The Hive Mind: Folksonomies and User-Based Tagging

by Ellyssa Kroski

There is a revolution happening on the Internet that is alive and building momentum with each passing tag. With the advent of social software and Web 2.0, we usher in a new era of Internet order. One in which the user has the power to effect their own online experience, and contribute to others'. Today, users are adding metadata and using tags to organize their own digital collections, categorize the content of others and build bottom-up classification systems. The wisdom of crowds, the hive mind, and the collective intelligence are doing what heretofore only expert catalogers, information architects and website authors have done. They are categorizing and organizing the Internet and determining the user experience, and it's working. No longer do the experts have the monopoly on this domain; in this new age users have been empowered to determine their own cataloging needs. Metadata is now in the realm of the Everyman.

User-Based Tagging and Folksonomies
Unlike social-networking sites such as LinkedIn and Friendster, which concentrate on developing relationships, social sites such as del.icio.us, 43Things and Flickr focus their attention on organizing data. Users organize their own or other's data in the public sphere and the social, or community, aspects arise from there as users share and seek out like-minded individuals.

del.icio.us is a social bookmarking site. What that means is that instead of saving a Webpage link in your "Favorites" (IE) or "Bookmarks" (Firefox) folder for future use, you save it to your del.icio.us page. The benefit being that you can then access the links from any computer since they are no longer only stored locally. As you add each page to del.icio.us, you are given the opportunity to add descriptive keywords, or tags to the link. This organizes your data by category or tag. You can look for what others have linked to and described by a particular tag or just browse others' collections of links.

43Things is like a giant, global to-do list. You can add all of those things that you have been meaning to do, such as write a book or become a pirate and see how many other people out there share the same goal. You can cheer others on in their endeavors or explore what other people in your hometown are doing. As with del.icio.us, you are offered the chance to add descriptive tags to each goal you add to your list. You can search goals by tag such as 'adventure' and 'professional' to view what others have listed as corresponding goals.

Flickr is a digital image storage and management website. It is a place for you to organize all of your photos into albums, tag them with descriptive keywords, and view others' images. Flickr allows navigation by tag or user as the previous two sites, as well as by group. Groups are places for users who share similar interests to post their images such as the Cats in Sinks or the Dogs Eating Potato Chips groups.

Technorati allows users to perform searches on blog content. Blogs are a medium that updates quite rapidly, unlike webpages. Performing a search here, as opposed to other search engines, will get you a results list with posts created minutes before. Technorati tracks over 21 million blogs as well as 3 million tags which have been created by users on other sites such as Flickr. Browsing by tag will get you not only blogs, but images from Flickr and Buzznet and websites from del.icio.us and Furl. Users can set up Watchlists by tag.

All of the above sites offer a view of the global "tag cloud". A tag cloud displays all of the most popular tags in use across a page with the more popular tags in larger sizes. In this way, users can tell at a glance what the most popular tags are with other users. In

del.icio.us, 43 Things and Flickr, you can also view a personal tag cloud which will show you which of your tags are more predominantly in use by you.

Users create keywords in order to be able to recall their information at a later time and often their tags are reflective of that. According to del.icio.us' Joshua Schachter, "It's basically a way to remember in public". For taggers, it's not about the right or the wrong way to categorize something and it's not about accuracy or authority, it's about remembering. In this way the categorization is customized for each individual while still serving all of them.

As users continue to add tags, a grassroots organizational scheme begins to emerge which has been dubbed by information architect Thomas Vander Wal, to be a folksonomy. A combination of "folks" and "taxonomy", the term has come to mean a non-hierarchical ontology that is created as a natural result of user-added metadata or tagging.

The Wisdom of Crowds
There has been much discussion in the information world about the concept of folksonomy as opposed to the traditional taxonomy with its controlled vocabularies and hierarchical nature. Likewise, there have been many advantages identified with respect to folksonomies and organizing web content.

Folksonomies are inclusive. While top-down taxonomies utilize a controlled vocabulary which is exclusionary by nature, folksonomies include everyone's vocabulary and reflect everyone's needs without cultural, social, or political bias. Because folksonomies include alternative views together with popular ones, they present a unique opportunity to discover "long tail" interests. The long tail, a phrase first discussed by Chris Anderson of Wired Magazine, consists of the interests of the minority that lie at the "tail" end of a power law, or statistical distribution, which charts the most popular topics. When combined, these non-mainstream, or niche interests far outnumber the popular ones. (16)

Folksonomies are current. Tagging-based systems offer a fluidity and currency which is not possible in a controlled, hierarchical taxonomy. Users create tags as quickly as they create content and they are immediately added to the ontology. This flexibility allows swift responses to changes in terminology and to world events. A large taxonomy such as the Library of Congress classification scheme, in which it could take years to add a date of death to an author's authority record, simply cannot compete with this rate of adaptability. In creating traditional classification schemes, the cataloger is put in the position of fortune teller, in that he/she must predict permanent categories in advance. The problem with this model is that things change; countries change names, computer technology expands, and sometimes groups of people change the way they refer to themselves, i.e. Blacks, Negroes, Afro-American, African Americans. And in the world of the Web, things change fast.

Folksonomies offer discovery. Hierarchical taxonomies are designed for finding specific resources whereas folksonomies are predisposed to discovering unknown and unexpected resources. These systems promote exploration and learning as users browse related topics, tags, and users. There is legitimate value in a discovery system as users have the opportunity to locate new resources that they might not ever have come across through searching. Oftentimes the user doesn't know exactly what resources that they are looking for at the start of their research. Discovery systems empower users to uncover alternative paths and related resources on their information journey.

Foksonomies are Non-Binary. In a traditional classification scheme, a controlled vocabulary must be made in advance in which one category term is selected which includes all related terms. When future objects are cataloged it must be determined that either they fit into a particular category or they do not. In a folksonomy, these items can fit into multiple categories. For example, an image on Flickr could have tags attached such as cat,

kitten, feline, tabby, cute. In a folksonomy the scheme is multi-faceted. As Clay Shirky points out in his discussion of this topic, there is "signal loss" when you merge multiple concepts into just one term. The Library of Congress subject heading for movies is 'Motion Pictures'. By reducing terms such as movies, film, and cinema to one all-encompassing category, the distinctive meanings of each term gets lost in the translation. While the cataloger might determine that they are similar enough to be considered one-and-the-same for the sake of the scheme, there are those who disagree. As Shirky perceptively points out; "the movie people don't want to hang out with the cinema people!" (27)

Folkonomies are democratic and self-moderating. Everyone has the opportunity to add something to the whole. Likewise, these systems are self-moderating. By their nature, these systems, encourage users from an individual standpoint to choose tags that appropriately describe items, which in turn helps them to remember them in the future. Similarly, since tagging is done in a public forum, the social dynamic sways users to choose relevant tags, according to Technorati's Dave Sifry.(19) When tagging a new item, many systems offer users a list of the most popular tags used for that resource. The idea is that the most popular terms tend to be the most relevant, just as a frequently cited article or book is considered to hold more authority in the academic realm. Gene Smith raises the concern that the idea of translating this practice to the web is often met with disdain by librarians. (30)

Folksonomies follow "desire lines". What we are witnessing with a folksonomy is an expression of the direct information needs and desires of the user. In a traditional classification scheme, catalogers must attempt to read the minds of the users, and make determinations based on their estimations of user needs. Since a folksonomy arises as a result of user tagging, it is reflective of the way that they categorize information.

Folkonomies offer insight into user behavior. Folksonomies give us a chance to observe how users tag their own resources as well as what kind of untraditional categories have surfaced. Among the new categorization types are functional tags such as to_read and to_watch. Both of these are temporary tags which are reflective of the adaptive nature of these bottom-up taxonomies. Although they are very subjective descriptors, they are also both useful tags to others who may want to view what others have on their reading lists. The self is now contained in various categories such as me, mine, and my_stuff.

Folksonomies also give us an opportunity to observe user behavior and tagging patterns. According to a study of the del.icio.us website conducted by the Information Dynamics Lab at HP Labs, a stable tag pattern emerges after the first one hundred bookmarks are placed for a particular website. They attribute this synchronization to user imitation of popular tags and to a common knowledge base shared by users of the site. As a result, alternative views exist alongside popular ones without being disruptive to the pattern.(11)

Since folksonomies follow desire lines as mentioned above, viewing user language and behavior in this way could help in the future development of top-down taxonomies.

Folksonomies engender community. There is a spirit of sharing and community in user-based tagging sites. Everyone has the common goal to catalog their own information, but also to share it with others. On sites such as 43Things.com, users cheer each other on to reach common goals, and on Flickr, cat-lovers can join up with dozens of groups dedicated to pics of the furry felines.

Folksonomies offer a low cost alternative. The cost of creating a traditional, hierarchical taxonomy with a controlled vocabulary is quite high. Expert catalogers or information architects are needed to determine the scheme as well as classify individual entities. User-based tagging provides a low cost alternative for cataloging Web resources.

Folksonomies offer usability. Top-down classification schemes require a trained or skilled user base. On the Web, that isn't realistic, as levels of user expertise and interest vary greatly. Folksonomies have a very small learning curve and are exceedingly comprehensible to the user. In his article on Authority, Peter Morville discusses the concept of anchoring as an information seeking behavior. From a psychological standpoint, people tend to be most influenced in decision-making by the first piece of information that they come across. Efforts are then made, on a sub-conscious level, to confirm this found information and avoid opposing viewpoints. (22) So, if users have a predisposition to anchor onto the most findable information out there, an appropriate response would be to provide the user with a system that is accessible or they will find it somewhere else that is.

Web 2.0 is about sharing and connectivity and participation. It is a user-centered era of the Web. We are moving away from expert-dictated, exclusionary models of information organization and toward inclusive, participatory ones.

Resistance is Futile. The fact of the matter is that the enormity of information that is now being published online through new mediums such as blog, wikis, etc., make a traditional taxonomy and controlled vocabulary an impossible solution. The cost and the amount of manpower required would just be too high. Folksonomies on the other hand are much more scalable from an economic standpoint. As Wikipedia co-founder Jimmy Wales comments "I wouldn't even want to think of what it would cost to replicate the Wikipedia categories with paid labor". In the absence of a professionally designed taxonomy, folksonomies are being viewed as a readily available, "better than nothing", stand-in. According to Clay Shirky, folksonomies are a "forced move", they are coming whether we like it or not. "It doesn't matter whether we "accept" folksonomies, *because we're not going to be given that choice*." (24)

The Hitch

In addition to the benefits to be gained from the collective intelligence intrinsic to folksonomies, there has been considerable debate concerning their flaws. As quickly as such drawbacks are identified, however, supporters of the organic ontologies have provided responses.

Folksonomies have no synonym control. In user-based tagging systems, there is no controlled vocabulary and therefore one authoritative term does not exist to describe a concept or entity. This is considered a shortcoming when different users describe assets using many different terms to presumably describe the same thing, i.e. cats, kittens, felines, etc. Because of the characteristic lack of control, there is also no way to regulate the use of plurals vs. singular, acronyms, etc.

Supporters respond that the lack of synonym control is a design choice rather than a weakness in these models. As with Shirky's movie and cinema example, the absence of restrictions allows users to choose words that precisely describe their digital assets without the loss of meaning resulting from a controlled vocabulary. (26) Additionally, many social tagging sites provide lists of "related terms" which encourage the use of "popular" synonyms. Nevertheless, this is a serious limitation of these systems that developers will need to grapple with as they grow in size.

Folksonomies have a lack of precision. As Bruce Sterling of Wired Magazine notes, "a Folksonomy is nearly useless for searching out specific, accurate information, but that's beside the point." (31) As mentioned earlier, folksonomies are discovery systems, without the powerful search capacity of a hierarchical taxonomy. Characteristically, they are going to have low precision rates.

As Lee of Headshift responds, "In practical usage scenarios the trade-off between simplicity

and precision makes sense."(15) Folksonomies are usable and accessible. Although precision is certainly important, it isn't everything. A traditional taxonomy, such as the Library of Congress classification system will allow users to precisely locate relevant resources concerning a topic such as World War II because of the strength of its controlled vocabulary, however, the user must know that the subject heading is 'World War 1939-1945' in order to reap the rewards of this system.

Folksonomies lack hierarchy. Folksonomies are flat systems. There are no parent-child relationships, no categories and subcategories. Hierarchy is a distinguishing trait of traditional taxonomies which are able to provide a deeper, more robust classification of entities. Such systems allow users a finer granularity in searching for resources.

According to Joshua Schachter of del.icio.us, adding hierarchy to the folksonomy model would decrease the level of usability and ease of access which are so valued in these systems.(25) As mentioned earlier, in order to maintain the merits of a folksonomy some sacrifices of functionality are made in favor sustaining a more usable and therefore useful model. From another perspective, Clay Shirky responds that "if hierarchy was a good way to organize links, Yahoo would be king of the hill and Google an also-ran service."(26)

Folksonomies have a "basic level" problem. Similar to the problem of synonym control is the concern that users will have different ideas about how to tag entities at a basic level as opposed to using a broader or narrower term. Golder and Huberman of HP Labs give the example; 'Perl' and 'JavaScript' vs. 'programming'. They raise the concern that "collective tagging, then, has the potential to exacerbate the problems associated with the fuzziness of linguistic and cognitive boundaries."(11)

As referenced with regard to synonym control, this may be a preference of folksonomies. There is no vocabulary control and as such, users can include all terms that may apply to the entity when tagging without concern for whether it is a basic, more general or more narrow term. Golder and Huberman themselves observe that because of the nature of these systems, people have an opportunity to learn from one another while tagging and categorizing.

Folksonomies have a lack of recall. Recall reflects the ability of a system to return all resources related to a topic. Because of the lack of synonym control, a folksonomy search will not effect a complete results list because of the use of similar tags. A search for 'cat' for example will not retrieve resources which have been tagged with kitten, feline, tabby, or even cats. This is a serious limitation of these systems.

Once again I will mention the concept of trade-offs and that although a user may not be able to locate every resource which has been organized in this fashion, the user will find nothing in a system which is too difficult or daunting to use. Flickr CEO Stewart Butterfield points out, "we'll have a million photos of Tokyo, and if the TOKYO tag only gets you 400k of them, it's OK. You're only going to look at 20 of them anyway."(25)

Folksonomies are susceptible to "gaming". Gaming is similar to spamming and involves an unethical user who propagates links, or in this case, tags in order to corrupt a system. The Blocklevel blog raises a valid point that "malicious users can purposely pollute the "Tag Sphere" by tagging every bit of content with every possible tag – effectively spamming the system."(15)

This is definitely a possibility with user-based tagging systems. Although popularized by invoking a spirit of cooperation among users, folksonomies are vulnerable as there are always those who don't play nice with others.

Libraries that Tag

Already there are libraries jumping on the tagging bandwagon hoping to provide their patrons with a user-friendly supplement to their existing systems.

La Grange Park Public Library, Thomas Ford Memorial Library, Rutland Free Library and the Delft Public Library, to name a few all use Flickr to display images of their libraries. Also taking advantage of the photo management site are 365 members of the "libraries and librarians" group and 78 members dedicated to displaying the humorous aspects of librarianship in the "library" group.

The University of Pennsylvania library has their own tagging system called "PennTags" which is based on the del.icio.us software and allows users to bookmark and tag websites as well as library cataloging records.

There are librarian groups tagging on CiteULike, a social bookmarking site for academic citations which allows users to export their libraries to Endnote or BibTeX.

The BBC has had a prototype software developed for them in order to have users tag their immense news archives. Based on popular social software systems, users can store bookmarked stories on their own page, view related stories and enjoy feeds from Flickr, Technorati and del.icio.us.

Summary

The advantages to top-down hierarchical taxonomies for library collections are without question. For cataloging the Web, however, they just aren't feasible. The new, "voice of the people" approach of folksonomies emerges at a time when attitudes about information organization and retrieval are shifting and the technology is developing to support them. The opportunities for learning about user behavior as well as the implications for improving and/or complementing existing taxonomies that these systems can provide are of no small import. We are on the cusp of an exciting new stage of Web growth in which the users provide both meaning and a means of finding through tagging.

References

1. "About." Del.icio.us. http://del.icio.us/about/ Viewed 10/30/05.
2. "About" Technorati. http://www.technorati.com/about/ Viewed 10/30/05.
3. Anderson, Chris. "The Long Tail." Wired, 12.10 October 2004. http://www.wired.com/wired/archive/12.10/tail.html Viewed 10/30/05.
4. Andrews, Robert. "Public could help BBC to index archive." Journalism.co.uk, November 1, 2005. http://www.journalism.co.uk/news/story1583.shtml Viewed 11/9/05.
5. Angeles, Michael. "Dan Brown on freetagging". Urlgreyhot, April 1, 2005. http://urlgreyhot.com/personal/weblog/dan_brown_on_freetagging Viewed 10/30/05.
6. Bryant, Lee. "Prototypes: BBC shared tags." Backstage BBC, May 12, 2005. http://backstage.bbc.co.uk/prototypes/archives/2005/05/bbc_shared_tags.html Viewed 11/9/05.
7. Burkeman, Oliver. "G2: Ideas: Folksonomy." The Guardian, September 12, 2005: p. 29.
8. Coates, Tom. (Weblogs and) The mass amateurisation of (nearly) everything…" Plasticbag.org, September 3, 2003. http://www.plasticbag.org/archives/2003/09/weblogs_and_the_mass_amateurisation_of_nearly_everything.shtml Viewed 11/10/05.
9. "Folksonomy." Wikipedia, October 26, 2005. http://en.wikipedia.org/wiki/Folksonomy Viewed 10/30/05.

10. "Folksonomy? ethnoclassification? libraries? wha?." Rawbrick, August 31, 2004. http://www.rawbrick.net/article/844/folksonomy_-ethnoclassification-libraries-wha Viewed 10/30/05.

11. Golder, Scott and Bernardo A. Huberman. "The Structure of Collaborative Tagging Systems." HP Labs. http://www.hpl.hp.com/research/idl/papers/tags/tags.pdf Viewed 11/9/05.

12. Harrington, CM. "Tag! You're it!." Blocklevel. http://blocklevel.com/weblog/information_architecture/tag_youre_it/ Viewed 10/30/05.

13. Lawley, Liz. "Social consequences of social tagging." Many-to-Many, January 20, 2005, Corante. http://www.corante.com/many/archives/2005/01/20/social_consequences_of_social_tagging.php Viewed 11/22/05.

14. Lee. "BBC Backstage prototype: social tagging." Headshift, May 12, 2005. http://www.headshift.com/archives/002498.cfm Viewed 11/9/05.

15. Lee. "Can social tagging overcome barriers to content classification?" Headshift, August 30, 2004. http://www.headshift.com/archives/002085.cfm Viewed 10/30/05.

16. "The Long Tail." Wikipedia, November 13, 2005. http://en.wikipedia.org/wiki/Long_Tail Viewed 11/13/05.

17. Mathes, Adam. "Folksonomies – Cooperative classification and communication through shared metadata." Adam Mathes, December 2004. http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html Viewed 11/10/05.

18. Maurer, Donna. "Findability vs. discoverability." DonnaM, March 8, 2005. http://www.maadmob.net/donna/blog/archives/000609.html Viewed 10/30/05.

19. Mercado, Andrea. "Tagging on Flickr & del.icio.us. Library Techntonics, October 24, 2005. http://www.librarytechtonics.info/archives/2005/10/tagging_on_flic.html Viewed 11/3/05.

20. Mieszkowksi, Katharine. "Steal this bookmark!" Salon, February 8, 2005. http://www.salon.com/tech/feature/2005/02/08/tagging/index_np.html Viewed 11/10/05.

21. Miller, Paul. "Web 2.0: Building the new library." Ariadne, 45 October 2005. http://www.ariadne.ac.uk/issue45/miller/ Viewed 11/3/05.

22. Morville, Peter. Ambient Findability. O'Reilly Media, Inc., Sebastopol, CA, 2005.

23. Morville, Peter. "Authority." Semantic Studios, October 11, 2005. http://semanticstudios.com/publications/semantics/000057.php Viewed 11/6/05.

24. Quintarelli, Emanuele. "Folksonomies: power to the people." ISKO Italia, June 24, 2005. http://www.iskoi.org/doc/folksonomies.htm Viewed 11/10/05.

25. Shirky, Clay. "Folksonomies are a force move: A response to Liz." Many-to-Many, January 22, 2005, Corante. http://www.corante.com/many/archives/2005/01/22/folksonomies_are_a_forced_move_a_response_to_liz.php Viewed 11/22/05.

26. Shirky, Clay. "Folksonomy, or how I learned to stop worrying and love the mess." O'Reilly Emerging Technology Conference, March 16, 2005, San Diego, CA. Transcript by Cory Doctorow. http://craphound.com/etech2005-folksonomy.txt Viewed 10/30/05.

27. Shirky, Clay. "Folksonomy." Many-to-Many, August 25, 2004, Corante. http://www.corante.com/many/archives/2004/08/25/folksonomy.php Viewed 11/12/05.

28. Shirky, Clay. "Ontology is overrated: Categories, links and tags." Clay Shirky's Writings about the Internet. http://shirky.com/writings/ontology_overrated.html Viewed 10/30/05.

29. Smith, Gene. "Tagging tags to make synonyms." Atomiq, October 31, 2005. http://atomiq.org/archives/2005/10/tagging_tags_to_make_synonyms.html Viewed 11/10/05.

30. Smith, Gene. "Folksonomy: social classification." Atomiq, August 3, 2004.
    http://atomiq.org/archives/2004/08/folksonomy_
    social_classification.html Viewed11/11/05.
31. Smith, Gene. "Peter Morville: the tagsonomy interview." You're It! October 19, 2005.
    http://tagsonomy.com/index.php/peter-morville- the-tagsonomy-interview Viewed
    11/15/05.
32. Sterling, Bruce. "Order out of chaos: What's the best way to tag, bag, and sort data?
    Give it to the unorganized masses." Wired, 13.04, April 2005. http://www.wired.com
    /wired/archive/13.04/ view.html?pg=4 Viewed 11/9/05.
33. Timothy. "Folksonomies in Del.icio.us and Flickr." Slashdot, January 4, 2005.
    http://slashdot.org/article.pl?sid=05/01/04/0117245 Viewed 10/30/05.
34. Todras-Whitehill, Ethan. "'Folksonomy' carries classifieds beyond SWF and 'For Sale'",
    New York Times, October 5, 2005.
35. Udell, Jon "Collaborative Knowledge Gardening." InfoWorld, August 20, 2004.
    http://weblog.infoworld.com/udell/2004/08/30.html
    #a1064 Viewed 11/20/05.
36. Vander Wal, Thomas. "Folksonomy Definition and Wikipedia." Vanderwal.net.
    http://www.vanderwal.net/random/entrysel.php? blog=1750 Viewed 11/11/05.
37. Willms, Jordan. "Gardened hierarchical folksonomy - the next evolution in semantic
    classification." Stream of Consciousness, June 27, 2005. http://www.jordanwillms.com
    /index.php/2005/06/27/
    gardened-hierarchical-folksomy-the-next-evolution- of-web-20/ Viewed 11/12/05.
38. Wright, Alex. "Folksonomies redux." Alex Wright, September 7, 2004.
    http://www.agwright.com/blog/archives/000905.html Viewed 11/15/05.

POSTED BY Ellyssa ON 12.07.05 @ 11:48 pm