Philip J. Binkowski. The effect of social proof on tag selection in social bookmarking applications. A Master's Paper for the M.S. in I.S. degree. December, 2006. 41 pages. Advisor: Barbara Wildemuth.

The growing popularity of social bookmaking applications like flickr and del.icio.us present new challenges to system designers because the effects of social psychological factors on users' tag choices have not been examined. The social psychological principle of social proof is particularly applicable to social bookmarking because it predicts that the tags applied by users will be more similar to each other if they are provided with a list of suggested tags. This study examines the effect of social proof on tag selection by comparing the degree of similarity between tags provided by a sample group and a collection of suggested tags provided to the treatment group. The results indicate that social proof can have an effect on users' tag selection. The conclusion briefly examines the beneficial effect of social proof on the quality of social bookmarking applications and other collaborative tagging applications

Headings:

Categorization/Psychology

Social Psychology

World Wide Web/Bookmarks

Indexing

Collaboration

THE EFFECT OF SOCIAL PROOF ON TAG SELECTION
IN SOCIAL BOOKMARKING APPLICATIONS

by
Philip J. Binkowski

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Caroline at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill

November, 2006

Approved by:

_____
Barbara M. Wildemuth (advisor)

Introduction

Perhaps one of the most common components of a modern individual's personal information space is their collection of bookmarks. An estimated 90% of web users have created bookmarks in their primary browser (Aula et al. 2005). Even before the World Wide Web existed, experimental hypertext systems provided a bookmark feature for helping users return to previously reviewed content (Bernstein 1988).

A typical web browser such as Microsoft's Internet Explorer allows the user to save bookmarks into a hierarchical collection of user-defined folders. The hierarchy model is probably, at least in part, an artifact of the way that people organize records in their personal file systems (Barreau & Nardi 1995). Both Barreau and Nardi (1995) and Boardman and Sasse (2004) observed a clear preference for location-based searches, suggesting that one reason for the continued support of folder hierarchies is that they provide a relatively simple information model that is easy to navigate in this fashion. The stability of hierarchies certainly facilitates retrieval – organizing the contents in an unambiguous fashion (Golder & Huberman 2006). However, this organization scheme does not support sharing of bookmarks with others particularly well.

Typical web browsers also support very little explicit metadata for describing bookmarks. Internet Explorer, for example, only allows the user to provide a label for a bookmark as well as storing the bookmark's creation and last modified dates. Instead, in organizing bookmarks, users rely on the implicit metadata of the bookmark's position in

their bookmark hierarchy. Because this implicit information is highly situational, it is of little use to other information consumers. Moreover, even if the metadata was useful to others, it would be difficult to articulate for the purpose of transmission as it is positional data for a particular file.

One possible solution to the problem of facilitating bookmark sharing is collaborative tagging. Collaborative tagging systems allow users to categorize their bookmarks with multiple user-defined tags. Such an organization system is sometimes referred to as a "folksonomy" – a term coined by Thomas Vander Wal from the words folk -as in non-professional users - and taxonomy (Smith 2004.). However, it should be noted that a folksonomy is not truly a taxonomy because this organization scheme does not incorporate such concepts as inheritance and exclusivity (Jacob 2004). Therefore, for the purposes of this article, such systems will be referred to using Golder and Huberman's (2006) term: *collaborative tagging systems*.

Collaborative tagging systems present new challenges to system designers because social and psychological factors may affect users' tag choices as well as their own information needs. One factor that is particularly applicable is the principle of social proof which indicates that people model their behavior on others (Cialdini 1995). Since users of collaborative tagging systems can see an aggregate list of tags provided by other users, it is logical to assume that they may view this list of tags as a surrogate for other people since previous research suggests that people can be manipulated to respond to a computer as if it is a person (Nass & Moon 2000).

If the principle of social proof does affect users' tag choices and all users are provided with the same list of suggested tags, this principle suggests that users will be

more likely to choose similar terms for their tags. A convergence of users' term choices

resulting from social proof would help address the proliferation of term variations, which

has been identified as a significant drawback of collaborative tagging systems (Golder &

Huberman 2006; Guy & Tonkin 2006).

Literature Review

Collaborative Tagging Systems

Using the most rigorous of definitions, collaborative tagging systems are not an

organization system but instead a grouping system, because they include neither a way of

establishing relationships between the categories nor a way of creating systematic order

of the categories (Jacob 2004). Advocates of collaborative tagging systems believe that

this very lack of structure is what makes such systems useful. Quintarelli (2005) suggests

that the strength of collaborative tagging systems is that they require neither a central

authority nor cataloging expertise but instead harness the domain expertise of the

collection of users. Moreover, Quintarelli contends, collaborative tagging may be the

only viable way of cataloging any significant amount of web content because traditional

classification schemes are too intellectually and financially expensive to encompass such

a large collection of information.

Collaborative tagging may also address one of the key issues with traditional

cataloging: neutrality. The problem is most succinctly expressed by Doctorow (2002) as

follows: "Any hierarchy of ideas necessarily implies the importance of some axes over

others" (paragraph 23). It is obviously impossible to create a truly neutral cataloging

scheme. However, a system that uses collaborative tagging allows a user to capitalize on

the work of others where appropriate and customize their organization scheme where

appropriate.

Until relatively recently, systems that allowed relatively naive users to collaboratively organize information resources were mainly experimental. The earliest known collaborative tagging system was "WebTagger," created by a team at NASA's Ames Research center (Keller et al. 1997). The goal of the project was to design a system that allowed bookmarks to be organized for sharing across groups. WebTagger included a relevance feedback mechanism to rank the tags applied to bookmarks based on previous relevance judgments made by the users. However the system differed from more recent collaborative tagging applications because it used a controlled vocabulary of tags instead of an open vocabulary.

More recent applications like Flickr (www.flickr.com) and Delicious (del.icio.us) use the open vocabulary that has become associated with collaborative tagging applications (Weiss 2005). These applications were more recently developed than WebTagger; Delicious, was founded in 2003 (del.ico.us 2006) and Flickr in 2004 (Ludicorp 2004). Unlike WebTagger, these systems are commercial ventures that are open to the general public and they have very much become part of general cultural knowledge. At the time of this writing Flickr and Delicious ranked 40th and 151st in total web traffic worldwide (Alexa.com 2006). Flickr is a web-based photo-sharing service that uses a tag-based system for organizing images. Delicious is a web-based bookmarking system – it incorporates a tagging system and the ability of users to share both their actual bookmarks and the tags they use to organize them.

When the user adds a new bookmark to a system like Delicious, tags are suggested based on an aggregate list of tags previously provided by other users (Weiss 2005).

Providing the user with the option of using existing tags reduces much of the cognitive load of organizing their bookmarks. Moreover, working from a shared categorization scheme allows bookmarks to be shared more readily between users, as a bookmark could be passed between users with the accompanying tags (Keller et al. 1997). This metadata accompanying the bookmarks would make shared bookmarks self-describing. Ideally, providing suggested tags will cause the users' disparate categorization schemes to converge such that the tags describing a particular bookmark would easily integrate with those of the user receiving the shared tags.

It is natural to question the utility of collaborative tagging systems because they may not converge as anticipated above. In the absence of a controlled vocabulary, tags specific to particular users might proliferate to the detriment of the overall system. However, user selection of tags in Delicious appears to follow a power law distribution similar to Zipf's law (Cattuto 2005). Thus, as with a more general corpus of linguistic utterances, a small number of tags are used most frequently as illustrated by an analysis of the distribution of tags for one of the most popular sites listed on Delicious, Slashdot.org (see Figure 1). Voss (2006) demonstrated that several other collaborative tagging applications, such as Flickr, have similar tag distributions.
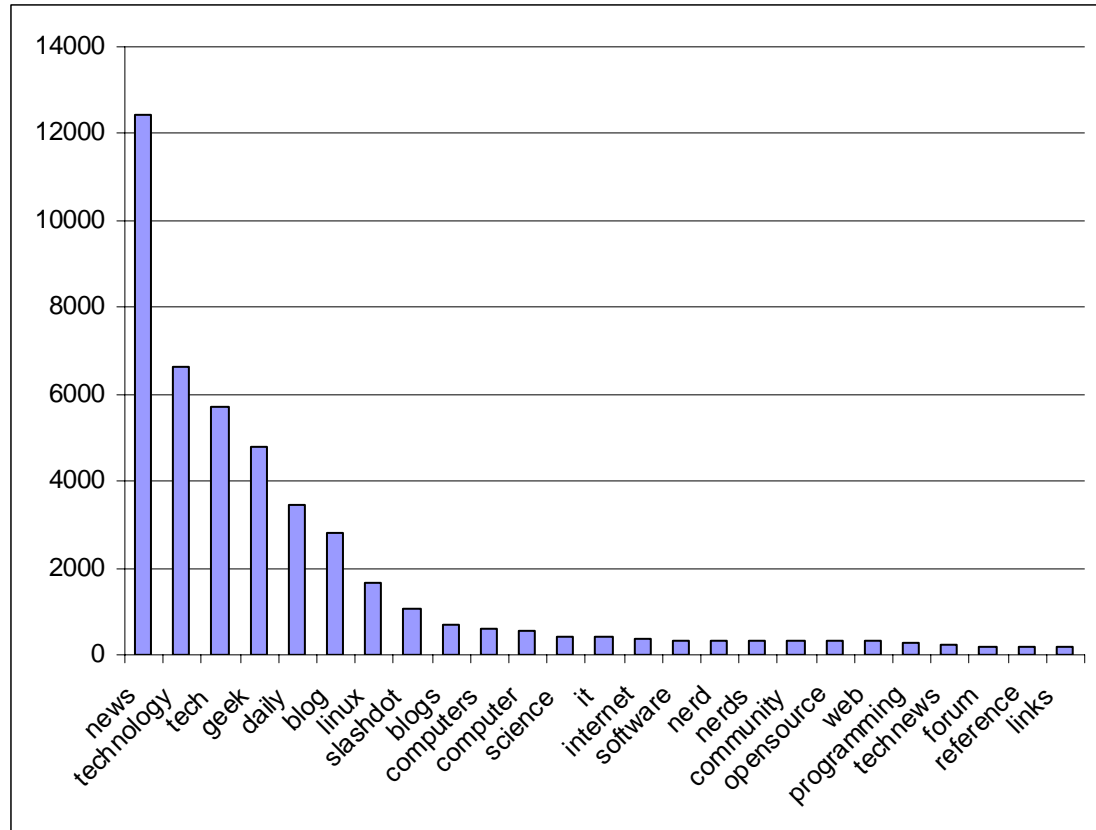
Figure 1- Frequency distribution of Slashdot.com tags

Golder and Huberman (2006) determined that, after approximately 100 Delicious bookmarks are applied to a particular URL, the proportion of tags applied to the URL stabilizes. Instead of the number of tags assigned to a bookmark proliferating indefinitely with the addition of users, "each tag's frequency is a nearly fixed proportion of the total frequency of all tags used" (p. 206).

In a study of the Technorati collaborative tagging application, Brooks and Montanez (2006) determined that Technorati users "are independently choosing distinct tags to refer to the same concepts" (p. 631). Thus, while we can expect a large degree of *conceptual overlap* between the sample used for this study and the general population, terms used for the tags may not overlap.

In Delicious, the first tag an individual uses has a strong tendency to be the highest overall rank of all tags assigned to a particular bookmark, with subsequent tags having decreasing overall ranks (Golder & Huberman 2006). Moreover, Kome (2005) discovered that "up to 90% of [Delicious] tags (per post) are members of hierarchies" (p. 17). From this information we can conclude that users tend to include tags in descending order of popularity as members of a hierarchy .

Social Proof

Particularly applicable to collaborative tagging systems is the social psychological principle of social proof. This principle suggests that: "We view a behavior as correct in a given situation to the degree that we see others performing it" (Cialdini 1993, p. 95). In the case of collaborative tagging applications this suggests that tags chosen for application to a particular bookmark will be more similar to those suggested by others than they are unique.

Social proof may particularly affect Delicious bookmarks in cases where users are unsure of how they should tag a particular URL, because uncertainty is known to contribute to the effect of social proof (Cialdini 1993). Thus, suggested tags may help novice users feel more confident about their tagging choices.

The amount of effort required to tag items may also affect an individual's decision to use tags suggested by others. If we start with the reasonable assumption that using suggested tags rather than one's own requires less effort, then Pirolli and Card's (1995) theory of information foraging suggests greater adoption of suggested tags because people adapt their behavior to optimize the information/effort ratio (Gattis 2002). Abrams et al. (1998) provide evidence of the effect of effort on bookmark files, concluding that

"many users cost-tune their archives by expending the least amount of effort needed to build up enough structure to support fast retrieval of their most useful bookmarks" (p. 44).

Since there is evidence that both social proof and the desire to reduce the amount of effort required to organize bookmarks both contribute to the way that people organize their bookmarks, it would be desirable to determine the degree to which these factors affect people's tag choices, with ultimate goal of encouraging the convergence of the users' tag collections. Greater tag convergence would allow easier sharing of tags because of better compatibility between the different users' tag collections.

Summary

Collaborative tagging applications show promise as a new way for individuals to organize their information. However, since such applications are relatively new, many questions remain unresolved in regard to their effectiveness. Because a significant distinction between collaborative tagging and traditional cataloging is the social effect of tags being applied in the context of an existing framework of tags provided by other users, one area where additional research is needed is the social factors affecting tag selection. A likely contributing factor is the social psychological principle of social proof. The principle indicates that users are more likely to behave in a manner in which they observe others behaving – especially when there is a sense of uncertainty about the correct action. Tag selection may be an activity for which users will seek social guidance.

Research Questions

The specific questions that this study examines are:

1. To what degree does the principle of social proof affect the user's tag choices

in a collaborative tagging application?

2. To what degree does the principle of social proof affect the user's satisfaction with their tag choices in a collaborative tagging application?

Study Methods

Overview

Study participants were recruited via a general campus mailing list. Volunteers were provided with a web survey containing ten websites to tag/categorize. The phenomenon of social proof was simulated by providing users with suggested tags from the Delicious site for half of the items (in a block) and not for the other half (in a second block). Following each block of websites, participants were asked to evaluate their satisfaction with the tags they applied. The participants' performance was evaluated in terms of the amount of overlap between the tags that they provided and that of the Delicious community. A detailed description of the study methods is provided here.

Study Participants

Study participants were recruited by means of an email sent to the university-wide mailing list (Appendix B). One hundred and sixty seven participants were recruited for this study, 139 of which provided usable data. The sample included students, staff and faculty members at the University of North Carolina.

Since a research goal was to acquire a relatively diverse sample representative of a wide range of Internet users, no particular selection criteria were applied to the sample. However, the participants were asked to describe their experience with the web in general and with web bookmarks in particular.

Item Selection

Website Selection

The ten websites tagged by study participants were randomly selected from three

sources: Alexa.com's top 100 sites, netcraft.com's most visited sites and populicio.us's list

of most popular del.icio.us sites. This selection method deliberately emphasized popular

sites, as such sites are more likely to resemble websites that the participants normally

visit. Multiple sources were used to compensate for potential biases within each

individual source.

Because they are potentially difficult to categorize, search engines and portals were

removed from the sample and replaced. Such sites could be considered to be related to an

infinite number of categories or no categories, depending on the individual user. Sites

containing pornography and potentially illegal content (such as BitTorrent tracker sites)

were removed to avoid exposing the participants needlessly to content they might find

controversial or upsetting. Sites in foreign languages (i.e. non-English) were also

removed from the list . The order in which the sites were presented was the same for both

groups and was determined via random selection.

Once the sites were selected, the links were normalized to use the "www." prefix

instead of the specialized prefixes used by some of the sites (e.g., "search.ebay.com" was

normalized to "www.ebay.com"). Using the "www." prefix ensured that the main index

page for each of the sites was used rather than a sub-index. By using the main index, a

larger pool of Delicious suggested tags for each site became available. For example, the

site address selected for "www.ups.com" was included on the original source list as

"ups.com". For this latter address, Delicious provided only three suggested tags while the

former was associated with 19 different tags. The exception to this rule was

craigslist.com. This site was unique within the sample because different content is

provided based on the user's geographic location. If a user links to "www.craigslist.com"

as a *new* visitor, the site defaults to San Francisco. If the user has previously visited

craigslist.com, the site will default to the last visited location. Therefore, the Raleigh, NC,

location was chosen to ensure that all users viewed the same content on the site.

After the sites were selected, they were placed into random order. Once ordered, the sites

were grouped into two blocks based on their order position. Thus, the first five sites were

included in block one and the second five sites in block two. The resulting list of sites, in

the order presented to study participants, is shown in Table 1.

Table 1. Site Collection

Block	Site URL	Site Description
Block 1		
	http://www.ups.com/	Corporate website for the United Parcel Service
	http://www.digg.com/	Technology and social networking site.
	http://www.typetester.maratz.com/	Web design tool for testing font faces.
	http://www.cnet.com/	Technology news and download site.
	http://raleigh.craigslist.org/	Classified ads website.
Block 2		
	http://www.ebay.com/	Internet auction site.
	http://www.flickr.com/	Photograph sharing site.
	http://www.forbes.com/	Forbes Magazine's website.
	http://www.microsoft.com	Corporate website for Microsoft, Inc.
	http://www.gizmodo.com/	Technology web log.

Suggested Tag Selection

To investigate the effects of social proof, tags were selected for each website. The

tags suggested were drawn from the most popular tags listed for each site on Delicious.

Popularity was determined by examining frequency counts of the applied tags using the Delicious history function. The 10 tags with the highest frequency of use in Delicious per site were used as the suggested tags

For some websites, there was a high degree of overlap in the tags supplied by the Delicious community. For example the following are the eight Delicious tags for the UNC homepage [www.unc.edu]:

1. unc
2. school
3. college
4. gradschool
5. universities
6. university
7. education
8. usa

In a situation like the above, where there is significant overlap between "universities" and "university", both tags were collapsed together as "universities." When collapsed, the plural form of the tag was consistently chosen. Tags were collapsed when the tags shared either the same word (varying by singular or plural) or diminutions of the same word (e.g. "tech" collapsed with "technology"). Tags were *only* collapsed when *all* of the words were shared by both tags. Thus, in the above example, "school" and "gradschool" would *not* be collapsed together despite overlapping on the word school. When tags were collapsed, the collapsed tags were placed in the slot held by the most popular version of the tag.

Several tags were ignored because they were obviously applied by either Delicious or a web browser. For example, the tag "imported" was ignored because it's automatically applied to all sites imported into Delicious from a web browser's bookmark collection.

For similar reasons, the "safari_import" and "imported_ie_favorites" tags were ignored.

A list of the selected sites with the suggested tags is included in Appendix A.

Research Design and Procedures

Study participants were recruited through an email sent to UNC's campus-wide

mailing list (Appendix B). Included in the email was a link to the survey. When a user

clicked on the link, he or she was randomly placed into one of two cohorts via a php

script. After receiving information on their rights as research subjects and providing

consent, the participants were asked some brief questions about their internet usage and

provided with instructions on how to use a collaborative tagging system (Appendix C -

Figure 1).

Both cohorts then received the first block of sites listed in Table 1. Cohort A

received the block of sites without any suggested tags (Appendix C – Figure 2) while

Cohort B received suggested tags (Appendix C – Figure 3) – from the list included in

Appendix A. In both cases, the participants were provided with a free text field for each

site to enter their tags. Users were given no guidance on how to delimit their tags as no

such guidance is provided by Delicious.

After tagging the first block of sites, participants were asked to describe their

satisfaction with their tags via three items adapted from Bailey and Pearson's (1993) user

satisfaction measure (Appendix C – Figure 4). This subset of the Bailey and Pearson

items were selected for their relevance to this study and focused on the following

concepts: *accuracy, completeness,* and *perceived utility.* Satisfaction with each of the

three concepts was measured via four semantic differential scales, for a total of twelve

scales.

Following the satisfaction items, both cohorts received the second block of websites. Reversing the previous block's treatment, Cohort A received suggested tags for this block while Cohort B did not. The second block of sites was followed by a repeat of the satisfaction items described above.

Dependent Variables and Data Analysis

To measure the degree to which social proof affected participants' tag choices, the Jacquard Similarity Coefficient was used to compare the amount of similarity between the participant's tags and the tags used by the Delicious community for a particular site

Comparisons were made between each participant's tags and three sets of the suggested Delicious tags for the same site: all of the suggested tags, the top 5 tags and top 3 tags. These top tags were determined by frequency of use on Delicious using Delicious' history utility. The mean similarity coefficients per site for the two groups were then compared using a one-way analysis of variance.

Participants' satisfaction with their tags was calculated first by averaging the four semantic differential ratings for each of the three items: accuracy, completeness and perceived utility. The difference between the groups of sites was evaluated for statistical significance using a t-test. For these t-tests and for all analyses of variance, the criterion for statistical significance was set at 0.01.

Results

Participant Demographics

The sample's description of their web usage indicates that they were collectively familiar both with the World Wide Web in general and, specifically, with web bookmarks. Participants reported that they spent an average of 14 hours browsing the

web per week (StdDev = 10.74) and had an average of 93 bookmarks saved in their web browser (StdDev = 188.22).

When asked, "How would you rate your experience with your web browser?", the majority (64%) indicated that they were either "Experienced" or "Very Experienced". Similarly, 67% of respondents indicted that they were either "Experienced or "Very Experienced" with using the bookmark features of their browsers. No users replied that they were "Not Experienced" in regard to either question.

Given the results above, it can be assumed that the sample's level of experience with the concepts of web browsing and bookmarking web pages did not prevent participants from understanding the study's tasks.

Participants were not asked about their experience with collaborative tagging applications nor were they asked whether they used the Delicious application.

Number of Tags applied by Site

When the analysis of the data commenced, it was immediately clear that one of the sites from block 1 would need to be excluded from the analysis. The typetester.maratz.com website did not load properly on Internet Explorer. Therefore, the responses for this website were provided by a different population consisting of only participants that did not use Explorer. Since Internet Explorer users comprise a significant percentage of overall web users, it was determined that inclusion of the typetester data would significantly skew results. Therefore, it was excluded from the analysis.

Overall, 4,198 tags provided by 139 respondents were analyzed. The average number of tags applied by each participant to a site was 3.40. The mean number of tags

for the two cohorts was compared using a one-way analysis of variance (ANOVA). As

described in Table 2, only one site, ups.com, showed a significant difference (F = 9.843,

p = 0.002) in the mean number of tags provided by participants that did receive suggested

tags and those that did not.

Table 2. Average Number of Participant Tags by Site

	Cohort A	Cohort B	F	p
Block 1				
ups	3.07	4.02	9.843	0.002*
digg	3.15	3.61	2.214	0.139
cnet	3.49	4.02	2.419	0.122
craigslist	3.55	3.95	1.093	0.298
Block 2				
Ebay	3.19	3.15	0.015	0.904
flickr	2.97	3.11	0.18	0.672
forbes	3.83	3.26	1.801	0.182
microsoft	3.29	3.60	0.633	0.428
gizmodo	3.08	3.23	0.163	0.687

Note: Shaded cells indicate cases where tags were suggested. Asterisk indicates a statistically significant difference with/without suggested tags.

A one-way analysis of variance was also performed on the number of tags applied by

both cohorts (combined) to each website. The number of tags did not vary based on site

(F = 2.024, p = .041).

The smaller differences between the cohorts in the number of tags assigned in block

2 suggests that participants may have changed their tagging style between blocks of

websites. This conclusion is supported by a one-way analysis of variance that indicates

the difference in number of tags across the site blocks is statistically significant (F =

6.761, p = .009). However, it should be noted that the actual difference in means between

the two blocks of sites is relatively small (3.58 for block one and 3.27 for block 2). Thus,

the observed effect on the number of tags has little actual impact on the number of tags.

Tag Overlap

Because of the wide lexical variations in tags, the similarity comparison was performed by hand. Plural and singular variants on tags were coded as matches. However, other variants on the same stem term, such as "mail" and "mailing", were not collapsed as in these situations the tags were considered separate terms representing slightly different concepts. The exceptions to this rule were tags that had been collapsed when the initial list of suggested tags was created. For example on the Cnet list of suggested tags, "tech" and "technology" were collapsed as "technology". Therefore, both the "tech" and "technology" participant tags were matched to technology during the analysis. Misspelled words were matched with their intended counterparts in the suggested tags unless the misspelling resulted in a completely different term (e.g., "mail" was not matched with "male").

Tag Overlap with All Suggested Tags

The amount of overlap between each user's tags for each site and the entire list of suggested tags for the site was evaluated with analysis of variance. Participant tagging with and without suggested tags was compared such that the mean Jacquard similarity coefficient for cohort A was compared to the corresponding Jacquard similarity coefficient for cohort B. All of the sites in block one exhibited a statistically significant effect based on the presence or absence of the suggested tags while none of the sites in block two did (see Table 3).

Table 3. Tag Overlap with All Suggested Tags

	Cohort A	Cohort B	F	p
Block 1				
Ups	0.169	0.261	15.445	< 0.000 *
Digg	0.118	0.209	12.515	0.001 *
Cnet	0.162	0.245	9.085	0.003 *
craigslist	0.105	0.203	19.780	< 0.000 *
Block 2				
Ebay	0.212	0.169	3.170	0.077
Flickr	0.153	0.144	0.227	0.635
Forbes	0.248	0.194	4.709	0.032
microsoft	0.199	0.153	5.161	0.025
gizmodo	0.167	0.126	4.064	0.046

Note: Shaded cells indicate cases where tags were suggested. Asterisk indicates a statistically significant difference with/without suggested tags.

A one-way analysis of variance was also performed on the overlap between all of the suggested tags and the tags applied by each participant to each website (cohorts combined, to investigate the website effect). The amount of overlap differed significantly based on site (F = 6.003, p < .000). A post hoc analysis was performed using the Least Significant Difference test, revealing that ups.com, cnet.com, ebay.com and forbes.com had significantly greater tag overlap than flickr.com and gizmodo.com. Ups.com, cnet.com and forbes.com, had significantly greater tag overlap than craigslist.com. Ups.com and forbes.com had significantly greater tag overlap than digg.com. Finally, forbes.com had significantly greater tag overlap than microsoft.com. The mean overlap between all suggested tags and the tags applied to each website by the full sample are summarized in Table 4.

Table 4. Mean Overlap Between Participant's Tags and All Suggested Tags by Website

Website	Block	Mean Overlap, Both Cohorts
forbes	2	.224
ups	1	.211
cnet	1	.200
ebay	2	.193
microsoft	2	.178
digg	1	.159
craigslist	1	.150
flickr	2	.149
gizmodo	2	.149

A one-way analysis of variance was used to evaluate the effect of website block on the overlap between all of the suggested tags and the tags applied by each user. The effect of block on the amount of overlap was not statistically significant (F= .009, p = .926).

Tag Overlap with Top 5 Suggested Tags

As with the overlap data for all suggested tags, an analysis of variance was used to evaluate the overlap between participants' tags and the top five tags, comparing those cases where tags were suggested with those cases where they were not. Participants' tags for two sites in block one, digg.com and craigslist.com, were observed to based on the presence or absence of suggested tags while, again, no sites in block two exhibited statistically significant differences in tag overlap (Table 5).

Table 5. Tag Overlap with Top 5 Suggested Tags

	Cohort A	Cohort B	F	p
Block 1				
ups	0.142	0.176	6.090	0.015
digg	0.083	0.134	14.242	< 0.000 *
cnet	0.123	0.145	1.658	0.200
craigslist	0.129	0.240	17.773	< 0.000 *
Block 2				
ebay	0.149	0.139	0.441	0.508
flickr	0.102	0.101	0.006	0.938
forbes	0.168	0.146	1.427	0.234
microsoft	0.146	0.118	3.232	0.074
gizmodo	0.128	0.106	2.042	0.155

Note: Shaded cells indicate cases where tags were suggested. Asterisk indicates a statistically significant difference with/without suggested tags.

The lower Jacquard similarity coefficients were expected because there were fewer tags on which to match (i.e. five tags instead of all 10). However, all of the sites in both blocks continued to exhibit mean similarity coefficients that were higher in the cases where suggested tags were present.

A one-way analysis of variance was used to evaluate the main effect of website on the overlap between the top five suggested tags and the tags applied by each participant. The amount of overlap differed significantly based on site (F = 8.844, p < .000). To evaluate the specific difference in overlap by site, a post hoc analysis was performed using the Least Significant Difference (LSD) test. It indicated that ups.com, cnet.com, craigslist.com, ebay.com, forbes.com and microsoft.com had significantly greater tag overlap than flicker.com. Ups.com, craigslist.com, ebay.com and forbes.com had significantly greater overlap than digg.com. Ups.com, craigslist.com, and forbes.com had significantly greater tag overlap than gizmodo.com. Finally, craigslist.com had significantly greater tag overlap than cnet.com, ebay.com and microsoft.com. The mean

overlap between all suggested tags and the tags applied to each website by the sample are

summarized in Table 6.

Table 6. Mean Overlap Between Participant's Tags and Top 5 Suggested Tags by Website

Website	Block	Mean Overlap, Both Cohorts
craigslist	1	.179
forbes	2	.158
ups	1	.157
ebay	2	.144
cnet	1	.133
microsoft	2	.133
gizmodo	2	.118
digg	1	.107
flickr	2	.101

A one-way analysis of variance was also performed to examine the effect of website

block on the overlap between the top five suggested tags and the tags applied by each

user. The effect of block on overlap was not statistically significant (F = 4.815, p =

0.028).

Tag Overlap with Top 3 Suggested Tags

A comparison was done on the overlap between participants' tags and the top three

suggested tags using analysis of variance. The findings from this analysis were parallel to

those from the previous analyses. For two sites there were statistically significant effects

based on the presence of suggested tags. These two sites, digg.com and craigslist.com,

were both from block one. As observed during the previous comparisons, none of the

sites in block two exhibited any statistically significant effects based on the presence of

suggested tags (Table 7).

Table 7. Tag Overlap with Top 3 Suggested Tags

	Cohort A	Cohort B	F	p
Block 1				
ups	0.112	0.124	1.216	0.272
digg	0.082	0.128	12.892	< 0.000 *
cnet	0.106	0.108	0.020	0.888
craigslist	0.086	0.133	7.294	0.008 *
Block 2				
ebay	0.129	0.121	0.249	0.618
flickr	0.089	0.098	0.476	0.491
forbes	0.132	0.127	0.132	0.717
microsoft	0.107	0.082	4.226	0.042
gizmodo	0.104	0.090	1.210	0.273

Note: Shaded cells indicate cases where tags were suggested. Asterisk indicates a statistically significant difference with/without suggested tags.

As previously observed, the similarity coefficient continued to be higher for almost all sites when suggested tags were present. The one exception, flickr.com , may simply be a statistical anomaly but, since this trend is not statistically significant, it is not possible to demonstrate.

A one-way analysis of variance was performed to investigate the effect of website on the overlap between all of the suggested tags and the tags applied by each participant. The amount of overlap differed significantly based on site (F = 3.612, p < .000). A post hoc analysis was performed using the Least Significant Difference test, revealing that the overlap for the digg.com website was significantly greater than forbes.com while flickr.com, microsoft.com and gizmodo.com had significantly smaller overlap than ebay.com and forbes.com. The mean overlap between all suggested tags and the tags applied to each website by the sample are summarized in Table 8.

Table 8. Mean Overlap Between Participant's Tags and Top 3 Suggested Tags by Website

Website	Block	Mean Overlap, Both Cohorts
forbes	2	.129
ebay	2	.125
ups	1	.117
cnet	1	.107
craigslist	1	.107
digg	1	.103
gizmodo	2	.097
microsoft	2	.096
flickr	2	.093

A one-way analysis of variance was used to investigate the effect of website block on the overlap between the top three suggested tags and the tags applied by each user. The main effect of block on overlap was not statistically significant (F= .009, p = .926).

Satisfaction Measure

Mean satisfaction ratings for the concepts of *accuracy*, *completeness* and *perceived utility* were compared across cohorts using a t-test such that the same were evaluated for the same block of sites. For all six pairs of means, no statistically significant difference was observed between groups (Table 9).

Table 9. Satisfaction Ratings by Cohort

	Cohort A	Cohort B	t	p
Block 1				
Accuracy	2.107	2.045	0.554	.581
Completeness	2.187	2.279	-0.760	.448
Perceived utility	1.910	1.934	-0.167	.867
Block 2				
Accuracy	1.986	1.975	0.083	.934
Completeness	2.105	2.237	-0.872	.385
Perceived utility	1.909	1.903	0.044	.965

Note: Shaded cells indicate cases where tags were suggested. Asterisk indicates a statistically significant difference with/without suggested tags.

Discussion

The results of this study indicate that social proof may have an effect on tag selection in collaborative tagging systems. For several sites, the presence of suggested tags resulted in higher overlap between the Delicious tags and those assigned by study participants. These differences in overlap were observed for digg and craigslist when the top three or five Delicious tags were used in the comparison; and these two sites were joined by ups and cnet when all the Delicious tags were used in the comparison. For the remaining sites, there was a consistent trend of higher tag overlap for the participants who received suggested tags. This latter observation suggests that, if further work was done in this area with a larger number of subjects, statistically significant results might be observed across a larger group of sites.

The strong effect of the websites themselves on the overlap data may indicate that the results are highly dependent on the content of the websites. For example the two sites for which social proof had a statistically significant effect on tagging, digg.com and craigslist.com, were both sites primarily driven by user content. These sites also have no, or at least very little, editorial supervision and include content on a wide range of subjects. Therefore, it is possible that the reason that users were more likely to use the suggested tags was related to their confusion over how to categorize these relatively eclectic collections of content.

The other two sites in block one, cnet.com and ups.com, are, in contrast, produced by a single entity with complete editorial control over the content. Moreover, the content on each these sites is generally part of a single domain. ups.com's content focuses entirely on providing shipping-related services over the web while cnet.com's content is focused entirely on technology. Therefore, it is not surprising that, as fewer suggested

tags were compared with the participants' tags, little difference was observed between the groups that did and did not receive the suggested tags. In the case of the cnet, for example, the domain of technology was easy to identify and tag – even in the absence of guidance.

However, the homogeneity of the content of these sites does not entirely explain the results because, for some sites, tags with relatively high frequency were observed that were not members of the suggested tags. For example, the tag "package" was used more often by the sample than most of the suggested tags for the site ups.com (Table 10). Moreover, "package" had never been assigned to ups.com by any Delicious users at the time of this writing. The difference in frequency rank between the tags provided by the sample and those provided by the Delicious population is also interesting to note. As shown in Table 10, the tag "business", which is currently ranked second in frequency of use on Delicious, is one of the tags least used by the sample. However, the difference in the number of tags applied by the sample (479) compared to the Delicious community (155) almost certainly affects this result.

Table 10. Comparision of Frequency of "Package" Tag to Suggested Tag Frequency for ups.com

	Delicious	Study sample	
	Rank	Frequency	Proportion of total tags
shipping	1	91	26%
mail	5	54	15%
ups	3	48	14%
package	N/A	42	12%
delivery	8	30	8%
postal	9	25	7%
services	10	23	6%
tracking	4	22	6%
business	2	16	5%
companies	7	3	1%
businessandfinance	6	1	0%

Note: Shaded cell indicates tag used only by the sample. Table excludes other sample-only tags.

This difference in the tags used by Delicious users and this study's sample may indicate that the suggested tags were not valid for the sample because there was not sufficient similarity between the composition of the sample and the Delicious population. Future work should consider recruiting a sample from the population of Delicious users or creating a collection of sample tags specifically for the targeted user population.

The presence or absence of suggested tags was not observed to have any statistically significant effect on the satisfaction level of the sample. This suggests that social proof plays little role in users' satisfaction with their tags. One explanation for this lack of effect may be that users are not particularly emotionally invested in their tag choices because the direct effect of their choices is deferred until they actually try to retrieve bookmarks using those tags.

Study Limitations

The methods used for this study should be reexamined in some depth. The lack of significant results for the second block of websites (across all comparisons of both tag overlap and number of tags applied) strongly suggests that participants' behavior changed in some way between the two blocks of sites. The most likely possibility is that both cohorts learned from the experience of tagging the first block of sites and so were less influenced by the suggested tags as they gained more experience with tagging. However, there is currently little evidence from this study to support this conclusion because no statistically significant main effect was observed between the blocks of sites for the tag overlap measure. There was a statistically significant difference ($F = 6.761$, $p = .009$) observed between the two blocks of sites regarding the number of tags with block one having a slightly higher greater number of tags (Mean, block 1 = 3.58; Mean, block 2 =

3.27). Thus, both cohorts of the sample learned to use slightly fewer tags between blocks of sites.

Another methodological issue concerns the collapsing of tags. The rule of matching plural and singular participant tags both to the plural suggested tag, and matching of diminutives of tags based on the way the tags were collapsed on the suggested tag list were consistently applied. However, as the author began to find a wide range of variations on some of the tags, such as "pics" and "pix" for "pictures" used for tagging the flickr site, it became apparent that the simple rules used to match tags were not sufficient.

The rules for collapsing tag variations were selected specifically to focus the tag overlap analysis on the variations in terms rather than the lexical variations within a single term. However, Guy and Tonkin (2006) identified several issues with Delicious tags, such as the proliferation of both singular and plural versions of a term and the misspelling of tags. While these issues were deliberately excluded from consideration in this study, they are certainly important and should be considered in future work. Such work should consider alternative rules for comparing tags such as only comparing exact matches or using a more complex stemming algorithm.

In comparing tags, another problem was the lack of explicit delimiters between tags. Brooks and Montanez (2006) have previously observed the problems with a lack of delimiters. In several cases, it was obvious from the context that multiple tags had been intended to refer to a single concept such as the tag "wall street" for Forbes.com and "digital photos" for Flickr. However, the author was unwilling to make inferences about the *meaning* of the tags, so multiple word tags were evaluated as individual words unless

an explicit delimiter (such as "," or "/") was used.

Conclusion and Suggestions for Future Research

This study has shown that the principle of social proof can influence the tag choices of collaborative tagging system users such that suggested tags are more likely to be selected. The effect of social proof appears to be most pronounced for websites where the content is varied enough to be confusing to users. This latter observation is consistent with prior work on social proof that suggests uncertainty is a contributing factor (Cialdini 1993).

Since social proof's effect is heightened by uncertainty, suggested tags may mitigate the anxiety of new users of collaborative tagging systems. To further explore this idea, future work will need to examine the strong effect of website on tag overlap. If some content is particularly difficult to tag because the primary categories aren't readily apparent, then studies of this content should produce results similar to those observed in this study for craigslist.com and digg.com.

The convergence of the terms used as tags, encouraged by social proof, has the potential to improve the usability of collaborative tagging systems. First, users that select more common terms for their tags will improve the ability of others to find their content as it is categorized by the preferred term. Moreover, the users themselves will improve their ability to find content similar to their own as more similar content will identified by the same tag.

Bibliography

Abrams, D., Baecker, R. & Chignell, R.M.(1998). Information archiving with bookmarks: Personal web space construction and organization. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, April 18-23, (pp.41-48). New York: ACM Press.*

Alexa.com (2006) *Top 500 Websites*. Retrieved October 23, 2006 from http://www.alexa.com/site/ds/top_sites.

Aula, A., Jhaveri, N. & Käki, M. (2005). Information search and re-access strategies of experienced web users. *Proceedings of the 14th International World Wide Web Conference,* May 10-14, (pp. 583-592). New York: ACM Press.

Bailey, J.E. & Pearson, S.W. (1993). Development of a tool for measuring and analyzing computer use satisfaction. *Management Science*, 29 (5), 530-545.

Barreau, D. & Nardi, B.A. (1995). Finding and reminding: File organization from the Desktop. *SIGCHI Bulletin*, 27(3), 39-42.

Bernstein, M. (1988). The bookmark and the compass: Orientation tools for hypertext users. *ACM SIGOIS Bulletin archive*, 9(4), 34 - 45.

Boardman, R. & Sasse, M.A. (2004). Stuff goes into the computer and doesn't come out': A cross-tool study of personal information management. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, April 24-29, (pp. 583-590). New York: ACM Press.

Brooks, C.H. & Montanez, N. (2006). Improved annotation of the blogosphere via autotagging and hierarchical clustering. *Proceedings of the Fifteenth International World Wide Web Conference*, May 23-26, *(*pp. 625 – 632). New York: ACM Press.

Cialdini, R. B. (1993). *Influence: Science and practice (3rd Edition).* New York: HarperCollins College Publishers.

Cattuto, C. (2006). Semiotic dynamics in online social communities. *European Physical Journal C*, 42(s02), 33-37.

Del.icio.us (2006). *Company Information*. Retrieved September 3, 2006, from http://del.icio.us/about/info.

Gattis, L. F. (2002). Planning and information foraging theories and their value to the novice technical communicator. *Proceedings of the 20th Annual International Conference on Computer Documentation*, October 20-23, 2002 (pp. 39-43). New York: ACM Press.

Doctorow, C. (2002). Metacrap: Putting the torch to seven straw men of the meta-utopia. *E-Learning Guru.* Retrieved September 5, 2006, from http://www.e-learningguru.com/articles/metacrap.htm.

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

Guy, M. & Tonkin, E. (2006). Folksonomies - tidying up tags?. *D-Lib Magazine*, 12(1). Retrieved November 21, 2006, from http://www.dlib.org/dlib/january06/guy/01guy.html

Jacob, E.K. (2004) Classification and categorization: A difference that makes a difference. *Library Trends*, 52(3), 515-540.

Keller, R.M., Wolfe, S.R., Chen, J.R., Rabinowitz, J.L. & Mathe, N. (1997). A bookmarking service for organizing and sharing URLs. *Computer Networks & ISDN Systems*, 29(8-13),1103-1115.

Kome, S. H. (2005). Hierarchical subject relationships in folksonomies. Unpublished master's paper, University of North Carolina at Chapel Hill, Chapel Hill, NC.

LiveJournal (2005). *New feature: Tags*. Retrieved September 5, 2006 from http://news.livejournal.com/86492.html.

Ludicorp (2004). *Flickr launches*!. Retrieved Septemper 5, 2006 from http://www.ludicorp.com/news_item_display.php?id=36.

Nass, C. & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1), 81-103.

Pirolli, P. & Card S. (1995). Information foraging in information access environments. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* May 7-11, (pp. 51-58). New York : ACM Press.

Quintarelli, E. (2005). *Folksonomies: Power to the people*. Retrieved August 15, 2006 from the International Society for Knowledge Organization Italy-UniMIB Meeting Archive at: http://www.iskoi.org/doc/folksonomies.htm.

Smith, G. (2004). *Folksonomy: social classification*. Retrieved August 15, 2006 from Atomiq at: http://atomiq.org/archives/2004/08/folksonomy_social_classification.html.

Weiss, A. (2005) The power of collective intelligence. *netWorker* 9(3), 16-23.

Voss, J. (2006). Collaborative thesaurus tagging the wikipedia way. Retrieved October 23, 2006 from the Collaborative Web Tagging Workshop Archive available at: http://arxiv.org/abs/cs/0604036.

Appendix A – Selected Websites with Suggested Tags

1) www.ups.com

UPS.com is the corporate website for the United Parcel Service (UPS). Site functions focus on tracking shipments, ordering UPS supplies, and requesting pickup of packages.

Delicious Tags
1. shipping
2. business
3. ups
4. tracking
5. mail
6. businessandfinance
7. companies
8. delivery
9. postal
10. services

2) www.digg.com

Digg is a weblog that is focused primarily on science and technology stories. Content is submitted and rated by site users. Ratings determine where on the site an item appears.

Delicious Tags:
1. news
2. technology
3. blogs
4. daily
5. web2.0
6. links
7. digg
8. web
9. social
10. community

3) typetester.maratz.com

Typetester is a typographic tool for web designers. The tool provides a way to compare the readability and aesthetic appearance of fonts on web pages.

Delicious Tags
1. fonts
2. typography
3. webdesign
4. design
5. css
6. font
7. web
8. tools
9. webdev
10. development
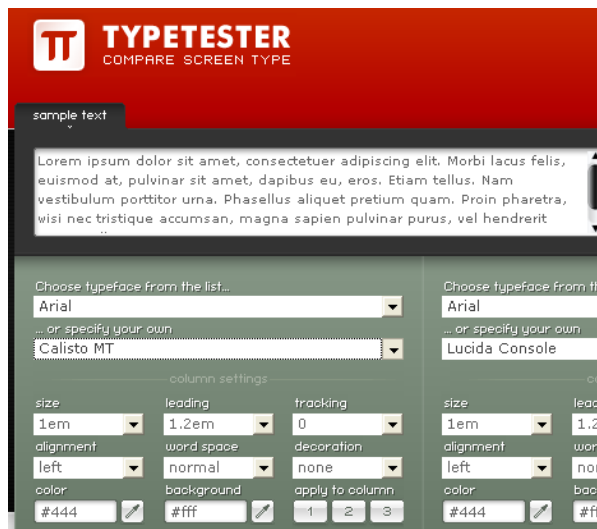
4) www.cnet.com

Cnetcom is a technology news site that also features reviews of computer hardware and software. A major component of cnet is the an archive of downloads including software demos and shareware downloads.

Delicious Tags:
1. technology
2. news
3. reviews
4. computers
5. gadgets
6. cnet
7. technews
8. software
9. shopping
10. reference
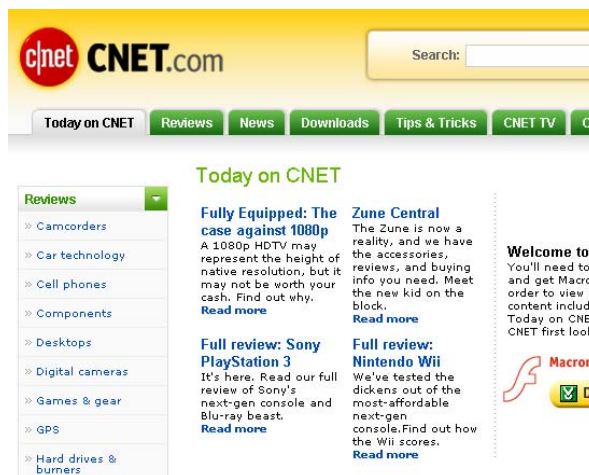
5) raleigh.craigslist.org

Craigslist is a classified ad site serving over 300 cities world-wide. The classifieds cover a wide range of topics.

Delicious Tags:
1. raleigh
2. classifieds
3. local
4. shopping
5. craigslist
6. community
7. reference
8. daily
9. durham
10. triangle

6) www.ebay.com

Ebay is an on-line auction and commerce site. Content primarily consists of listings for auction items.

Delicious Tags:
1. shopping
2. ebay
3. auctions
4. buy
5. search
6. store
7. sell
8. online
9. books
10. deals

7) www.flickr.com

Flickr is a photograph sharing site that organizes content using a collaborative tagging sytem.

Delicious Tags:
1. photography
2. flickr
3. sharing
4. web2.0
5. images
6. blog
7. web
8. community
9. social
10. pictures

8) www.forbes.com

Forbes.com is the website for the business magazine, Forbes. Like the print publication, Forbes.com's content focuses on business news.

Delicious Tags:
1. business
2. news
3. finance
4. magazines
5. investing
6. financial
7. money
8. forbes
9. economics

9) www.microsoft.com

Microsoft.com is the corporate website for the software developer, Microsoft, Inc. Content includes product information, support information and developer resources for Microsoft products.

Delicious Tags:
1. microsoft
2. software
3. windows
4. software_developers
5. computers
6. company
7. technology
8. os
9. office
10. downloads
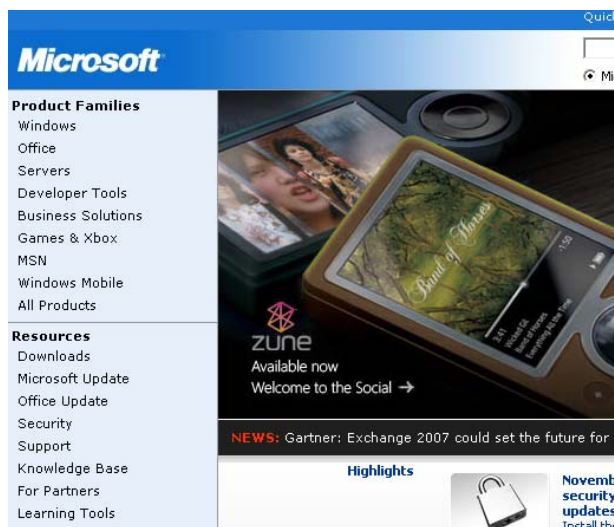
10) www.gizmodo.com

Gizmodo is a weblog that posts technology-related items including product reviews and news articles.

Delicious Tags:
1. gadgets
2. blogs
3. technology
4. news
5. daily
6. geek
7. hardware
8. cool
9. design
10. electronics

(Gizmodo image altered slightly for clarity)

Appendix B – Recruitment Email

Subject: Participants wanted for UNC Web Bookmarking Study - $25 prize

As a graduate research project, I am conducting a study on how to improve people's ability to share information about websites with others. Currently websites like http://del.icio.us and http://www.flickr.com allow people to share websites and pictures with one another. Users on these sites organize their information by keywords or "tags". What I am interested in is how people decide which tags to use.

Participants will be directed to a web survey where they will be asked to view a series of websites and answer some brief questions about them. The survey should take between 30 and 45 minutes to complete. Two participants will be randomly drawn to win $25 gift certificates from Student Stores.

Prior experience with tagging or a specific website is not necessary.

If you're interested in participating in this study, detailed information and instructions are provided with the web survey at {survey url removed}.

Please email Phil Binkowski at {email address removed}.unc.edu if you have questions about the survey.

Thanks,

Phil Binkowski
MSIS Candidate
School of Information and Library Science
{email address removed}

Barbara Wildemuth (advisor)
Professor
School of Information and Library Science
{email address removed}

All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the Institutional Review Board at 919-966-3113 or by email to IRB_subjects@unc.edu.

Appendix C – Survey Screen Shots

Figure 1. Tagging Instructions

Web Bookmarking Survey

Many people currently organize their web bookmarks into folders to make it easier for them find their bookmarks. Another way to organize bookmarks is to use a set of keywords or "tags" to label the bookmarks. Then the bookmark owner can find their bookmarks by looking at the list of tags they have used. One of the main differences between folders and tags is that many tags can be assigned to a single bookmark while typically a bookmark only gets filed in one folder.

Following you will see links to two groups of five websites. Please click on each of the links and examine the website for a few minutes (the website will appear in a separate window). After you view the site, please write down the tags that you think best represent the website - that is, the words that would be most meaningful to you if you wanted to find the site again. You can use as many tags or a few as you wish. For some of the sites you will be provided with suggested tags and for others you will not.

Example: www.unc.edu

Some Possible Tags: school college gradschool universities university

Example: www.google.com

Some Possible Tags: search google searchengine web internet tools

Figure 2. Block of Sites without Suggested Tags

Please review this site: [www.ups.com] and add the tags you think are appropriate.

Please review this site: [www.digg.com/] and add the tags you think are appropriate.

Please review this site: [typetester.maratz.com] and add the tags you think are appropriate.

Please review this site: [www.cnet.com] and add the tags you think are appropriate.

Please review this site: [raleigh.craigslist.org] and add the tags you think are appropriate.

Figure 3. Block of Sites with Suggested Tags

Please review this site: [www.ups.com] and add the tags you think are appropriate.

Possible Tags: shipping business ups tracking mail businessandfinance companies delivery postal services

Please review this site: [www.digg.com] and add the tags you think are appropriate.

Possible Tags: news technology blogs daily web2.0 links digg web social community

Please review this site: [typetester.maratz.com] and add the tags you think are appropriate.

Possible Tags: fonts typography webdesign design css font web tools webdev development

Please review this site: [www.cnet.com] and add the tags you think are appropriate.

Possible Tags: technology news reviews computers gadgets cnet technews software shopping reference

Please review this site: [raleigh.craigslist.org] and add the tags you think are appropriate.

Possible Tags: raleigh classifieds local shopping craigslist community reference daily durham triangle

Figure 4. Satisfaction Scales for Accuracy, Completeness, and Perceived Utility

For the next questions you will be asked to describe how you felt about applying the tags *to the five websites you just viewed*.

Please rate the tags you applied in terms of their accuracy or correctness:

Accurate				Inaccurate
○	○	○	○	○

High				Low
○	○	○	○	○

Consistent				Inconsistent
○	○	○	○	○

Sufficient				Insufficient
○	○	○	○	○

Please rate the tags you applied in terms of their completeness or comprehensiveness:

Complete				Incomplete
○	○	○	○	○

Consistent				Inconsistent
○	○	○	○	○

Sufficient				Insufficient
○	○	○	○	○

Adequate				Inadequate
○	○	○	○	○

Please rate the tags you applied in terms of their utility or their ability to serve your needs given the amount of time/effort required:

High				Low
○	○	○	○	○

Positive				Negative
○	○	○	○	○

Sufficient				Insufficient
○	○	○	○	○

Useful				Useless
○	○	○	○	○