Incentives for Contribution in del.icio.us: The Role of Tagging in Information Discovery

Rick Wash
School of Information
University of Michigan
rwash@umich.edu

Emilee Rader
School of Information
University of Michigan
ejrader@umich.edu

ABSTRACT
We describe a qualitative study of the social bookmarking website del.icio.us. We arrive at the surprising finding that it is the identity of the user who bookmarked a web page, not the tags (keywords) associated with the web page that provide the most benefit for information discovery. Individuals' choices about what to bookmark and which tags to apply have implications that extend far beyond their own private motivations. The degree of consensus among users in their selection of tags influences the "signal to noise" ratio when browsing by a given tag. Unfortunately, the incentive structure of del.icio.us is such that, given the vocabulary problem[5] and the individual, private benefits of bookmarking and tagging, consensus tags do not emerge. Without consensus tags, information discovery by tag is inefficient. However, a single user's bookmarks reveal information about his or her interests that is useful in ascertaining similarity to one's own interests. Browsing the bookmarks of such users results in more efficient information discovery than browsing by tag.

Author Keywords
Social computing, social bookmarking, unintended consequences, incentives, del.icio.us, tagging, browsing

ACM Classification Keywords
H.5.3 [Information Interfaces and Presentation]: Group and Organizational Interfaces --- Computer-supported Cooperative Work, Collaborative Computing; H.4.3 [Information Systems Applications]: Communications Applications --- Information Browsers

INTRODUCTION
Del.icio.us is an online tool for "social bookmarking" that provides the capability for users to bookmark web pages and associate user-generated metadata, or tags, with them[12, 21]. Users' bookmarks are stored online, and can be viewed by other users of the system and accessed from any computer with an Internet connection.

The literature investigating user behavior and overall usage patterns in tagging systems like del.icio.us is growing rapidly. A widely-held belief is that in such systems, tags are the primary means of both organizing one's own bookmarks, and discovering new information that has been bookmarked by others with similar interests. It is also commonly believed that users of systems that support publicly viewable tags are motivated both by private, personal outcomes such as an organized bookmark collection, and public, group outcomes such as the emergence of an overall organization scheme, or folksonomy, from the accumulation of users' individual tagging actions[12, 22].

There has been much focus on the different kinds of tags, the functions they might serve, and whether tag usage might "converge" on conventions or a consensus vocabulary. For example, many researchers make a distinction between tags that describe the contents of a bookmarked webpage, and tags with a functional purpose related to expected future use of that bookmark[7, 19, 23]. Sen et al.[19] found that differences in suggested tags provided by the user interface can affect tag choices. Xu et al.[22] proposed an algorithm for suggesting tags based on how they have been previously applied to bookmarks in such a way as to encourage convergence on a common vocabulary.

Tag convergence could greatly benefit users interested in discovery by providing a consistent keyword vocabulary. But, what if tag convergence isn't really happening? The vocabulary problem[5] predicts that any two randomly selected users should not choose the same tag for the same bookmark more than 20% of the time. For convergence to be taking place users would have to explicitly and intentionally choose consensus tags, or be somehow influenced to do so. Most conclusions published thus far supporting the existence of tag convergence generally arise from an analysis of data scraped from web pages or from system logs. A notable exception is Yew, Gibson and Teasley[23], who found through interviews with users of an educational system that supported blogs and tagging that conventions did in fact emerge among members of a class. Analyzing log data in the absence of talking with users is

like finding a broken clay pot at an archaeological dig: we can guess that it was smashed as part of a salt-making process, but it could simply be that people are clumsy.

This paper extends previous work by interviewing users to determine the motivations behind their bookmarking and tagging choices, and exploring the consequences of those choices for other users of the system. We wanted to know, are tags are actually useful for discovery, or do users of del.icio.us go about discovery in some other way?

We assume that behavior is motivated and intentional, and users perform only those actions that yield benefits. For example, one benefit users derive from tagging their bookmarks is the feeling of being organized and able to find their bookmarks again in the future, should they need them. In this study, we identified the actions that our respondents performed regularly with del.icio.us, the motivations that led to these actions, and the benefits received from these actions. We found that as respondents chose bookmarks and tags in such a way that provided them with the most private benefit, their choices had an impact on the usefulness of the bookmark and tag information for others' discovery of new information via del.icio.us. Respondents found that information about who bookmarked a web page was more valuable for information discovery than the keyword or category information (tags) associated with it. We explain this by showing that the structure of benefits for bookmarking and tagging results in the selection of tags with little regard for their use by others.

This explanation has implications for systems with user-contributed content, such as Digg.com or Wikipedia.com. Such systems rely on users providing content that is interesting or useful to everyone else on the system. But users, acting in their own self interest, provide content in a way that is most beneficial to them. Del.icio.us has succeeded in aligning some incentives; a user bookmarking for his own benefit is doing so in a way that is useful to others. Unfortunately, incentives for tagging are not as well aligned. A user who chooses tags in a manner useful for private findability causes problems for other users who then try to use those tags for discovery.

DESCRIPTION OF DEL.ICIO.US

Users can post web pages to del.icio.us, creating bookmarks that are stored and displayed to the user on a page generated for them. Additionally, when creating a bookmark, the user can enter keywords (called tags) and a text string (called a note) that is stored with the URL. The system allows users to easily restrict the bookmarks they view to those associated with a specific tag, by clicking on that tag in a list that appears on the web page.

Del.icio.us makes all of this information publicly available. This means that users are capable browsing by each other's bookmarks and tags. The system will also display all the bookmarks associated with a given tag, or will display the most popular or the most recent bookmarks with a given tag. Users who wish to be notified when new bookmarks are posted can "subscribe" to users, or tags, or a combination, and bookmarks associated with these will appear on dynamically generated web pages.

	Mean	Std. Dev.
N	12	
Avg. # Bookmarks	950	1030.77
Avg. # Tags	400	356.075
Avg. # User Subs	14.6667	14.227
Avg. # Tag Subs	0.5	0.798
Avg. # Combo Subs	0.416	1.443
Percent with User Subs	1.00	
Percent with Tag Subs	0.3333	
Percent with Combo Subs	0.08333	

Table 1: Summary data for our respondents

Del.icio.us supports subscribing to another user, called adding the other user to your "network". If user A subscribes to user B, every time B posts a new bookmark it will show up in A's del.icio.us account. A user can also subscribe to all bookmarks associated with a given tag, or can subscribe to all bookmarks by a given user with a given tag (combo subscription).Finally, there is a mechanism in the system that takes the form of a special tag where one user can send a URL to another user. This URL appears on a third page, called "Links for you."

METHOD AND PARTICIPANTS

For this study we conducted a series of twelve 1.5 hour semi-structured interviews with regular users of del.icio.us over the summer of 2006. All these users had used del.icio.us for multiple months, and use del.icio.us regularly (at least once a week). Five of the twelve users are or recently were masters students at a local university, three are PhD students, one is an undergraduate, and three are information technology professionals. Respondents were self-selected by responding to fliers around campus and to internet postings on del.icio.us. Table 1 provides some simple statistics on their use of del.icio.us. The high standard deviations indicate that respondents exhibited a wide variety of usage. For example, the number of bookmarks by the respondents varied from a low of 60 to a high of approximately 3000.

The interviews included three phases. In the first phase, the interviewer asked general questions about the respondent's use of del.icio.us: how often do you use it, what do you bookmark, how do you choose tags, and similar high-level questions. The second phase consisted of ten search tasks. We had ten printouts of web pages found in del.icio.us, five bookmarked by the respondent and five bookmarked by other people, and we asked the respondent to find these using only del.icio.us. Respondents were asked to think

aloud during this task. Finally, the interviewer looked through the respondent's bookmarks and asked specific questions, such as "What were you doing when you bookmarked that? Why did you bookmark it? How did you choose these tags for it?" The interviewer also looked through the inbox and network and asked detailed questions about the subscriptions and usefulness of those pages.

The interviews were recorded, transcribed and coded in Atlas.ti[1]. The analysis was conducted in a similar fashion to Miles and Huberman[14]. The informal coding began with a list of classes of behavior we were looking for, and the code list developed as we proceeded. We focused on identifying the stated motivations of the respondents, the actions undertaken on del.icio.us, and the benefits received from using del.icio.us. Summary matrix displays were built showing which users had which motivations, undertook which actions, and received which benefits[2]. Causal chains were identified, and a theory was constructed to explain the observations. All names in this paper are pseudonyms, and statements have been anonymized to protect the identities of our respondents.

MOTIVATION, ACTION, AND BENEFIT

In this paper we build causal incentive chains for del.icio.us. First, a user starts with a need, or motivation, that they suspect can be satisfied by del.icio.us. For example, respondents indicated that they were motivated to keep track of interesting web pages so they could be easily returned to when needed. Next, a user undertakes one or more actions based on this motivation. Continuing the example, respondents would use the del.icio.us posting interface to bookmark a web page they found interesting. Finally, a user receives some benefit as a result of this action. The respondents were able to return to the interesting web pages when needed.

Roles and Actions

Users of del.icio.us can play two different roles with respect to the information stored in the system, and these roles are defined by distinct sets of actions. The first role is that of the information *producer*, who bookmarks web pages and associates tags with them. The second role is as information *consumer*, who uses the information in del.icio.us when searching or browsing the system by tag or person[11]. A user of del.icio.us performs both of these roles at different points in time. The bookmarking actions (producer) occur when a user is browsing other websites and decides to save the web page they are viewing in del.icio.us. The tagging actions (also producer), occur while users are in the process of saving the bookmarks. Tagging can be thought of as *packaging* the information for later re-use[11], because tags are one of the major methods the system provides for users to interact with content. Finally, discovery actions (consumer) include searching or browsing to discover new information among all of the bookmarks that other users have stored in del.icio.us.

A person can be both producer and consumer of their own content, in the case where they are concerned only with the private benefits derived from organizing their bookmarks, and storing them online where they can be accessed from anywhere. They can consume content that is produced by others, which happens whenever a user looks at a bookmark someone else has stored in del.icio.us. By default, all users produce content for others' consumption when they bookmark and tag a webpage, because the bookmarks and tags are all public unless explicitly specified as private by the user.

INCENTIVES IN DEL.ICIO.US

Privately-Motivated Bookmarking

The primary motivation that all respondents had for bookmarking websites was to keep track of web pages they found useful or interesting. Respondents wanted to be able to go back and access the web pages again if needed. As one respondent, Fred, said: "Any web page that I see, I basically ask myself this question: Would I ever have a need to find this again? And if I do I just bookmark it." On these topics, respondents tended to bookmark web pages that were particularly interesting to them, "the cream of the crop" as Fred called them. The action of bookmarking leads to the benefit of easy access.

A second personal reason to bookmark web pages that seven respondents found important was the ability to access their bookmarks from multiple computers. Zoe liked del.icio.us "because I'm working on so many computers and so many different places, it's just made me so much more efficient. It's a lifesaver."

The websites that respondents bookmarked can generally be divided into a number of categories. The most common type of website bookmarked was one that contains information on a topic of specific interest to the respondent. When asked what type of web pages he or she bookmarks, all respondents typically began by listing a number of topics their bookmarks cover. For example, Alice bookmarked PhD programs she was interested in, Bob bookmarked library-related links, Charlie liked web pages on sustainability, Oscar looked for programming skills, and Marvin was into community informatics. There is a different list of topics for every respondent in our study.

[1] http://www.atlasti.com/

[2] It is not immediately clear how one can make best use of del.icio.us. Indeed, all twelve respondents talked of an exploratory period where they attempted to figure out how to make del.icio.us useful for them. Actions that were undertaken during this period and never since (only once or twice, and in the past) are excluded from the totals reported below. We are primarily concerned with how people regularly use del.icio.us, not this learning period. The effects of this learning period are an open question.

Another type of webpage that eight respondents liked to bookmark was web pages that they intend to finish reading sometime in the future. Bob described this well: "Umm, I get to that situation where I have eight different tabs open in Firefox and I don't really have the time to read them all. Go up, get up and do something, and so I'll bookmark a bunch of them so that I will go back and in theory read them later." Half of these respondents expressed discontent over rarely actually returning to read these web pages.

A third type of web page consists of reference information or internet tools. These web pages generally helped these nine respondents perform other actions, such as finding information. Examples include new search engines (Zoe), manuals for the Perl programming language (Oscar and Eve), and collaborative text editors (Eve).

A final type is novelty or funny web pages. These are web pages with no specific purpose but are interesting for their novelty value. Fred described these well: "something funny. Like a video of a monkey sniffing itself or something. […] Or if something is just, oh wow cool, a new story or an amusing rant or a blog post, those get added […] as well." Another example was when Victor bookmarked a web page "because I thought the title was so ridiculous. Who would name a winery Rosechambeau winery?"

By bookmarking these diverse types of web pages, the respondents received the benefits of easy access to interesting or useful information, and access to that information from multiple computers.

Privately-Motivated Tagging

The primary motivation respondents reported for tagging was to organize their bookmarks. Alice said, "I like to be able to organize things in different ways." Marvin liked "to be able to group [bookmarks] logically." Peggy felt that "it's really nice to be able to categorize things the way that you see them."

One of the main methods ten of twelve respondents used to achieve this feeling of organization was to attempt to reuse tags he or she had used before. "I will not add a new tag until I have a group of things that I think it goes with" said Zoe. Respondents received multiple benefits from this practice. Reusing old tags made bookmarks easier to find by reducing the length of the list of their own tags a user must visually search on their personal page. Respondents wanted to avoid losing track of their bookmarks. Victor described this problem: "One of my friends, his tag section goes way down below the fold […] I'm like 'How on earth do you sort through all these?' And he said, 'I don't.'"

Respondents often reported that they had created mental rules or definitions for some of their tags. For example, Peggy described some of her rules about tags related to blogs: "So 'blogs' are usually other people's blogs. 'Blogging' would be something that's about usually research about blogging. And then if it's something like Blogger for instance or LiveJournal then that would be a 'bloggingtool'."

Another related benefit of tagging is making it easier to find bookmarks in the future. Seven respondents said they chose tags by trying to guess what terms they will search on to find the bookmark. Eve described her thought process:

> **Interviewer:** So, on the librarian video, how did you choose the tags that you have?
> **Eve:** […] If I were looking for this again […] I'd be like "What was that video about the girl in the library with that guy?" But girl and guy is not very helpful, so library and video won.

One major problem with tags is that multiple different tags can have the same meaning, both in terms of synonyms, and plurals or parts of speech. People in general have a hard time being consistent within themselves with the tags they use[7], and our respondents were no different. Trent described how he handles this problem: "I've been sloppy in the past about 'collaborative' and 'collaboration,' so this one got tagged as both. Just to make sure that I got coverage." Eve did the same thing: "So, apparently I'm using funny and humor interchangeably. And not reliably. So I should remember that when I'm looking for something funny I also label it humor. […] Apparently at the time I was using both funny and humor." Respondents also had problems with singular/plural tags and with misspelled tags, all of which create the situation where multiple tags have the same logical meaning. Respondents with this problem speak of their tags as "dirty," and the occasional act of fixing this as "cleaning" their tags.

Five respondents in our study used tags to group bookmarks they had saved in del.icio.us that they felt were related to a specific project. Whenever they bookmarked a web page that is related to the project, one of the tags applied to that web page was the project name. Marvin said this was "so I can just type in [the project name] and the things related to that project should show up if I did it right."

An interesting way ten of twelve respondents chose tags was for purely personal purposes, similar to the "functional purpose" tags of Golder and Huberman[7] and the personal tags of Sen et. al.[19]. These personal tags were intended to only have meaning for the person who applied them, often as some kind of reminder similar to those seen in other personal information management studies. Rather than placing icons in a specific location on the desktop to serve as reminders, as described by Barreau and Nardi[1, 3], respondents were using special tags as a kind of location within del.icio.us. "toread" is one example, used by at least four of the respondents. Fred had a "wishlist" for items he would like to purchase. Both Alice and Oscar used the tag "research" to refer to web pages that might be useful for their respective research projects. However, the meaning of all of these tags is highly personal, and can only be correctly interpreted and understood by the person who applied it, or people with whom they are close.

		Novelty	Topical	Social
By Tag	Browsing specific tags	.	....	
	Subscribing to a tag		:	
	Searching del.icio.us	.	:	
By Person	Looking at "Saved by X others"	..	:.....	
	Subscribing to someone you know	::::...	::::	::::…
	Subscribing to another user		:.	
	Subscribing to a famous person	..	:.	.
	Being referred to bookmarks	:		
Sys	Browsing Popular	:..	..	
	Browsing Recent	:	:	

Table 2: Actions used for discovery. Bottom row of dots is the number of respondents who use the action at all. Top row is the number who use it regularly.

Socially-Motivated Producers

In addition to the private motivations, respondents were motivated to bookmark and tag web pages for other people. The benefit they received for these actions was social recognition. The most common action of this kind was direct and intentional sharing with known others. Eight respondents talked about bookmarking web pages for the purpose of sharing them with specific people, in a variety of ways: using del.icio.us's built-in sharing ('for:x' tag), using a previously agreed upon tag, through knowing that someone specific subscribes to the respondent's bookmarks, or just by telling the person to go look for it.

Five of those eight, however, only rarely created bookmarks for the purpose of sharing them with others. As Trent explained, he does this "in lieu of, once upon a time you would have sent an e-mail with a link in it." Two respondents used del.icio.us as a tool to share links with one or more other people as they worked together. Oscar described how this helped with deciding where to publish a research paper:

> So, I'm publishing it with Professor [xxx] And instead of writing it out or emailing him, when I met him I just, you know, pulled up delicious and I just clicked the tag journal […] It lists the journals that I've come across chronologically, so the top few journals are the ones I wanted him and I to talk about. […] It was used as a shared discussion tool.

A second type of socially-motivated production is sharing bookmarks with unspecified others, or the public. Four respondents mentioned consciously trying to build a collection of links on a specific topic that would be useful to others. Alice said, "I tag everything on [topic of interest] I can find. I was so frustrated when I started working with this stuff that I just couldn't find information about it. […] There aren't many places for it so I have probably collected one of the larger lists out there." Respondents who did this usually seemed to want to be known as an expert on the topic of the collection. Additionally, the notes field on bookmarks often served as a general remark to others. Peggy saw the notes field as "almost like a blogging entry. I hardly ever post links to my blog anymore. If I have some sort of a comment about a link then I do [notes field]."

A third socially-derived benefit came from using del.icio.us to drive traffic to one's own web page. Two respondents posted links to their own web pages on del.icio.us hoping that other del.icio.us users would find these and click on them. Also, one respondent mentioned knowing that some search engines index del.icio.us and so bookmarks on posted there can increase their Google PageRank[2] or other similar measures. A related use was to add some of one's own bookmarks to one's web page through the del.icio.us RSS feeds. Two respondents discussed having all of their own bookmarks with a given tag appear in a side box on their web page.

The respondents were only motivated for these actions if they were aware of other users looking at their bookmarks, and this level of awareness varied widely. One respondent seemed to have no knowledge at all of others looking at his bookmarks. Seven respondents had directly told other people to look at their bookmarks, either indicating a specific bookmark ("I couldn't remember the Dog Judo link but I wanted him to check it out, so I sent him a thing that said Go to Dog Judo on my del.icio.us" from Eve) or directing them to a certain tag.

Six respondents were aware of other people who subscribed to their bookmarks[3]. Often this awareness came from friends mentioning during a conversation having seen their bookmarks. Isaac was aware of his friends subscribing to him because "Like every so often [a friend] will say I noticed you bookmarked that or [another friend] will say that. Certainly in person and maybe email." Half of the respondents mentioned a general awareness that others can see their bookmarks. But this awareness rarely affected their actions: "I do make a conscious decision of whether or not I want it to be available for everybody but 98% of the time I don't care." (Charlie) "Even though [my use of del.icio.us] is oriented primarily towards myself, the awareness that it is public never goes away totally." (Trent)

[3] Del.icio.us now makes this information available in the "your fans" section on the network page, but this was a relatively new addition at the time of the study.

Figure 2: Example Bookmark from Del.icio.us

Socially-Motivated Consumers

From the interview data, the primary way respondents used other del.icio.us users' bookmarks and tags was for discovering new information that they did not previously know. There are a number of actions on del.icio.us that lead to discovery: browsing by tag, subscribing to a person, subscribing to a tag, and clicking on the "Saved by X other people" link (see Figure 1). The motivations for discovery can be divided into three categories based on the type of information sought: novelty information, topical information, and social information. Table 2 shows various actions used for these classes of discovery.

Novelty discovery occurred when a respondent wanted to "look for something entertaining" (Alice). Basically, the user browsed del.icio.us for web pages he or she has not yet seen, but does not have a specific topic or idea in mind. This has been called "serendipity browsing", and is often unstructured, and undirected[4]. Often this type of discovery was for entertainment: "I check out [my friend, he] always goes to really interesting places." (Eve) The most common action undertaken by users for novelty discovery was subscribing in del.icio.us to someone he or she knew personally. As Bob said, he subscribed to a friend because "I like to pick his brain for cool stuff."

Our respondents were also motivated to use del.icio.us to seek information relevant to specific topics they found interesting through "general purpose browsing"[4], for *topical discovery*. Relevance is a relationship between the user's interpretation of the information in front of her, and her search context, which includes her goals, needs and assumptions at that moment. This means relevance is subjective, and situational. Users are able to make personal judgments of relevance throughout their information seeking activities based on the information in front of them; this is what guides them when selecting links to follow in del.icio.us[16, 18].

Respondents' general purpose browsing for topical discovery was either a one-time seeking behavior, or it was due to a continuing interest in the topic. For one-time seeking, the most common action was to click on the "Saved by X other people" link. See Figure 1 for how this looked in del.icio.us. Bob described this reasoning:

> "If I've got something bookmarked myself and it says 'Saved by X other people,' then its more intriguing to me if there are very few people who have saved it. Because that means I belong to this elite group of people who actually find this stuff interesting. […] And then maybe take a look at what else they've bookmarked because if they are interested in something that I'm interested in maybe they've got other stuff that I'd be interested in"

For continuing topical discovery, by far the most popular method was to subscribe to someone, usually someone the user knows in real life but occasionally someone who is either famous, or someone they found on del.icio.us by looking at "Saved by X other people." As Zoe discovered, "I found that certain people tend to tag the same things I'm interested in." Only one respondent (Trent) subscribed to any tags, and he was careful to block users (using a built-in blocking mechanism in del.icio.us) who post too many bookmarks for which "none of it fit my definition."

A less common way in which respondents went about one-time topical discovery was browsing del.icio.us by looking at specific tags. Only four respondents out of twelve mentioned this behavior, and they reported browsing by tag only rarely. This illustrates an interesting trend: respondents chose to browse by people rather than tags for information discovery. Presumably, this is because respondents found that browsing in this way satisfied their information need. In other words, they received greater benefit from browsing by user than by tag. A number of respondents specifically mentioned dissatisfaction in trying to use tags as filters for discovery.

The information contained in the web pages a user chooses to bookmark reveals information about his interests, and it is also evidence for what has been going on in his life. Some respondents picked up on this, and used del.icio.us to keep tabs on what specific individuals had been doing lately (*social discovery*). "It's just interesting to see what it is they're up to. So like my friend Matt who's not in the area anymore, I think I get a sense of what it is he's doing" as Peggy said. Mostly, this occurred as subscribing to friends. Charlie gives a good example:

> "One of my friends […] just got, you know, just got a job in San Francisco. I believe she went out to interview and then all of a sudden there's like 50 links to apartment search in San Francisco, and a few days later she tells me oh I got the job in San Francisco and I was like I know."

The next section attempts to explain the failure of tags for information discovery as an unintended consequence of the incentive structure for choosing bookmarks and tags.

BOOKMARKS, TAGS, AND DISCOVERY

So far, the only effects we mentioned have been direct benefits for the user. However, bookmarking and tagging actions have effects that stretch far beyond these motivating benefits. Specifically, these actions have effects that cross users, and roles. The producer action of bookmarking a web page in del.icio.us is like an endorsement, and those endorsements constrain what is available in the system for consumers to discover. If a web page is not bookmarked in del.icio.us, it cannot be accessed by browsing del.icio.us. In

PRODUCER ACTION	BENEFITS TO CONSUMER: Self-Produced	Produced by a Friend	Produced by General Public
Assign Tags	1. Feel Organized 2. Reminders	1. Receiving specific information via the 'for:x' tag	NONE. Due to the vocabulary problem, no consensus forms
Post Bookmarks	1. Re-finding web pages[3] 2. Available from any computer 3. Social recognition from friends 4. Visibility to del.icio.us users	1. Social discovery, i.e., keeping up with friends	1. 'Endorsements' by someone with similar interests 2. More content available in system for discovery (novelty, topical, social)

Table 3. Consumer benefits by type of producer action. Should be read as in, "Producers' actions of posting bookmarks benefit themselves by making their bookmarks available from any computer."

the same fashion, producers' selections of tags to apply to web pages they have bookmarked influence the effectiveness of those tags for consumers' discovery. If a tag is applied inconsistently to a number of different topics, it will not filter out enough irrelevant information, and the consumer must work much harder to satisfy their information need.

When trying to discover new novelty, topical, or social information, respondents generally looked for collections of bookmarks that were likely to contain the sought-after information. Collections are temporary sets of URLs that are returned by del.icio.us by applying a filter to the set of web pages posted to del.icio.us. These filters can operate by person (web pages bookmarked by person X) or by tag (web pages associated with tag Y). The filters can be either one-time searches, or ongoing subscriptions.

Useful collections generally have three properties. First, useful collections have a *reasonable number of bookmarks* to sort through. This is particularly important for ongoing discovery through subscriptions, where the rate of incoming bookmarks depends on how prolific others are; too many bookmarks make it difficult to keep with the influx of web pages. When Oscar subscribed to a very heavy del.icio.us user, it "sort of cluttered [his] stream" of incoming URLs. Second, useful collections include a *large number of relevant bookmarks* (recall). This means that they cover a large portion of the bookmarks that the user would want to see. And finally, useful collections include *few irrelevant bookmarks* (precision). Fred said this about a collection with poor precision: "I'd go through it, but I end up skipping a lot." He ultimately stopped using those subscriptions altogether.

Tags and Discovery

First, we discuss discovery through filtering by tags. This includes actions such as browsing the system by looking at specific tags, and subscribing to a tag. This also includes searching del.icio.us with the search box, which primarily uses tags and titles to return results. As we mentioned above, discovery using these methods is relatively rare, presumably because users receive insufficient benefits from these actions to continue using them. We conjecture that this use is rare because collections based on tags have none of the three properties listed above.

The collection of bookmarks under a given tag is not bounded in quantity. As del.icio.us becomes more popular and more users join the system, an increasing number of users might apply a given tag when bookmarking web pages. This leads to individual tags being associated with a large and increasing number of bookmarks. Indeed, three respondents mentioned trying to subscribe to tags, until they began receiving too many web pages to keep up with. When choosing tags for a bookmark, none of the respondents expressed concern about increasing the number of bookmarks system-wide already associated with a tag.

The vocabulary problem[5] causes two additional difficulties with respect to tags. First, people rarely agree when selecting single words (tags) for the same concept. This means that a single tag rarely covers most of the interesting web pages related to a given topic in del.icio.us. This is a problem of poor recall; a user browsing by a tag is likely to miss a lot of relevant information. Eight respondents noticed this problem just within their own bookmarks! Trent mentioned being "sloppy in the past about 'collaborative' and 'collaboration'" and Charlie "at one point had 'recipe' and 'recipes'."

Second, different people often have different concepts in mind that they end up representing with the same word when they apply a tag to a bookmark. Fred subscribed to the 'security' tag, but as a security expert he found many of the bookmarks with that tag too basic for him. However, a novice user would likely find the web pages he bookmarked under 'security' too advanced. This is a precision problem – there are many irrelevant web pages returned by the tag, whether you are a novice or an expert!

Users and Discovery

The other method for discovery involves filtering the bookmarks on del.icio.us by a specific user. This can be accomplished by clicking on "Saved by X other people," subscribing to other del.icio.us users, and being told to go look at a user's bookmarks. The respondents in our study found these actions to create more benefits than browsing by tag for discovery. Collections generated through filtering by user generally have all three properties of useful collections.

The set of bookmarks created by a single user is generally small, with our respondents ranging from 1 bookmark per week to a high of 13 bookmarks a day, and averaging about two bookmarks per day. This is a much smaller number than the hundreds of bookmarks a day that some tags (like "music") have associated with them. Users are motivated to store their bookmarks in a manner that makes it easy to find them in the future; limiting the number of bookmarks in their collection aids in this goal. Additionally, users only have so much time in the day for bookmarking, further limiting their quantity of bookmarks. These two factors combine to make the quantity and frequency of bookmarks belonging to most users reasonably small. This property is useful for discovery, because users are not overwhelmed by too many bookmarks.

Users are also motivated to bookmark web pages in a limited topic area related to their interests[21]. The majority of bookmarks for all of the respondents were from a small set of topics the respondent is interested in. Assuming that the user will continue to bookmark URLs on the same topics, then looking at a del.icio.us user's current pattern of bookmarking is a reasonable predictor of how beneficial subscribing to that user might be. Five respondents reported looking on del.icio.us for other users with similar interests, and then subscribing to those users.

Fred indicated some level of awareness that other users may subscribe to him. He described this well:

> **Interviewer:** Do you ever think of anyone when you're posting, like you're posting for this person or that person?
> **Fred:** Sometimes I do but it's mostly just a general sense of the network, of people. If someone tracks the stuff that I post enough, then I assume that they basically care about the eigenvalues of the things that I like. And so I figure if I like it then they'll like it, and if I don't like it then they don't like it.

The largest problem with subscriptions like this is the prediction that people will continue bookmarking web pages on the same topics they have shown interest in. Often users' interests will change, and then they will no longer be a good match, a good subscription. Fred talked about unsubscribing from people when they went "on bookmarking tangents." Bob experienced this with one of his subscriptions: "he was bookmarking some cool stuff but he stopped." Zoe decided this about one of the users she was subscribed to: "the stuff that she had that was interesting to me must have been a blip in her life." This changing of topics, however, is what makes social discovery so interesting.

Users choose a relatively small number of bookmarks on a focused topic area. This makes filtering by user quite effective. On the other hand, users choose tags without considering the number of bookmarks associated with them, and without compensating for the vocabulary problem. This causes collections filtered by tag to have a low signal / noise ratio.

RELATED LITERATURE REVIEW

Furnas et al.[5] found that random pairs of people use the same label for an object at most 20% of the time. The implications of these findings for social bookmarking systems are dire: if two random users create a tag for the same URL, they are far more likely to choose different tags than the same tag. Similarly, if an information consumer attempts to imagine what tags might be applied to the information he is looking for, chances are low that he will end up using for the correct tags. This robust tendency results from humans' imprecise and flexible use of language in conversational settings, where meaning is determined by the surrounding context and complex communication processes. Information producers in del.icio.us who are motivated by private benefits are unlikely to spend the time and effort that would be required to conform to a common, public vocabulary. Markus[11] referred to the challenge of designing appropriate incentives for packaging content for the use of others as the "discretionary database problem".

A closely related literature comes from economics. In economics, a public good is anything which is "nondiminishable," meaning that one person using it does not reduce the availability for others, and "nonexcludable," which basically means it is available to the public and no one is excluded[13]. One of the distinguishing features of these goods to economists is that most of the time individuals are insufficiently motivated to contribute to public goods relative to what would be societally efficient. The standard solution is to have the government provide the good, as is the case for the provision of national defense (the army). Voluntary provision of public goods by individuals is an open and interesting problem[20].

The bookmark and tag data on del.icio.us is a public good. One person browsing users or tags does not noticeably diminish the ability of others to do the same, and the data is publicly available so no one is excluded. Del.icio.us encourages users to provide this data to the system by providing them the private benefits we discussed above. However, this incentive is not perfect; we have discussed how the voluntary provision of tags on del.icio.us does not provide a high-quality public good that is useful for discovery.

One finding in this paper is that del.icio.us works similarly to recommender systems like GroupLens[17] or MovieLens[8]. Recommender systems work by algorithmically finding other users with similar interests and recommending items those users liked. Del.icio.us does not have this algorithmic method of finding similar users; most users either know people through other means or find similar users by looking through the "Saved by X other people" links. The same principle holds, though, that the best recommendations come from other people with similar interests. In del.icio.us, bookmarking a web page is like 'voting' for that web page, or giving it an endorsement, that other like-minded users can benefit from. Maltz and Ehrlich[10] presented a related system for collaborative filtering that relied on explicit sharing of "pointers" to documents. It, however, did not have the private benefits that del.icio.us has to encourage contribution.

DISCUSSION AND CONCLUSIONS

It is important to understand how users of del.icio.us use the system for personal benefit, because these actions and choices affect the bookmarks and tags that are available for others to use for discovery. Producers choose web pages to bookmark in such a way that consumers' filtering by user can prove beneficial, but choose tags in such a way that makes it difficult for consumers to use them to discover new information.

This understanding can be applied in other situations. For example, Amazon.com recently implemented tagging on their website. Users can apply tags to various products on the website. However, Amazon already has its products organized. We conjecture that the apparent problems with Amazon's tagging system[6] are due to two observations. First of all, adding tags to products that are already organized provides little additional private organizational benefit. This means that few users will choose to apply tags for their own use. Secondly, tags themselves are not naturally useful for discovery due to the vocabulary problem.

Other systems similar to del.icio.us are being built for different environments. The Dogear[15] system from IBM is a social bookmarking tool designed for enterprises. One of the stated goals of the Dogear system is to support discovery of relevant information, and support the formation of communities of practice. The system should provide benefits to users for bookmarking in such a way that the resulting bookmarks will be useful for discovery. Following the del.icio.us model is likely to lead to effective discovery by user, but not by tags. However, in many enterprises users already know and talk to the other users with similar interests and thus the usefulness of Dogear might be limited. (Engineers tend to prefer sources of information that are nearby, and oral, to internal documents and library sources[9].)

More generally, similar concepts apply to any system with user-generated content that is made public and used by others. Systems such as Digg.com (a community generated news site) and Wikipedia.com (a community generated encyclopedia) rely on user contributions for their content. We argue that incentives for private benefit must be aligned with the needs and interests of other users of the system, in order for the system to be a success.

For social bookmarking systems, we make two untested recommendations based on our findings. First, since filtering by user seems to be the most useful method of discovery, provide technological systems to help consumers find other users with similar interests. Right now del.icio.us has the "Saved by X other people" link. We can imagine augmenting this with an algorithm that attempts to guess which users have the most similar interests and recommending these users as a source of new information.

Second, the primary reasons that tags fail is due to quantity concerns and to the vocabulary problem. If we can properly motivate the producers to choose tags in such a way to avoid the vocabulary problem and choose "better" tags, then filtering by tag may become more useful. Suggesting tags, like del.icio.us' primary posting interface does currently, is a step in the right direction. However, these suggested tags are rarely used because, as Zoe put it, "They don't match what I'm trying to do." Suggested tags interfere with users private benefits of using del.icio.us. Is there any way to have producers benefit from choosing tags that consumers find useful? To answer this, a related question must also be answered. What would "better" tags look like? What properties of tags would consumers find useful?

ACKNOWLEDGMENTS

Thanks to Jeffrey MacKie-Mason, Judy Olson, Jude Yew, Yong-mi Kim, Stacy All, Lian Jian, and the rest of the ICD lab group (Anna Osepayshvili, Greg Gamette , John Lin, Kil-Sang Kim, Ben Stearns, and Benjamin Chiao) for numerous helpful discussions.

REFERENCES

1. Barreau, D. and Nardi, B.A. Finding and reminding: File organization from the desktop. SIGCHI Bulletin, 27, 3 (1995), 39-45.
2. Brin, S. and Page, L., The Anatomy of a Large-Scale Hypertextual Web Search Engine. in 7th International World Wide Web Conference, (1998).
3. Bruce, H., Jones, W. and Dumais, S., Keeping and Re-Finding Information on the Web: What Do People Do And What Do They Need? in ASIST 2004: Proceedings of the 67th ASIST annual meeting, (2004).
4. Chang, S.-J. and Rice, R.E. Browsing: A multidimensional framework. Annual Review of Information Science and Technology, 28 (1993), 231-271.
5. Furnas, G.W., Landauer, T.K., Gomez, L.M. and Dumais, S.T. Statistical Semantics: Analysis of the Potential Performance of Key-Word Information Systems. The Bell System Technical Journal, 62, 6 (1983), 1753-1806.
6. Golbeck, J. The Tagging Hall of Shame: Amazon. Retrieved September 27, 2006 from

http://www.oreillynet.com/xml/blog/2005/12/the_tagging_hall_of_shame_amaz.html
7. Golder, S. and Huberman, B.A. The Structure of Collaborative Tagging Systems. Journal of Information Science, 32, 2 (2006), 198-208.
8. Harper, F.M., Li, X., Chen, Y. and Konstan, J. An Economic Model of User Rating in an Online Recommender System 10th International Conference on User Modelling, Springer-Verlag, Edinburgh, UK, 2005.
9. Hertzum, M. and Pejtersen, A.M. The information-seeking practices of engineers: searching for documents as well as for people. Information Processing & Management, 36, 1 (2000), 761-778.
10. Maltz, D. and Ehrlich, K., Pointing the way: active collaborative filtering. in Proceedings of the SIGCHI conference on Human factors in computing systems, (1995), 202-209.
11. Markus, L.M. Toward a Theory of Knowledge Reuse: Types of Knowledge Reuse Situations and Factors in Reuse Success. Journal of Management Information Systems, 18, 1 (2001), 57 - 93.
12. Marlow, C., Naaman, M., boyd, d. and Davis, M., Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead. in WWW 2006 Collaborative Web Tagging Workshop, (2006), Publisher.
13. Mas-Colell, A., Whinston, M. and Green, J. Microeconomic Theory. Oxford University Press, 1995.
14. Miles, M.B. and Huberman, M. Qualitative Data Analysis: An Expanded Sourcebook. Sage Publications, Inc., 1994.
15. Millen, D.R., Feinberg, J. and Kerr, B., Dogear: Social bookmarking in the enterprise. in Proceedings of the SIGCHI conference on Human Factors in computing systems, (2006), ACM Press
16. Pirolli, P. Rational Analyses of Information Foraging on the Web. Cognitive Science: A Multidisciplinary Journal, 29, 3 (2005), 343-373.
17. Resnick, P., Iacovou, N., Sushak, M., Bergstrom, P. and Riedl, J. GroupLens: An open architecture for collaborative filtering of netnews. Computer Supported Collaborative Work, ACM, 1994.
18. Schamber, L. Relevance and Information Behavior. Annual Review of Information Science and Technology (ARIST), 29 (1994), 3-48.
19. Sen, S., Lam, S.K., Rashid, A.M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F.M. and Riedl, J., tagging, communities, vocabulary, evolution. in CSCW 2006, (2006).
20. Smith, V. Experiments with a decentralized mechanism for public goods decision. American Economic Review, 70 (1980), 584-599.
21. Udell, J. Managing Metadata, InfoWorld.com, 2005.
22. Xu, Z., Fu, Y., Mao, J. and Su, D., Towards the Semantic Web: Collaborative Tag Suggestions. in WWW 2006 Collaborative Web Tagging Workshop, (2006).
23. Yew, J., Gibson, F.P. and Teasley, S.D. Learning By Tagging: Social tagging as a means of group knowledge formation, University of Michigan, Ann Arbor, MI, 2006.