

2007

Why People Tag? Motivations for Content Tagging

Oded Nov

Polytechnic University, New York, onov@poly.edu

Chen Ye

University of Illinois at Chicago, cye1@uic.edu

Follow this and additional works at: <http://aisel.aisnet.org/sighci2007>

Recommended Citation

Nov, Oded and Ye, Chen, "Why People Tag? Motivations for Content Tagging" (2007). *SIGHCI 2007 Proceedings*. 13.
<http://aisel.aisnet.org/sighci2007/13>

This material is brought to you by the Special Interest Group on Human-Computer Interaction at AIS Electronic Library (AISeL). It has been accepted for inclusion in SIGHCI 2007 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Why People Tag? Motivations for Content Tagging

Oded Nov

Polytechnic University, New York
onov@poly.edu

Chen Ye

University of Illinois at Chicago
cye1@uic.edu

ABSTRACT

Tagging, or using keywords to annotate images, bookmarks, and blogs, is gaining much popularity. Since tagging is seen as an important change in the way images are organized and shared, we need to understand what drives this behavior. We draw on taxonomy of individual-level motivations for tagging, and research on the impact of social presence on tagging, and examine the drivers of tagging. We develop a scale of tagging motivations, which distinguishes between motivations stemming from three categories of intended audience: the taggers themselves, their family and friends, and the general public. Using multiple sources, including a survey and independent system data, we find that the levels of the Self and Public motivations, as well as social presence factor are positively associated with tagging level, and that the family & friends motivation is not associated significantly with tagging level. Implications of the research are discussed.

Keywords

Tags, motivations, photo sharing, social presence, Flickr.

INTRODUCTION

Tagging, or using keywords in order to add metadata to shared content (Golder and Huberman, 2006), is gaining much popularity in recent years (Cattuto et al., 2007; Golder and Huberman, 2006; Marlow, et al., 2006). Tags (e.g. “Christmas”, “talk”, “Canada”) are used to annotate various types of content, including images, bookmarks, and blogs, through web-based services such as Flickr, del.icio.us, and Technorati, respectively (Shneiderman et al., 2006). The popularity of tagging is attributed, at least in part, to the benefits users gain from effectively organizing and sharing very large amounts of information (Cattuto et al., 2007; Ames and Naaman, 2007).

Recently, researchers have focused their attention on image tagging in online communities such as Flickr. With more than 3 million users, who have uploaded more than 130 million photos (Quittner, 2006), Flickr is a prominent example of a collaborative photo sharing system. Tagging is seen as an important change in the way images are organized and shared, enabled by the transition from analog to digital photography (Shneiderman et al., 2006). Image tagging in Flickr is done by annotating them with

tags, or unstructured textual labels, mostly by the user who uploaded the images (Marlow, et al., 2006). These tags make the images searchable by the uploading user, as well as by others (Ames and Namman, 2007).

Each Flickr user has a photostream, which includes the images he or she uploaded, and he or she can make each image viewable by other users, or alternatively only by self, or by designated friends and family. In addition, each user can join any number of groups, which are normally formed around a shared area of interest (e.g., the Fishing group <http://www.flickr.com/groups/fishing/>). Images presented in the group’s Flickr page are the images shared by the group members, and in addition, there is an online discussion space available for members.

BACKGROUND

To understand what underlies tagging, we need to find out what motivates taggers. Research so far (e.g. Ames and Naaman, 2007; Wash and Rader, 2007) has focused on individual-level motivations. Other studies (e.g. Lee, 2006) have looked solely at the social level, focusing on the social presence as a driver of tagging without looking at individual-level motivations.

Individual-level motivations: In their study of the motivation for tagging on Flickr, Ames and Naaman (2007) draw the distinction between motivations stemming from three categories of intended audience of the tags. The categories are: *self*, *family & friends*, and the general *public* of Flickr users.

Within each category, another division is based on the function of the tags, or the tagger’s intended use. Here, Ames and Naaman distinguish between the function of *Organization* and the function of *Communication*. The *Organization* function involves organization and future retrieval of images, while the *Communication* function involves communication of additional context to viewers of the image.

The *Self* category involves the organization function, emphasizing organization and order, which are intended to facilitate future search and retrieval, and the communication function, which involves adding context to the image, for example, by tagging with the names of the people that appear in the photo, or the name of the place where it was taken.

The *Family & Friends* category also involves the organization function, which is intended in this case to facilitate future search and retrieval by friends and family. In addition, the category also involves the communication function, which in this case emphasizes adding context to the image not only by including names of people and places, but also by adding details known only to the intended viewers, inside jokes and nicknames.

The *Public* category involves the organization function, which is intended to help the general public of Flickr users find the images. It also involves the communication function, which in this case emphasizes signaling photographic abilities, giving the photographer the satisfaction of knowing that his or her photos are getting attention, and gaining reputation in the general Flickr community.

Social presence as a driver of tagging: According to social psychology research, behavior is affected by presence – actual, imagined, or implied – of others (Allport, 1968). The effect of social presence exists also when the presence was computer mediated (Savicki, et al, 1999). Moreover, perceived social presence was found to have a positive effect on tagging in del.icio.us, an online bookmark management system in which tagging is used extensively (Lee, 2006).

Much of the research on tagging motivations to date has been qualitative (e.g., Wash and Rader, 2007; Ames and Naaman, 2007), and therefore provides a useful conceptual background but no quantitative assessments of the motivations. Other, quantitative, research focused only on the social level (Lee, 2006). Moreover, no study so far used multiple, independent sources, and combined user reported data with system data.

What seems to be missing is a rigorous analysis of taggers' motivations, at both the individual and the social level, using independent sources such as taggers reports about their motivations, coupled with system data of their tagging behavior. This is the subject of the present study.

RESEARCH MODEL

Based on their qualitative analysis of Flickr user interviews, Ames and Naaman (2007) suggest that the *Public* category is the predominant motivation for tagging. Therefore, we would expect to find a strong correlation between this category and a user's tagging activity. Weaker motivations, according to Ames and Naaman (2007), are the *Self* and the *Family & Friends* motivations, and we expect to find a correlation between these motivations and the number of tags a user has.

Social presence is manifested on Flickr in group membership – when a user joins a group or adds people to his contact list, the user implicitly accepts that his images will be exposed to members of the group or the user's contacts, thereby leading to a perception of social presence, and possibly affecting the user's tagging behavior. Therefore, we expect to find a correlation

between the number of groups a user is a member of (a proxy for the user's level of perceived social presence) and the number of tags the user has.

Other potential drivers of tagging, which can serve as control variables, are the number of images a user has and the number of years a user has been on Flickr. We expect that the more photos a user has, and the longer he or she has been on Flickr, the more tags he or she will have.

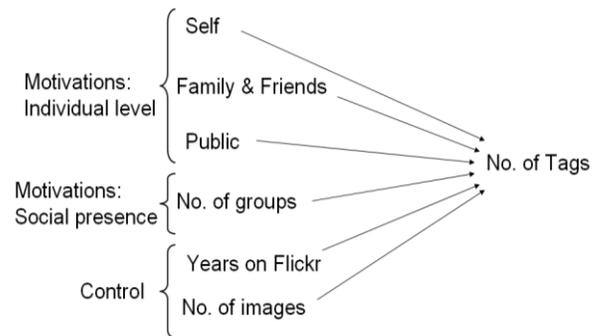


Figure 1. Research model.

METHODOLOGY

A web-based survey was administered to Flickr users, using a combination of user-reported data and independent system data of the actual tagging and image uploading behavior of the user.

To measure tagging motivations, we have developed a scale based on Ames and Naaman's (2007) qualitative work. The scale includes three constructs, representing the three categories of intended users of the tags as perceived by the user: *Self*, *Family & Friends*, and *Public*. For each construct, we have included items representing both the communication and the organization functions. All of the motivation items in the questionnaire were presented as statements to which taggers were asked to state how strongly they agree, on a scale of 1 to 7.

An initial set of fourteen items for each construct was formed based on the definitions of the three categories of motivations, and user responses reported in the interviews conducted by Ames and Naaman's (2007). To ensure face and content validity, these initial items were reviewed by three researchers familiar with tagging and scale development, and regular Flickr users. As a result, wordings of some items were revised. The next step involved a structured sorting exercise (Moore and Benbasat 1991). Eight individuals, including both researchers and lay persons, were asked to rearrange a randomized set of initial items, written on cards, into the categories they were intended to measure. Consequently, seventeen items were dropped either because more than two judges placed them in an unintended category, or because some judges considered them to be ambiguous. This gave us 8, 9, and 8 items for the *Self*, *Family & Friends*, and *Public* categories, respectively. In the next

step, a pilot study was carried out among randomly chosen Flickr users ($N = 193$) to validate the scale. An exploratory factor analysis using principle component analysis (PCA) was carried out and resulted in a three-factor solution. Items showing factor loading higher than 0.6 and cross-loadings lower than 0.4 were retained, and others were dropped. The retained items were then subject to another exploratory factor analysis which showed satisfactory factor loadings for all items. In addition, each of the three constructs showed at least a 0.8 Cronbach's alpha, indicating good reliability. The final scale contains

4, 6, and 6 items for Self, Family & Friends, Public, respectively, and was used in the survey. Table 1 contains examples of the questionnaire items used in the survey.

Since we are interested in tagging, only users with a minimum of ten tags listed on their Flickr pages were approached. This way, we tried to avoid getting data from users whose tagging was an isolated, unrepeated experience. In addition, we avoided approaching users who tagged in languages other than English, to ensure that respondents understand the questions.

<i>Motivation</i>	<i>Item</i>
Self (organization)	"When I tag my Flickr photos I make it easier for myself to find my photos later."
Self (communication)	"When I view my photos on Flickr and try to recall where and when they were taken, it's useful if I tagged them earlier."
Family & Friends (organization)	"When my friends or family search in my Flickr photos, it's easier for them if I tagged these photos earlier."
Family & Friends (communication)	"Tagging is a way for me to describe my photos to my friends or family."
Public (organization)	"When I tag my photos I make it easier for other Flickr users to find my photos."
Public (communication)	"Tagging is a way for me to provide details about my photos to other Flickr users."

Table 1. Examples of motivations and questionnaire items

System data, such as number of photos or tags per user, is available via Flickr's API (Application Programming Interface) system upon log in. The Flickr API allows third party websites to communicate with Flickr and exchange information. Respondents were asked, at the end of our web-based survey, to log in via the survey website to their Flickr account. This way, Flickr data about the respondents who logged in was automatically extracted using the Flickr API and recorded together with the respondents' responses to the questionnaire.

To measure social presence we used the number of groups to which a user belongs, as reported by the respondents. As for control variables, the number of photos a user has is extractable via the Flickr API and is therefore an independent, system generated measure. The number of years on Flickr, on the other hand, is not available via the API and therefore the respondents were asked to report it as part of the questionnaire.

One methodological issue in interpreting results from survey studies is common method bias (Straub et al. 2004). As suggested by Straub et al. (2004), when independent variables are measured using perceptual scales, one way to avoid common method bias is to measure the dependent variable using objective data. In our study, users' tagging level was measured using system log data retrieved directly through the Flickr API, and therefore, common method bias should not be an issue in interpreting our results. This is one of the main strengths of our study.

Randomly chosen 1205 Flickr users were emailed an invitation to participate in our web-based survey. A total of 208 valid responses were received, representing a 17.3% response rate. 55.1% of the respondents were male, and their median age was 32.

RESULTS

To ensure the validity of our measures on the final sample, we conducted a principle component analysis (PCA) with varimax rotation using SPSS. The PCA produced a three-factor solution, as expected. The three-factor solution explained 68% total variance in the PCA. Table 2 presents the mean, standard deviation, and factor loading of each measurement items.

To further assess factor validity, we also calculated the average variance extracted (AVE) for each measure (Fornell and Larcker 1981). As illustrated in Table 3, each factor has an AVE above the .50 threshold, and the square root of AVE is higher than the correlation with other factors, demonstrating discriminant and convergent validity (Chin 1998; Straub et al. 2004). In addition, all constructs have Cronbach's alpha values above 0.70, indicating satisfactory reliability (Straub et al. 2004).

		Mean	SD	1	2	3

1	SELF1	5.41	1.51	0.807		
	SELF2	5.49	1.46	0.834	0.327	
	SELF3	5.30	1.51	0.841		
	SELF4	5.11	1.55	0.796	0.314	
2	FF1	5.39	1.33		0.790	.300
	FF2	5.47	1.24		0.750	
	FF3	5.30	1.43		0.705	
	FF4	5.04	1.60		0.684	
	FF5	5.38	1.32		0.793	
	FF6	5.40	1.39	0.335	0.724	
3	PUBLIC1	5.87	0.99			0.784
	PUBLIC2	5.70	1.25		0.342	0.729
	PUBLIC3	5.64	1.23	0.319	0.339	0.685
	PUBLIC4	5.86	1.19			0.838
	PUBLIC5	5.97	1.04			0.883
	PUBLIC6	5.87	1.08			0.771

Table 2. Item Means, Standard Deviations, and Factor Loadings

Note: factor loadings below .300 are suppressed

We conducted a hierarchical linear regression to test our model. As expected, it was found that the levels of the Self and Public motivations, as well as the social presence factor (i.e. the number of groups), and the number of photos and years on Flickr, were positively associated with tagging level (see Table 4). Moreover, as expected, it was found that the Public motivation was strongest, the Self motivation weaker, and the Family & Friends motivation not associated significantly with tagging level.

Construct	Mean	SD	α	1	2	3
1. Self	5.33	1.31	.90	.820		
2. Family & Friends	5.31	1.10	.88	.607*	.742	
3. Public	5.75	0.98	.87	.278*	.367*	.784

Table 3. Construct Means, Standard Deviations, Reliability, Intercorrelations, and AVE

The diagonals are the sq. root of the AVE of each factor

** Significant at the 0.01 level, two-tailed test

CONCLUSIONS

Given the growing popularity of tags as means of facilitating the organization and sharing of large amounts of information (Cattuto et al, 2007), collaborative content sharing systems such as Flickr, or YouTube may benefit from understanding what motivates users to tag. To understand why users tag, we looked at individual level motivations, using a newly-developed scale, based on the work of Ames and Naaman (2007), as well as a social presence driver (the number of groups). We controlled for the number of photos users have and number of years they

have using Flickr. A strength of the present study is the use of data from multiple, independent sources. The data included survey responses as well as system log data retrieved directly through the Flickr API, to measure the dependent variable. This enabled us to overcome potential common method bias (Straub et al. 2004).

The preliminary findings of this research in progress suggest that, as expected, both social presence and individual level motivations affect users' tagging level, with the exception of the Family & Friends motivation. The latter finding should be viewed in light of Ames and Naaman's (2007) qualitative work, which suggests that the Family & Friends category would be a relatively weak motivation since family and friends follow the user's images anyway. The social presence driver was found to be stronger predictors of users' tagging, which comes as no surprise given the collaborative, public nature of websites such as Flickr. The number of photos a user has is also a predictor of tagging level, as expected.

Assuming that the correlations found also involve causality, it is advised that managers of collaborative content systems seeking to increase tagging activity focus their communication and marketing efforts on those factors that have a strong impact on the level of tagging. For example, the motivation of tagging photos for public users who are not friends or family has a positive effect on tagging level. Therefore, it may make sense for organizers of content systems to focus their cultivation efforts in this area, by highlighting to such users the possibility of being exposed to other, unknown users.

In line with findings from previous research on other collaborative systems (Lee, 2006), social presence proved to have a positive effect on tagging in the present study. It

would therefore make sense for organizers of content-sharing systems to focus efforts in this area as well, by exposing users to the benefits of being in a group and encouraging users join groups that focus on users' interests. In addition, the social presence effect on tagging lends support to the recommendation for designing

collaborative systems in such ways that they provide opportunities for social presence in order to boost tagging.

This research is still in progress and we plan to continue our data collection, so that we can develop and test our research model with a larger sample size.

		Step 1			Step 2		
		B	t	p	β	t	p
Results of individual predictors	Independent Variables						
	Control						
	Constant		2.352	.020		-2.865	.005
	Years on Flickr	.125	1.906	.058	.087	1.759	.080
	Number of Images	.326	4.980	.000	.284	5.764	.000
	Motivations: Individual level						
	Self	-	-	-	.137	2.223	.027
	Family & Friends	-	-	-	-.076	-1.190	.235
	Public	-	-	-	.152	2.903	.004
Motivations: Social presence							
	Number of Groups	-	-	-	.588	12.054	.000
Results of the overall model	R ²	.134			.528		
	Adjusted R ²	.126			.514		
	F	15.939 (df = 2, p < .001)			37.654 (df = 6, p < .001)		

Table 4. Hierarchical Linear Regression Results

A larger sample would also allow us to conduct further data analyses, including factor validation and model testing, using Structural Equation Modeling tools. Further research may also be helpful in understanding how different motivations influence contribution in different content sharing systems. The present study, addressing one of the prominent collaborative content sharing systems, is hopefully a useful step in this direction.

ACKNOWLEDGMENTS

The authors thank Gudmundur Kristjansson for helpful research assistance.

REFERENCES

- Allport, G. (1968) The Historical Background of Modern Social Psychology. In Lindzey G. & Aronson, E. (Eds.) *The Handbook of Social Psychology* Addison-Wesley.
- Ames, M. and Naaman, M. (2007) Why We Tag: Motivations for Annotation in Mobile and Online Media, *Proceedings of CHI 07*, San Jose, CA.
- Cattuto C, Loreto V, Pietronero L. (2007) Semiotic dynamics and collaborative tagging, *Proceedings of the National Academy of Sciences* 104, 1461-1464.
- Chin, W. (1998) The Partial Least Squares Approach for Structural Equation Modeling. In Marcoulides, G. (Ed.) *Modern Methods for Business Research*, Lawrence Erlbaum.
- Fornell, C. and Larcker, D. (1981) Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research* 18, 39-50.
- Golder, S., and Huberman, B. (2006) Usage patterns of collaborative tagging systems, *Journal of Information Science* 32, 2, 198-208.
- Lee, K. (2006) What Goes Around Comes Around: An Analysis of Del.Icio.U.S as Social Space, *Proceedings of the 2006 Conference on Computer Supported Cooperative Work*, Banff, Canada.
- Marlow, C., Naaman, M., Davis, M. and Boyd, D. (2006) Tagging Paper, Taxonomy, Flickr, Academic Article, To Read, *Proc. of the 17th ACM Conference on Hypertext and Hypermedia*. Odense, Denmark.
- Moore, G. and Benbasat, I. (1991) Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation, *Information Systems Research* 2, 3, 192-222.
- Quittner, J. (2006) Picturing the Web's Future, *Time Magazine* 167, 19, p.121.
- Savicki, V., Kelley M and Oesterreich, E. (1999) Judgments of Gender in Computer Communication, *Computers in Human Behavior*, 15, 2, 185-194.
- Shneiderman, B., Bederson, B., and Drucker S. (2006) Find That Photo! Interface Strategies to Annotate, Browse, and Share, *Communications of the ACM* 49, 4, 69-71.
- Straub, D.W., Boudreau, M.-C., and Gefen, D. (2004) Validation Guidelines for IS Positivist Research, *Communications of AIS*, 13, 380-427.
- Wash, R. and Rader, E. (2007) Public Bookmarks and Private Benefits, *American Society for Information Science & Technology Annual Meeting*, Milwaukee, WI.