

Tagging, Sharing and the Influence of Personal Experience

Chei Sian Lee, Dion Hoe-Lian Goh, Khasfariyati Razikin, Alton Y.K. Chua

Division of Information Studies, Wee Kim Wee School of Communication & Information

Nanyang Technological University, Singapore

Email: {leecs, ashlgoh, khasfariyati, altonchua}@ntu.edu.sg

Abstract

Social bookmarking or tagging is the process of assigning and sharing among users, freely selected terms to resources. This approach is a form of user-generated metadata and allows users to locate new resources through the collective intelligence of other users. Social tagging offers a new avenue for resource discovery as compared to taxonomies and subject directories created by experts. While social tagging has its advantages, one possible drawback is that tag creators, who come with different preferences, experiences, and beliefs, among other factors, may view the same document differently and therefore apply different tags even though they may have the same goal of content organization and sharing. In this paper, we argue that familiarity is an important issue to be investigated in social tagging systems, and our goal is to examine the influence of the level of familiarity with social tagging on the effectiveness of tags for content sharing. We found that high familiarity with the concept of tagging, Web directories, and social tagging systems are significantly and positively associated with high tag effectiveness for content sharing. Implications of our findings and opportunities for future work are discussed.

Keywords

Social tagging, Web 2.0, user study, familiarity, effectiveness

1. Introduction

Social computing or Web 2.0-based applications are empowering users to create, publish and share content such as text, images, video and other multimedia. As this new avenue for content-creation becomes increasingly popular, the resulting information explosion requires new techniques to manage, search and access such content.

Social tagging is one such approach for managing and discovering content on the Web, document repositories or digital libraries. It allows users to annotate links to useful resources by assigning uncontrolled keywords (tags), facilitating their future access by the tag creator (Macgregor & McCulloch, 2006). Put differently, tags are a form of user-generated metadata used to organize content, and because they may be shared by other users of the social tagging system, they in effect provide an alternative way of discovering and accessing content apart from search engines and subject directories. In addition, through tags, a user can potentially locate like-minded users who hold interests in similarly-themed resources, leading to the creation of social networks (Lee, 2006). An example of a social tagging system is the popular site

del.icio.us, which is used for tagging and sharing of Web sites. Besides such purpose-built applications, social tagging has also been used in blogs, wikis, media sharing, document repositories and digital libraries because they have become an accepted way of managing and discovering content. Examples include Flickr, last.fm, Connotea, and YouTube.

Advocates of social tagging have argued that this user-centered approach to content organization and sharing has advantages over conventional methods of categorization such as those based on taxonomies and controlled vocabularies. For example, hierarchical taxonomies may in some instances be too rigid to organize resources that contain a diversity of topics, and the non-hierarchical nature of tags might be better suited for this purpose (Morville, 2005). In addition, because traditional classification methods tend to rely on specialists such as trained catalogers to organize and describe information, they may use terms that are specific to a specialized community, resulting in under-accessed resources (Bowker & Star, 1999). In contrast, rather than depending on experts to categorize resources, tags harness the tacit knowledge of possibly large numbers of ordinary people (Lakoff, 1987). This presumably better describes resources in such a way that users are able to find relevant information more effectively. Tags are therefore also known as “folksonomies”, short for “folk taxonomies”, suggesting that they are created by lay users, as opposed to domain experts or information professionals.

On the other hand, other researchers have contended that there are shortcomings to social tagging for content sharing. First, the counter-argument of the flexibility afforded by free keyword assignment is the resulting ambiguity of tags due to a lack of controlled vocabulary (Macgregor, & McCulloch, 2006). This leads to the problem of vocabulary mismatch (Furnas et al., 1987) between tag creators and users due to the intrinsic properties of natural language that includes polysemy and synonymy (Golder & Huberman, 2006; Olsen et al., 1998). For example, the tag “java” may refer to an island in Indonesia, a programming language or coffee (polysemy), while a document about cars could be tagged as “car”, “automobile” and “vehicle” (synonymy). Further, documents may be tagged with subjective or ego-centric terms (e.g. “cool”, “todo”, “me”, “toread”) that have meaning only for the tag creator or a select few within a group of users. Next, tags may sometimes be driven by the tag creator’s self-serving agenda (Chua, 2003), lead to problems such as tag spamming where popular but unrelated tags are deliberately used to attract traffic to certain Web sites (Koutrika et al., 2007). Taken together, these issues may hinder the effective use of tags for organizing and sharing content.

As the use of social tagging continues to grow in popularity, there is an emerging body of research that explores their effectiveness for content organization and sharing. For example, work has been conducted on examining the ability of tags to classify blogs using text categorization methods (Sun et al., 2007), investigating the effectiveness of tags to classify Web resources in del.icio.us (Razikin et al., 2007), and comparing the use of tags against author assigned index terms in academic papers (Kipp, 2006). From a user perspective, work has also been conducted on motivations on behind tagging (Ames & Naaman, 2007), and on tagging dynamics and usage (Farooq et al., 2007). While such work has shed much light on tag effectiveness, a key component that has been relatively under-researched pertains to individual characteristics of the tag creator and their influence on the choice of tags used. Clearly, people with different preferences, experiences, and beliefs, among other factors, may view the same

document differently and therefore apply different tags even though they may have the same goal of content organization and sharing (Sen et al., 2005).

In this paper, we focus on a user's familiarity with social tagging and its influence on the choice of tags applied to a Web document. Users who have high familiarity with social tagging are considered as experienced tag creators, and they are likely to select tags that come from a common vocabulary shared by a community of users in a social tagging system through a process of learning and exploring of the content and interacting with other users (Golder & Huberman, 2006). In contrast, novice tag creators (i.e. low level of familiarity with social tagging), because of their unfamiliarity with tagging concepts or the community of users, may apply inappropriate terms or those that have meaning only to themselves (Marlow et al., 2006) despite having intentions of sharing content. These arguments find support in information retrieval and seeking literature that show differences between novices and experts in selecting terms for browsing or executing searches (Borgman, 1996; Sutcliffe et al., 2000). Taken together, we argue that familiarity is a fundamental issue to be investigated in social tagging systems, and our goal is to examine the influence of the level of familiarity with social tagging on the effectiveness of tags for content sharing. The remainder of this paper is organized as follows. In the next section, research related to this work is reviewed. A description of our experimental methodology and the results are then presented. We then provide a discussion of the implications of our findings and conclude with opportunities for future work in this area.

2. Related Work

The increase in popularity of social tagging on the Web, document repositories and digital libraries has correspondingly attracted much research from a variety of perspectives including the architecture and implementation of systems (e.g. Hammond et al., 2005; Puspitasari et al., 2007), usage patterns in tagging systems (e.g. Golder & Huberman, 2006; Marlow et al., 2006), user interfaces (e.g. Dubinko et al., 2006; Li et al., 2007), and the use of social tagging in digital libraries and information retrieval systems (e.g. Farooq et al., 2007; Yanbe et al., 2007) among others. Here, we focus our review on related literature that investigates effectiveness of tags was a means for organizing and sharing content.

Studies have been conducted that compare tags with controlled vocabularies to determine how tags differ from keywords assigned by experts. Lin et al. (2006) evaluated tags from Connotea and Medical Subject Heading (MeSH) terms and found that there was only 11% similarity between MeSH terms and tags created by the users. The authors argued that this is because MeSH terms serve as descriptors while tags primarily focus on areas that are of interest to users. Similarly, Kipp (2006) compared tags with author supplied tags from Cite-U-Like and indexing terms from INSPEC and Library Literature to determine the overlap in terms of usage. Results showed that approximately 21% of the tags were the same as the indexing terms. The reason for the divergence between tag creators and indexers was attributed to the different emphases placed on an article by these two groups. For example, tag creators may consider time management information (e.g. "todo", "toread", "maybe") to be important as a tag for articles to indicate a desire to read them in the future, while such information will be disregarded by expert indexers. Taken together, these findings suggest that experts and users (non-experts) employ vocabularies

that have little overlap, potentially causing access problems not only in systems using controlled vocabularies (e.g. MeSH), but possibly in systems employing social tagging as well.

Effectiveness has also been studied using automated, machine learning approaches. An early work done on automatic text categorization in social tagging systems was conducted by Brooks and Montanez (2006) using blog articles. The authors used 350 popular tags from Technorati and from these, obtained 250 of the most recent articles from the collected tags. Clustering was done on these articles using TFIDF and the cosine similarity measure. Results of the clustering procedure indicated that tags were able to organize articles in the broad sense, but not as effective in indicating the specific content for an article. Sun et al. (2007) focused on the effectiveness of tags in classifying entire blogs. The dataset consisted of 52709 descriptions and 161 tags obtained from BlogFlux. Automatic text categorization using Support Vector Machines (SVM) was adopted and the study compared the classification results of the blogs based on tags only, tags and the description of blogs, and descriptions only. It was found that tags and descriptions had the best classification results in terms of precision, recall and F-measure values, while tags alone were effective at classifying than using blog descriptions alone. In short, the results suggest that that tags can help users find relevant information.

Apart from blogs, Razikin et al. (2008) studied the effectiveness of tags to classify Web content in del.icio.us. The corpus consisted of 100 tags and 20210 documents. Using SVM, experiments were run on two feature sets: document terms only, and document terms plus tags. Surprisingly, results indicated that using document terms only produced better classification results in terms of F-measure than using terms plus tags. Nevertheless, both F-measures from the experiments were relatively low at 0.59 and 0.56, suggesting that not all tags were effective at content discovery, and that the performance of the SVM classifier was likely to be influenced by the tag creator's motivations, and his/her interpretation of the document content. Next, Levy and Sandler (2007) investigated social tags as a source for metadata to describe music. Using 236974 tags collected for 5722 tracks from last.fm and MyStrands, a Correspondence Analysis was performed to visualize a two-dimensional semantic space defined by the tags. Findings from their work suggest that tags were effective in capturing music similarity, and could be used to describe mood and emotion in music.

While the above studies focus on tags and their usage on the aggregate, to the best of our knowledge, little work has also been conducted employing a user-centric approach that examines individual characteristics of users and their influence on effective tag use. For example, a model of tag vocabulary evolution in the MovieLens recommender system was proposed by Sen et al. (2005). The model consisted of three constructs: personal tendency (e.g. experiences, knowledge, preferences), community influence (e.g. behavior of other tag users), and the tag recommendation algorithm. A transaction log analysis was conducted on 3366 users, 3263 tags and 11443 tag applications within a one month period. Among the findings relevant to our present study, the authors found that an individual's habit of applying a certain set of tags, and investment in creating his/her own personal tag vocabulary would influence the types of tags created in the future. As well, the tags created by the community in a social tagging system would also influence an individual's choice of tags such that a gradual alignment between the user's and the community's vocabulary takes place.

While our present study shares the goal of investigating tag effectiveness with the above work, our research is differentiated in several important ways. Firstly, the studies by Lin et al. (2006) and Kipp (2006) were limited to scholarly articles while Brooks and Montanez (2006) and Sun et al. (2007) used blogs. Here, the context of the medium of communication differs in our study. The purpose of an academic article is to disseminate information in a formal and objective manner, and typically caters to a limited audience. In contrast, blogs contain commentaries and sentiments, cater to a more diverse readership, and offer a wider variety of topics. The pages that are tagged in del.icio.us are diverse, and not limited to ordinary Web pages, but also includes blogs and academic articles. While the work by Razikin et al. (2008) employed del.icio.us data, our approach to investigating effectiveness differs as well. In particular, the literature that we have reviewed analyzed effectiveness given a collection of tags and documents, and individual characteristics of users were not considered. In this sense, our present study bears similarity with the MovieLens work (Sen et al., 2005) in which personal tendencies were investigated. However, that study did not consider the influence of familiarity, and employed a transaction log analysis methodology which falls short of providing an understanding of users' perceptions, intentions, satisfaction, and behavior, among other factors (Lau & Goh, 2006). Taken together, the present study is therefore timely as we adopt a survey methodology to study the influence of familiarity on the effectiveness of tags in content sharing.

Familiarity is often described as the experience with the *what*, *who*, *how* and *when* of what is happening (Gefen et al., 2003). It also reduces social uncertainty through increased understanding of what is happening in the present (Luhmann, 1979). Hence, past studies have indicated that familiarity plays an important role in people's judgments and decision-making (e.g. Winter, 1973; Johnston & Hawley, 1994; Garcia-Marques & Mackie, 2001; Smith et al., 2006). In particular, Winter (1973) argued that the more familiar people were with their materials, the more positive outcomes and attitudes were expected. Within social psychology research, multiple studies have indicated that people tend to spend less effort to process familiar messages than the same message they encountered for the first time (e.g. Claypool et al., 2004; Garcia-Marques & Mackie, 2001). Similar findings were also reported in educational research. Specifically, students were found to spend less effort to solve highly similar to previously presented problems as they tend to try to retrieve the answer from their memory (Reder & Ritter, 1992).

When applying the concept of familiarity in our current context of social tagging, we contend that users' familiarity with different concepts of social tagging is likely to influence how users select keywords which ultimately affect their effectiveness for content sharing. Golder and Huberman (2006) noted that users choose tags based on their experiences with other tagging systems, their interests and knowledge, as such the choice of tags by new users are likely to be different from the more experienced ones. Further, Sen et al. (2005) also elaborated that as users interact with the tagging systems, their preferences for choosing tags may change. Hence, these studies indicate that users who have high familiarity with social tagging may choose tags differently from users who are less familiar with social tagging, and thus affect the effectiveness of the tag for content sharing. However, to our knowledge, none of the studies has specifically examined the influence of individual's familiarity on tagging outcomes. Specifically, the present study aims to explore the following research question: *What is the nature of the relationship between familiarity and the effectiveness of tags for content sharing?*

3. Method

3.1. Selection of Tags

The present study utilizes the same del.icio.us dataset from Razikin et al. (2008). Eight tags (i.e. “free”, “funny”, “java”, “3d”, “interesting”, “re”, “library”, “economics”) from the dataset were randomly selected based on their diverse characteristics and representativeness (Razikin et al., 2008). Using these eight tags, three random documents associated with each tag from the del.icio.us dataset were selected, resulting in an evaluation sample of 24 documents. There were no overlaps in the 24 documents selected. In other words, each selected document could only be mapped to one of the eight tags. The documents were then randomly grouped into four different sets, where there were six documents in each set.

3.2. Participants

A total of 262 students from a large university in Singapore participated in the study. Participation was voluntary and anonymous, and required the completion of a questionnaire. Students, both undergraduates and postgraduates, participated in the study. The first part of the questionnaire consisted of questions relating to the demographic profiles, familiarity with regards to social tagging, search engines, Web directories, and social tagging systems. Demographic information of the participants is presented in Table 1.

Table 1. Demographic Profile of Participants (N=262).

1.	Age	Count (Percentage)
	18-20	110(42.0)
	21-25	91(34.7)
	26-30	35(13.4)
	31-35	18 (6.9)
	>35	8(3)
2.	Gender	Count (Percentage)
	Female	155 (59.2)
	Male	107(40.8)
3.	Type of Participants	Count (Percentage)
	Full-time Students	189 (72.1)
	Working Professionals (i.e. part-time students)	73 (27.9)

The second part of the questionnaire consisted of questions relating to the selected documents and their associated tags. Specifically, each participant was randomly assigned to a set of six documents as described above. For each document, four tags were listed and participants were asked to select the most appropriate tag for that document given its content. Figure 1 shows an example of a document that was meant to be tagged as “3d”. Since none of the documents had been tagged by more than one of the four listed tags, we ensured that there was only one correct option out of the four given.

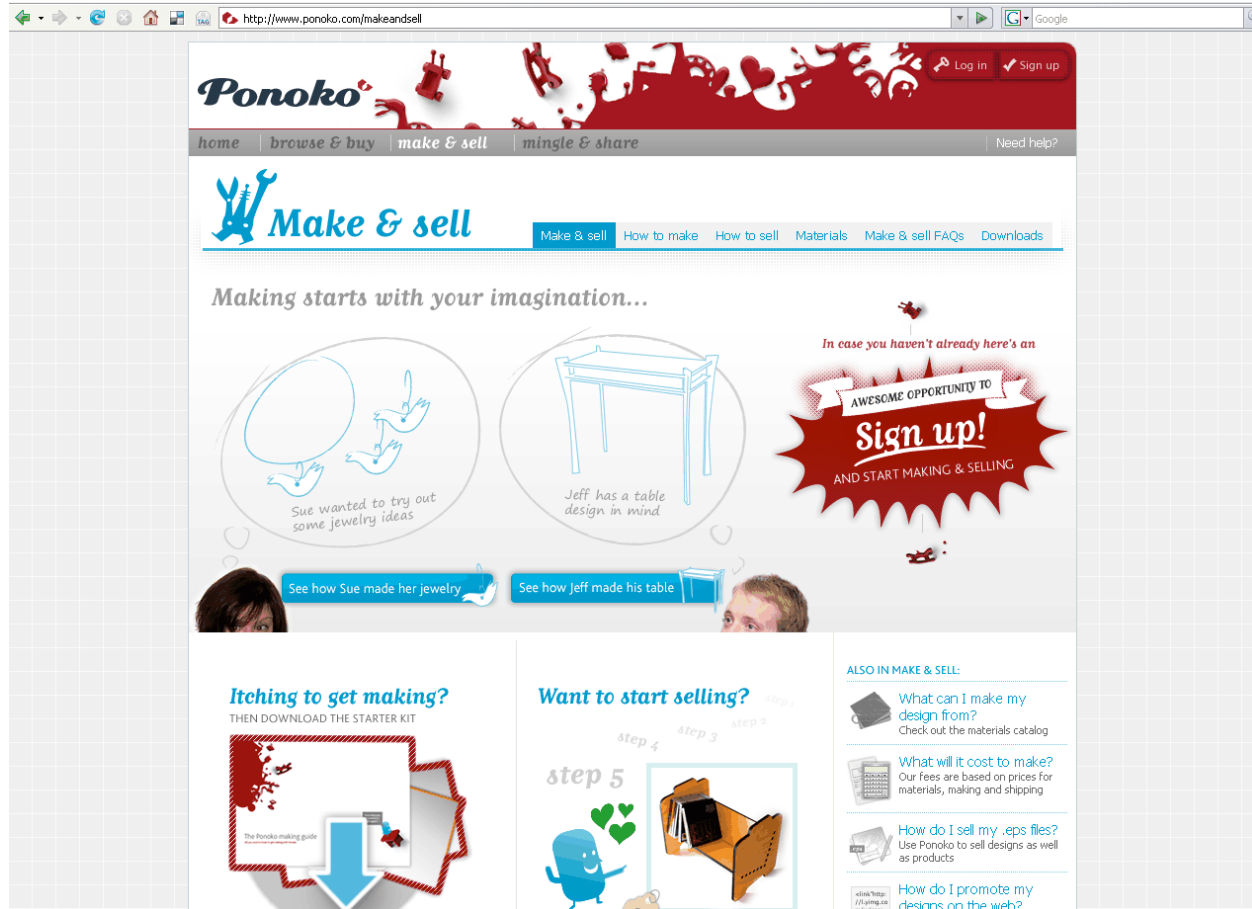


Figure 1. A document tagged with “3d” by the tag creator.

3.3. Tag Effectiveness for Content Sharing

We measured the effectiveness of tags for content sharing by the number of correct tags indicated by the participants. Specifically, a correct tag means that the answer given by the participants agree with the original tag on the selected document. Such agreement indicates that tags could be potentially used to share the selected document. Since all participants had to answer six questions, the maximum score for tag effectiveness was six (i.e. participant answered all the 6 questions correctly), and the minimum score for tag effectiveness was zero (i.e. participant answered all the six questions incorrectly).

3.4. Familiarity

We examined four dimensions of familiarity which include familiarity with the concept of tagging, search engines, Web directories and social tagging systems. Participants assessed their familiarity with these four items on a scale of one to five in increasing level of familiarity. A scale of one indicates a complete ignorance of the concept. A scale of five on the concept of tagging indicates the frequent use of tags to annotate and access Web content. A scale of five on search engines denotes the ability to use query terms, various techniques and advanced search functionalities on a variety of search engines. Likewise, a scale of five on Web directories

indicates the ability to browse through Web directories as part of an information seeking strategy. A median split was carried out on these three variables to divide participants into either high familiarity or low familiarity with social tagging, search engines, and Web directories. Lastly, the participants indicated their familiarity with social tagging systems by indicating if they used social tagging systems (e.g. del.icio.us, Cite-U-Like, etc.) High familiarity refers to users who have visited one or more social tagging systems while low familiarity refers to users who have not visited any social tagging system at all.

4. Results and Analyses

4.1. The Effect of Familiarity

A series of ANOVA tests was conducted to explore the relationship between familiarity and tag effectiveness for content sharing (shown in Table 2). Specifically, we wanted to explore if the number of correct tags was significantly different between users with high familiarity and low familiarity with the concept of social tagging, search engines, Web directories, and social tagging systems. The statistical analysis was conducted using SPSS. Our results indicated that the level of familiarity with the concept of social tagging ($p < 0.01$), Web directories ($p < 0.05$) and social tagging systems ($p < 0.01$) had significant effects on the number of correct tags. However, the level of familiarity with search engines was not significant. Stated differently, our results showed that high familiarity with the concept of social tagging (mean = 3.42), Web directories (mean = 3.48), and social tagging systems (mean = 3.32) were significantly positively associated with the number of correct tags. We also note that the means for the number of correct tags were significantly lower for low familiarity with the concept of social tagging (mean = 2.98), Web directories (mean = 3.09) and social tagging systems (mean = 2.5).

For the purpose of comparison between correct and incorrect tags, we also ran a series of ANOVAs using the number of incorrect tags as the dependent variable (see Table 3). As expected, familiarity with the concept of social tagging, Web directories, and social tagging systems had significant positive effects on the number of incorrect tags while the effects of familiarity with search engines were not significant. We note that low familiarity with the concept of social tagging (mean = 3.02), Web directories (mean = 2.91), social tagging systems (mean = 3.5) were significantly positively associated with the number of incorrect tags. It should also be noted that the means for number of incorrect tags were significantly lower for high familiarity with the concept of social tagging (mean = 2.58), Web directories (mean = 2.52), and social tagging systems (mean = 2.68). The means plot for the number correct tags and incorrect tags across the four familiarity dimensions are shown in Figures 2 and 3 respectively.

Table 2. Differences between high and low familiarity on number of correct tags.

Dependent Variable = Number of Correct Tags
High Familiarity (HF) Vs Low Familiarity (LF)

Familiarity with Tagging			Familiarity with Search Engines			Familiarity with Web Directories			Familiarity with Social Tagging Systems		
HF Mean (S.D.)	LF Mean (S.D.)	F-value	HF Mean (S.D.)	LF Mean (S.D.)	F-value	HF Mean (S.D.)	LF Mean (S.D.)	F-value	HF Mean (S.D.)	LF Mean (S.D.)	F-value
3.42 (1.26)	2.98 (1.24)	7.39**	3.35 (1.18)	3.13 (1.37)	1.18	3.48 (1.12)	3.09 (1.35)	6.02*	3.32 (1.24)	2.5 (1.43)	7.12**

*p<.05; **p<.01;

Table 3. Differences between high and low familiarity on number of incorrect tags.

<i>Dependent Variable = Number of Incorrect Tags</i> High Familiarity (HF) Vs Low Familiarity (LF)											
Familiarity with Tagging			Familiarity with Search Engines			Familiarity with Web Directories			Familiarity with Social Tagging Systems		
HF Mean (S.D.)	LF Mean (S.D.)	F-value	HF Mean (S.D.)	LF Mean (S.D.)	F-value	HF Mean (S.D.)	LF Mean (S.D.)	F-Value	HF Mean (S.D.)	LF Mean (S.D.)	F-value
2.58 (1.26)	3.02 (1.24)	7.39**	2.65 (1.18)	2.87 (1.37)	1.18	2.52 (1.12)	2.91 (1.35)	6.02*	2.68 (1.24)	3.5 (1.43)	7.12**

*p<.05; **p<.01;

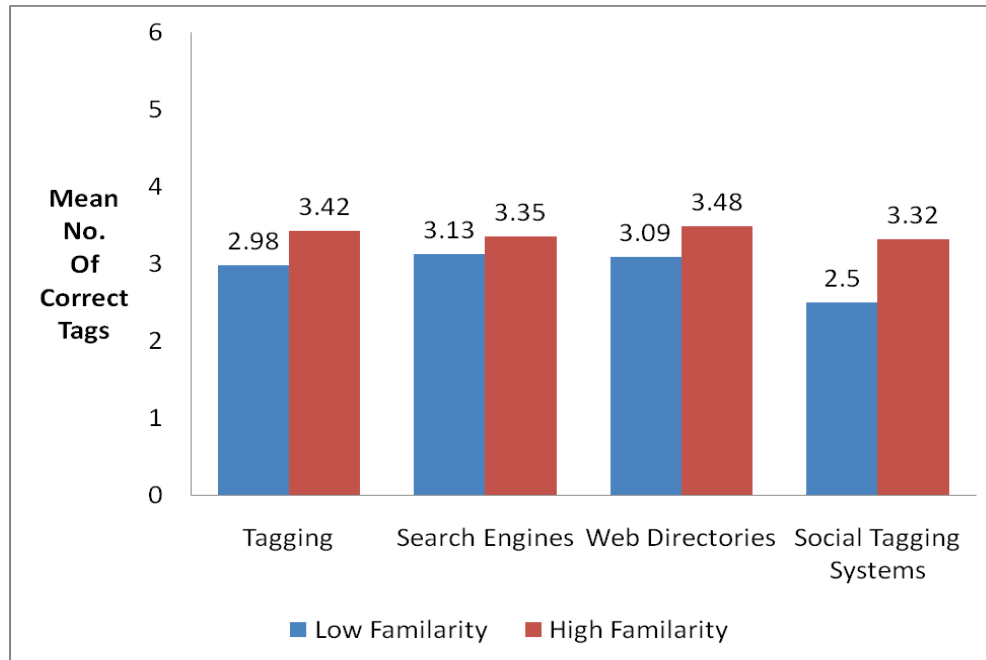


Figure 2. Mean number of correct tags versus levels of familiarity.

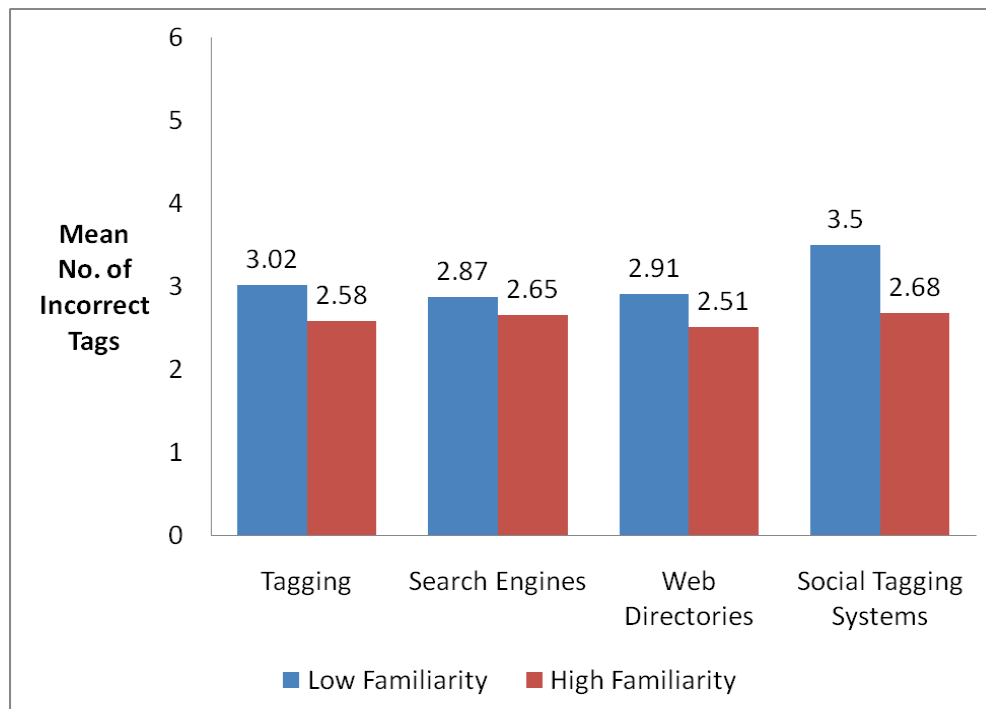


Figure 3. Mean number of incorrect tags versus levels of familiarity.

4.2. Discussion

To reiterate, this research aims to explore the research: *What is the nature of the relationship between familiarity and the effectiveness of tag for content sharing?*

Overall, our study shows that people with high familiarity (i.e. concept of tagging, Web directories, and social tagging systems) perform better in terms of creating effective tags for content sharing. Apparently, this does not seem to fully support the Wisdom of Crowds theory which suggests that the quality of tags created by a community is thought to be better than that provided by an expert (Surowiecki 2004). Specifically, our overall finding suggests that experts (i.e. high familiarity) are likely to perform better than novices (i.e. low familiarity) in terms of using more effective tags for content sharing. However, we want to emphasize that this finding does not necessarily mean that our study does not agree with the notion of “folksonomies” (i.e. created by lay users, as opposed to domain experts or information professional). Rather, our findings suggest that there are different types of lay users and in particular, lay users who are less familiar with different aspects of social tagging may not be suitable as tag creators as the tags they apply may be less effective for content sharing (Sen et al., 2005).

Not surprisingly, we found that people with high familiarity with the concept tagging and social tagging systems tend to be associated with identifying effective tags for content sharing. This implies that the more tagging experience individuals accumulate, the better they will be in terms of coming up with tag descriptors which others agree and understand. This could be due to the fact that familiarity with tagging helps individuals to increase their knowledge of common and less ambiguous vocabulary terms, enabling them to use better descriptors (Macgregor & McCulloch 2006). Additionally, our findings also suggest that such familiarity is likely to help individuals avoid the pitfall of choosing subjective and ego-centric tags (e.g. ‘toread’, ‘me’, ‘todo’) that have meaning only for the tag creator or a select few within a group of users. Finally, high familiarity could also mean that such users better understand the tagging community and therefore align their vocabulary with the larger group (Lee, 2006).

Interestingly, we found that familiarity with search engines has little effect on the effectiveness of the tags for content sharing whereas familiarity with Web directories has a significant effect on the quality of the tags. We argue that searching using search engines and creating tags have different objectives and require different skill sets even though both are related to information access. In particular, the experience gained using Web directories may help to enhance a set of skills that will be useful for creating tags. That is, when users are creating tags, the focus is on organizing the information, and possibly sharing the content with other users. Here, the experience gained from browsing Web directories and the familiarity with the way information is organized systematically may be helpful in guiding the creation of tags for sharing content (Garcia-Marques & Mackie, 2001). In contrast, when users employ search engines, the terms used in queries are meant to access and retrieve information (e.g. Ellis, 1989; Marchionini, 1995). The notion of using these terms as tags for content organization to facilitate future access may therefore not be present.

5. Conclusion

The objective of this study is to investigate the nature of the relationship between familiarity and the effectiveness of tags for content sharing. We found that high familiarity with tagging, web directories and social tagging systems are positively associated with higher tag effectiveness for content sharing. Conversely, low familiarity with tagging, Web directories, and social tagging systems are positively associated with lower tag effectiveness for content sharing. Hence, our results suggest that familiarity with social tagging plays an important role in the creation and usage of effective tags for content sharing.

Our research has provided several important contributions. One of the major contributions is in highlighting the relevance of the notion of familiarity among users towards systems that employ social tagging, including document repositories and digital libraries. This finding has important practical implications for such systems as well. Specifically, to reduce the number of poor quality tags or to curb tag spamming that impede content sharing, social tagging systems may want to restrict tag creation to more experienced users who are familiar with the concept of tagging rather than allowing any new and inexperienced users to create tags. Alternatively, if such a restriction is not practical, other options could be considered. The first involves introducing some form of online training for new users to familiarize them with the different aspects of social tagging. The second option involves a semi-automated tagging approach in which the system analyzes a resource such as a Web page and suggests possible tags, but leaving the user the freedom to make his/her own selections (e.g. Farooq et al., 2007; Muller, 2007). A third alternative could employ reputation mechanisms (Wathen & Burkell, 2002) for tags and/or tag creators to assist users in making a decision on whether to use a particular tag or not. Next, to the best of our knowledge, no other studies have attempted to examine the influence of individuals, particularly on the different levels of familiarity, on tagging outcomes. As discussed, research has tended to focus on the usage of tags on the aggregate. Here, we provide a different direction and employ a user-centric approach that examines individual characteristics of users and their influence on effective tag use. Thus, from this work, we have taken an important first step to expand our knowledge on the relevance of an individual's influence on tagging effectiveness for content sharing.

This is an ongoing work that provides opportunities for future research. First, we did not separate tags into different categories. For example, Golder and Huberman (2006) proposed categories such as tags that identify what the content is about, those that identify who owns the content, those that are task organizing, and so on. It is likely that effects of familiarity may be different for different categories of tags. Hence, an important area of research will be to expand this study to examine the effects of familiarity on the outcomes of various categories common in social tagging systems. Next, our study assumed that the tags collected were effective for content sharing, but research suggests that tag creators have many motivations for tagging documents (e.g. Ames & Naaman, 2007; Thom-Santelli et al., 2008) that may not be apparent to others or effective for content sharing. Future work may want to focus on collecting tags that are created by experts for the purpose of content sharing and used that as the benchmark for tag effectiveness. Nevertheless, we contend that since the tags in our dataset are publicly available to other users, they have the potential to be used for effective content sharing. Additionally, it is likely that such motivations for tagging documents are likely to be related to users' risk perceptions (Lee et al. 2007) of the system, other users, or documents. Here, a potential area of future work could attempt to focus on tags that have been explicitly targeted for public access

and examine the relationship between tag motivations and users' risk perceptions. Finally, this study employs a cross-sectional design. Ideally, such research should be conducted by using longitudinal designs with time lags appropriate to the variables involved to better capture the level of familiarity as new users become more experienced as they spend more time interacting with the systems and other users. Hence, an interesting line of future research is to study the tagging patterns of actual users of a social tagging system over time.

Acknowledgements

This work is partly funded by NTU research grant number RG25/05.

References

- Ames, M. and Naaman, M. (2007) "Why we tag: Motivations for annotation in mobile and online media". In *Proceedings of the 2007 SIGCHI Conference on Human Factors in Computing Systems* (New York: ACM Press), 971-980.
Available at: <http://www.dlib.org/dlib/april05/hammond/04hammond.html>.
- Borgman, C. L. (1996) "Why are online catalogs still hard to use?" *Journal of the American Society for Information Science*, Vol. 47, No. 7, 493-503.
- Bowker, G. C. and Star, S. L. (1999) *Sorting Things Out: Classification and Its Consequences* (Cambridge, MA: MIT Press).
- Brooks, C. H. and Montanez, N. (2006) "Improved annotation of the blogosphere via autotagging and hierarchical clustering". In *Proceedings of the 15th International Conference on the World Wide Web* (New York: ACM Press), 625-632.
- Chua, A. (2003) "Knowledge sharing: A game people play". *Aslib Proceedings*, Vol. 55, No. 3, 117-129.
- Claypool, H. M., Mackie, D. M., Garcia-Marques, T., McIntosh, A., and Udall, A. (2004) "The effects of personal relevance and repetition on persuasive processing". *Social Cognition*, Vol. 22, No. 3, 310-335.
- Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P., and Tomkins, A. (2006) "Visualizing tags over time". In *Proceedings of the 15th International Conference on the World Wide Web* (New York: ACM Press), 193-202.
- Ellis, D. (1989) "A behavioral approach to information retrieval design". *Journal of Documentation*, Vol. 45, No. 3, 171-212.
- Farooq, U., Kannampallil, T. G., Song, Y., Ganoë, C. H., Carroll, J. M., and Gilles, C. L. (2007) "Evaluating tagging behavior in social bookmarking systems: Metrics and design heuristics". In *Proceedings of the 2007 International ACM Conference on Supporting Group Work* (New York: ACM Press), 351-360.
- Furnas, G.W., Landauer, T.K., Gomez, L.M., and Dumais, S.T. (1987) "The vocabulary problem in human-system communication". *Communications of the ACM*, Vol. 30, No. 11, 964-971.
- Garcia-Marques, T. and Mackie, D. M. (2001) "The feeling of familiarity as a regulator of persuasive processing". *Social Cognition*, Vol. 19, No. 1, 9-34.
- Gefen, D., Karahanna, E., and Straub, D. W. (2003) "Trust and TAM in online shopping: An Integrated Model". *Management Information Systems Quarterly*, Vol. 27, No. 1, 51-90.
- Golder, S. A. and Huberman, B. A. (2006) "Usage patterns of collaborative tagging systems". *Journal of Information Science*, Vol. 32, No. 2, 198-208.

- Hammond, T., Hannay, T., Lund, B., and Scott, J. (2005) "Social bookmarking tools (I)". *D-Lib Magazine*, Vol. 11, No. 4.
- Johnston, W. A. and Hawley, K. J. (1994) "Perceptual inhibition of expected inputs: The key that opens closed minds". *Psychonomic Bulletin and Review*, Vol. 1, No. 1, 56-72.
- Kipp, M. E. (2006) "Exploring the context of user, creator and intermediate tagging". Paper presented at the *7th Information Architecture Summit*. Available at: http://www.iasummit.org/2006/files/109_Presentation_Desc.pdf
- Koutrika, G., Effendi, F. A., Gyöngyi, Z., Heymann, P., and Garcia-Molina, H. (2007) "Combating spam in tagging systems". In *Proceedings of the 3rd international Workshop on Adversarial information Retrieval on the Web* (New York: ACM Press), pp. 57-64.
- Lakoff, G. (1990) *Women, Fire, and Dangerous Things* (Chicago, IL: University Of Chicago Press).
- Lau, E. P. & Goh, D. H. (2006) "In search of query patterns: A case study of a university OPAC". *Information Processing and Management*, Vol. 42, No. 5, 1316-1329.
- Lee, C. S., Watson-Manheim, M. B. and Ramaprasad, A. (2007) "Exploring the relationship between communication risk perception and communication portfolio". *IEEE Transactions on Professional Communication*, Vol. 50, No. 2, 130-146.
- Lee, K. J. (2006). "What goes around comes around: An analysis of del.icio.us as social space". In *Proceedings of the 2006 ACM Conference on Computer Supported Cooperative Work* (New York: ACM Press), 191-194.
- Levy, M. and Sandler, M. (2007) "A semantic space for music derived from social tags". In *Proceedings of the 8th International Conference on Music Information Retrieval, ISMIR 2007*. Available at: http://ismir2007.ismir.net/proceedings/ISMIR2007_p411_levy.pdf.
- Li, R., Bao, S., Fei, B., Su, Z., and Yu, Y. (2007) "Towards effective browsing of large scale social annotations". In *Proceedings of the 16th International Conference on World Wide Web* (New York: ACM Press), 943-952.
- Lin, X., Beaudoin, J. E., Bui Y., & Desai, K. (2006) "Exploring characteristics of social classification". Paper presented at the *17th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research*. Available at: <http://dlist.sir.arizona.edu/1790/>.
- Luhmann, N. (1979). *Trust and Power* (Chichester. England: John Wiley & Sons).
- Macgregor, G. and McCulloch, E. (2006). "Collaborative tagging as a knowledge organization and resource discovery tool". *Library Review*, Vol. 55, No. 5, 291-300.
- Marchionini, G. H. (1995) *Information Seeking in Electronic Environments* (Cambridge, England: Cambridge University Press).
- Marlow, C., Naaman, M., Boyd, D., and Davis, M. (2006) "HT06, tagging paper, taxonomy, Flickr, academic article, to read". In *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia* (New York: ACM Press), 31-40.
- Morville, P. (2005) *Ambient Findability* (Sebastopol, CA: O'Reilly Media).
- Muller, M. J. (2007) "Comparing tagging vocabularies among four enterprise tag-based services". In *Proceedings of the 2007 International ACM Conference on Supporting Group Work* (New York: ACM Press), 341-350.
- Olsen, K. A., Sochats, K. M., and Williams, J. G. (1998) "Full text searching and information overload". *International Information and Library Review*, Vol. 30, No. 2, 105-122.
- Puspitasari, F., Lim, E.P., Goh, D.H., Chang, C.H., Zhang, J., Sun, A., and Theng, Y.L. (2007). "Social navigation in digital libraries by bookmarking". In *Proceedings of the 10th*

- International Conference on Asian Digital Libraries ICADL 2007, Lecture Notes in Computer Science 4822* (Berlin, Germany: Springer-Verlag), 297-306.
- Razikin, K., Goh, D. H., Cheong, E. K. C., and Ow, Y. F. (2007) "The efficacy of tags in social tagging systems". In *Proceedings of the 10th International Conference on Asian Digital Libraries ICADL 2007, Lecture Notes in Computer Science 4822* (Berlin, Germany: Springer-Verlag), pp. 506-507.
- Razikin, K., Goh, D. H., Chua, A. Y. K., and Lee, C. S. (2008) "Can social tags help you find what you want?" In *Proceedings of the 12th European Conference on Research and Advanced Technology for Digital Libraries ECDL 2008, Lecture Notes in Computer Science 5173* (Berlin, Germany: Springer-Verlag), pp. 50-61.
- Reder, L. M. and Ritter, F. E. (1992) "What determines initial feelings of knowing. Familiarity with question terms not with the answer". *Journal of Experimental Psychology: Learning, Memory and Cognition*, Vol. 18, No. 3, 435-451.
- Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhous, J., Harper, F. M., & Riedl, J. (2006) "Tagging, communities, vocabulary, evolution". In *Proceedings of the 2006 ACM Conference on Computer Supported Cooperative Work* (New York: ACM Press), pp. 181-190.
- Smith, E. R., Miller, D. A., Maitner, A. T., Crump, S. A., Garcia-Marques, T., and Mackie, D. M. (2006) "Familiarity can increase stereotyping". *Journal of Experimental Social Psychology*, Vol. 42, No. 14, 471-478.
- Sun, A., Suryanto, M.A., and Liu, Y. (2007) "Blog classification using tags: An empirical study". In *Proceedings of the 10th International Conference on Asian Digital Libraries ICADL 2007, Lecture Notes in Computer Science 4822* (Berlin, Germany: Springer-Verlag), pp. 307-316.
- Surowiecki, J. (2004) *The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economics, Societies, and Nations* (New York: Doubleday).
- Sutcliffe, A. G., Ennis, M., and Watkinson, S. J. (2000) "Empirical studies of end-user information searching". *Journal of the American Society for Information Science*, Vol. 51, No. 13, 1211-1231.
- Thom-Santelli, J., Muller, M. J., and Millen, D. R. (2008) "Social tagging roles: Publishers, evangelists, leaders". In *Proceedings of the 2008 SIGCHI Conference on Human Factors in Computing Systems* (New York: ACM Press), pp. 1041-1044.
- Wathen, C. N. and Burkell, J. (2002) "Believe it or not: Factors influencing credibility on the Web". *Journal of the American Society for Information Science and Technology*, Vol 53, No. 2, 134-144.
- Winter, F. W. (1973). "A laboratory experiment of individual attitude response to advertising exposure". *Journal of Marketing Research*, Vol. 10, No. 2, 130-140.
- Yanbe, Y., Jatowt, A., Nakamura, S., and Tanaka, K. (2007) "Can social bookmarking enhance search in the Web?" In *Proceedings of the 2007 Conference on Digital Libraries* (New York: ACM Press), pp. 107-116.