Structures in Collaborative Tagging
An Empirical Analysis

A. Fani Marvasti1  D.B. Skillicorn2

1 Mathematics and Computer Science Department
Baha’i Institute for Higher Education, Iran Email: amin.fani@bihe.org

2 School of Computing
Queen’s University, Canada
Email: skill@cs.queensu.ca

Abstract

It is widely believed that users choose meaningful tags; that the combinations of these tags from multiple users helps to elicit meaning for tagged objects; and that implicit communities can be discovered among the users who use a particular tag or tag a particular object. We examine data on tagging practices from the Delicious web site (delicious.com) and find that there is little support for any of these beliefs. Although users individually do seem to have a small set of tags that they use in a controlled and effective way, they also use very large sets of other tags in a much more haphazard way. These poorly managed tags obscure much of the collective sense making and implicit community structure. We derive some suggestions for improving collaborative tagging systems.

Keywords: social computing, tagging, information retrieval, collaborative sensemaking

1 Introduction

Collaborative tagging systems allow users to label any web object, typically a page but possibly also an image, audio, or video file, with a set of strings, called tags. Usually, these labels are then visible to others and the objects can be retrieved by the original tagger, and usually by others, using tags as search terms.

The use of tags allows objects to be categorized in any way that is useful to the individual doing the tagging. In contrast, hierarchical organization of information requires decisions about the best single place in the hierarchy where that information should be categorized. Tags are therefore sometimes considered as a substitute for taxonomies.

There are a number of widely-held assumptions about collaborative tagging, some of them embedded in the business plans of the businesses that provide collaborative-tagging sites. These assumptions are:

• Users label documents with carefully-chosen and meaningful tags that reflect some semantic aspect of the documents, or of the relationships between the individual users and the documents. Each user is creating a mapping from the document, and its content and properties, to an appropriate set of strings, in such a way that the user can, at some future time, generate the inverse function with high probability. In other words, the tag is chosen so that, when the user wants to retrieve the document again, s/he will be able to infer or remember one of the tags required to find the desired document.

This implies that the set of tags should be parsimonious and that individual tags should be well-separated in meaning. If they are to add functionality above that of search engines, they should also reflect aspects of the documents or their properties that are orthogonal to the content, and perhaps also hard to infer automatically. In other words, the greatest potential benefit is when a tag captures a human decision about the ‘meaning’ of an document. (Rashmi (Sinha 2005), for example, has proposed a cognitive model that describes the advantages of tagging over categorization, and proposes how humans might choose tags.) Users receive recompense for making such decisions by exploiting the tags of other users to improve their own retrieval success.

• Users act collectively to make sense of the content and meaning of each document. Each user makes a small intellectual contribution by choosing ‘good’ tags and, in return, is able to use the well-chosen tags of others to find information that might be difficult to discover by conventional search. In some systems, tags already used to label a particular document are displayed so that a user can reinforce the existing opinions about tags that are good; or suggest new ones to the community of those interested in this document. Smith (Smith 2008) explains tagging as a way of bridging personal and community knowledge: Chi (Chi 2008) claims that users choose tags both for themselves and for others; and Weick et al. (Weick et al. 2005) describe tagging as collective sense making. For this collective process to work, each user must maintain and use a mental model of how to label a new document effectively, as described in the previous point, and these models must generally agree.

• The set of users who tag a particular document, and the set of users who use a particular tag have other similarities, so that these common connections allow the detection of communities among users that might be hard to discover in other ways. The implicit graph connecting users, tags, and documents can be used to carry out clustering, query expansion, and enhanced search experience, for example by generating recommendations for one user based on the knowledge supplied implicitly by others. John and Seligman (John & Seligmann 2006) and Rashmi (Sinha 2005) build social networks from the way in which documents have been tagged.

Copyright ©2010, Australian Computer Society, Inc. This paper appeared at the Thirty-Third Australasian Computer Science Conference (ACSC2010), Brisbane, Australia. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 102. B. Mans and M. Reynolds, Eds. Reproduction for academic, not-for profit purposes permitted provided this text is included.
Here we report empirical studies of tagging behavior at one of the most popular collaborative-tagging sites, Delicious (delicious.com). In general, our results provide little support for these assumptions about how collaborative tagging is done. In particular, we observe that:

- Users tag in a haphazard way. The tags used often have little connection to the document being tagged, individuals use very large numbers of similar tags, and users make many mistakes in tagging. This suggests that they do not use a parsimonious mapping of content to tags.

- Users do not appear even to reuse their own tags successfully. They tend to generate tags similar to those they have already used, apparently having forgotten the precise form of their previous tags. Clearly retrieval cannot be effective when this happens.

- Given that individual users fail to reuse their own tags well, it is not surprising that there is little evidence that one user’s tags help other users to search more effectively.

- Although use of the same tag or tagging of the same object obviously indicates some kind of common interest, the implicit communities this suggests seem to be very diffuse. This is partly because of the use of multiple similar tags which diffuse what might be a single community into a multitude of weaker communities. Discovering any communities that do exist seems difficult.

Our tentative conclusion is that collaborative-tagging systems are ineffective for most users, even for their individual information-retrieval needs. As a form of collaborative sense making, it is even less clear that collaborative tagging achieves very much. There seems to be some users who are tagging documents with many tags, presumably altruistically to make them accessible to as many other users as possible. However, there is little evidence that this pays off in better access by other users.

It is not clear whether this failure of collaborative tagging is related to novelty, so that tagging quality will come as users gain experience; or whether it is related to interface design, so that better information would produce better tagging; or whether tags are so individual and taste-based that they do not generalize well.

There is a need for interview-based analysis of tagging behavior to validate these tentative conclusions. Meanwhile, it seems unlikely that deeper understanding of social processes can be derived from today’s collaborative-tagging data.

2 Related Work

Several empirical studies of tagging systems and tagging behavior have been carried out. Golder and Huberman (Golder & Huberman 2006) examined the different kinds of tags in use and classified them as having the following functionalities: identifying what a document is about; identifying what kind of document it is; identifying who owns it; refining existing categories; identifying qualities or characteristics; self-referencing by the tagger; and organizing tasks. Smith (Smith 2008) also suggests that some tags are intended to attract search engines. Golder and Huberman also analyzed the temporal properties of tags, showing that some tags remain popular while others enjoy bursts of popularity. They also show that the distribution of tags associated with a particular document tends to become stable, so that the relative frequency of each tag becomes constant (and novel tags are then rarely applied to it).

Wetzker et al. (Wetzker et al. 2008) analyzed Delicious between 2004 and 2008 and observed that the site is heavily biased towards technology-related content. They also showed there is a power-law relationship between users and documents tagged. The top 1% of users generate 22% of all tagging events, and the top 10% generate 62%. There is also a power-law relationship between tags and documents tagged: 700 tags account for 50% of all tagging events. They also indicated that they suspected that the most active Delicious users were actually spamming, that is tagging with the intent of driving traffic to certain web pages or sites, but we find no evidence to support this.

Chi and Mytkowicz (Chi & Mytkowicz 2008) also used data from Delicious, and examined tags and documents using entropy-based measures. Their most significant finding was that the mutual information between tags and documents is linearly decreasing with time, so tags are becoming less and less useful as search terms. They also find that the rate of increase in the total number of tags in use has flattened, something that we do not observe for individual users.

Halpin et al. (Halpin et al. 2007) did a similar analysis over a longer time period, and found that power laws are characteristic of activity in this domain. They also showed that tag distributions become stable, supporting the results of Golder and Huberman, and showed how these distributions evolve in the period before stability.

Zeng and Li (Zeng & Li 2008) experimented with using the structure of the tripartite graph of users, tags, and documents to draw conclusions about user interests that could be used to make recommendations. They conclude that tags can be useful to improve user similarity calculations, but doing so requires very sophisticated clustering techniques.

Begelman et al. (Begelman et al. 2006) used cooccurrence similarity to cluster tags and were able to show that some plausible clusters could be obtained. One example, the cluster based around “health” includes: “shopping”, “research”, “nutrition”, “food”, “diet”, “fitness”, “workout”, “running”, “article”, “science”, “esport”, “sport”, “product”, “life”, “lifehacker”, “howto”, “gtd”, “reference” and “tip”. While some of these tags are clearly health-related, many of the others have only a tenuous connection to it, suggesting that tag similarity is quite diffuse.

Santos-Neto et al. (Santos-Neto et al. 2006) analyzed tagging behavior in CiteULike and Bibsonomy, tagging systems that specialize in bibliographic records. They found a significant correlation, for each user, between number of tags and number of tagged objects, perhaps because finding a bibliographic reference requires a unique key. They also observed a large number of singleton users with no overlap of tags or tagged objects with anyone else.

3 Properties of Tags

We carried out a number of empirical studies using data collected from the delicious.com website during the second half of 2008. Delicious has more than 5 million registered users who have tagged more than 150 million objects.

Delicious allows registered users to tag any web object they encounter. When this happens, sets of popular and recommended tags are displayed from which the user can choose, but arbitrary new tags may also be created. Each tag is constrained to be a single (blank delimited) string. Various communication mechanisms within the site are possible. For
Table 1: Statistical properties of tags and keywords extracted from the chosen documents.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std</th>
<th>Dev</th>
<th>Max</th>
<th>Min</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tags T</td>
<td>46.97</td>
<td>11.39</td>
<td>15</td>
<td>20</td>
<td>4</td>
<td>41</td>
</tr>
<tr>
<td>Keywords K</td>
<td>24.92</td>
<td>9.52</td>
<td>37</td>
<td>2</td>
<td>3</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 2: Statistics of Jaccard similarity between tags and extracted keywords

The simplest functionality of such a site is that it replaces individual browser bookmarking, with the added advantage that tagged objects are accessible from anywhere on the Internet, not only the browser from which they were first encountered. It seems clear that many users use Delicious only for this purpose.

Tags, and the objects they label are public on the Delicious web site, unless explicitly hidden, so anyone, not just a registered user, can search using a tag (to get the corresponding objects) or a url (to see the corresponding tags). Lists of the most popular and recent tags over the whole site are also available.

Delicious is just one of a number of popular sites that permit tagging. For example, Youtube (www.youtube.com) specializes in video, Flickr (www.flickr.com) in images, Last.fm (last.fm) in music, and CiteULike (www.citeulike.org) in bibliographic references.

### 3.1 The relationship of tags to content

A dataset consisting of 100 documents and their tags was extracted from the Delicious.com web site. Each document had been tagged by at least 50 users, so that the set of tags applied is likely to have converged to a relatively stable set (as predicted by (Golder & Huberman 2006)). The documents were also required to be primarily textual so that, for example, pages of links to video or audio were excluded. The process was seeded by beginning from users who had used tags on the popular list.

Each document was processed to remove ephemeral content, for example inserted advertisements, and the html markup was removed to leave plain text. The plain-text documents were then processed using Yahoo's Term Extraction web service, which extracts keywords and phrases. (It is also advertised as an automatic way to construct tags from documents.) The Yahoo service extracts phrases as well as single words, while Delicious tags are only allowed to be single strings, which causes some discrepancies, although minor.

Statistical properties of the sets of tags and keywords extracted from this set of documents are given in Table 1. It is clear, and slightly surprising, that the number of tags is substantially larger and more variable than the extracted keywords. Note especially that the minimum number of tags for any of these documents is 20.

We compute the similarity between the set of tags and the set of keywords for each document using Jaccard similarity:

\[
\text{sim}(T, K) = \frac{|T \cap K|}{|T \cup K|}
\]

The results are given in Table 2. It is clear that the overlap between tags and keywords is extremely low, that is tags are not simply descriptions of the tagged document’s content.

Table 3: Percentage distribution of non-keyword tags

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>Purpose or type words such as “reference”, “howto”, “tips”, “tools”, “tutorial”, and “free-ware”.</td>
</tr>
<tr>
<td>19%</td>
<td>Modified keywords. This includes both plurals and other modifications of actual keywords, or strings built by concatenating words (because of Delicious’s limitations), for example, “word-pressplugins”</td>
</tr>
<tr>
<td>11%</td>
<td>Acronyms and non-English words. This includes many common abbreviations, for example “ux” and “aussie”; as well as words from other languages that were not recognized as keywords by the term extractor</td>
</tr>
<tr>
<td>7%</td>
<td>URLs and formal terms including file types such as “mp3” or “pdf”</td>
</tr>
<tr>
<td>4%</td>
<td>Times and dates, for example when a product was released or a book published</td>
</tr>
<tr>
<td>3%</td>
<td>Process words, for example “toread”, “todo”</td>
</tr>
<tr>
<td>3%</td>
<td>Adjectives, for example “interesting”, or “cool”</td>
</tr>
<tr>
<td>2%</td>
<td>Punctuation and function words, for example “””, ““”, and words such as “and” and “or”. Some of these are clearly caused by misunderstanding Delicious’s rules for forming tags</td>
</tr>
<tr>
<td>1%</td>
<td>Misspellings of plausible keywords</td>
</tr>
</tbody>
</table>

One possible explanation is that users do not tag documents with content-based tags because searching on content is straightforward with mainstream search engines. However, at present there is no way to search on both content and meta-information encoded in other kinds of tags, so it would seem reasonable to use content tags extensively.

### 3.2 The role of tags

The typically 90% of tags that did not correspond to keywords were examined manually and categorized as shown in Table 3.

The most striking thing, of course, is how rare it is to label documents with tags that could be considered abstractions of their content, such as hypernyms. Golder and Huberman (Golder & Huberman 2006) suggest that this is a common form of labelling, but we find little evidence for this.

Labelling with purpose tags is extremely useful since inferring purpose algorithmically is difficult. This category is by far the most useful for sense making. Providing tags that describe document content in other languages and using common abbreviations is also useful. However, many abbreviations and variations are somewhat idiosyncratic. It is conceivable that algorithmic techniques such as stemming might help to condense the tag sets in use, but our impression is that the variety is so great that this would be extremely challenging.

It is also clear that users would prefer a richer language for tagging that would allow multiple-word
tags. However, addressing this deficiency would not necessarily be an improvement because the number of possible multiple-word phrases to describe a particular concept is large, making it even harder to get agreement among users.

Process and adjectival tags show a lack of awareness of the global context of tagging, because such tags rely on context to be useful. Labeling a document as “toread” is useful to an individual only if it is also tagged with a unique identifier, and generating such a unique identifier in an unknown global environment is difficult. There is little sign that such identifiers are being used.

Similarly, an adjective is an individual opinion which is unlikely to have useful weight in an environment with 5 million users.

The results suggest that people often use tags as a way of adding personal notes or non-obvious details about the content for their own consumption – they use tags for browsing and organizing, rather than information retrieval, similar to the way they put their personal photos in an album and add marginal notes.

4 How Individuals Tag

We now consider tagging behavior in more detail by looking at the set of tags used by an individual.

4.1 How individuals use tags

We extracted data for 100 random users, each of whom had used at least 100 tags. The frequencies of use for each of the tags for each user were then plotted, sorted in descending order of frequency of use. Results for four users are shown in Figure 1, truncated after the 100 most-frequent tags. The distributions are very similar for all of the users investigated so we did not investigate a larger set of users.

The discontinuities in frequencies, indicated by the circles, separate tags with qualitatively different usage patterns. There is typically a small set of tags, typically less than 10, that are heavily used. There are then one or two sets that are used at lower frequencies. The broad shape of these curves is consistent with the power-law behavior observed by others; but the discontinuities suggest that a refinement is needed. The boundaries between these regions suggest that there may be two different mental mechanisms at work for handling tags – some concept of a working set of well-remembered tags that users maintain and which perhaps captures their primary interest(s); and a much larger set of tags that are recollected and applied in a more diffuse way.

4.2 How individuals tag documents

We extracted data for 500 random users and considered how many documents they each tagged. The results are shown in Figure 2. There are a number of users with both very large numbers of tags used and very large numbers of documents tagged. The bottom left-hand corner of the figure is expanded in Figure 3, showing that users use many tags and label many documents. In general, for a given user there are more tagged objects than there are tags, but not by much. This suggests that tags are reused to label multiple objects, but not much on average. Given the distribution of tag frequencies observed in Figure 1, this suggests that some tags are used heavily, but most are used only a very few times.

It is clear from this figure that many users are extremely active taggers. The number of tags used by an individual ranges from 6 to 23,887, with mean 2009; and the number of documents tagged by an
It has been suggested that users with either high rates of tag use, or tagging large numbers of documents are spamming, possibly algorithmically. We investigated this question by examining, manually, the tags and documents of the individuals at the high end of the spectrums of both. There are 15 individuals who have labelled more than 10,000 documents. In all cases, the tags they used appeared to relate to the content of the pages they tagged, and the sets of documents tagged were not from a limited set of domains. In other words, it seems implausible that they were tagging with any malign purpose such as directing traffic to any particular documents or sites.

There are several discernible characteristics to the tag usage of these heavy-tagging individuals:

- They use personalized tags, acronyms, and numbers heavily;
- They tend to include urls and domains as tags;
- They use many redundant tags (for example, “weblog”, “blogs”, “blog”, “blogging”, “blogger”, and “weblogs” for the same document);
- They make many typographical and spelling errors; and
- They use both English and non-English words.

It seems implausible in the extreme that individuals maintain a mental list of thousands of tags that they believe will be useful for their own retrieval. However, it is possible that these users regard themselves as acting altruistically, labelling documents in as many ways as they can imagine so that they can be retrieved by other users in many different possible ways. This provides some evidence for a social dimension to tagging. Nevertheless, the actual usefulness of this approach is limited by the number of errors made in the tags themselves, and the low specificity of the tags used. The potential benefit seems small compared to the amount of effort made.

4.3 Document similarity based on users

To compute the similarity of documents tagged for each user we use the same set of 500 users whose total number of tagged documents were considered earlier. For each user, we select the 50 most-recently tagged documents. For each pair of documents, similarity is computed as the Jaccard similarity using both the extracted keywords and tags used for each document. This captures both the inherent similarity (based on the content of each pair of documents) and the user views of similarity (based on the tags applied to each pair of documents). The average similarity between the documents tagged by each user are shown in Figure 4, together with the standard deviations (sorted into increasing order of average similarity).

This figure shows that the average overlap among documents tagged by any user is rather small; in other words, users have diverse sets of interests. For 90% of the users, the similarity among documents is below 4%, while for the remaining 10% it only reaches a maximum of 14%. This suggests that knowledge of which users have tagged a document provides little useful information for finding, for example, related documents.

Some examples of the similarity matrices for particular users are shown in Figure 5, with lighter shades reflecting greater similarity, and with rows and columns alternately sorted to the top right-hand corner. For User 54, there is a great deal of weak similarity among tagged documents, but very little clustered structure. In contrast, User 209 has two well-defined clusters of documents, but the remaining documents are even less related than those of User 54. User 128 lies in between these extremes.

The maximum overlap between documents tagged by each user, seen in Figure 6, shows some interesting structure. There are three groups of users in the figure with identical similarities (flat regions of the curve in the figure): a block of 40 users with maximum document similarity of 33%, a block of 22 users with maximum document similarity of 66%, and a block of 50 users with maximum document similarity of 100%. The similarity of documents is computed using both the tags and keywords associated with each document. For most documents, the size of the tag set is roughly twice the size of the keyword set. The most likely explanation is that, for the first group,
there are documents where the keywords match but the tags do not; for the second, there are documents where the tags match but the keywords do not; and for the third group, there are documents where both match.

4.4 Tagging over time

To examine the way in which users tag more deeply we now consider how an individual uses tags over time.

We collected a dataset of two sets of users: ten who have tagged 100 or fewer documents ever; and ten who have tagged more than 1500 documents. The users were selected randomly from among those who used tags on the popular bookmark list. Users of the first kind are assumed to be naive or beginning users, while those of the second kind are assumed to be experienced.

We then retrieved, for each user, data about the tags used on consecutive days when any activity was recorded, and computed the Jaccard similarity between successive days. The results, for a typical inexperienced and experienced user are shown in Figure 7 (note the different scales on the axes in the two graphs).

The typical usage of an inexperienced user is that there is either a lot of overlap between two consecutive days, suggesting that an area is being explored; or very little overlap, suggesting that the user has moved on to a new topic of interest. It is not surprising that such a user can remember and reuse tags effectively, since they have only a small set of tags in use.

In contrast, experienced users have little overlap from one day to the next. This seems surprising, since we might expect that such users have a well-defined set of interests, and would continue to tag new documents related to these interests, so their tag sets would be stable and overlaps from day to day would be large. Manual inspection of such users suggests that the low overlap is the result of imperfect recollection of which of a possible set of tags they have previously used. For example, a single user used “addon”, “addons”, “add-on”, and “add-ons”; while another used “answer”, “answers”, “answers,” and “answers)”, both within a short time period. This reinforces the earlier result suggesting that users maintain a small list of tags that they can remember and use appropriately, and a much larger set that they remember only hazily and so use with less control.
5 Communities of Users

We now consider the question of whether either tags or documents provide any extra structure that connects users into communities. In other words, are the users who all use a particular tag similar in other ways. The existence of such communities derived from tagging practices could be used to provide recommendations, or to expand queries by including tags known to be related (via communities) to the tag being used for search.

To compute the similarity of users using a particular tag, we collected the top 500 most-popular tags, and selected the 50 most-recent users to have used each of these tags. We now consider how similar each of these groups of 50 users is.

There are several useful measures of user-user similarity that could be computed for such a set of users: the similarity of the set of documents they have tagged; the similarity of the content of the set of documents they have tagged; or the similarity of the tag sets they have each used. The first is problematic because there is not a 1-1 relationship between documents and the urls that describe them (for example, because of the use of tiny urls); and the second is expensive and limited by the effectiveness of content extraction. So we use the third, and compute the Jaccard user-user similarity for each of the 50 users of each tag.

Figure 8: Average and standard deviation of user-user similarities for the first 500 most-popular tags

The average similarity for the 500 tags is shown in Figure 8, together with the standard deviations. The list of tags is sorted into ascending order of user-user similarity. For approximately 90% of the tags, the user-user similarity is low, so there is little evidence of user communities associated with these tags. For the remaining 10% of tags, the user-user similarity is as high as 42%, but the standard deviations of these similarities also increase substantially, suggesting that there is not a systematic relationship – the similarity is highly tag-dependent. The maximum similarities per tag were also computed and, for all but one tag, there are at least two users with identical tag sets.

Some of the actual user similarity matrices are shown in Figure 9. As before, these matrices have been sorted alternately by row and column to move the largest entries to the top right-hand corner.

Figure 9: User-user similarity matrices for a selection of tags; lighter shades mean greater similarity, coloured bars indicate (weak) clusters

There are only a few small clusters among the users who have used each tag; the only exception we discovered is the tag “book” which has one large cluster. Note especially that purpose tags, which might be expected to be useful to others, do not have substantially more structure than content tags. These figures suggest that there is useful knowledge about implicit user communities present in tagging data,
but that it will be hard to extract. This is partly because the useful data is overwhelmed by less-useful frequent data. Just as the presence of highly advertised bestsellers makes collaborative recommendation difficult, the presence of popular but non-useful tags may make implicit community discovery difficult.

6 Conclusions

The business plans of commercial collaborative tagging sites, and much academic work on collaborative tagging has assumed that tags are chosen carefully based on a desire and expectation to use them for later retrieval; that tags can synthesize information encoded in existing tags with each new user’s view of an object to produce a collective synergy of understanding; and that common use of tags, and tagging of common documents reflects the presence of underlying implicit communities of interest.

We find very little evidence, at present at least, to support these assumptions. The primary problem seems to be the way in which tags are actually chosen. Far from being carefully chosen to reflect the content, meaning, and context of an object, tags seem to be selected on the fly, with little thought. There is some evidence that users maintain a very small list of tags (of size perhaps no larger than 10) that they are able to use in a careful, effective, and consistent way. However, they then seem to use very large sets (thousands) of tags in a haphazard way, labelling similar objects encountered even very close in time with different tags. This makes it highly implausible that users are able to retrieve previously tagged objects even using their own tags.

The collaborative aspect of tagging also seems to be only weakly supported. It is clear that many users are using tags to provide information that is not directly present in objects such as meta-information or translations of the content into other languages. What they do not appear to do is to use tags to describe the content at different level of abstraction (e.g., hyponyms of the content words). The potentially useful information from tagging is largely spoiled, or at least diffused, by poor and inconsistent choices of the actual tags. It seems that some users are labelling with very large numbers of tags in an effort to make objects as easy as possible to find, but it is not clear that this effort is worth the improved access.

This suggests that tagging performance might be improved by restricting both the form and number of tags that a user could apply to any object. For example, forcing a maximum number of tags (say 5) would encourage clearer thought about the best tags for a given object, and would also encourage conventions such as always using singular rather than plural noun tags. More forceful presentation of existing tags would also help to develop conventions about what makes a good tag. (On the other hand, Chi and Mytkowicz suggest that users be presented with existing tags and encouraged to choose novel tags.) Thus interface design in tagging systems might well improve the performance and usefulness of such systems.

We also found little evidence of user communities among those who use any given tag, and little evidence of similarities among the documents tagged by a given user. It appears that each individual tends to have wide interests, as expressed in what they tag. Trying to find communities via these commonalities will only be effective if powerful techniques are used, since the communities are so weak.

These results suggest that both technology and social constraints can be used to improve collaborative tagging. Better interfaces that force quality of tags rather than quantity would probably help. More direct feedback just as how others are using tags would also probably improve the quality of tagging. There is clearly a demand for richer tagging languages, but it is not clear that this would improve the situation because the increased expressiveness also creates more ways to express essentially the same tag.

It is conceivable that the deficiencies we have observed are because tagging is a new phenomenon, and users have not yet become accustomed to how to use it well. However, Delicious is dominated by technologically sophisticated users, so it would be surprising if this were the only issue.

References


