Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices

Margaret E.I. Kipp and D. Grant Campbell

Faculty of Information and Media Studies
University of Western Ontario
mkipp@uwo.ca
gcampbel@uwo.ca

Abstract

This paper analyzes the tagging patterns exhibited by users of del.icio.us, to assess how collaborative tagging supports and enhances traditional ways of classifying and indexing documents. Using frequency data and co-word analysis matrices analyzed by multi-dimensional scaling, the authors discovered that tagging practices to some extent work in ways that are continuous with conventional indexing. Small numbers of tags tend to emerge by unspoken consensus, and inconsistencies follow several predictable patterns that can easily be anticipated. However, the tags also indicated intriguing practices relating to time and task which suggest the presence of an extra dimension in classification and organization, a dimension which conventional systems are unable to facilitate.

Introduction

The rise of collaborative tagging systems has aroused the interest of information designers and information architects. Some argue that the tagging patterns emerging on such sites as del.icio.us, Flickr.com, 43things.com and Connotea constitute an entirely new method of information organization, one which can scale in ways that conventional systems cannot. Others argue that collaborative tagging systems flaunt too many standard principles of conventional indexing to be a viable replacement, no matter how far we lower our standards. This paper attempts to shed light on this debate by looking closely at the patterns of tagging in the del.icio.us bookmarking system, and assessing how the descriptors applied by individual users in this popular tool both resemble and differ from descriptors that professional indexers would apply.

Background

By allowing individual users to apply their own verbal descriptors to digital objects, personal tagging systems constitute a revolutionary form of verbal indexing. Rather than enforcing a pre-conceived controlled vocabulary, either by employing professional indexers or forcing authors to tag their articles from an established thesaurus, systems like del.icio.us assume that users can create useful results when given absolute freedom. This assumption has its roots in the development of the Web as a system which operates "as close as possible to no rules at all" (Berners Lee & Fischetti, 1999, 15). Intellectually, the assumption echoes the widespread belief that the World Wide Web is a complex, adaptive system, of the sort described in complexity theory (Waldrop 1992, 11; Johnson 2000, 18). In a complex system, there is no pacemaker, of the sort we see in libraries, with their centralized indexing and cataloguing systems. Instead, each unit, following a set of very simple rules, contributes to a spontaneously self-organizing pattern, in which the whole becomes greater than the sum of its parts. This view gets some support from the success of the Google PageRank system, in which individual votes for a site's quality, manifested as links, combine to create large-scale patterns of consensus about site quality and relevance that can drastically improve search engine performance. If the same holds

true for tagging, then personal tagging systems offer the chance to create indexing systems that scale in ways which conventional organization schemes cannot, thereby making them appropriate for large, dynamic systems (Shirky, 2005). Just as the ranking system of Google becomes more illuminating as the number of links grows, so do the patterns and clusters of user-centered tags, as more and more people adopt tagging practices. Early research has shown regularities in user tagging activity, regularities which could enable us to predict stable tagging patterns. Furthermore, early research suggests that when a URL acquires a certain number of taggers, the most common terms tend to remain stable (Golder & Huberman, 2006).

Skeptics of tagging systems, many of them in the field of Information Architecture, argue that personal tagging merely constitutes "mob indexing" (Morville, 2006). Because users are untrained in indexing methods, they cannot create tags which cohere into useful patterns, at least not to the extent that would justify dispensing with controlled vocabularies and faceted browsing schemes.

Before we can resolve this controversy, we must find ways to analyze the tagging data that is currently available to us, in order to measure the extent to which tagging represents a useful subject access mechanism. And because user-generated tags flaunt so many rules of indexing consistency, we cannot evaluate them by the usual standards of thesaurus construction. Untrained users will not, of themselves, produce rigorously-designed thesaural structures; we need to determine whether the results they do create are useful anyway.

If user tags are forming useful patterns, co-word analysis provides a possible means of detecting them, based as it is on the assumption that the co-occurrence of words in a particular field in two or more documents is a measure of the strength of the relationship between the co-occurring words (Wilson 1999). Co-word analysis has many applications in Web environments (Leydesdorff & Vaughan 2006); many of these applications emerge from the historical use of co-word analysis to study changes in fields of study over time (Courtial, Callon & Sigogneau 1984; Courtial 1994). In particular, co-word analysis is sometimes used to enhance the effectiveness of traditional indexing, by providing a map for indexers to related words and synonyms (Courtial, Callon & Sigogneau 1984).

A co-word analysis of tagging practices, therefore, could provide insight into the patterns that are emerging through these practices, and the extent to which they are consistent with, and supportive of, conventional indexing and classification practices. Equally important, such an analysis might well show important differences between user tagging and conventional indexing: differences which indexers would do well to notice.

Our study, therefore, posed three research questions:

- What patterns of consistent user tagging activity emerge through analyses of tagging frequency and co-word analysis?
- To what extent do these patterns of tagging support and enhance some of the other traditional ways of classifying documents?
- To what extent do these patterns defy these traditional methods, suggesting viable and promising alternatives to traditional subject access tools?

Methodology

This study was conducted using data collected from the del.icio.us bookmarking service. The basic

component of del.icio.us is the bookmark entry or post made by each user upon encountering a website of interest. A user typically has many posts which bookmark different sites; in addition, a website can be bookmarked with del.icio.us posts by many different users. A del.icio.us post is composed of a set of elements linking the URL being bookmarked to a specific user. Users are encouraged to assign tags to each bookmark and a majority do. Each del.icio.us post consists of a URL, an optional extended description, a set of tags, a username and the date on which it was posted. Posts may be displayed by user, by tags, or by URL. For this study, posts were examined based on the URL.

Because co-word frequency graphs tend to be inverse J shapes with many co-word pairs occurring infrequently and few co-word pairs appearing more often (Wolfram 2003, 140), data collected were to be selected from URLs which had been tagged by as many people as possible. Data were selected from the del.icio.us popular feed (http://del.icio.us/popular) for January 30th and 31st, 2006, under the assumption that a popular URL would be highly tagged. This proved true for the most part; however, because del.icio.us determines popularity by current activity rather than by cumulative activity, many of the sites (those which had only recently attracted interest) had relatively few tags. Therefore, an additional set of links was selected from articles tagged with the terms 'health', 'productivity' or 'programming' which had been tagged by at least 500 users.

The raw data was collected from del.icio.us based on the previously described set of URLs using a patched version of delicious.py from http://del.icio.us-py.berlios.de/.

An analysis package was created in Python to analyse data drawn from del.icio.us by delicious.py. This package is called delicious-analyse.py. This package uses the full set of posts for one URL gleaned from delicious.py to produce descriptive statistics, coword matrices and frequency data for the tags assigned by users to the given URL.

Descriptive statistics included:

- the total number of posts created by users for all of the URLs;
- the total number of tags created by all of the posts;
- the total number of unique tags created by all of the posts;
- the total number of unique tags with a specific frequency for each URL;
- the number of users with a specific number of tags for each URL;
- a count of the total number of tags and total number of unique tags for each URL..

The frequency data was collected directly from the posts themselves. Tags were matched alphabetically and a list of all unique tags was generated along with a count of the number of appearances of this unique tag in the full set of posts. Only alphabetically identical tags were merged; spelling variations, plurals and tags distinguished only by additional punctuation were all counted as unique tags. For example, 'toread' and 'toread,' are counted as two separate tags since this is how the del.icio.us interface views them. While users may view these tags as the same or so similar as to be basically identical, the system views them as different and will therefore display them separately and treat them as separate entities.

Co-word analysis examines the number of times each co-word pair occurs together and enables the researcher to speculate and form hypotheses about the relationship between words that co-occur frequently. In the data from del.icio.us, this translated to counting the instances of the co-word pair being used together in a post. In other words, if users A, B, and C had all tagged the same URL with

tags X and Y, then X and Y would be a co-word pair with 3 occurrences. Due to the large volume of low frequency tags occurring, the number of tags included in the co-word analysis was restricted to no more than the 40 most commonly used tags.

The resulting co-word frequencies form a symmetrical co-occurrence matrix which can be analysed using multi dimensional scaling (MDS) to produce a graphical visualisation of the strength of the relationships between the tags. In MDS, the distances on the graph represent relative similarities and differences, respectively, between terms, according to their occurrence in the group. Terms which are used frequently together would thus be closer together on the graph. The resulting graphs can be analysed for clusters of terms which are frequently used together, which in traditional co-word analysis would be considered to be related terms.

In this study, the frequency graphs and co-word graphs were analysed for trends in tag usage including clusters of related tags in the co-word graphs, patterns of drop-off in the frequency graph, the use of synonyms and spelling variations, tagging practices which fall outside the traditional boundaries of classification. Analysis of the co-word graphs was qualitative, rather than statistical in nature: the visual representation of tags, clustered by similarity or commonness of co-occurrence, was used to suggest possible patterns of use that could be analyzed in future studies, using methods that support more intensive statistical analysis.[1]

Analysis

Descriptive Statistics

The total number of URLs examined was 64. The total number of posts examined in this study was 58728. The maximum number of posts for a URL was 5172 for www.43folders.com. The minimum number of posts was 53 for www.burningcam.com. The average number of posts per URL was 917. The total number of tags in the study was 165831, with a maximum of 13,809 tags applied to a single URL and a minimum of 49 to a single URL. However, the total number of unique tags was 18904 (max: 1252, min: 23), indicating a high degree of overlap between users tagging a URL (as well as between URLs since many URLs in the sample were about productivity tools or web programming). The average number of tags used for a single URL was 295.

While a majority of users do tag content, a certain percentage of users 3462 (6 %) in this study did not assign any tags to the URLs they selected. Additionally, the number of users assigning multiple tags dropped very quickly. In fact, 65% of users assigned between 1 and 3 tags to a URL.

Certain tags occurred with a high frequency across all categories. The top ten tags across the entire sample are listed in table 1. Two interesting tags which had a high frequency of use across categories were 'toread' at 939 uses and 'cool' at 906. A total of 3049 unique tags were identified as being time or task related. These time and task related tags constitute 16% of all tags. Such tags do not conform to the traditional classification schemes and will be discussed later in the paper.

Tag	Frequency
productivity	6989
gtd	5782
lifehacks	4753
tools	4672
web2.0	4077
web	3876
organization	3469
css	3382
blog	3320
wiki	3187

Table 1: Top Ten High Frequency Tags

Number of uses of tag	Number of tags at each frequency
1	5593
2	1307
3	496
4	296
5	163
6	125
7	107
8	82
9	51
10	47

Table 2: Low-Frequency Tags (those appearing between one and ten times only)

A number of tags were used only once in the entire sample. Since they were only used once they do not appear in del.icio.us' frequency listing unless users examine all the other entries posted for a given URL. Despite their low frequency of occurrence, some of these tags are quite interesting. A number are time related including numerous variations on 'toread' and 'todo'. Among the low frequency tags, a total of 5593 unique tags (30%) were used only once across the entire sample. (See Table 2)

Discussion

In classical classification, terms fall into specific relationships such as the thesaural relationships of broader and narrower terms. Synonymy and homonymy are handled by the use of a controlled vocabulary for specific concepts. A rigorously designed hierarchy of words allows indexers to situate a document in the field of knowledge.

Tagging begins with a mass of ordinary words assigned to a particular item. At first, this system seems chaotic, but with simple visualisations of the tags, patterns begin to appear.

Graphing tags with their frequencies of use for a particular URL shows a definite shape, a power law, in which the frequency of tag use drops off relatively quickly (see Shirky, 2005, Hammond, et al., 2005). However, a classic power law shows a much steeper drop off than is apparent in all the tag graphs. In fact, some graphs show a much gentler slope at the beginning of the graph before the rapid drop off (see figure 1).

Shirky notes that this drop off does not always occur immediately and suggests that this shows that user consensus views such a site as being about more things than a site whose tags experience a sharper
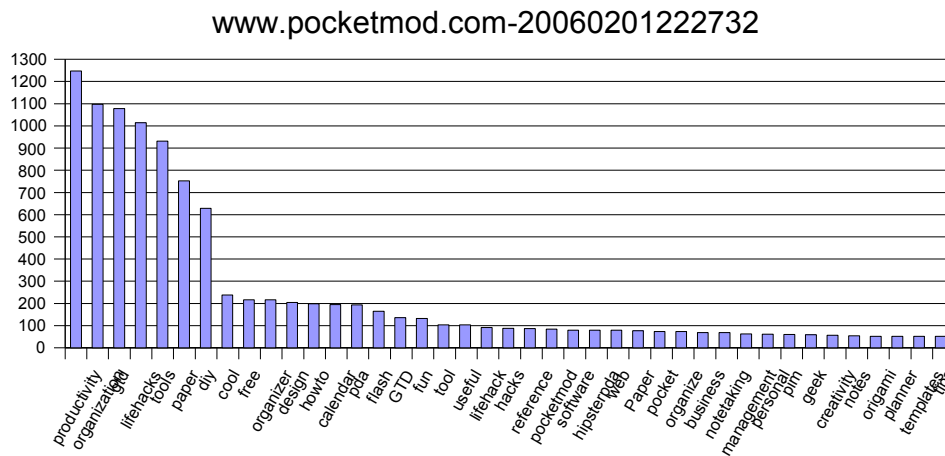

Figure 1: Tag frequency graph with slow drop off

drop off in frequency. (Shirky 2005) In figure 1, the drop off is much more pronounced after the first 7 tags, suggesting that users in the large feel that the first 7 tags are important parts of the description of this site. At first it might be reasonable to assume that this pattern shows a site which has not been highly tagged and that consensus has not yet been reached, but www.pocketmod.com had been tagged by 4171 users at the time the data was gathered from del.icio.us and the four most common tags had been used over 1000 times each. Judging by the average number of tagging users per URL in this sample (917), this is a large number of users. This suggests that, at the time of sampling, consensus had settled on these 7 tags as most representative of the aboutness of the document. A number of other graphs in the sample exhibit this less rapid drop off suggesting that users may tend to settle on a cluster of related terms which they feel best describes the aboutness of the URL.

Synonyms, acronyms and spelling variations occur frequently in the tag lists for a single URL. For the URL www.iasummit.org, the four most common tags are: conference, ia, IA, and information_architecture. The last three tags all refer to the same concept 'information architecture' and do not really add to the aboutness description of the document, although the combination of information_architecture and its acronym could be of potential use to a newcomer to the field.

The frequency graphs suggest a relative consensus on terms, but a visual analysis of the co-word graphs shows that although high frequency tags do appear to be used more often with other high frequency tags, there is no definite connection between the position on the frequency graph and that on the co-word graph. This suggests that while high frequency tags are used frequently, they are not necessarily used most frequently with other high frequency tags. We may be seeing indications that taggers apply tags according to a mix of communal and individual notions of aboutness and usefulness. The del.icio.us interface makes it easy to both add tags that are frequently used, and to add tags that are part of the user's individual tag cloud. We can expect, therefore, tagging patterns to follow both emerging consensus among users, and emerging trends within the user's own tagging system, however idiosyncratic.

As in the frequency graphs, spelling variations were common. Spelling variations consisted of capitals vs lower case initial letters on words, singulars vs plurals, and variations between British English and American English spellings. On the co-word graphs, these variations were often quite distant from

each other. This seems to imply that distances on the co-word graphs do not signify lack of similarity, but instead lack of concurrent usage. This makes sense intuitively as users who use the British English spelling would likely not also use the American English spelling in their list of tags. However, in a large group of users, one would expect both variations to appear. Similarly, singular and plural forms of words both occur commonly in tags, where a standard form would be selected in classical classification. In the co-word graphs, these terms are also generally separated showing that they are not used together.

The use of punctuation in tags is common and is often used to compensate for the inability of the del.icio.us system to accept multi word keywords (only single word tags are acceptable, a space serves as separation between tags). Many users tag URLs with multi word concepts, either because the language they are using does not contain a single word for the concept, or because they wish to connect two or more concepts explicitly as is often done in classical classification with pre-coordinate headings such as Libraries--United Kingdom. A prominent example of this phenomenon is shown in figure 2 with the use of terms such as socialsoftware, socialbookmarking and social_bookmarking. Each refers to a multi word concept describing various initiatives on the web to share user input. It is notable that the terms socialbookmarking and social_bookmarking, which refer to the same concept, are separated on the MDS graph and do not in fact belong to the same clusters.
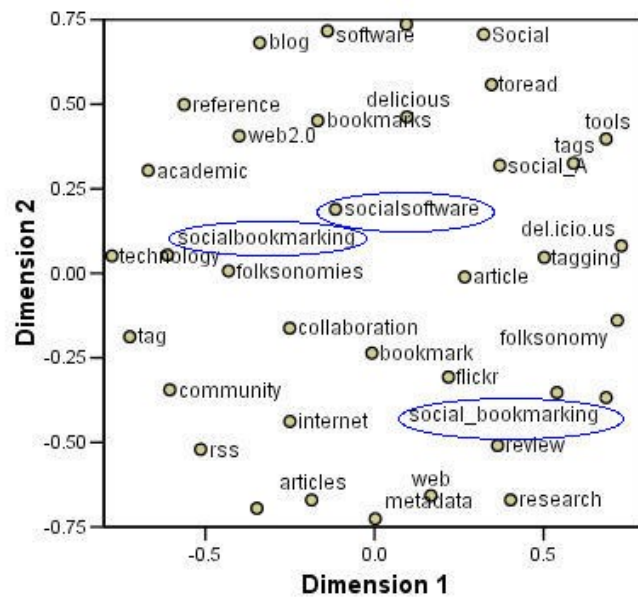

Figure 1: Co-word graph for Hammond et al. 2005 (normalised raw stress = 0.15764)

A similar phenomenon appears when examining synonyms and highly related concepts that in a thesaurus would qualify as related terms, or possibly broader or narrower terms. In general, such related terms do not cluster together on the co-word graph. Again, as for the spelling variations and emulated multi word groupings, this is intuitively logical as users of one synonymic term are less likely to include additional synonymic terms in their personal tags. It is only in concert that such synonyms are added to the full list of tags for a URL. For example, in figure 3, the co-word graph includes terms such as downloads, filesharing, filetransfer, download, upload, ftp and p2p. All of these terms are related in that they refer to the transfer of files, however, only downloads, filesharing and filetransfer

are clustered together. Intuitively, one might expect the term p2p, with its close connection in the media with filesharing, to be a part of this cluster, but it is not. Similarly, in figure 4, one might expect terms such as nutrition and diet to cluster together, but they do not. In fact, neither nutrition nor Nutrition cluster with diet, although Nutrition does form a cluster with food.
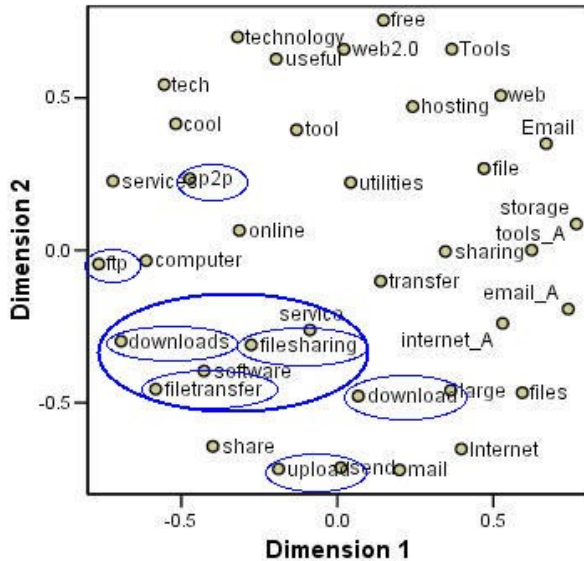

Figure 3: Co-word graph for www.yousendit.com(normalised raw stress = 0.15805)
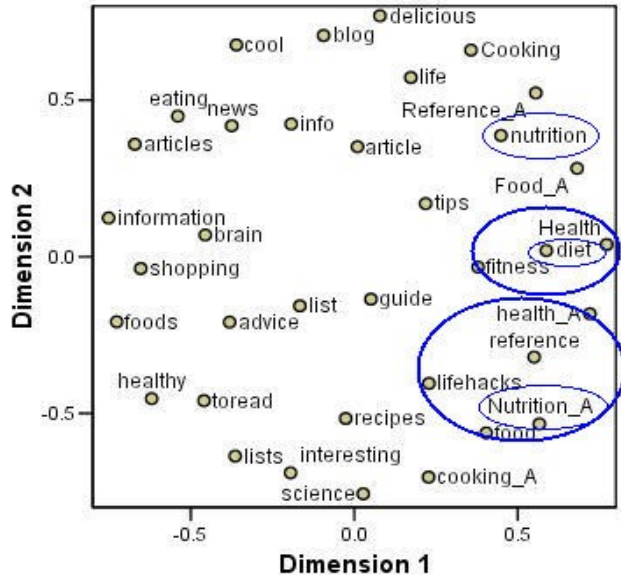
Figure 4: Co-word graph for www.bellybytes.com/articles/29foods.shtml(normalised raw stress = 0.15758)

Thus, it often appears that in the co-word graphs of tags, similarity is not indicated by lack of distance. In fact, similar terms are often widely separated on the graph, as indicated by the divergences between terms separated only by the fact that one version is capitalised or plural.

One possibility is that the different variations on terms show the existence of differing user groups and their differing views of what terms are the most commonly used within that user group. A possible example of this is shown in figure 4 where the terms Health, diet and fitness are clustered together in one grouping and the terms health, reference, lifehacks, Nutrition and food are clustered together. The term lifehacks is a specific term for 'changes made to ones life in order to achieve a potential benefit' and thus indicates a user group at work, which, while having similar goals to the user group interested in diet and fitness, uses differing terminologies.

In any case, many of these phenomena show evidence of the growth of an entry vocabulary to a subject where similar terms all occur together, allowing users to work their way into a subject via common words.

While the majority of co-word graphs appeared similar to figures 1-4, two items in the study demonstrated unusual co-word graphs (see figures 5, 6).
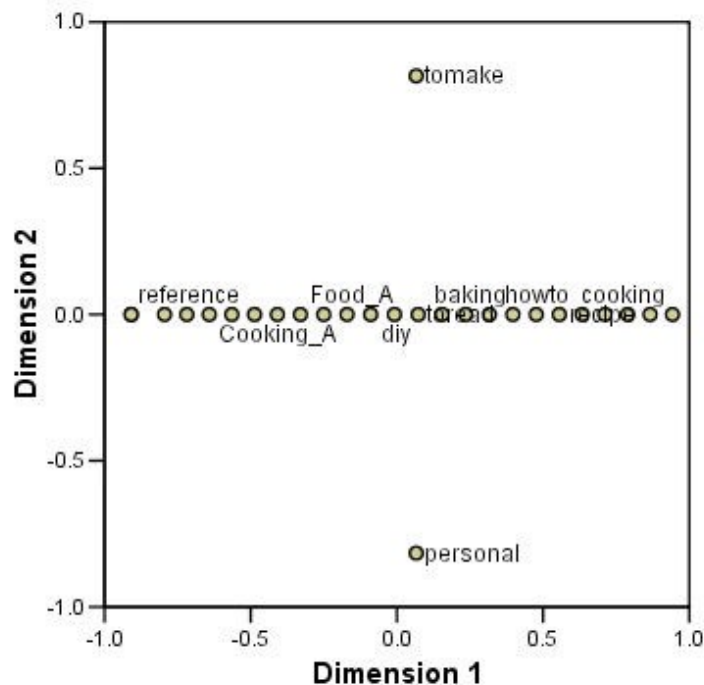
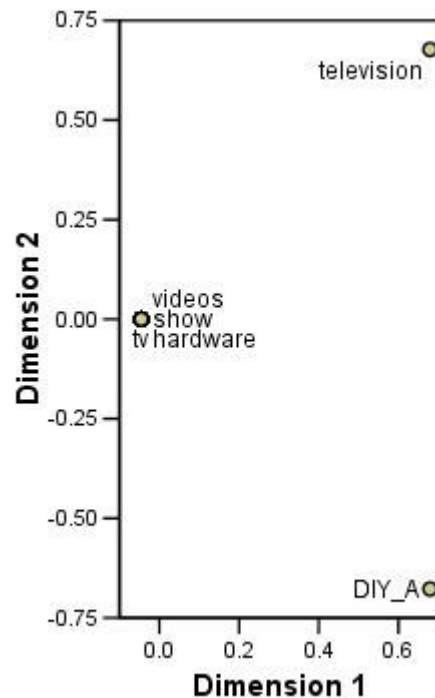**Figure 5: Co-word graph of www.101cookbooks.com/archives/001199.html (normalised raw stress = 0.26631)**

Figure 6: Co-word graph of www.gidut.tv (normalised raw stress = 0.27616)

In Figure 5, the majority of references are clustered in a long line across the graph with two outliers, both terms which would not appear in a traditional classification system. In this instance, it appears that terms in the line appear frequently with nearby terms. In Figure 6, all terms included in the co-word analysis occur in a very tight cluster at coordinates (0.04, 0) along with videos, show, tv and hardware. It is interesting to note that the term television is not part of this cluster at all, although it is the long form of tv.

Conclusion and Future Directions

While the tagging practices prevalent on del.icio.us would never be mistaken for professional indexing practice, some of the differences may be less significant than they at first appear. The fact that not all tag frequency graphs show an immediate rapid drop off does not actually indicate a substantial departure from classical classification trends. In fact, it it has commonly been accepted that documents are about multiple things and that the indexers' task is to assign appropriate descriptors to describe this aboutness. Thus, the difference in the shapes of the frequency graphs seems to indicate the perceived breadth of the aboutness of the document.

The data suggest that while the relationships between tags do not always follow the co-word clusters, they often follow relationships of synonymy that system designers can anticipate. Similarly, the presence of considerable consensus on certain terms suggests that emerging patterns from the tagging data are consistent, to some degree at least, with conventional indexing concepts of aboutness. In this sense, the data both supports earlier studies that detect predictable patterns (Golder & Huberman 2006), as well as suggestions that tagging patterns provide a useful "fast layer" of subject access that can filter down to slower, more deliberate layers (Morville 2006, 140).

However, the presence of time-related tags suggests that users are also doing something else with their tags: something that conventional subject access systems have always avoided. Tags such as "GTD," "To Read," and even "Cool" defy traditional subject analysis in several ways:

- They express a response from the user rather than a statement of the aboutness of the document;
- They are intrinsically time-sensitive;
- They suggest an active engagement with the text, in which the user is linking the perceived subject matter with a specific task or a specific set of interests.

These temporal tags could hardly be integrated into thesauri. The tag "To Read" expresses nothing about the document's subject. Nor is it likely to remain relevant for any length of time; the user will eventually read it, or never get around to reading it, making the tag an anachronism either way. They do, however, provide interesting evidence of the energy that an individual user throws against a knowledge structure.

Classification systems are traditionally visualized in two dimensions, offering coordinate relationships across an array, or hierarchical relationships along a chain. These two dimensions appear either as conventional tree-structure graphs, or as patterns of indentation in sequential lists of terms in thesauri or classification manuals. The temporal dimension of such systems manifests itself in the regular revisions of the system, generally through regular updates that are then consolidated into full-scale revisions.

User tags such as "to read" suggest that the user brings a new temporal dimension to a classification: one related not to long-term shifts in terms and their relationships, but rather to short-term needs and enthusiasms, which, by relating to a specific interest or a specific task, place the document in a set of relationships that, while not expected to endure, pull documents into idiosyncratic relationships.

If temporal tags were to become more sophisticated, their effect on subject access systems might be transformative. Vector space modeling offers us a method of modeling document similarities based on the presence or absence of terms by means of a hypercube: an n-dimensional space that offers dimensions corresponding to every term (Salton & McGill 1983; Meadow, Boyce & Kraft 2000). If temporal tags add a third dimension to classification systems, it is possible that some transformation of the vector-space model might allow us to find new ways of modeling the passing effects of specific tasks and interests on document relationships in a dynamic way that serves as a foundation for new forms of visualizing document relationships.

If this is true, we may well, as systems designers, want to be wary of solving problems too quickly. If we think of these tags as manifestations of this user energy, then the co-word visualizations acquire fresh implications. The distance between synonymous terms suggests the mutual exclusivity we expect from controlled vocabularies. But that doesn't necessarily mean that the system designer should replicate the controlled vocabulary by collapsing the two terms together. The difference between "information architecture" and "ia" could represent significant cultural differences, with "information architecture" representing an academic formality and "ia" representing the convenient acronym of a consultant working within the professional information architecture community. As Shirky reminds us (2005), there may well be cases in which people who watch films don't want to talk to people who watch movies.

This study does not definitively resolve the controversies surrounding collaborative tagging systems. Frequency and co-word analysis, however, provide some illuminating insights into the way tagging patterns emerge. They reveal that closely-related terms are not necessarily revealed through co-occurrence; they also reveal that users employ a wide variety of conventions in constructing tags: conventions which they apply inconsistently. Time will tell whether consistent conventions emerge. While clustering is present in many cases, it is not always clearly marked, suggesting that tagging, like many other indexing methods, resorts to multiple terms to describe the aboutness of documents. These results suggest two important directions for future research. First, they suggest that there is continuity between conventional indexing and user tagging: a continuity that could form the basis for a complementary system of subject access that could enrich conventional indexing rather than crowding it out. Second, the differences suggest that user tagging extends beyond the traditional objectives of subject access, and expresses a dynamic relationship between document and user, and between subject and task, which may lead to new ways of modeling subject access.

Acknowledgements

Notes

[1] It should be noted that the graphs produced by the multidimensional scaling have somewhat high normalized stress values. In co-occurrence studies (co-word, co-author etc.), this does not necessarily indicate a poor fit for the data. In fact, the qualitative aspects of the visual representation may be more important since multi-dimensional scaling is the process of projecting a multi-dimensional pattern on a space with fewer dimensions. A better picture of the closeness of the MDS to the actual geometric relationships between the tags would require combining the MDS data with other forms of statistical analysis such as factor analysis (see Leydesdorff and Vaughan, 2006).

References

Berners-Lee, T., & Fischetti, M. (2000). Weaving the Web: the original design and ultimate destiny of the World Wide Web. New York: HarperBusiness.

Bhattacharya, S, & Basu, P.K. (1998). Mapping a research area at the micro-level using co-word analysis. Scientometrics 43(3): 359-372.

Courtial, J.P. (1994). A coword analysis of scientometrics. Scientometrics 31(3): 251-260.

Courtial, J.P., Callon, M., & Sigogneau, M. (1984). Is indexing trustworthy? Classification of articles through co-word analysis. Journal of information science 9: 47-56.

Golder, S. & Huberman, B.A. (2006). The structure of collaborative tagging systems. Journal of the American Society for Information Science and Technology. (Forthcoming)

Johnson, S. (2001). Emergence: the connected lives of ants, brains, cities and software. New York: Scribner.

Leydesdorff, L., & Vaughan, L. (2006). Co-occurrence matrices and their applications in information science: extending ACA to the Web environment. Journal of the American Society for Information Science and Technology. (Forthcoming)

Meadow, C., Boyce, B.R. & Kraft D.H. (2000). Text information retrieval systems. 2nd ed. San Diego: Academic Press.

Morville, P. (2006). Ambient Findability: what we find changes who we become. Cambridge: O'Reilly.

Salton, G. & McGill, M.J. (1983). Introduction to modern information retrieval. New York: McGraw-Hill.

Schiffman, S.S., Reynolds, M.L. & Young, F.W. (1981). Introduction to Multidimensional Scaling: Theory, Methods, and Applications. New York: Academic Press.

Shirky, C. (2005). Ontology is overrated: categories, links and tags. [Online] [http://www.shirky.com/writings/ontology_overrated.html]

Waldrop, M. (1992). Complexity: the emerging science at the edge of order and chaos. New York: Simon & Schuster.

Weinberger, D. (2005). Taxonomies to tags: from trees to piles of leaves. Release 1.0 February 2005. [Online] [http://www.edventure.com/release1/abstracts.php?counter-4579726]

Wilson, C. (1999). Informetrics. Annual review of information science and technology 34: 107-247.

Wolfram, D. (2003). Informetrics and IR system content. Applied informetrics for information retrieval research. Westport, CT: Libraries Unlimited.