

Tagging: Emerging Trends

by Gene Smith

Information Architecture

Last year there were a handful of popular blog posts about how tagging had gone stale. The initial rush of excitement over tagging, created by social bookmarking site Del.icio.us [1] and photo-sharing Flickr [2], had given way to a kind of malaise.

Tagging became popular in 2003 when, along with open APIs and user-generated content, the Web 2.0 phenomenon captured the attention of web designers, developers, information architects and entrepreneurs. Interest in tagging was stoked by pundits like Clay Shirky, David Weinberger and Tim O'Reilly. They celebrated its openness, its scalability and its responsiveness to the needs of real users.

But by 2007 a number of people noticed that tagging seemed to be stuck. In the blog post that started this conversation, Matt Mower wrote [3] the following:

“I have been surprised, disappointed and excited that, despite the widespread adoption of tagging across many applications, the state of the art in tagging seems firmly wedged in 2003...

“Tagging in 2007 seems to have advanced no further than a means by which one or more users of a site (or application) can group content around a loose framework of concepts.”

Just as this conversation was happening, I was in the middle of writing a book on tagging. I had made a point of reviewing the academic research as well as taking a detailed look at the tagging systems being built by entrepreneurs, Web 2.0 startups and established software companies.

And what I found could hardly be called stale.

Gene Smith is the author of *Tagging: People-Powered Metadata for the Social Web* (2008, New Riders). At nForm User Experience he advises a variety of clients on their IA, design and social media strategies. He can be reached by email at gene.smith@nform.ca. He also blogs at <http://atomiq.org>.

There had been significant innovation in tagging over the past few years. But it wasn't happening at Flickr or Del.icio.us – their tagging systems hadn't changed much since 2005.

Still, because they were so popular and because their data was used for much of the academic research on tagging, Flickr and Del.icio.us had come to represent all tagging systems. And while they're still excellent examples of large-scale popular tagging systems, they're no longer the *beau ideal* when it comes to tagging.

In this article I'll discuss four trends that point toward tagging's future.

1. **More structure.** Uncontrolled vocabularies are being replaced by tagging systems that understand the difference between Polish and polish.
2. **Automanual folksonomies.** Some tagging systems combine algorithmic and manual approaches, closing the gap between what we might call traditional information structures and the emergent structure of Flickr and Del.icio.us.
3. **Leveraging communities.** Some systems have their users help reduce the noise and eliminate meaningless duplication in their tags.
4. **User-generated innovation.** Tags have developed into a cheap and easy way for people to innovate on top of a web application.

These trends aren't discrete and standalone. Designers are combining them to create unique applications unlike simple, open systems we've known.

More Structure

In the first wave of folksonomies, people saw the flexibility and openness of tagging as an advantage. You might recall that what followed the popularity of tagging was a wave of criticism of more traditional information structures.

Taxonomies and controlled vocabularies were drubbed for being too restrictive, too slow to adapt and inefficient. Worse, they imposed a particular

world view on their users. Clay Shirky, for example, noted that there are important differences between some apparently synonymous terms [4]:

Even closely related terms like movies, films, flicks and cinema cannot be trivially collapsed into a single word without loss of meaning and of social context. (You'd rather have a Drain-O® colonic than spend an evening with people who care about cinema.)

In other words, if you treat *movies* and *cinema* as synonyms you're ignoring what we might call their sociosemantic differences.

The great thing about tagging is that it allows – even enables – these kinds of differences. There is no right or wrong way to tag a bookmark or photo. In a tagging system the movie people don't need to meet the cinema people.

This isn't the first time these criticisms have surfaced. Cory Doctorow in his popular 2001 essay “Metacrap” said that taxonomies “denuded the cognitive landscape”[5]. Early tagging systems were pretty well aligned with these kinds of libertarian ideas. But in the last few years innovative tagging systems have emerged that introduce more structure without sacrificing many of the features, like openness or sociality, that make tagging valuable.

There are certainly dozens of consumer web applications that use some form of structured tagging. Three examples will demonstrate the diversity of innovation in this area:

- Zigtag [6] is a social bookmarking service that maps tags to concepts, letting you distinguish between “apple” the fruit and “Apple” the computer manufacturer when you tag. Zigtag built its database of concepts, which has millions of entries, by mining publicly available data sources.
- Wesabe [7], a personal finance planner, imports your banking records to help you understand how and where you spend your money. Tags are your primary tools for organizing your transactions in Wesabe. The service makes tagging your bank statements significantly less tedious through “sticky tags,” tags that are attached to a merchant for every future transaction you have with them.

- Buzzillions [8] is a product reviews site that integrates tags, facets and taxonomies in a seamless way. Normally product reviews are unstructured text, and they're great for helping you decide if you should buy product. Buzzillions use tags instead of unstructured text for its reviews, which lets them use the reviews themselves as kind of filter-and-find navigation.

Most interesting of all, these aren't experimental systems or tech demos. They are real products, suggesting that there's a market demand for structured tagging.

Leveraging Communities

Tagging isn't usually an explicitly collaborative activity. In most cases users don't discuss or negotiate which tag to apply to a web page or bookmark the way they might discuss the contents of a Wikipedia article.

But when we look at Del.icio.us, as an example, we can see the outlines of a community. There are shared interests expressed through the tag cloud – Linux, JavaScript, design, Google, web2.0, travel.

And in at least one case – LibraryThing [9] – users are collaborating to help improve their tag collections. LibraryThing is a social cataloging application that lets you add and tag books from your personal library. It then helps you find other books you might like and people who have similar collections.

Tags are one of the primary ways LibraryThing users catalog the 25 million books they have added to the system. LibraryThing's problem – one shared by any system with a sufficiently large set of tags – is that there are many tags that say essentially the same thing.

The situation isn't like Shirky's example of *movies* and *cinema*. We're talking about cases like “WWII” and “World War 2,” or “science fiction” and “sf.” That is, cases where the sociosemantic delta is zero.

So LibraryThing has added a clever community-driven controlled vocabulary for their tags. Users can make any two tags equivalent, and when they're equivalent the more popular tag becomes the preferred term. LibraryThing also keeps a history of which tag equivalencies have been

made. Any tag equivalency can be easily unmade by any LibraryThing member. Tag equivalencies are subject to community negotiation in much the same way as an article on Wikipedia.

LibraryThing's founders guide this process through a simple philosophy – only combine tags that are virtually identical in meaning. But even this allows for significant reduction in noise; for example dozens of variations on “world war 2” have been collapsed into a single preferred tag “wwii.”

The community can also decide when two tags that are seemingly identical contain those important sociosemantic differences. Take *humor* and *humour* as an example. The American spelling is more often used with American authors and American-style of humor. The British spelling is more often associated with the dry, British-style humour. But there is some overlap – Douglas Adams's books are tagged with both versions of *humor*.

Through this fairly simple approach, LibraryThing users eliminate meaningless duplication of tags and create a bottom-up community-driven controlled vocabulary.

Automanual Folksonomies

Other websites mix a little bit of top-down structure with their bottom-up tags, creating an “automanual” folksonomy. They work well in situations where some aspects of tagging are desirable – open-ended and social – while others are not – like its unpredictability.

Consider Etsy [10], a marketplace for hand-made goods – kind of like eBay for knitters, crafters and other folks who make one-of-a-kind items. Etsy uses an automanual approach to create part of its site navigation. Etsy asks sellers to choose from a set of pre-defined tags for each item they sell. They then provide suggestions for each additional tag. Sellers can pick from the suggested tags or enter their own.

Etsy's pre-defined tags form the top-level category navigation on the website. The suggested tags are actually sub-categories for each of the main categories. While users are nudged toward these suggested tags, they can still enter their own tags. Through this approach, Etsy creates a fairly stable navigation system that remains responsive to the needs of users.

LibraryThing has also created an automanual system called TagMash.

TagMash is a kind of search where you combine tags to create a list of matching books. TagMash has a simple weighting feature that lets you de-emphasize or negate a tag from your query. If you like books about World War II set in France, but only fiction, you can create a TagMash to find all the books with those tags. (You also get the benefit of the tag equivalencies that LibraryThing users have created.)

The folks at LibraryThing have also experimented with using TagMash to emulate the Library of Congress Subject Headings and the subject taxonomy used by bookstores like Amazon.

Mapping a TagMash to a taxonomy branch or subject heading creates an evergreen listing of books for a category – as long as people keep adding tags to books. This is, in effect, a very cheap maintenance system for a classification system. It probably won't produce the same levels of accuracy and consistency as a professional cataloguer, but it leverages a community of interest to achieve similar results.

Peter Van Dijck, an information architect who has experimented with automanual techniques, said this about mixed bottom-up tagging with top-down structure:

“I notice a hesitance toward hard-coded semantics and manual work – people think these things won't scale. I learned to mix it up... a small amount semantics on top of minimal structure can work wonders.” [Personal communication, August 10, 2007]

The success of LibraryThing and Etsy suggest that this could be a fruitful technique for information architects who want the benefits of tagging but can't cede control entirely to their users.

User-Generated Innovation

So far this article has focused on tags as part of an organizational system. And in fact this focus is typical of the discussions on tagging that have happened over the last few years.

But there's another side to tagging that's important to appreciate: tags are one easy way for people to hack, mash-up and innovate on top of a web service or application.

SMITH, continued

Let's consider Flickr as an example. Flickr lets you add any text string as a tag, and it creates an RSS feed for every tag entered in the system. So if you tag a photo as "obstreperous," you'll find an RSS feed for everything tagged "obstreperous" and your photo will be in that feed.

This is, in effect, a very simple read/write system for metadata. Even though it's quite primitive, it lets users experiment with new features and services. Flickr's geotagging feature emerged from this kind of experimentation. One active Flickr user, Dan Catt, came up with a simple method for placing photos onto a Google map. He started with a marker tag – "geotagged." This created an RSS feed for every geotagged photo. Then he added two machine tags for the latitude and longitude coordinates. (Machine tags are special tags that take the form "namespace:key=value" and they can be used to encode just about any sort of metadata in a tag [11].) He was able to find all the "geotagged" photos using the RSS feed and then parse out the values of the machine tags to place them on a map.

Initially, people had to geotag their photos manually. They would enter the marker tag, look up the latitude and longitude coordinates of their photos and then enter that data as machine tags. But this was enough of a system that Flickr's emerging geotagging community could build basic tools to display the photos. Later, people created interfaces to make geotagging photos a lot easier.

Flickr eventually hired Dan Catt and now supports geotags natively. They developed a better interface for geotagging photos, as well as a machine tag search feature into their API.

What makes this story of tag-enabled innovation remarkable is that it's not unique. Other systems that use the same architecture – one where a data feed of tagged objects is available for every single tag – have seen similar innovation. Connotea [12], a social bookmarking application for scientists, IBM's internal social bookmarking engine Dogear and Del.icio.us have all had applications, mash-ups and experiments built off their tags and feeds combination.

This is an interesting dimension of tagging that's usually subordinate to classification and information structure. But it's important because it suggests tagging's value is partly in how it allows people to interact with information – both tags and the thing that's being tagged – and to change their information environment to better fit their needs.

Conclusion

These four trends show that tagging continues to evolve. Now that hype around tagging is muted, designers, developers and product managers are using tags to solve their problems, improve their products and help their customers.

For long-time information architects, these new approaches may seem unusual. They directly rely on user contributions, they leverage active communities, and they freely mix top-down structure with bottom-up innovation. They point to a future where information architects work at the edges, managing the emerging properties of folksonomies alongside the semantic relationships of taxonomies and controlled vocabularies.

But these trends are also great news for the discipline of information architecture. Three years ago some pundits suggested that folksonomies might replace IA altogether. Today we're seeing tags, taxonomies and facets intermingling to create new and valuable information structures. Most importantly, we're seeing tags being used to solve the classic problems of IA – helping people find and use information, making meaning from the tangle of language and reducing the cognitive and economic costs of ambiguity. ■

Resources Mentioned in the Article

- [1] Flickr [website]. <http://flickr.com>
- [2] Del.icio.us [website]. <http://del.icio.us>
- [3] Mower, M. (2007, August 23). *To the trough of disillusionment we go!* Retrieved July 2, 2008, from <http://matt.blogs.it/entries/00002618.html>
- [4] Shirky, C. (2004, August 25). Folksonomy. Retrieved July 2, 2008, from <http://many.corante.com/archives/2004/08/25/folksonomy.php>
- [5] Doctorow, C. (2001). *Metacrap: Putting the torch to the seven straw-men of the meta-utopia*. Retrieved July 2, 2008, from www.well.com/~doctorow/metacrap.htm
- [6] Zigzag [website]. <http://zigzag.com>
- [7] Wesabe [website]. <http://wesabe.com>
- [8] Buzzillions [website]. <http://buzzillions.com>
- [9] LibraryThing [website]. <http://librarything.com>
- [10] Etsy [website]. <http://etsy.com>
- [11] For more about machine tags see www.flickr.com/groups/api/discuss/72157594497877875/
- [12] Connotea [website]. www.connotea.org