags that are virtually meaningless to anybody except their creators (idiosyncratic abbreviations, truncated, or "judgmental" tags such as "new", "checklater", "interesting", "cool", "things-to-read", etc.). In such cases, folksonomies turn into fauxsonomies, so to speak, since they lose their functionality, keeping the appearance of a tool for categorization.

The goal of this study is to identify the relative weight of the tags other than content-based in a set of sites on a particular topic. Fifty web sites on social bookmarking tagged by del.icio.us'

users (total number of users 33,229), have been analyzed. 186 common tags assigned to those sites were divided into the three above-mentioned groups.

The ratio between the content-based, format-based, and user-specific tags indicates that the shares of format-based and user-specific tags are quite noticeable - 15% and 9% respectively. In other words, almost one fourth of the tags associated with the topic in question does not make sense for any users other than their author or can be useful only within individual collections.¹ This conclusion is supported by the fact of the low frequency of their occurrence on the lists of common tags. The latter suggests that the current stage of social tagging has a potential to evolve in two opposite directions: toward further socialization, or toward deeper individualization and disintegration of online communities. Further research is necessary to show which trend is stronger.

References

Mathes, A. (2004, December). Folksonomies: Cooperative classification and communication through shared metadata. Retrieved from <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>.

Guy, M., and Tonkin, E. (2006, January). Folksonomies: Tidying up tags? *D-Lib Magazine* 12(1). Retrieved from <http://www.dlib.org/dlib/january06/guy/01guy.html>.

Notes

¹ The principles of the grouping tags for the purpose of this study will be explained and illustrated in greater detail and graphical representations of data will be provided.

SHARED, PERSISTENT USER SEARCH PATHS: SOCIAL NAVIGATION AS SOCIAL CLASSIFICATION

Robert J. Sandusky <sandusky@utk.edu>

University of Tennessee, Knoxville, TN, USA

The cognitive effort expended by users of bibliographic and other networked information systems as they develop and execute search strategies, review search results, and make relevance judgments, is a valuable source of social interaction that is not made available to other users of these systems. Searching for, reviewing and selecting system objects represent a user's *implicit classification of these objects into evaluated sub-collections* culled from the entire collection. However, networked information systems are typically designed such that users work in isolation from each other – that is, they have no awareness of other users, past or present, and no awareness of what those users are doing or may have done in the past. This poster builds upon the author's experience in designing and building – but not evaluating – a prototype user activity and search path recording and retrieval system as part of the Interspace system (Sandusky, Powell, & Feng, 1998) and explores the implications of designing networked information systems in a way that supports creation and use of stored, shared representations of users' search paths – as an implicit form of social classification – in order to improve the effectiveness of current users' interactions with the system.

While networked information systems (e.g., digital libraries; digitized collections of scholarly articles) reduce barriers of time and place, empirical research indicates that there are important social aspects related to the use of traditional library collections and services that are ignored as services are digitized (Twidale et al., 1997; Procter et al., 1997; Nardi & O'Day, 1996). Socially-based mechanisms for support of collaboration and communication are needed in order to compensate for the loss of direct face-to-face interaction suffered following the adoption of networked information systems (Ackerman, 1994; Levy & Marshall, 1995; Twidale et al., 1997).

Social navigation is defined as using information from other people to help make information-related decisions (Dieberger et al., 2000). Some systems provide real-time social information, such as the indicators of who's logged into chat or instant messaging systems. One online example is the common e-commerce recommendation affordance: other people who bought *From*

Gutenberg to the Global Information Infrastructure also bought *Digital Library Use: Social Practice in Design and Evaluation*.

Four dimensions for a social navigation capability in a networked information system can be identified:

1. *Synchronicity*: ranges from synchronous, real-time interaction (e.g., Twidale et al., many CSCW systems) to asynchronous interaction (e.g., read / edit wear, user search paths).
2. *Locale*: ranges from co-located (people collaborating on searching at a physical library) to distributed interaction (e.g., virtual reference services).
3. *Explicitness*: the degree to which users make explicit decisions / take explicit action to record, mark, categorize information in support of information sharing, ranging from explicit (e.g., tagging) to implicit (e.g., automatic recording of user search paths).
4. *Privacy*: the degree to which the identity of either the creator or user of information supporting social interaction is revealed in the context of system use.

Recording, persisting, indexing, and making user search paths available to later system users is a specific application of the notions of social navigation and history-enriched digital objects (Hill & Hollan, 1993). Consistent with Dieberger et al.'s notion of social navigation, user search paths are neither "pre-planned aspects" of the networked information system nor a result of the system's designers' notions of correct use, but are instead emergent, reusable representations of the system's natural use. In terms of the four dimensions noted earlier, a user search path system would be asynchronous, distributed, and implicit (recording and construction of the persistent indexed paths is automatic). The level of privacy can vary depending upon the details of the implementation: most likely user search paths would be anonymous, but it's conceivable that user search paths could be a means to direct social interaction between the path creator and the path finder.

Benefits to implementing user search paths in networked information systems include:

- *Reduce users' cognitive effort*: (1) users make direct use of previous users' searches, including seeing, reusing, or adapting search terms and combinations of search terms, thus improving their own search strategies; (2) users see previous users' results sets, including which items from the results set the previous user selected for viewing;

- activities such as viewing the surrogate record, viewing the full text, and saving or marking surrogate records can be distinguished separately
- *Provide social learning affordance:* exposing other users' searches may help later users learn about how to use the system effectively by revealing how other users have constructed searches (Dieberger et al., 2000)

While some research-oriented social navigation systems have been constructed and evaluated (see Dieberger et al., 2000), functional shared user search paths have yet to be built and evaluated by the LIS community. Many important research questions remain:

- Are representations of other users' search paths valuable or do they merely add to information overload?
- What are the alternative techniques for recording, representing, and indexing users' search activities, and which are most effective?
- What are the most effective way to integrate user search paths into user interfaces for information retrieval?
- Which kinds of collections benefit most from applying this approach? Large collections of objects with some degree of topical coherence or general purpose Internet search engines?
- How does the nature of the user's task affect the utility of this approach?
- What are the risks to privacy of construction of a search path recording and retrieval system a privacy risk?
- Can user search paths be used in combination with other social classification techniques, and to what effect?

References

Ackerman, M. S. (1994). Providing social interaction in the digital library. *Digital Libraries '94: Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries* (College Station, TX, June 19-21, 1994).

Dieberger, A., Dourish, P., Höök, K., Resnick, P., and Wexelblat, A. (2000). Social navigation: Techniques for building more usable systems. *Interactions* 7(6): 36-45.

Hill, W., and Hollan, J. (1993). History-enriched digital objects. *Third ACM Conference on Computers, Freedom and Privacy* (San Francisco, CA): 917-920.

Levy, D. M., and Marshall, C. C. (1995). Going digital: A look at assumptions underlying digital libraries. *Communications of the ACM* 38(4): 77-84.

Nardi, B., and O'Day, V. (1996). Intelligent agents: What we learned at the library. *Libri* 46: 59-88.

Procter, R., McKinlay, A., Goldenberg, A., Davenport, E., Burnhill, P., and Cannell, S. (1997). Enhancing community and collaboration in the virtual library. *Proceedings of the First European Conference on Research and Advanced Technology for Digital Libraries* (Pisa, Italy).

Sandusky, R. J., Powell, K., and Feng, A. (1998). Design for collaboration in networked information retrieval. In B. M. Wildemuth, K. Liberman, and D. H. Sonnenwald (eds.), *Collaboration across boundaries: Theories, strategies and technology: Proceedings of the 1998 ASIS Mid-year Meeting* (Orlando, FL, May 16-20, 1998). Retrieved June 3, 2006, from <http://www.asis.org/Conferences/MY98/Sandusky.htm>.

Swigger, K. M., and Hartness, K. (1996). Cooperation and online searching via a computer-supported cooperative problem solving environment. *Journal of the American Society for Information Science* 47(5): 370-379.

Twidale, M. B., Nichols, D. M., and Paice, C. D. (1997). Browsing is a collaborative process. *Information Processing & Management* 33(6): 761-783.

THE USE OF COLLABORATIVE TAGGING IN PUBLIC LIBRARY CATALOGUES

Louise Spiteri <louise.spiteri@dal.ca>
Dalhousie University, Halifax, NS, Canada

In recent years, significant developments have occurred in the creation of customizable user features in public library catalogues. These features offer clients the opportunity to customize their own library web page and to store items of interest to them, such as book lists. Client participation in these interfaces, however, is largely reactive; clients can select items from the catalogue, but they have little ability to organize and categorize these items in a way that reflects their own needs and language. Digital document repositories such as library catalogues normally index the subject of their contents via keywords or subject headings. Traditionally, such indexing is performed either by a librarian, or else is derived from the authors of the documents. In contrast, collaborative tagging, or folksonomies, allows anyone to freely attach keywords or tags to content. Folksonomies have the potential to add much value to public library catalogues by enabling clients to: (a) organize personal information spaces, i.e., to store, maintain, and organize items of interest in the catalogue using their own metatags; and (b) supplement existing controlled vocabularies, such as Library of Congress Subject Headings (LCSH), with tags that reflect more closely and specifically their information needs. For example, if clients use the LCSH heading *Motion pictures* to categorize items of interest, they could use their own tags, such as *Cult movies*, to sub-divide this category if no parallel terms exist in LCSH.

In order to understand more fully these applications, it is important to examine how folksonomies are structured and used, and the extent to which they reflect user needs not found in existing lists of subject headings. The purpose of this proposed research is thus to examine:

1. The structure and scope of folksonomies. How are the tags that constitute the folksonomies structured? To what extent does this structure reflect and differ from the norms used in the construction of subject headings such as LCSH? What are the strengths and weaknesses of folksonomies (e.g., reflect user need, ambiguous headings, redundant headings, etc.)?

2. The extent to which LCSH headings reflect user-derived folksonomies. How much overlap exists between LCSH headings and popular tags? How well do LCSH headings mirror user-derived tags for similar concepts?

The folksonomies of three popular bookmark manager sites will be examined: Del.icio.us (<http://del.icio.us>), Furl (<http://www.furl.net>), and Technorati (<http://www.technorati.com>). These sites provide daily logs of the most popular tags that have been assigned by their members on a given day. Tags will be acquired over a one-month period and analyzed in relation to section 6 of the NISO guidelines for thesaurus construction, looking specifically at: (a) the form of tags (single/multiword terms; types of concepts); (b) the grammatical form of tags (nouns, noun phrases, verbal nouns adjectives, and adverbs); and (c) the selection of preferred forms of the tags (acronyms, abbreviations, full spelling, and spelling variations). The tags obtained from each site will then be compared to the LCSH authority files; notice will be made of exact matches, equivalent terms used to express the same concept, and the absence of equivalent LCSH headings.

USING SOCIAL BOOKMARKS IN AN ACADEMIC SETTING: PENNTAGS

Jennifer Erica Sweda <jesweda@pobox.upenn.edu>

University of Pennsylvania, Philadelphia, PA, USA

PennTags is a social bookmarking tool for locating, organizing, and sharing favorite online resources, developed by librarians at the University of Pennsylvania. The Penn Community can collect and maintain URLs from the open Web, links to journal articles, and records in Franklin (online catalog) and VCat (online video catalog). Users can organize posts by assigning tags and/or by grouping them into projects. PennTags offers unique development, implementation and research opportunities because it is one of the only such tools available specifically for an academic audience and is integrated with other Penn library products (Thomas & McDonald, 2005; Lippincott, 2005). It constitutes an experiment in tagging behavior, patron interaction with the library, and the research process and is yielding data in those areas.

Based on experience/feedback, we addressed the issue of “sloppy tags” (Guy & Tonkin, 2006) by allowing users to exercise more control in cleaning up their tags. PennTags allows for a true patron-created folksonomy for describing library resources. Individual tagspaces have their own tagclouds and we changed the posting page to offer a “My Recent Tags” option, so users can strive toward consistent *personal* folksonomies. They may also execute a “Bulk Change” for specific *personal* tags. A future release may include an “Other Users Have Used ...” feature to promote system-wide consistency, but possible disadvantages of this approach (Shirky, 2005) will need to be discussed. We also include a section in our FAQ on tagging tips and examples to improve tagging practice, but we are not pursuing librarian intervention.

Users can tag items in our OPAC and the keywords used will be viewable along with authorized subject headings assigned by professional catalogers. Tags and users are hyperlinked so that users can investigate other resources sharing those terms/names, adding another layer of searching to the catalog and increasing its use as a discovery/recall tool (Voorhees & Harman, 1999; Shirky, 2005). The breadth and quality of patron-supplied metadata in the OPAC may encourage us to consider offering co-searching of PennTags and Franklin or, in the case of a 0-hit OPAC search, the option for a re-direct to PennTags content (in library resources or the open Web).

While the future addition of an influence-ranking system for searching will form a social matrix that can be measured, PennTags is not just a social networking tool but also a content management system. Users can search the open Web and library products and compile an easy-to-manipulate, easy-to-share collection of resources (reference guides, course materials, homework assignments). Because these projects become stand-alone resources marked by assigned subject terms and the ability to be searched, they each function as a catalog of sorts. PennTags marks a new opportunity for studying how patrons interact with library-owned and open Web resources. We can see what interests them, what they collect, annotate, collate and winnow to create research projects (Thomas & McDonald, 2005). PennTags provides librarians with a mechanism to see the arc of the online research process and its final results and may impact collection policies and library services.

References

Guy, M., and Tonkin, E. (2006, January). Folksonomies: Tidying up tags? *D-Lib Magazine* 12(1). Retrieved September 8, 2006, from <http://www.dlib.org/dlib/january06/guy/01guy.html>.

Lippincott, J. (2005, March/April). Net generation students & libraries. *EDUCAUSE Review* 40(2): 56-66.

Shirky, C. (2005). Ontology is overrated: Categories, links and tags. Retrieved September 5, 2006, from http://www.shirky.com/writings/ontology_overnated.html.

Thomas, C. F., and McDonald, R. H. (2005). Millennial net value(s): Disconnects between libraries and the information age mindset. In M. Halbert (ed.), *Free Culture and the Digital Library Symposium Proceedings* (pp. 93-105). Atlanta, GA: MetaScholar Initiative at Emory University.

Voorhees, E. M., and Harman, D. (1999). Overview of the Eighth Text Retrieval Conference (TREC-8). Retrieved September 4, 2006 from http://trec.nist.gov/pubs/trec8/papers/overview_8.pdf.