Collaborative and Social Tagging Networks

Emma Tonkin, Edward M. Corrado, Heather Lea Moulaison, Margaret E. I. Kipp, Andrea Resmini, Heather D. Pfeiffer and Qiping Zhang gather a series of international perspectives on the practice of social tagging of documents within a community context.

Introduction

Social tagging, which is also known as collaborative tagging, social classification, and social indexing, allows ordinary users to assign keywords, or tags, to items. Typically these items are Web-based resources and the tags become immediately available for others to see and use. Unlike traditional classification, social tagging keywords are typically freely chosen instead of using a controlled vocabulary. Social tagging is of interest to researchers because it is possible that with a sufficiently large number of tags, useful folksonomies will emerge that can either augment or even replace traditional ontologies. As a result, social tagging has created a renewed level of interest in manual indexing [1]. In order for researchers to understand the benefits and limitations of using user-generated tags for indexing and retrieval purposes, it is important to investigate to what extent community influences tagging behaviour, characteristic effects on tag datasets, and whether this influence helps or hinders search and retrieval.

This article reports on research presented on a panel at The American Society for Information Science & Technology (ASIS&T) 2007 annual conference which investigated the use of social tagging in communities and in context. The panel was co-sponsored by SIG-TAG, a special interest group of ASIS&T that is interested in the study of social tagging, the Special Interest Group on Knowledge Management (SIG-KM), and the Special Interest Group on Classification Research (SIG-CR).

Panel participants described studies around the world that explore to what extent and in what manner users, consciously or unconsciously, take into account their communities of practice when assigning tags. Through the use of these tags, community members may retrieve and view relevant Web sites and online videos. Each study examines how different communities use social tagging to disseminate information to other community members in the online environment. Two of the studies are exploratory in nature, investigating tagging within specific communities of practice. One of these looks at the Code4lib community which has a primarily North American focus while the other examines the use of tagging on video sharing sites used by a community of French teenagers. Other studies described here explore cognitive and social factors of tagging in China, collaborative classification practices in social tagging tools, and the many roles of terminology in a collaborative work environment. Additionally, this article describes FaceTag, a working prototype of a collaborative tool conceived for social tagging of
specialised domains.

There are many possible approaches toward examining the panel's theme, of which these studies form a representative sample. Tags may be examined 'by hand', which is time-intensive but information-rich. Various forms of automated statistical analysis may be applied, trading accuracy against scale. The latter has the benefit of suggesting that practical application of the approach may be plausible. Models of user behaviour may be examined through development of tools designed to test hypotheses or elicit feedback from users - such as FaceTag.

Several models of tagging behaviour, aimed at describing the ways in which people tag, are invoked in these studies along with metrics such as the number of tags given, tag co-occurrence and measured frequency. This reflects an ongoing dialogue between researchers; some apply methods from social network analysis, some from the many subfields of linguistics, knowledge management and classification research. Tagging practice is generally from a known stance, such as metadata, keyword or thesaurus provision, a matter of situating the relevance of the concept to known disciplines. Here too, we begin with a familiar theme; writ large, this panel examines certain facets of contextuality in information retrieval.

Social Tagging in the Code4Lib Community

The first case study explores the social tagging practices of the Code4Lib community of practice. Social tagging can be described as 'the collective assignment of keywords to resources' [2]. The free-form nature of social tagging tools allows users to assign their own verbal descriptors [3]. It has been suggested that once enough user-generated tags have been supplied, social tagging could lead to interesting folksonomies with benefits similar to a formal controlled vocabulary [4]. However, people have varying motivators for tagging [5] and often use tags for purposes other than assigning subject terms. Kipp and Campbell [6], for example, found that people often use time and task-related tags such as "toread" -- tags are sometimes used to describe intended action rather than semantic content. This case study began from the hypothesis that an individual applying tags for themselves might make use of these tags in a manner that differs from the the approach taken by an individual who is consciously targeting their annotation at a community’s collective attention.

Online communities of practice may use a specific tag on social bookmarking sites that enable resources to be “advertised” within that community. Through the study of these tags, it is possible to investigate the community’s influence on the individual’s choice of tags, and to what extent community members consider community while tagging. If members of a community do not tag differently for the community than they do for themselves, is it truly social tagging?

Code4Lib is an organic community consisting of librarians and library software developers. One way Code4Lib shares information is by bookmarking items in del.icio.us [7], a popular social tagging Web site, with the tag ‘code4lib’. Once an item is tagged with ‘code4lib’, it is shared in three ways: on a Web page created through the del.icio.us site, on the Planet Code4Lib blog aggregator, and on the Code4Lib Internet Relay Chat (IRC) channel. It is assumed that members of the Code4Lib community want to share an online resource with the community if the set of tags applied to the bookmark includes the tag ‘code4lib’. Conversely, it is assumed that community members are bookmarking resources for their own personal use when they do not include the ‘code4lib’ tag in the set of social tags they assign.

Tags of fifteen Code4Lib members who bookmarked at least five items with the tag ‘code4lib’ on del.icio.us were reviewed. All users whose tags were reviewed are active community members and are aware that items tagged with ‘code4lib’ are shared with the community. Ten recent bookmarks tagged with ‘code4lib’ (community) were analysed. If community members tagged less than 10 items with ‘code4lib’, all bookmarks with the tag were reviewed. Ten items bookmarked by these members that did not include the ‘code4lib’ tag (personal) were also examined. All tags (n=872) associated with bookmarked resources were analysed according to Golder and Huberman’s [6] seven categories of tags. Sets of tags were separated by user and were placed into categories based on the inclusion or exclusion of the ‘code4lib’ tag within the set. Both the overall number of tags and the numbers of tags in each category were analysed using the Wilcoxon Signed-Ranked tests to determine if there was
any statistical difference in kinds or number of tags used. While casual observation shows differences in how some individuals tagged for themselves (set of tags which did not include 'code4lib') versus for the community (set of tags which included 'code4lib'), overall, there was no significant difference in types of tags used in each set.

There was a significant difference ($\alpha = 0.01$) found in the number of tags applied for the two sets. The average number of tags used in a set when 'code4lib' was included as a tag was 3.70 compared to only 2.97 tags when 'code4lib' was not included. A larger number of tags was applied when tagging for the community. This may indicate that community members do indeed tag differently for a community than they do for themselves. However, when the tag code4lib is excluded from the count of tags for these resources, the difference does not turn out to be statistically significant. This suggests that the only difference is the inclusion of the community tag.

The results of this study call into question the idea that people tag differently when they tag for communities as opposed to when they tag for themselves. One explanation is that, regardless of whether they are actively tagging for a community, community members always take into account how other people use their tags. Conversely, community members may never take into account how the community of practice will use tags, even when those members are actively sharing bookmarks with their peers.

**Social Tagging in France: Communities of Practice**

In France, an engaged community of social taggers has emerged in online video sharing sites. Current French teenagers are very comfortable with new communication-rich technologies through the use of social networking resources like blogs and cell phone text messaging. They are now applying their social networking skills to tagging and sharing community-based uploaded videos. Certain French secondary school students are engaging in dance battles in a movement called Tecktonic Killer [9]. Members of this counter-culture community of dancers stereotypically have Mohawk haircuts, listen to techno music, and try actively to out-perform their peers in these dance battles. The merging of the online and offline world is a part of the French net-generation reality [10]. Beside the analogue competitions, dancers make videos of themselves in their homes and upload them to video-sharing Web sites. Postings include videos, social tags, and descriptions. Community members then view and comment on each other's videos, recreating the competitive environment that exists in the analogue world.

For the current study, twelve videos of Tecktonic Killer dancing that encouraged community interaction by using tags and enabling comments were selected from three video-sharing sites used by the French teenagers. Reasonable attempts were made to select home-made videos from each of four categories: average videos, popular or highly rated videos, beginners, and parody videos.

Use of the tags by this community reflect established social tag usage in the Anglo world [6]. A total of 77 tags were assigned to the twelve videos. Tags included numerous proper nouns such as personal names of dancers, video site usernames, dance club affiliations, and city/country names. There were also 28 generic terms designed to help users navigate to the pages. Of the name of the dance mentioned in tags, there were 13 variants of spelling and terminology, many in the same tag set. There were a total of seven terms that used English-language words including dance, fake, and fashion. The parodies averaged 12 tags per video; the serious ones averaged between four and five tags, with many having only two or three tags.

Community members who post videos are actively tagging for the community and not strictly speaking for themselves. They are in search of the notoriety that comes from a well-received and often-viewed video, and actively solicit viewers and comments. Viewers contribute to the community process by adding comments, marking videos as favourites, rating the quality of the dancing in the video, or making the poster a 'friend'. By participating in the feedback loop, viewers actively contribute to the success or failure of a dancer's video.

Implications for future study specific to this community are many-fold. Studying the Tecktonic Killer tags in other Web 2.0 environments as identified by Technorati [11] would also serve to show how these users act in different social tagging environments. Using only the video posting sites as the
focus of study, one community member’s tags, comments, and favourites could be studied, or the interaction of users with others’ videos and other tags would also be considered and compared. As the phenomenon becomes mainstream and perhaps if it moves outside mainland France, its adaptation and migration can also be studied via the use of community-based social tags.

Metadata Tagging in China: Three Models of Tag Use

With the development of Web 2.0, where multi-author contribution will be common, collaborative tagging becomes a popular concept. The tagging metadata contributed by users are used for individual-based activities like searching, filtering, navigating, and group-based social networking like finding people with common interest and so on. Tagging data will help to reveal the knowledge sharing and topic networks based on the relations among tagging words. Tagging data will also constitute a social interaction among taggers. It was seen how tagging fosters social information organisation and collaboration, and how to leverage the explosive volume of metadata. Tagging can also promote information by making it seem more important, i.e. by raising its profile. Tags can be seen as a way to ‘vote’ on the relevance or level of interest of content. As seen in the image below, from the size of the tag one can infer its social popularity.

![Figure 1: Research Web Page](image)

There are three models of how tagging is used in China as metadata [12]. The first way is where the user employs tagging to link information through the tag to the user. This is tagging being used as a way to indicate the relationship of the user to the information, showing how the user perceives ('reads') that information object.
The second model portrays users that are connected together through their use of tags. This is where real social networking comes in, as users are tagging to relate their concept of information to another user's concept of some piece of information. This type of tagging is also heavily misused, either consciously ('tagspam') or unintentionally. It may be used inconsistently, by tagging in order that other users see desired information despite the fact that the information is not really classified under their expected concept of that tag.

Within the final model, tags are used to link banks of data (or information) to other information. The
tags are acting as metadata to allow search engines to know which information is related to other information. This type of tagging is used greatly within ontologies.

![Diagram](Image)

**Figure 4: Model 3: 'Information - Tag - Information'**

Tags have only recently become popular in the Chinese language and are used mainly in the second model form. There are currently several terms in use on Chinese Web sites: What Do People Call Tags in Chinese?

- **Label** (标签)
  - Bokee, Sina Blog, Zhuaxia, Douban
- **Category** (分类):
  - Baidu音乐掌门人, 百度搜索
- **Folksonomy** (大众分类)
  - BlogBus
- **Keyword** (关键词)
  - Yculblog.com(歪酷博客)
- **Sign** (标识)
  - 56.com(我乐网)

**Figure 5: Examples of Tags in Chinese** 

A brief inspection shows that tag use varies between keywords, like 'Music' and 'Novels', to opinions, like 'dislike', personal descriptions such as 'I am happy today', or 'Very tired', and finally dates or date information with respect to events in a person's life. They may be used as a means of introducing oneself. Tags may contain complete sentences (written Chinese does not have to contain spaces).

Users of tagging systems in China do not generally have an in-depth understanding of tagging; similarly, service providers generally do not recognise the importance of tagging systems as a
resource or a service. They are, however, widely used; blog search, wikis, social bookmarking, photo and video sharing, and social networking sites all provide a mechanism.

Tag systems are still quite new to China, and research in Chinese-language tagging is equally in its infancy. However, it is an interesting area that is likely to merit further research in the future.

**Collaborative Classification Practices in Social Tagging Tools**

The previous three studies examined coherent user groups and their approaches to tagging. The following three studies examined social bookmarking and tagging in various contexts as they compare to more traditional hierarchical classification systems used in libraries. The first study examined tag usage for highly tagged sites on del.icio.us exploring the convergence of tag usage and the co-occurrence of terms in tag clouds. The second study examined journal articles tagged on CiteULike [13] and compared these terms to professionally assigned index terms and author keywords associated with the same article. Common themes in both studies led to a third study examining time and task management and affective tag use on del.icio.us, CiteULike and Connotea [14] [6][3][15].

While del.icio.us is commonly known in the Web 2.0 community, CiteULike and Connotea have garnered a smaller audience with their focus on academics and academic articles. Like del.icio.us, however, both raise interesting issues concerning the organisation of knowledge where personal organisational systems developed by users to keep track of their own materials are aggregated on the Web to form a novel way to organise and store material.

These tools lead to a series of interesting questions about the social organisation of information:

- What does tag convergence on frequency graphs and co-word usage suggest about the utility of tagging?
- Are categories emerging in social tagging that will complement those developed through professional methods?
- What implications does the use of affective or time- and task-related tags have for the organisation of information?

Each study examined one of these questions.

The first study, using del.icio.us, examined tag usage for highly tagged sites. Specifically, it examined the frequency of use of specific keywords on these sites and how they were used in conjunction with each other.

Tags were collected from del.icio.us for 64 highly tagged sites and graphed by frequency of occurrence to see whether a consensus would form in the use of tags to describe an item. A consensus appeared in all cases in the frequency graph with a curve showing a rapidly decaying level of usage. This is commonly described in terms of Zipf’s law (a power-law distribution); however, many graphs showed a much steeper drop off than Zipf’s law would suggest.
Figure 6: Tag frequency for pocketmod.com in early 2006

This suggests that consensus does exist in the aggregate. The terms that are most common tended to provide a reasonable description of the content of the site and remained constant over time. Data collection in November 2007, almost two years after the initial collection, showed that the seven most frequent terms remained constant over time.

Figure 7: Tag frequency for pocketmod.com in late 2007

Co-occurrence graphs of tags, examining the frequency of use of groups of tags together, showed interesting clusters of terms. Some graphs had what appeared to be different user groups with differing priorities. Clusters of similar terms provided an adequate description of the site using differing terminology.
The second study, involving CiteULike, studied tag use and the types of tags used on academic articles compared to author keywords and subject headings assigned to those same articles by their authors or by professional indexers. Mathes [16] suggested that since users and professional indexers have different contexts in which they relate to the document, it would be worthwhile to study their differing use of terminology.

The comparison was done using standard thesaurus terms -- equivalence (EQ), synonym (SYN), narrower terms (NT), broader terms (BT), related terms (RT), etc.) based on a modification of a system used by Voorbij [17]. His categories included all the common thesaural relations (SYN, NT, BT, RT) and additional categories were added for 'related but not in the thesaurus' and not related [3]. The most common relationship found between the three groups of terms was 'related but not in the thesaurus.' The next most common was RT and then equivalence (equivalence included plurals and variant spellings but did not include acronyms which were listed as synonyms). The third study examined commonalities found between the two studies. Many studies have discovered similarities in common tag characteristics (e.g. spelling variations, sesquipedalians -- compound words, acronyms, etc) [8][5][6][3] and remarked on the use of time- and task-related tags (e.g. 'toread', 'todo') or affective tags (e.g. 'cool' and 'interesting'). However, most studies tend to dismiss these terms as noise in the tag cloud despite studies showing that time, task or project labels factor heavily in personal information management [18][19]. In the first study of del.icio.us, these tags formed ~16% of the total number of unique tags and the second study of CiteULike showed high usage of these tags as well. The tags 'cool', 'funny' and 'toread' appear in the main del.icio.us tag cloud.
This suggests that these tags may form an important part of tagging and are worthy of further study. The third study examined these tags and found that ‘toread’ and ‘fun’ were popular on all three sites studied [15]. Affective terms appeared more frequently on Citeulike and Connotea than expected considering their focus on academic articles. In the study, biology articles were most often listed as ‘toread’; while maths and physics articles were listed as ‘fun’[15].

Studies show that tagging has all the problems of free-text search/indexing, though tag groups tend to converge on a useful set of terms after a threshold number of users. Though new terminology may alter this profile. While users do use some terms from thesauri and subject heading lists, users also use terminology which is rare or completely absent from subject heading lists (e.g. time and task tags) or is not part of a formal thesaurus.

Distributed user tagging demonstrates need for examination of how users associate groups of documents (time and task management). What is the effect of personal and subjective terms such as ‘cool’, ‘fun’ and ‘toread’ in a social bookmarking system? What happens when these terms are aggregated? Amazon and Google use personal information to generate popularity or relevance indicators, do non-subject tags offer any similar advantages?

While consensus exists in the aggregate, as seen in frequency graphs of del.icio.us tag usage, there are suggestions in the individual tag lists and in coword graphs that there are different but equally descriptive clusters of terminology in the tag clouds which are worthy of further study.

The differing terminology use in tag lists suggests that tagging may be a working example of Vannevar Bush's associative trails. He argued that associative trails better represented how users actually work with their documents: as a holistic process of association closely tied to themselves and their work rather than by categorisation. [20]. This suggests that user tagging could provide additional access points to traditional controlled vocabularies and provide users with the associative classifications necessary to tie documents and articles to time, task and project relationships as well as other associations which are new and novel.

**FaceTag**
FaceTag is a collaborative tagging tool conceived for bookmarking resources in a given domain. Its main goal is to show how in communities of practice the flat keywords space of user-generated tags can be effectively mixed with a richer faceted classification scheme to improve the information architecture of a social tagging system. FaceTag was designed to take advantage of the capability of tags to match the user’s language, which was expected to carry a number of advantages. The low cognitive cost of tagging items is partially resultant from the fact that no formal structure or vocabulary is involved, and in many cases those searching for objects also benefit from the idiosyncratic nature of tags, since specialised or dialect uses of terminology are represented as well. Tags have great value in serendipitous browsing -- that is, unstructured or undirected searches through which interesting objects are found, although the user may not have set out to find that particular object [21].

However, despite their low cognitive cost, their capability of matching the user’s language, and their great value in unstructured browsing, folksonomies are haunted by a number of important issues related to intrinsic language variability and imprecision: polysemy, homonymy, plurals, synonymy and basic level variation are linguistic issues which do not appear easy to solve, and which can dramatically reduce the effectiveness and benefits brought on by the very use of a tagging system [22].

FaceTag introduces a number of correctives to address these issues:

- It allows tag hierarchies. Users have the opportunity to organize their resources by means of parent-child relationships.
- By means of the underlying, domain-related faceted classification structure [23], tag hierarchies are semantically assigned to established facets. These are used to section and navigate the resource domain flexibly.
- Tagging and searching can be mixed to maximize findability, browsability and user-discovery.

For the test implementation, the reference domain was Information Architecture (IA). Although the final facets have been reviewed to fit into a more semantic perspective, the initial choices in the FaceTag system are directly based on the Classification Research Group (CRG) findings, which postulate 11-13 general categories. Table 1 shows the matching between CRG standard categories and IA-related categories that were used to define the facets, while Table 2 shows actual tags or other terms used in those facets. This is being constantly revised upon users’ feedback, as guidelines and not as a restrictive framework. User needs are paramount and should be employed to determine the most valuable facets for every system and community of practice.
Table 1. FaceTag facets definition by CRG standard categories. (Facets in brackets have been considered of secondary importance and discarded.)

<table>
<thead>
<tr>
<th>CRG</th>
<th>FaceTag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thing</td>
<td>[Documents, resources]</td>
</tr>
<tr>
<td>Type</td>
<td>Resource Types</td>
</tr>
<tr>
<td></td>
<td>(e.g. online report, case study...)</td>
</tr>
<tr>
<td>Part</td>
<td>--</td>
</tr>
<tr>
<td>Property</td>
<td>Language</td>
</tr>
<tr>
<td>Material</td>
<td>Themes</td>
</tr>
<tr>
<td>Process</td>
<td>--</td>
</tr>
<tr>
<td>Operation</td>
<td>[Activities]</td>
</tr>
<tr>
<td>Product</td>
<td>Deliverables</td>
</tr>
<tr>
<td>Byproduct</td>
<td>--</td>
</tr>
<tr>
<td>Patient</td>
<td>Purposes, Markets</td>
</tr>
<tr>
<td></td>
<td>(e.g. Industry, Health...)</td>
</tr>
<tr>
<td>Agent</td>
<td>People</td>
</tr>
<tr>
<td>Space</td>
<td>[Country]</td>
</tr>
<tr>
<td>Time</td>
<td>Date</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Facet</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resource Types</td>
<td>white paper, case study, blog &gt; enterprise web,</td>
</tr>
<tr>
<td>Language</td>
<td>predefined values (based on ISO Standard ISO 639-2)</td>
</tr>
<tr>
<td>Themes</td>
<td>competitive analysis, classification &gt; facets, web 2.0 &gt; folksonomies, information design &gt; navigation design &gt; breadcrumbs</td>
</tr>
<tr>
<td>People</td>
<td>weinberger, e. reiss, morville</td>
</tr>
<tr>
<td>Purposes</td>
<td>industry, public administration, health, software &gt; companies &gt; google, education &gt; conferences &gt; www2006</td>
</tr>
<tr>
<td>Date</td>
<td>date of publication (not insertion)</td>
</tr>
</tbody>
</table>
Usability studies show that information seekers in domains with a large number of objects prefer that related items be in meaningful groups to enable them to understand relationships quickly and thus decide how to proceed: without any means to explore and make sense of large quantities of similar items, users feel lost and fail. Providing ways to generate and navigate such groups from a flat set of objects is a challenge, as both clustering and faceted classification have been proposed in the past. FaceTag works differently. At the heart of FaceTag is a zooming engine which maintains a history of all tags and searches a user is requesting (engaging). When users engage a tag, the interface adjusts and zooms in providing a filtered view, a subset of all available resources based on this active selection. By selecting more tags from different meaningful facets (as they are editorially chosen for the current domain), users can effectively shrink the results they obtain until they find either what they are looking for, or the last possible result set arising from the intersection of all active tags. This feature set enables considerably enhanced, berrypicking-like [24] search strategies: since tags can be disengaged freely and in whatever order suits the users, opening up new unpredicted result sets, users might decide to pursue another search and enter new different subsets following some other information scent (that is, a theme or term that appears promising), or start anew by clearing up their history completely, never finding themselves facing dead-ends which produce no results at all. The term 'berrypicking' is often used in information science to refer to models in which, instead of a search that completes in a single query-response, the user tries a set of (usually evolving) search terms, finding many different pieces of relevant information.

Table 2. Actual FaceTag facets and examples of tags. Hierarchical groups of tags are set off with commas. Tags following a greater than (>) symbol are on the next lower level of the hierarchy. The language facet is an exception, since it uses a predefined list of languages in the ISO 639-2 notation.
Figure 10: Using FaceTag: refining and zooming in by engaging tags from different facets: 5 bookmarks in FaceTag
Figure 11: Using FaceTag: refining and zooming in by engaging tags from different facets: 3 bookmarks in FaceTag

Figure 12: Using FaceTag: refining and zooming in by engaging tags from different facets: 1

Main Articles: 'Collaborative and Social Tagging Ne... http://www.ariadne.ac.uk/issue54/tonkin-et-al/
Currently, research is underway to pinpoint both a stable set of facets in the IA domain and an effective UI. Samples of IA-related bookmarks from the IAI Library and from del.icio.us are being collected in order to perform iterative card sorting tests with different user groups. The purpose of such tests is to map some of the mental models by which users represent the IA knowledge domain. These results will provide the basis to tune the facet architecture. Similarly, the user interface is being designed through documented heuristics and patterns and verified at each iterative step by small usability tests. More extensive user research will involve the use of think-aloud protocol sessions with more than five testers for each session. Scenarios will include storing bookmarks and retrieving them. Looking at preliminary results, a critical task addressed by the application is the assignment of new bookmarks and the association of tags to relevant facets. The current interface is rather simple but other options, leveraging advanced tag suggestion and tag/facet association, are beginning to be evaluated.

FaceTag began from the assumption that bottom-up and top-down classification methods are complementary: tagging is inclusive, simple and matches the user's real needs and use of language, whilst formal methods provide coherent, consistent systems for organising data but are expensive and do not have the advantage of corresponding to the user's use of language. The system that was built as a result combines both approaches, to mutual benefit. As evaluation is now underway it is hoped that results will become available later this year.

**Social Networking and Annotation**

Language, it can be said, acquires associations through use. As Chalmers writes, ‘The individual and their prejudice are changed through the use of language, and the language changes through its use by individuals’. Language shifts, evolving along with the context of use; as language shifts, so does perception. Language is often used characteristically within a community, which has led to the description of the ‘speech’ or ‘discourse’ community both as a means to characterise the terminology and language use shared between participants; it is also possible to look on the speech community as a community to which one claims membership by attaining that set of characteristics in speech.

Swales describes discourse communities – ‘spatially dispersed, formed around sociorhetorical functions, and mainly mediated by texts’ – as sharing:

- common public goals
- mechanisms for intercommunication
- participatory mechanisms to provide members with feedback and information
- discourse expectations reflected in genres
- specialised terminology, and
- a critical mass of experts

However, the reality is more complex than a simple social network. According to researchers in the area, the idea of a large, stable meta-structure containing a linguistically stable group of individuals can be discounted; Prior quotes Marilyn Cooper as arguing for ‘seeing discourse communities... as the products of continual hermeneutic work, as social phenomena where varied values and practices intersected, as ways of being in the world, not narrow intellectual commitments.’ Studies have found that in academic and disciplinary writing, where one might expect to find a discourse community, one finds complex spaces that display apparently stable emergent characteristics resultant from many local interactions, discourses, practices and identities.

The various studies discussed during this article would seem to bear out this assertion. A Chinese study has shown that the existence of a tagging system can itself constitute a social statement, situating its publishers as progressive; by comparison, functionality of the service comes a remote second. From the evidence of the American study of 'code4lib', a community brought together by a joint occupation and interest, showing no evidence for the 'social' in 'social tagging', to the description of French youths applying community-specific terminology in a declarative act, establishing an identity...
as a member of a community, it is clear that, as Ellis & Larsen-Freeman state [31], 'there are many agencies and variables that underpin language phenomena.'

Characteristics of tags depend on specifics of the user interface, the use case and user population, and the user's own motivation for using the interface. Tags are 'language-in-use', which is to say that only a subset are formal or explicit (as for example are keywords); many or even most tags are informal, contain limited explicit information, transient, and are intended for a limited audience. If semantic annotation and literal annotations are taken to be linked, then it follows that semantic annotations may be no more formal than any other marginal note, implying that frameworks such as Marshall’s dimensions of annotation [32][33] have descriptive power in analysing tag sets.

We are left with several preoccupations: the context in which a tag is written, and the context in which it is read; the community interactions which underlie both processes; untangling the confusions that result from the many and varied uses of the terms 'context' and 'community. A fourth expression can be added to this list, to form a quartet of interests: caution. Without investigation and analysis, seeing named social entities in a dataset may simply reflect our own preconceptions.

**Conclusions and Future Work**

In these studies, the uses made of tags in several communities was examined. We have seen many diverse definitions and uses of the term ‘community’: language communities, speakers of Chinese, English or French; nations; discourse communities, or communities of practice, united by shared discussion, participation and terminology; task groups, held together by a shared aim. Some of these uses are more functional in practice than others. Some metrics show no effect; others suggest more. In discussing these results, our interest was drawn collectively to linking the various phenomena described and the approaches applied within these studies, feeling that these present studies have only scratched the surface. Open questions abound: to which fields might one look for inspiration, applicable methods, or illustrative results? In general, we believe that there is further interesting work to be done in the various ways in which tagging systems are used - whether as an indexing resource, a status symbol for a Chinese startup, a central resource for community organisation and development or as a channel for communication between members of a community.

The SIG-TAG mailing list [34], an informal resource for discussion and joint work for researchers and those with an interest in the fields brought together in this article, may be joined online.

**References**

7. del.icio.us http://del.icio.us/
American Association of Applied Linguistics Conference.
34. SIG-TAG mailing list http://mail.asis.org/mailman/listinfo/sigtag-l

Author Details

Emma Tonkin
Interoperability Focus
UKOLN

Email: e.tonkin@ukoln.ac.uk
Web site: http://www.ukoln.ac.uk/

Edward M. Corrado
Systems Librarian
The College of New Jersey

Email: ecorrado@ecorrado.us
Web site: http://www.ecorrado.us

Heather Lea Moulaison
Doctoral Student
Rutgers University

Email: heather.moulaison@rutgers.edu
Web site: http://www.eden.rutgers.edu/~moulaiso/

Margaret E. I. Kipp
Instructor
College of Information and Computer Science
Long Island University

Email: margaret.kipp@gmail.com
Web site: http://www.eskimo.com/~meik/

Andrea Resmini
Information Architect
CIRSFID, Università di Bologna

Email: andrea.resmini@unibo.it
Web site: http://www.resmini.net

Heather D. Pfeiffer, PhD
Computer Systems Analyst and Researcher
New Mexico State University

Email: hdp@cs.nmsu.edu
Qiping Zhang, PhD
Assistant Professor
College of Information and Computer Science
Long Island University

Email: qiping.zhang@liu.edu
Web site: http://myweb.cwpost.liu.edu/qzhang/

Article Title: "Collaborative and Social Tagging Networks"
Author: Emma Tonkin, Edward M. Corrado, Heather Lea Moulaison, Margaret E. I. Kipp, Andrea Resmini, Heather D. Pfeiffer and Qiping Zhang
Publication Date: 30-January-2008 Publication: Ariadne Issue 54
Originating URL: http://www.ariadne.ac.uk/issue54/tonkin-et-al/