Social Tagging and Communities of Practice
Two Case Studies

Abstract
In investigating the use of social tagging for knowledge organization and sharing, this paper reports on two case studies. Each study examines how two disparate communities of practices utilize social tagging to disseminate information to other community members in the online environment. Through the use of these tags, community members may retrieve and view relevant Web sites and online videos. The first study looks at tagging within the Code4Lib community of practice. The second study examines the use of tagging on video sharing sites used by a community of French teenagers. Uses of social tagging to share information within these communities are analyzed and discussed, and recommendations for future study are provided.

Introduction
This paper investigates the use of social tagging in two disparate communities of practice. Social tagging is the act of applying tags in a social setting. Tags, as defined by Guy and Tonkin are “any word that defines a relationship between the online resource and a concept in the user’s mind” (2006). Social tagging, therefore, can be described as “the collective assignment of keywords to resources” (Trant 2006). The studies reported on in this paper look at how much the people tagging take into context the other people in their social space as defined by communities of context. The first study looks at the social tagging practices in a community of practice consisting of people who develop and implement computer software for and in libraries. The second study investigates the use of tagging on video-sharing sites by French high school students that participate in dance battles in a movement known as Tecktonik Killer. These two studies of disparate groups hope to add to the still developing body of Library and Information Science (LIS) literature on social tagging.

Etienne Wenger, who coined the term along with Jean Lave, defines communities of practice on his Web site as “groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly” (Wenger n.d.). In order to be considered a community of practice, three characteristics are crucial. The first characteristic is that there must be an identifiable domain of interest. For example, Linux would be the domain of a group of people the meet monthly to discuss and learn about the Linux computer operating system. The second characteristic is a need for community. This could be a physical or virtual community but it is important that the community of practice engage in discussions and joint activities in order to help each other and share information. Third, members of a community of practice are practitioners and not merely observers. A group of people who regularly attend the opera together would not be considered a community of practice even though they have a shared interest.

Social Tagging in the Code4Lib community
The first case study explores the social tagging practices of the Code4Lib community of practice. The Code4Lib community consists mostly of systems librarians and software developers who work for libraries. Community members use various online methods to
share information and to learn from each other including Internet Relay Chat (IRC), the Planet Code4Lib blog aggregator, and an online journal.

Online communities of practice may use a specific tag on social bookmarking sites that enable Web-based resources to be shared, or “advertised,” within that community. By studying these tags, it is possible to investigate what influence, if any, the community has on the individual’s choice of tags, and to what extent the members consider community while tagging. If members of a community of practice do not tag differently for the community than those outside of the community, they may be using a social tagging tool, but any community or social benefit may simply be a by-product of personal tagging. Code4Lib is one such community that shares information by using a social tagging tool. Items bookmarked in del.icio.us, a popular social tagging Web site, with the tag ‘code4lib’ are shared in three ways with the community: 1) on a Web page created through the del.icio.us site, 2) on the Planet Code4Lib blog aggregator, and 3) on the Code4Lib IRC channel.

This study builds on a previous study that investigated tags made by fifteen Code4Lib members who bookmarked at least five items with the tag ‘code4lib’ on del.icio.us. Bookmarks tagged with ‘code4lib’ (community) were compared with items bookmarked by these members that did not include the ‘code4lib’ tag (personal) in order to see if different types of tags were used for the community. The tags were analyzed according to Golder and Huberman’s (2006, 203) seven mutually exclusive semantic categories: 1. Identifying what (or who) the item being tagged is about; 2. Identifying what the item being tagged is; 3. Identifying who owns the tag (ownership); 4. Refining categories; 5. Identifying qualities or characteristics of the item being tagged; 6. Self reference; and 7. Task organizing. Sets of tags were separated by individual user and were placed into categories based on the inclusion or exclusion of the ‘code4lib’ tag within the set. The total number of tags and the numbers of tags in each of Golder and Huberman’s categories were investigated to determine if there was any difference in kinds or number of tags used. While casual observation showed differences in how a few of the individuals tagged for themselves versus for the community, overall there was no significant difference in types of tags used in each set. The one statistically significant difference ($p < .01$) found was the number of tags applied in each set. The average number of tags used when the ‘code4lib’ tag was included was 3.70 compared to only 2.97 when the tag was not included. However, it was notable that when the tag ‘code4lib’ is excluded from the count of tags for these resources, the difference did not turn out to be statistically significant. This may mean the only difference was the inclusion of the community tag. (Tonkin et al. 2008)

**Method**

The tag ‘code4lib’ was searched for in the social bookmarking site del.icio.us in order to identify Web sites that someone, presumably community members, tagged with the term. Each Web site tagged was reviewed to see if it was about the Code4Lib community in some way (e.g. the Code4Lib conference Web site). If it was determined to be about the community, the site was excluded from the study. If it was not about Code4Lib, it was assumed that the site was tagged with ‘code4lib’ in order to share the site with other members of the community. Up to ten sets of user tags assigned to the Web sites that where identified this way where investigated. The most recent five sets of tags containing the term ‘code4lib’ were investigated. When there were less than five sets of tags containing ‘code4lib’ all of the tag sets containing the tag were examined. An equal number of tag
sets for the same resources from users that did not assign the tag ‘code4lib’ were also examined. For example, if there were four users who tagged a resource with ‘code4lib’ then four sets of tags from users that did not assign that tag were also explored. All sets of tags were manually investigated to see if any trends or differences could be observed. All totalled, one hundred sets of tags assigned to twenty-three separate resources were reviewed.

Results
The group of tag sets containing the ‘code4lib’ tag included a total of 288 tags (5.76 tags per set) while the group without the tag had only 221 total tags (4.41 tags per set). The difference between the numbers of tags by community taggers and non-community taggers was statistically significant when compared using a t-test (\( p < .05 \)). Because the cases studied are not assumed to have a normal distribution, the Wilcoxon non-parametric test was also performed and yielded identical results. Despite a higher number of total tags in the Code4Lib group, the number of different tags was almost equal. The Code4Lib group assigned 123 unique tags while the non-community groups assigned 121 unique tags. Between the two groups, 189 unique tags where used. The tags were investigated to see if any difference in the types of tags used could be found by assigned each tag to one of Golder and Huberman’s seven kinds of tags. However, this did not produce any interesting results.

When the community identifier tag is excluded, the Code4Lib group still assigned, on average only 0.3 additional tags per item. T-tests and a Wilcoxon non-parametric test were performed on the data excluding the ‘code4lib’ community tag from the results. Without the community tag, no statistical significance was found. This result is in accord with the results found in the earlier study of tagging in the Code4Lib community of practice that found the only statistical difference may be explained by the inclusion of the community tag.

Discussion
One observation was that the Code4Lib community group was less likely to use a tag that defined something as ‘opensource’ (or related tags such as ‘open_source’ or ‘open-source’). ‘Opensource’ was one of four tags that were repeated the most times in the non-community group (eight times). The tag, or one of its variants, was only used three times by the Code4Lib community group. Considering the larger number of tags assigned by the Code4lib group this is an interesting observation. One possible explanation for this is that the Code4Lib community of practice works mostly with Open Source Software and thus, the observation of something being an Open Source project is not as worthy of a tag delineating that fact. Future studies may want to consider looking into the degree of specificity of tags within a community of practice.

The twenty-three Web sites reviewed all related to computer programming or libraries, and most related to both. One possible explanation for only a slightly larger number of tags assigned by the Code4Lib community group members do not feel any more ownership of these types of resources than the other taggers. Since they do not have ownership, they may not have an incentive to supply additional tags. Ownership, as we will see in the study of the Tecktonik killer community below, appears to imply a greater level of commitment to a resource, even in the online world. Thus, one of the challenges for organizations or communities of practice that hope to take advantage of social tagging will be to
somehow install a sense of ownership or some other incentive to get people to tag items more thoroughly. It appears that simply being a community of practice member does not necessarily imply a sense of ownership. As Lorcan Dempsey (2008) points out in his Weblog, one cannot simply add social networking tools to a Web site “and expect it to work well.” For example, think of all those empty Web-based forums that have been installed on sites over the years. Web sites with a larger user base such as del.icio.us may be more immune to the issue of ownership and incentive because of sheer scale, but it is unlikely that small communities can reach critical mass to develop interesting trends or folksonomies without cultivating a sense of ownership.

**The Tecktonik killer community**

Some counter-culture French high school students engage in dance battles in a movement called “Tecktonic killer” (Lazimi 2007). These teenagers dance to techno music and have a shared identity among members of the community. Along with competitions in the offline world, community members make videos of themselves dancing and post them online in video-sharing sites. Other community members can then comment on the technique in the videos, “favourite” the videos, and bookmark them using social bookmarking utilities like delicious. The competitive environment can remain, and community members can learn moves, vie for attention, and share thoughts with other community members. As a way of studying the extent to which the community of practice surrounding the Tecktonik killer dance phenomenon in France is using social tagging, tags supplied by both creators and users (those who posted videos and those who bookmarked them in del.icio.us) were investigated. Comments made by both groups were also collected, in part to understand better participation by French-speaking taggers, and in part to compare the degree of investment on the part of the two groups.

Research previous to the current study was carried out in October in 2007 right after the movement had started to be reported in French news sources. When the prior case study was performed, twelve videos posted to three video sharing sites used by the French (YouTube, Daily Motion, and video.fr) were investigated. Four types of videos were selected: exemplary, average, beginner, and parody. The tags for each type of video were counted and compared. The results of the study implied that French teenagers were familiar with social networking technologies and were participating in tagging activities in a way that was similar to other taggers. The French community in question has adapted to Web 2.0 in ways similar to other online communities. (Tonkin et al. 2008)

Author indexing and user indexing are traditionally perceived as being different approaches to the indexing task. The current research seeks to understand if taggers of Internet resources tag more when they perform the online equivalent of author indexing. “Endo-tags” or tags of one’s own material are assumed to form more robust tag sets due to the advertising function they incorporate. When supplying tags for their own content, endo-taggers are advertising their work to other members of the community. More tags make non-text resource more findable. In this study, we expect to see a relatively large number of endo-tags in video-sharing sites such as YouTube and Daily Motion.

“Exo-tags” or tags created by users of posted content will form less robust sets in comparison. Exo-tags are sufficient if they guide users back to a previously-found resource. It is assumed, for this research, that in social bookmarking sites like del.icio.us, exo-taggers are tagging the content of others for their own personal use. Exo-tags do not carry the
same advertising function as endo-tags, but rather serve as personal signposts in the online environment.

By using bookmarks, exo-tags, exo-comments, endo-tags and endo-comments for freely available video content, the current study seeks to understand more about the use of tags types in these online environments.

**Method**

Two alternate spellings of the dance phenomenon (tecktonic, tecktonik) were searched in the del.icio.us social bookmarking site as a means of identifying relevant resources. There is no well-known French-created social bookmarking site, but prior experience shows del.icio.us to be a site likely to be used for bookmarking French content. Search results were examined manually to ascertain the nature of the resource being bookmarked. To maintain continuity with the prior study, videos were the only bookmarked resources retained. The activity for a total of fifty-three del.icio.us users was collected. In some instances, multiple users had bookmarked a single URL in del.icio.us. All of the exo-tags, exo-comments, and data about date entered were harvested from del.icio.us and entered in a spreadsheet for computation.

On the video-sharing site for the corresponding video resources, the creator-supplied endo-tags and endo-comments, the number of times the resource was viewed, and the date it was posted (for purposes of disambiguation) were collected. Fifteen videos posted to video sharing sites were examined for this study.

**Results**

Endo-taggers supplied an average of 10.8 tags per video. Bookmarkers who supplied exo-tags gave an average of three per video. Eight bookmarkers did not supply tags; they bookmarked a URL for a Tecktonik killer video without supplying any exo-tags at all. The difference between average numbers of supplied endo- and exo-tags is extremely statistically significant when compared using a t-test ($p < .001$). Because the cases studied are not assumed to have a normal distribution, the Wilcoxon non-parametric test was also performed and yielded identical results.

**Discussion**

In his book *The Long Tail*, Chris Anderson (2006) describes the reputation economy that has developed online. Although they do not receive money for their efforts, generators of content who contribute to the online body of knowledge are trying to improve their own reputation. By tagging heavily, endo-taggers are seeking status while contributing to the online community of practice, and doing their best to draw attention to their contribution.

Conversely, exo-taggers in del.icio.us supply less tags. In part, this may be the result of being less invested in the creation of the content. Exo-taggers may also perceive that they are tagging for themselves instead of tagging for the community.

The issue of language also presents relevant cultural and community-based questions. Many of the del.icio.us bookmarks (n=15) supplied comments that were not in French. Some seemed to be written by native speaker of English who were discovering the originally French movement: “what happens when you mix Breakdance & Techno? This!” Other comments were less idiomatic and may not have been written by native speakers of English. Five del.icio.us exo-taggers supplied comments in French. One of these exo-
taggers also supplied the English-language tag “music” to the set. The use of English, the lingua franca of the Internet, in the bookmarks implies an understanding of the international aspect of the movement. This is less obvious in the video-sharing sites where all but one of the comments was in French.

The importance of the community and the difference between core members and casual observers in this study should also be acknowledged. Endo-taggers are likely to be core members of the community who speak French, participate in offline dance battles, and compete in the online arena as well. The Paris TechnoParade of September 15, 2007 popularized the Tecktonik Killer style of dance and introduced it to international visitors and casual outside observers in France. We note that exo-taggers in this study largely tagged their resource in September 2007 or after, signalling a late-comer status in terms of the movement. One del.icio.us exo-comment in French calls the dance “ridicule” and another English comment describes it as “French people dancing in interesting ways to electronic.” These are not the comments one would expect from core members of the community who feel invested. Furthermore, exo-taggers are not necessarily invested in the advertisement of the movement and this may in part explain their reluctance to tag robustly.

Conclusion
While investigating the community approaches to tagging in case studies is useful, it does not provide an overview of the phenomenon. To see trends more clearly, it is advisable to look at larger data sets. However it should be remembered that with small or emerging communities, it is not always possible to get a larger data set to investigate and sometimes it is necessary to use smaller data sets in an attempt to understand what is happening at that time inside of the community.

These two different studies on disparate communities of practice show similarities in the way social tagging can be used for knowledge organization and information retrieval. The Code4Lib and Tecktonik Killer communities both take advantage of tagging techniques in order to share information with other community members. They both use tags not only to describe a document, but also to assign ownership and promote sharing. The Code4Lib community does this using Golder and Huberman’s (2006) ownership tags while Tecktonik Killer participants tag items with generic terms and proper nouns that advertise affiliations and community. Further study on larger communities would help determine how people consider their community of practice when tagging. Lastly, community taggers could themselves be interviewed and studied to understand better their motivations and their strategies of use.

References

