

Corrine Harkcom

LIS 2001 – Spring 2006

Final Project

Social bookmarking and tagging sites have been increasing in popularity along with the notion of “Web 2.0”. In a social bookmarking system, users are able to store internet-based information packages (web sites, digital photographs, etc.) and create basic metadata, or tags, to describe and organize the content. These sites have created a model that could possibly serve as a way to organize the information in the World Wide Web, a model that strays from the traditional view of an index. In order to understand how these tags effect the organization of information, it is important to first understand where and how social bookmarking and tags are created and used within the online environment.

There are many sites that are allowing users to create a folksonomy through the implementation of tag, but there are four main sites that will be mentioned throughout: del.icio.us, Flickr, Library Thing, and PennTags. In all cases, the tags are user-created and can use the tags to search for other content with the same assigned tag; however the sites focus on different types of content. Del.icio.us allows users to store and tag links to online content, whereas Flickr is used to store and tag digital images. Library Thing allows users to create a personal booklist, importing available metadata from other sites (ex. Amazon) and libraries (ex. Library of Congress), which can include call numbers, Library of Congress Subject Headings, and even the MARC record. PennTags

allows users to create “projects” and save a list of University of Pennsylvania library material that is related to the project and tag the content.

There is most certainly merit to the creation of a folksonomy as a way of organizing content. The value of a tool based on voluntary collaborative information creation has been shown in Wikipedia (ignoring momentarily the debate surrounding quality), and allows for a broad coverage of materials at little or no cost. Because anyone can add to or edit out the information in Wikipedia, when the information on the topic is static, it demonstrates a passive agreement that there is nothing to be changed or added, and thus “correct”. With social bookmarking, a site tagged by thousands of people all using the exact same tag, has entered the same passive agreement of the assigned tag being “correct” in the way it describes the content.

Other attempts at organizing the web are failing both in usability and breadth. Professional indexes to the web, such as the [Librarians’ Internet Index](#), are losing funding and manpower¹, and are unable to index the entire web (let alone individual images or books), and require one to drill down through too many categories to get to actual content. Tagging allows for free and unlimited man power, which allows for broader and deeper coverage of the web. The question of what to index becomes irrelevant; with enough users everything of value would eventually be tagged. Commercial indexes like the one available from Yahoo! have similar problems, and even the volunteer-based [Open Directory Project](#) has the same problematic site model and a

1 American Library Association (2006). *California Halves Funding for Librarians’ Internet Index*. Retrieved April 23, 2006 from <http://www.ala.org/ala/online/currentnews/newsarchive/2006abc/march2006ab/liiholve.htm>

human editorial review board². A similar review process for tagging might weed out tags that were viewed as too personal, irrelevant, or repetitive.

Furthermore, because tagging bypasses review, the metadata is at the very least available for users to search. Search engines may or may not use the meta tags within a document's HTML in their search algorithms, but their use in the algorithms do not matter if the meta tags are not present. Many sites do not use meta tags, with the page author either viewing them as passé or being ignorant of their existence. Tagging sites like del.icio.us are essentially creating these meta tags for each site in a way that is searchable through the del.icio.us interface.

With digital images, inherent metadata is not searchable and does not include information relevant to the content of the image. Information such as the time, date, and the type of camera with which the image was taken can be extracted from code, but is rarely useful from an information organization standpoint. Additionally, since the image itself is not searchable, search engines must rely on other data such as adjacent text, alternative display text, and file names. However, few people take the time to rename their image files from their cameras or scanners leaving messy irrelevant numbers as the file name, and even less bother to define an alternative text display within the HTML. Adjacent text may be completely irrelevant, misleading, or insufficient in describing the information package. With Flickr, intelligent human analysis is done on the subject of the image, and tags unlimited in the number of terms that can be associated with the image, its contents, or events that it may represent. This allows

² Who we are and what we do (2004). Retrieved on April 23, 2006 from <http://www.dmoz.org/help/geninfo.html>

images to be organized and grouped together into a searchable database, with a better method of assigning metadata than image search services such as that available from Google.

While libraries and books do not have quite the same problem with a lack of metadata, the present metadata is relatively inflexible and impersonal. Limitations on the possible number of assigned subject headings does not allow for every possible use, regardless of the exhaustivity of assigned subject headings. However, if a person views something as relevant to a topic that is not covered with subject headings and tags it as such, another user may find the materials as well. In the case of PennTags, another user could view a project with a topic similar to their own, and be able to see how another user has organized library and online materials that are deemed relevant by to the project by its owner. Additionally, PennTags allows for inclusion of all resources available to U. Penn affiliates, web content can be tagged in addition to materials owned or subscribed to by the library.

Unfortunately the largest problem facing social bookmarking as a way to organize web content, from a library science perspective, is the lack of authority control. In an attempt to make tagging as simple as possible, most sites do not allow spaces (though this is slowly changing). This has forced users to be creative about tagging for concepts traditionally described as two or more words, such as the "United States of America". While a common workaround is the use of the underscore in place of the space ("United_States_of_America"), other alternatives could be `unitedstates`, `united-states`, `usofa`, `America`, `usa`, and `us` (which can be easily confused with the

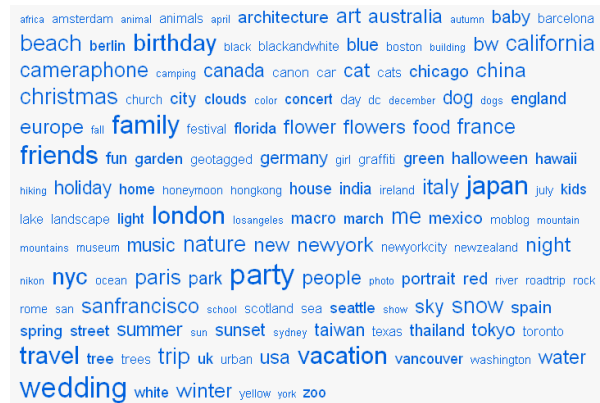
concept of “us” as in “myself and another person”). Additionally, a single item may have several names, and within any one of those names there may be several more ways to type the name. This leaves little room for authority control, and the average tag creator would likely have no understanding of or tolerance for authority control.

Another issue social bookmarking faces is the personal nature of a tag. While some users may want to do the best job possible of tagging their saved content by creating tags (whether for their own organizing and retrieval purposes or those of others as well) that wholly capture the contents of a book, article, image, or website, others use the tags to organize the materials in a way that is meaningful to only themselves. Similar to the way someone may organize files on a computer for the most efficient personal retrieval, tags may be specially created so that the most important tag floats to the top of the alphabetically arranged tag list. Thus a person may create the tag ‘_jobs’ so that it would always be at the top of the list (as opposed to simply ‘jobs’, since the underscore is filed before “a”). A tag might also be used to create a personal list, so the tag ‘to_do’ would mean very little to anyone other than the individual who created the tag.

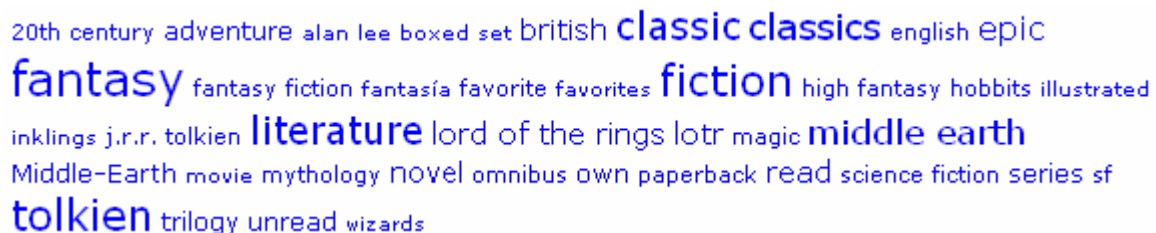
Users may have trouble identifying the content of an information package in order to accurately tag the contents. If a Flickr user finds an image of a beautiful city skyline on the web that they would like to keep in their online photo book, they may do so without knowing specifically which city is portrayed in the photograph. The user may simply tag the image ‘pretty_pictures’, ‘skyline’, and ‘city’. The average user is also likely

to have trouble with subject analysis of an online document, unfamiliar with concepts like exhaustivity and specificity and subject headings.

Finally, there is no hierarchy or organization to the tags. Some sites, such as Flickr, have attempted to create clusters of related tags (which are plagued with all the problems associated with the lack of authority control). Some sites use tag clouds (like the one from Flickr shown to the right), which are simply a list of tags in alphabetical order, with the most popular tags appearing the largest in the cloud. For Flickr, these tags are generally not meaningful for organization purposes, other than they are the broadest topic associated with events, places, and things people take pictures of.



However, the cloud tags from Library Thing are especially interesting from perspective of information organization as related to libraries. This tag cloud (below) is obviously for Tolkien's *The Lord of the Rings*. Again, the largest, most easily visible



words are those that the most people have associated with this item. This is what Golder and Huberman³ refer to as tag stability. Their research concluded that stable

³ Golder, S.A. & Huberman, B.A. (2006) Usage patterns of collaborative tagging systems. *Journal of Information Science*. 32:2. 198-208.

patterns emerge in tagging, with the tag consensus being capable of organizing web documents in an interactive way.

What makes this particularly interesting for librarians (and is unique to Library Thing) is that one can also view the MARC record for the item. A comparison between the MARC record and the tag cloud

show evidence of the collective wisdom of collaborative tagging efforts in the similarities between the Library of Congress Subject Heading and the most popular tags for the items. Noticeably absent is in this example is the LCSH for

Library of Congress (20 copies)

```
001 4060046
005 20041112131703.0
008 711013s1965 nyub 000 1 eng
035 $9 (DLC) 66000953
906 $a 7 $b cbc $c orignew $d u $e ocip $f 19 $g y-gencatlg
010 $a 66000953
040 $a DLC $c DLC $d DLC $d OCoLC $d DLC
050 00 $a PZ3.T576 $b Lo3 $a PR6039.O32
100 1 $a Tolkien, J. R. R. $q (John Ronald Reuel), $d 1892-1973.
245 14 $a The lord of the rings, $c by J. R. R. Tolkien. With a new foreword by the
260 $a New York, $b Ballantine Books $c [1965]
300 $a 3 v. $b maps. $c 18 cm.
505 0 $a Pt. 1. The fellowship of the ring.--Pt. 2. The two towers.--Pt. 3. The return
650 0 $a Baggins, Frodo (Fictitious character) $x Fiction.
650 0 $a Middle Earth (Imaginary place) $x Fiction.
650 0 $a Fantasy fiction, English.
740 0 $a Fellowship of the ring.
740 0 $a Two towers.
740 0 $a Return of the king.
920 $a Pt. 2 and pt. 3 missing; do not replace through purchase per recommendin
991 $b c-GenColl $h PZ3.T576 $i Lo3 $p 00013546976 $t Copy 1 $v 1 $w BOOKS
```

“Baggins, Frodo (Fictitious character)”, with most users opting for the more general ‘hobbits’. One may try to argue that Library Thing attracts more librarians, and hence has users who are more informed than the general population, the comparison is nevertheless intriguing.

The tag cloud system in this example also reflects the personal nature of the tagging. Tags such as ‘favorite’, ‘own’, ‘read’, and ‘unread’ are included, and give little meaningful organization to anyone beyond the individual creating the metadata. Golder and Huberman argue that the personal tags can benefit others, suggesting that one person’s ‘toread’ list might be of interest to another, however I feel the personal tags offers little to the organization of information packages.

Simply put, social bookmarking and tagging was created and is meant to organize information on a personal basis. The issues faced by catalogers in a professional capacity are similar to those faced in a nonprofessional capacity; the difference is that most users of social bookmarking tools are not trained professionals. Programming authority control and tag collocation into the system could allow a person searching for a t-shirt to simultaneously search all items tagged as 't-shirt', 't_shirt', 'tee-shirt', and 'tee', which would improve the overall structure and allow users to continue as they are comfortable.

Controlled thesauri could possibly lead users to assign more specific terms more accurately, but I question whether this would be something users want. Many users are unfamiliar with the logic behind information organization and its application. While tagging seems on the surface to be a solution, there are still many problems with this type of system as the solution to organizing the tangled web of online information.

References:

Who we are and what we do (2004). Retrieved on April 23, 2006 from

<http://www.dmoz.org/help/geninfo.html>

American Library Association (2006). *California Halves Funding for Librarians' Internet*

Index. Retrieved April 23, 2006 from

<http://www.ala.org/ala/online/currentnews/newsarchive/2006abc/march2006ab/iihalve.htm>

Golder, S.A. & Huberman, B.A. (2006) Usage patterns of collaborative tagging systems.

Journal of Information Science. 32:2. 198-208.