Collaborative Tagging and Information Management: Influences on Tag Choices in del.icio.us

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

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

ABSTRACT

Collaborative tagging systems have the potential to produce socially constructed information organization schemes. However, their effectiveness depends on how users choose tags. Using data from del.icio.us, a popular collaborative tagging system for organizing web bookmarks, we quantitatively test three hypotheses concerning users' tag choices: 1) Users imitate other user's choices of tags, 2) Users choose tags from the organization of their own personal collection, and 3) Users choose tags recommended by the system. We find evidence for strong influence of a user's existing organization and little evidence to support the other hypotheses.

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces – measurement; H.5.3 Information Interfaces and Presentation: Group and Organization Interfaces – web-based computing, asynchronous interaction

Author Keywords

collaborative tagging, social bookmarking, information management, logistic regression, quantitative analysis

INTRODUCTION

User-contributed metadata, also known as *tagging*, is increasingly receiving attention as a tool for digital information management. Tagging provides a means for users to associate personally salient keywords or labels with content items, enabling them to find the content later via information they are predisposed to recognize or recall [5]. Tagging helps users package information for future information seeking and reuse [7]. Tagging has not only been applied to personal information management; many so-called *collaborative tagging* systems have appeared in recent years. Collaborative tagging systems such as del.icio.us and citeulike.org publicly expose individual users' associations between content items and tags, thereby providing visibility into words others have used to tag similar items. Grudin [3] suggests that collaborative tagging can be a low-effort solution for shared or group information management, because it does not require that users try to conform to a controlled vocabulary or organization scheme. However, in other shared information management contexts, research has shown that such effort is necessary for effective information reuse [7].

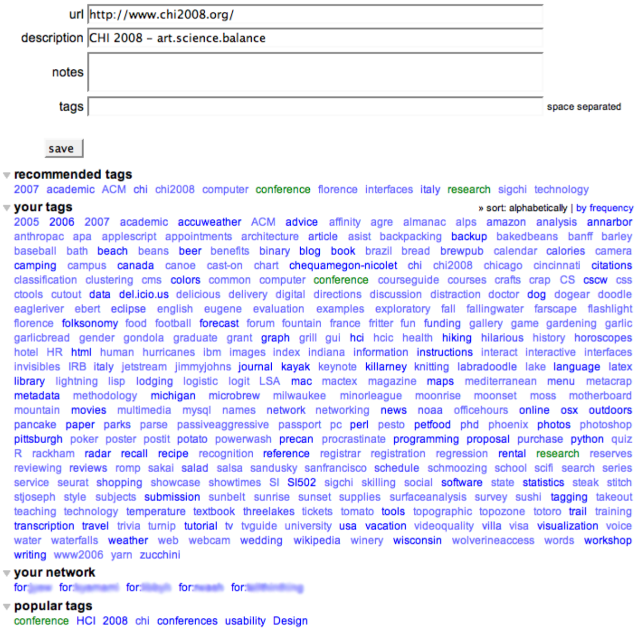
In a collaborative tagging system, users navigate large amounts of information via tags. Users interested in viewing content tagged a certain way by others can browse the system by clicking on tags. Tags provide the "information scent' [10] that connects users with information; tags are the infrastructure upon which information organization and seeking takes place. This has interesting consequences when one considers information retrieval effectiveness. If a given tag is applied in an inconsistent manner among many users, more variability exists in the content items displayed when a user browses to a particular tag [6]. For example, users tend to use high-level tags like "technology" and personal tags like "to read", as well as words like "apple" that can refer to a computer or a fruit, and "photos" or "pictures" which are synonyms. Influences and patterns in how users make tag choices could affect not only their own use of a particular tagging system for personal information management, but also impact the utility of the system as a whole for the information seeking of others [11].

In this paper, we focus on the social bookmarking website del.icio.us, as a case study of a collaborative tagging system supporting both personal and shared information management. del.icio.us is an online application that allows users to save and tag their own web bookmarks so they are accessible from any networked computer. It is an interesting case for several reasons. The bookmark and tag histories for over one million users are public and can be viewed (and analyzed) by anyone. del.icio.us has recently received attention in the research literature as the canonical example of a collaborative tagging system for information management [2] (in contrast with the photo sharing website flickr.com, which incorporates tagging but has a different overall purpose). Finally, other researcher suggests [2, 4] that a socially constructed shared vocabulary might emerge on del.icio.us.

We wanted to look for evidence of a social process affecting tag choice, using a quantitative analysis of tagging data from del.icio.us. Golder and Huberman [2] speculate that users might be imitating each others' tag choices; in other words, tag choices might be influenced by tags that had been previously applied to the same web page by other users. However, it is reasonable to assume that there might be other sources of influence on users' tag choices having to do with personal information management goals. For example, a user interested in del.icio.us only for organizing and re-finding their own bookmarks might strive for consistency within their own "controlled vocabulary", to maintain a shorter list of tags [11]. Or, users might desire to expend as little effort as possible when choosing tags, and simply select the tags the system recommends when they create a new bookmark.

If imitation or another social process is at work, we should be able to detect it by analyzing similarities between users' tag choices when bookmarking a new webpage, and all tags

CHI 2008, 5 April - 10 April, 2008, Florence, Italy.



url http://www.chi2008.org/

description CHI 2008 – art.science.balance

notes

tags space separated

recommended tags

2007 academic ACM chi chi2008 computer conference florence interfaces italy research sigchi technology

your tags » sort: alphabetically | by frequency

2005 2006 2007 academic accuweather ACM advice affinity agre almanac alps amazon analysis annarbor anthropac apa applescript appointments architecture article asist backpacking backup bakedbeans banff barley baseball bath beach beans beer benefits binary blog book brazil bread brewpub calendar calories camera camping campus canada canoe cast-on chart chequamegon-nicolet chi chi2008 chicago cincinnati citations classification clustering cms colors common computer conference courseguide courses crafts crap CS cscw css ctools cutout data del.icio.us delicious delivery digital directions discussion distraction doctor dog dogear doodle eagleriver ebert eclipse english eugene evaluation examples exploratory fall fallingwater farscape flashlight florence folksonomy food football forecast forum fountain france fritter fun funding gallery game gardening garlic garlicbread gender gondola graduate grant graph grill gui hci hcic health hiking hilarious history horoscopes hotel HR html human hurricanes ibm images index indiana information instructions interact interactive interfaces invisibles IRB italy jetstream jimmyjohns journal kayak keynote killarney knitting labradoodle lake language latex library lightning lisp lodging logistic logit LSA mac mactex magazine maps mediterranean menu metacrap metadata methodology michigan microbrew milwaukee minorleague moonrise moonset moss motherboard mountain movies multimedia mysql names network networking news noaa officehours online osx outdoors pancake paper parks parse passiveaggressive passport pc perl pesto petfood phd phoenix photos photoshop pittsburgh poker poster postit potato powerwash precan procrastinate programming proposal purchase python quiz R rackham radar recall recipe recognition reference registrar registration regression rental research reserves reviewing reviews romp sakai salad salsa sandusky sanfrancisco schedule schmoozing school scifi search series service seurat shopping showcase showtimes SI SI502 sigchi skilling social software state statistics steak stitch stjoseph style subjects submission sunbelt sunrise sunset supplies surfaceanalysis survey sushi tagging takeout teaching technology temperature textbook threelakes tickets tomato tools topographic topozone totoro trail training transcription travel trivia turnip tutorial tv tvguide university usa vacation videoquality villa visa visualization voice water waterfalls weather web webcam wedding wikipedia winery wisconsin wolverineaccess words workshop writing www2006 yarn zucchini

your network

for:[redacted] for:[redacted] for:[redacted] for:[redacted] for:[redacted]

popular tags

conference HCI 2008 chi conferences usability Design

Figure 1. Screen capture of the del.icio.us bookmark posting interface

previously associated with that webpage. If, however, users are influenced more by personal information management goals, their tag choices should be more similar to tags they had previously used when saving other bookmarks than tags previously associated with the webpage. We can test these predictions of the alternative hypotheses using the large amount of data available online, to determine whether one or more are significant determinants of tagging behavior.

METHOD

By default, bookmarks and tags in del.icio.us are public information. Each new bookmark has the following metadata associated with it: the username of the person saving the bookmark, the tags selected by that user, and the date and time the bookmark was created. Users browsing del.icio.us view subsets of bookmarks delimited by metadata such as a particular username, tag, or user-tag combination. For example, clicking the tag *library* in the list of popular tags on del.icio.us displays all webpages bookmarked by any del.icio.us user having the tag *library* associated with them. Clicking on a username displays webpages bookmarked by a particular person. The metadata for a given webpage can also be displayed including the usernames of all the users who bookmarked it, and all the tags ever associated with it. The Library of Congress home page has been bookmarked in del.icio.us by 1962 different users, and tagged "library" by 874[1].

When a user creates a new bookmark, the interface (Figure 1) displays *recommended tags* selected automatically by the system, *your tags* which are all tags chosen in the past by that user, and *popular tags* for that particular webpage.

The Dataset

Over two weeks in January 2007, we downloaded the entire bookmark and tag history for approximately 20,000 different webpages in del.icio.us. The webpages were chosen by periodically sampling the "recently posted" and "popular" del.icio.us pages. We randomly chose 30 webpages from our sample that had been bookmarked by at least 100 users. Then, in June 2007 we downloaded the complete public bookmark histories for all of the approximately 12,000 users who had ever bookmarked any of these 30 webpages. In other words, our dataset contains the complete tag histories for 30 webpages bookmarked in del.icio.us, as well as tag histories for all users who ever bookmarked any of those 30 webpages.

Model and Data Setup

We set up a logistic mixed model regression[1] to evaluate the influence of three predictors on users' tag choices:

1. Tags previously associated with a webpage by other users (the *imitation* hypothesis)
2. Tags a given user had applied before on other web pages at the time they bookmarked the web page (the *organizing* hypothesis)
3. Tags recommended by the system (the *recommended* hypothesis) [2]

If the *imitation* hypothesis has a strong influence, we can assume a social process is at work, and a socially constructed vocabulary is truly emerging. If tagging behavior is determined more by *organizing* than by *imitation*, then we might expect to see different tagging patterns. For example, word frequency counts follow a power-law frequency distribution (Zipf's Law) in a variety of contexts. It could be a fundamental statistical principle of language use; identifying what causes this pattern is an open research question in linguistics [9]. Finally, if the *recommended* hypothesis is true, users' tag choices are influenced by the del.icio.us recommended tags algorithm. In this case, either the algorithm is doing a great job of predicting users' tag choices, or people are lazy.

We model the dependent variable — the choice of a single tag — as a yes/no choice. Because we have no record of which potential tags a user considered and rejected, we make a simplifying assumption that the list of observations for each user consists of a yes/no choice for all tags applied to the particular webpage at the time our data was collected. We attempt to estimate the probability of saying "yes" to each tag as a function of three different factors included in the model as predictors. First, if *imitation* is shown to have strong influence on a particular tag choice by a particular user, then the probability that a tag is chosen should be higher if the word has been used previously as a tag. This would be reflected in the model as a large, positive coefficient for the "used.onsite" predictor. Second, if *organizing* is shown to have strong influence, the probability that a word is chosen should be higher if the word has been previously used by that user as a tag for a different webpage. This would be reflected by a large, positive coefficient for "used.byuser". For the *recommended* hypothesis, the recommendation algorithm is not publicly known; however, some experimentation with del.icio.us has led us to believe that a tag is much more likely to be recommended if it has both been applied previously to that webpage and used previously by the user. Therefore, we approximated the *recommended* hypothesis by including an interaction term that

[1] As of Sept. 16, 2007

[2] It is difficult to concretely specify this hypothesis because del.icio.us does not reveal its method for choosing recommended tags, and the method may have changed multiple times.

is 1 when both $\text{used.onsite} = 1$ and $\text{used.byuser} = 1$. We believe the interaction is a adequate proxy for the recommendation algorithm.

The model also includes several controls for other factors that may influence the probability of choosing a tag. Some tags seem to "fit" the webpage better than others (i.e., *library* for the Library of Congress home page). Since the data include repeated measures for each tag, it is important to control for per-tag variability using fixed effects. This is represented in the model by "tag_dummys". Finally, some users tend to assign more tags to their bookmarks than others. For example, one user in our sample always chose the maximum of ten tags for every bookmark, while another user always chose exactly one. Our data also include repeated measures for user, so we controlled for within-user variability using random effects. The model is set up as follows:

$$\begin{aligned}\text{tag\_chosen} = f(&\text{used.onsite}, \text{used.byuser}, \text{interaction},\\ &\text{tag\_dummys}, \text{random\_effect}(\text{user}))\end{aligned}$$

RESULTS

We estimated the model using maximum likelihood estimation, separately for each of the 30 sites in the study. This allowed us to compare webpages and determine whether an overall pattern exists.[3] We summarize the estimates for the model coefficients in Table 1.

Interpreting the Odds Ratios

In logistic regression, the dependent variable is dichotomous, meaning it takes only two possible values. The model is used to estimate the probability of the dependent variable taking on the value 1, given a set of predictors. This probability is represented in the form of *odds*. For example, a probability of 50% can be represented as 1:1 odds, and 2:1 odds translates to a 66% probability. The coefficients for the predictors in a logistic regression model are the natural logarithm of odds *ratios*, or the ratio of the odds of one possible outcome divided by the odds of another outcome. In the model, our predictors are dummy variables that can be either 1 or 0. Therefore, the coefficient represents the natural logarithm of the ratio between the odds that a tag will be chosen when the value of the predictor is 1 to the odds when the predictor is 0. If the coefficient is positive, then the probability of a tag being chosen is greater when the value of the predictor is 1 (or true). If the coefficient is negative, the probability of a tag being chosen is greater when the predictor is 0 (or false).

Consider the first row in Table 1, "A List Apart". For the tag *webdesign* as applied to this site, the odds are about 1:10 (9.1% probability) that an average user would choose *webdesign* as a tag if no one has used it on the site before, and that particular user hasn't used it as a tag for another bookmark[4]. The coefficient (log-odds ratio) for used.onsite is -0.1303, yielding an odds ratio of $e^{-0.1303} = 0.878$. Therefore, the odds of choosing *webdesign* if it has been used on this site before are 0.0878:1, or about 8.1%. (0.0878:1 / 1:10 yields an odds ratio of 0.878.) Since this coefficient is negative, the odds and the probability are decreased. Similarly, the odds ratio for the used.byuser coefficient is $e^{3.773} = 43.5104$. The odds of choosing *webdesign* if the user has previously used it to tag a different bookmark are approximately 4.35:1, or 81%. Finally, if the user has both used *webdesign* before, and it has been associated with this webpage before, the odds ratio is $e^{-0.1303+3.773-0.6507} = e^{2.992} = 19.93$. All three coefficients are included, because all three predictors are 1 (true). The odds that the user will choose *webdesign* are 1.993:1, or approximately 67%.

Hypotheses

The three hypotheses stated above can be operationalized in the model as follows:

1. Users choose tags by *imitation*, copying them from others who have bookmarked the webpage. $(\text{Used.onSite} > 0)$
2. Users choose tags for *organizing* their own collection of bookmarks, preferring tags they themselves have used in the past. $(\text{Used.byUser} > 0)$
3. Users choose the *recommended* tags that del.icio.us provides. $(\text{Interaction} > 0)$

A Wald test can be done on each parameter estimate, similar to the standard t-test used in Ordinary Least Squares (OLS) regression. It compares the Null hypothesis that the true value of the parameter is 0 with the alternative hypothesis that the parameter is not 0. The stars in Table 1 show the statistical significance of these Wald tests. The *imitation* hypothesis is only supported for one webpage at the 5% level. Although the Wald test for the used.onsite predictor is significant for 11 webpages, 10 have a negative parameter estimate, which does not support the *imitation* hypothesis. From this we reject Hypothesis 1. However, the *organizing* hypothesis is supported for 29 of the 30 webpages at the 5% level. The parameter estimates are generally quite high, indicating a strong effect. From this pattern of results, we conclude that Hypothesis 2 is supported. Finally, the *recommended* hypothesis is supported for 3 of the 30 webpages at the 5% level (the other 11 estimates that are significantly different than 0 are negative). However, because we are uncertain how well this proxy approximates the recommended tags algorithm in del.icio.us, we hesitate to draw any conclusions about this hypothesis. One webpage showed the same pattern of parameter estimates but none of these estimates could be established as statistically significantly different than zero, because the model fit produced very large standard errors for the estimates. This is a known problem in logistic regression when the data set is too sparse. We believe that the data matrix we have for this webpage is insufficient to extract reliable estimates.

When fitting a complicated model, it is important to compute some diagnostic goodness-of-fit statistics. In OLS regression, the F statistic is a statistical test that the model actually fits the data. Technically, it is a hypothesis test that the specified model fits the data better than the simplest possible model – the mean of the data. For logistic regression, the G_m statistic is analogous to the F statistic. It compares the specified model to the mode of the data, which is the simplest explanatory statistic for a binary variable. The G_m test is statistically significant at the 0.1% level for all 30 models. The OLS R^2 statistic represents how much of the variability in the data the model is able to explain. It is a substantive, rather than statistical test of significance. The R^2_L statistic is the logistic equivalent of R^2 [8], and represents the percentage of the likelihood explained by the model. For our models, R^2_L indicates that this model explains about 50% of

[3] Combining the data for all 30 sites into one large dataset proved computationally infeasible.

[4] The odds vary by tag. Due to space constraints, tag-specific results are not reported here

Title	*Users*	*Used.onSite*		*Used.byUser*		*Interaction*		$G_m(df)$			R_L^2
A List Apart: Articles: Alternative Style	395	-0.1303		3.773	***	-0.6507	**	5892	31	***	0.5378
London Underground History	369	-0.3896	*	3.132	***	-0.01643		6325	32	***	0.5028
Haiku — Desktop Operating System	161	-0.6901	*	2.466	***	0.953	*	2025	18	***	0.5306
affiliates homepage — Spread Firefox	214	-1.112	***	2.708	***	0.6126	*	2107	21	***	0.4528
PayPalSucks.com	122	-0.1500		3.33	***	-0.4669		1065	15	***	0.4587
OS X Maintenance And Troubleshooting	282	-0.6088	**	3.064	***	-0.1555		3633	26	***	0.4784
The Library of Congress	552	-0.4761	***	3.698	***	-0.2488	.	7561	37	***	0.4553
GDI+ FAQ main index	114	-0.2459		3.475	***	-0.5984		1271	19	***	0.4714
MetaGer	174	-0.3399		4.58		-1.191		1215	16	***	0.4123
eHomeUpgrade	270	-0.1015		3.608	***	-0.5326	*	3305	33	***	0.427
Getting started with SSH	938	0.0692		3.345	***	-0.4176	*	17976	41	***	0.5744
err.the_blog	457	0.434		3.684	***	-0.4777		7557	30	***	0.5562
Beer Advocate - Respect Beer.	489	0.04503		3.203	***	-0.2474		6737	24	***	0.5486
Old Computers	258	-0.2638		4.01	***	-0.6646	**	3651	26	***	0.4917
Snipplr - Code 2.0	1137	-0.05543		3.302	***	0.1691		26150	95	***	0.4949
DotNetNuke	714	-0.0992		3.672	***	-0.6925	***	11712	52	***	0.4837
BibDesk — Home	303	-0.3209	.	4.047	***	-0.4422	*	5921	33	***	0.5218
Tiny Icon Factory	819	-0.04774		2.878	***	0.5023	**	15580	56	***	0.5106
Mint: A Fresh Look at Your Site	560	-0.1584		3.543	***	-0.2467		10333	42	***	0.4787
101 Cookbooks	1200	-0.02729		4.288	***	-1.07	***	20459	43	***	0.6172
Telegraph newspaper online	447	-0.5028	**	4.256	***	-0.7313	***	4711	19	***	0.5316
GlimpsesThe Uncanny Valley	166	0.05604		2.928	***	0.1534		2174	34	***	0.3752
DVDStyler - Home	157	-0.8417	**	2.525	***	0.5488		2422	18	***	0.5122
digg labs / swarm	501	-0.3542	*	2.933	***	0.5093	**	9392	52	***	0.4884
Flickr: The HDR Pool	596	-0.3129		3.226	***	-0.3512		9000	27	***	0.5665
Sxip Identity	496	-0.2242		3.974	***	-0.8316	***	8009	37	***	0.4919
Many Eyes	466	0.3673	*	3.04	***	-0.1294		9219	52	***	0.4774
Obscure Sound - Indie Music Blog	116	-0.4379		3.067	***	0.1129		1047	13	***	0.5509
JotSpot Wiki (dojomanual)	218	0.01734		3.757	***	-1.134	**	3049	27	***	0.5128
BasKet Note Pads	124	-0.7433	**	3.225	***	-0.3192		2170	23	***	0.466

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘’ 1

Table 1. Logistic Regression Results

the likelihood.

DISCUSSION

The results in Figure 1 have a clear pattern; of our three explanatory variables, the strongest influence is users' previous tag choices. The coefficients on used.byuser consistently indicate a much larger influence than that of used.onsite or the interaction term. While user variability and individual tag 'fit' (represented by control variables in the model) play an important role in the choice of tags, the data indicate that users' desire for personal organization is also important.

This analysis also casts doubt on the imitating hypothesis and the recommended hypothesis. We were only able to detect an influence of these hypotheses in 1 and 3 sites, respectively, and in these instances the influence was small. If there is a social process at work promoting a socially constructed vocabulary, we doubt that it takes the form of direct imitation. We are less sure about the effect of recommendation because we do not have a compelling measurement of this explanatory variable.

We believe ours is the first quantitative study of how users of del.icio.us choose tags to compare these hypotheses from the literature. This work provides evidence that can be used to understand how users of del.icio.us choose words to use as tags.

REFERENCES

1. A. Agresti. *An Introduction to Categorical Data Analysis*. Wiley, second edition, 2007.
2. S. Golder and B. A. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2):198–208, 2006.
3. J. Grudin. Enterprise knowledge management and emerging technologies. In *HICSS '06*, 4-7 January 2006.
4. H. Halpin, V. Robu, and H. Shepherd. The complex dynamics of collaborative tagging. In *WWW '07*, 2007.
5. M. Lansdale. The psychology of personal information management. *Applied Ergonomics*, 19(1):55–66, 1988.
6. G. Macgregor and E. McCulloch. Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review*, 55(5):291–300, 2006.
7. L. M. Markus. Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success. *Journal of Management Information Systems*, 18(1):57 – 93, 2001.
8. S. Menard. *Applied Logistic Regression Analysis*. Quantitative Applications in the Social Sciences. Sage University Press, 2002.
9. M. E. J. Newman. Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46:323–351, 2005.
10. P. Pirolli. Rational analyses of information foraging on the web. *Cognitive Science*, 29(3):343–373, 2005.
11. R. Wash and E. Rader. Public bookmarks and private benefits: An analysis of incentives in social computing. In *To appear at ASIS&T '07*, 2007.