@toread and Cool: Subjec-
tive, Affective and Associa-
tive Factors in Tagging

Margaret E.I. Kipp
margaret.kipp@gmail.com
Palmer School of Library and Information Science
College of Information and Computer Science
Long Island University

Abstract

This paper examines the use of non subject related tags in social bookmarking tools. Previous studies of tagging determined that many common tags are not directly subject related but are in fact affective tags dwelling on a user’s emotional response to a document or are time and task related tags related to a users current projects or activities. These tags have been analysed to examine their role in the tagging process.

1. Introduction

Social tagging is increasingly a subject of interest in library and information science (and related fields) as social tagging tools such as del.icio.us, Flickr and LibraryThing have become increasingly popular. Some argue that the use of tags and appropriate tag visualisations to support the process of organisation and search in environments where other classification or subject indexing is absent. However others suggest that such terms do not aid in search or organisation due to the ample evidence of such symptoms of mob indexing as spelling variants and lack of synonym or vocabulary control.

Through simple visualisations of tags, such as sorting tags by frequency or displaying tag clouds, in which tag size denotes popularity, tagging systems form interesting new taxonomies or folksonomies of related terms that are broader in scope and terminology than those created by controlled vocabulary thesauri (Kipp and Campbell 2006; Hammond et al 2005). Additionally, social tagging can provide support for the storage of trails of associations as users progress through the search or browse process, mirroring Vannevar Bush’s associative trails (Bush 1945).

Previous studies of social tagging systems (Del.icio.us: Kipp and Campbell 2006; Golder and Huberman 2006 and CiteULike: Kipp 2006) all report that while most tags are subject related, there is often a small but significant core of tags which are not subject related at all, but in fact related to time, task, project or affect (emotional response). These tags raise interesting questions about the nature and purpose of tagging. While subject related tags can be seen to have obvious comparison to traditional controlled vocabularies and indexing systems, these non subject tags are more difficult to place within the framework of universal knowledge organisational systems. In fact, these tags suggest that users are doing more than just classifying material and are in fact engaged in a more holistic process of relating their tagged items to the entire context in which they are being used, not merely the subject.

This study examines the nature and use of non subject tags in three social tagging systems one aimed at the general public, Del.icio.us, and two aimed at academics, CiteULike and Connotea.

2. Background

As the personal computer has become, more and more, a part of everyday life, people have increasingly begun to store data electronically. Stored e-mail, personal documents and photographs all quickly grow to the point that it becomes difficult to find a particular item without a good organisational system.

This organisational problem is extremely familiar to library and information scientists and the solution initially has been to replicate the traditional hierarchical systems for information management (prior to faceted classification) on the computer in the form of the file system. While personal information organisation tends to be a smaller problem than organisation on the web, it will become of interest to indexers and classificationists as tagging systems become more popular. Tagging systems rely inherently on the same type of organisational skills that people apply (or do not apply) to the organisation of their own personal information.
Early research in the area of personal information management looked at how people organise documents in their offices. Malone (1983) found that people organised their desks not just to enable retrieval, but also to remind them of what they were working on. Additionally, people found that classifying documents was a cognitively difficult process, which made it easier to simply pile documents. Malone suggested that computer systems should provide automatic classification of files (by date at minimum) as well as space for piles of unclassified material (Malone 1983). This research is corroborated by other researchers who note that users find it easier to find things by recognising them than by searching for or remembering them (Bewley et al. 1987, 662; Sellen and Harper 2002, Kwasnik 1991). This phenomenon certainly explains the piles of project related material found on most desks. Bowker and Star discuss this phenomenon and also remark on the highly task-oriented folk taxonomies people develop for organising the things on their desk (Bowker and Star 1999, 2-3).

More recent work has concentrated on how people organise electronic documents using folders or labels on the computer. The Keeping Found Things Found Project (University of Washington) explores how people organise information on the computer in support of projects. Their study showed that folders were more than just a method of organising for later retrieval, which replicates Malone’s finding that people organise things for more than just findability. Folders also allowed people to break down a project into parts. They also found that folders showed a distinct tension between organising information for current use and later reuse (Jones et al. 2005). This suggests that classification actually helps people to understand the full extent of a project and organise its sub tasks.

Thus, research into personal information management shows that users want to do more than organise information by subject. They want and need to have a lot of contextual information about what they want to do with the information, what they did do with the information or even what they think they will do with the information. This contextual information describes the users’ interactions with the information and their thoughts about how it impinges on their lives and is associative in the sense used by Vannevar Bush (1945) rather than classificatory. In fact, evidence of time and task based management tags appears frequently in existing social bookmarking services (Kipp 2006a; Kipp 2006b; Kipp and Campbell 2006; Golder and Huberman 2006). Studies of social tagging tools suggest that there are differences between indexing as created by users versus trained indexers. Kipp (2006) examined tag use in CiteULike, a social bookmarking service for academics. While a majority of tags chosen by CiteULike users were indexing terms, related to indexing terms, a surprising number of these terms were not subject related at all. Terms such as toread and fun showed up in the sample. (Kipp 2006) These terms do not describe the aboutness of the document and would seem at first glance to be noise in the tag cloud. A study of Del.icio.us by Kipp and Campbell (2006) found similar results. While a majority of tags were subject related and, in fact, bore some evidence of the development of a decent consensus on the aboutness of the studied URLs, over 16% of the tags in this study were found to be non subject related. The majority of these non subject tags can be classified into two broad groups: affective tags and time, task or project related tags (a small subset of tags consist of prepositions, conjunctions and other parts of speech from tag phrases which were separated by the system into individual tags).

Affective terms consist of words that describe an emotional state. Rubin, Stanton and Liddy (2005) discuss the use of affective terms in text to discern the emotional slant of a text. Their work attempts to classify subjective evaluative terms in the text into positive or negative affect categories. Examples of positive affect terms are enthusiastic and excited. Examples of negative affect terms are dull and unhappy. (Rubin, Stanton and Liddy 2005) Time and task related tags consisted of compound words such as ‘toread’ and ‘todo’ and appeared to indicate a desire to combine information about tasks and activities with subject classification terms. Many of the time and task related tags examined in this study are of the form ‘toread’, ‘todo’, ‘tobuy’ and especially the many potential spelling variations associated with the term ‘toread’. These tags appear to indicate a desire on the part of users to more closely associate the task of classifying a subject and tying it to a concrete project or task. An analysis of these time and task tags along with affective tags and other non subject tags could shed additional light on the tagging phenomenon. As well, such an analysis could provide invaluable information on how users classify and organise information.

3. Research Questions

1. What patterns of user tagging activity emerge on examination of affective or time and task related tags?
2. How do users use time and task related tags or affective tags to indicate the value they see in a document?

3. What implications do the use of affective or time and task related tags have for the organisation of information?

4. Methodology

This study examines the use of non subject tags in three social bookmarking tools which do not fit the mould of traditional cataloguing and classification. These tags include two major categories: affective (emotional) tags and time, task or project related tags.

The three social bookmarking tools chosen for this study were Del.icio.us, CiteULike and Connotea. Del.icio.us is a social bookmarking service oriented towards any user. No special features are provided to encourage any particular group or the bookmarking of any specific type of item. CiteULike is a social bookmarking service designed for use by academics who wish to bookmark academic articles for later retrieval. Connotea is a social bookmarking service designed, like CiteULike, for academics. While CiteULike was originally quite strict in only allowing academic journals, Connotea allowed academics to store less scholarly material from the beginning.

Data was collected from Del.icio.us, CiteULike and Connotea. Posts in a social bookmarking tool consist of at minimum a title, URL and associated user name. A majority of posts (94% in Kipp and Campbell 2006) will have associated tags. A minority of posts will also contain a written description or note. The list of affective and time and task related tags used for this study was assembled from a number of sources. First, a study by Kipp and Campbell (2006) which examined patterns in tagging. Analysis of this data showed approximately 16% of tags were time and task related. Time and task or affective tags were located in multidimensional scaling graphs of cotag (coword) data. (Kipp and Campbell 2006) Additional tags were collected from a pilot study by Kipp (2006) examining the similarities and differences between descriptors, author keywords and user tags assigned to academic articles bookmarked in CiteULike. Despite the scholarly nature of this social bookmarking site, affective tags were located in the sample and time and task related tags were also part of the population. Additional affective tags were collected from Rubin, Stanton and Liddy (2005) on the subject of techniques for natural language processing of affective terms in text. This list is not an exhaustive list of either time and task or affective tags, but does provide a good preliminary examination of the phenomenon. Examples of affective tags include interesting, fun and cool. Examples of time and task related tags include @toread, todo, and tobuy. The full list of tags examined is in the Appendix.

Posts were collected from all three social bookmarking sites between October 20th and October 31st. Posts from each social bookmarking tool were collected in a single collection sweep lasting from 5-6 hours for CiteULike and Connotea to 40 hours for Del.icio.us. All posts using the tags from the list were collected and stored for later analysis.

5. Analysis and Results

5.1 General Results

A total of 78 tags were examined in this study. Of this number, 48 fell into the category of time, task or project related tags and 30 were affective tags. A majority (73) of the tags were in English; 5 tags were in French (lire, alire, @lire, acheter, amusant). A total of 1831 posts were collected from CiteULike, 2891 from Connotea and 198630 from Del.icio.us. This gives a total of 203352 posts in all from all three sites. Since the number of posts obtained from Del.icio.us is several orders of magnitude larger than the other two sites (del.icio.us has over a million users), data was normalised for comparisons.

A number of the tags in this study are very popular and appear on the respective popular or frequently used tag cloud pages for their sites. As of April 18th, 2008, the tags 'cool', 'daily', 'fun', 'funny', 'toread' and 'work' appear in Del.icio.us tag cloud, the tag 'and' appears in Connotea's cloud and the tag 'of' appears in CiteULike's cloud. Many of the affective terms were only lightly used in CiteULike and Connotea but appeared in Del.icio.us, no doubt due to the size of the respective populations and the nature of the different sites. Only one of the affective terms from Rubin et al (2005) was not used at all.

<table>
<thead>
<tr>
<th>Citeulike</th>
<th>Connotea</th>
<th>Del.icio.us</th>
</tr>
</thead>
<tbody>
<tr>
<td>fun, ToRead, todo, interesting, cool</td>
<td>fun, ToRead, important, unread, funny</td>
<td>fun, ToRead, funny, cool, interesting</td>
</tr>
</tbody>
</table>

Table 1: Most Popular Tags (top 5)

ToRead and fun were popular tags on all sites. The presence of the tag 'fun' on CiteULike and Connotea was initially a surprise, however, Connotea allows scholars to bookmark non scholarly materials and in any case it is certainly reasonable to expect...
dedicated scholars to find some scholarly material
fun or interesting.

5.2 Time, Task or Project Related Tags

The majority of time, task and project related tags
in the sample are variations on the 'toread' tag. This
is due to the relative difficulties in collecting data on
true project related tags, which could have different
meanings to different people (e.g. course codes).
Many of the variations on toread have very low usage
numbers on CiteULike and Connotea, in fact many
variations which are quite popular in Del.icio.us are
not used at all. The diversity of toread type tags in
Del.icio.us versus that in CiteULike and Connotea
does suggest that users of delicious are more highly
divided on how to write 'toread'. CiteULike provides
additional logic for tagging an item as toread and pro-
viding an interest marker of how interested you are in
reading it. Neither of the other services offers this as
a possibility. This may account for the relatively high-
er total number of toread type tags in Connotea.

Initially, the 'toread' seems to be a tag with very lit-
tle value outside of a single person's personal organi-
sational system, but collective patterns of interest
have been used in a number of situations to enhance
retrieval or access to systems. Google's PageRank
algorithm relies on user hyperlinks for its indexing
and ranking while Amazon's recommendation system
has shown that collective information about buying
patterns can be very useful for users who are inter-
ested in finding material that is like the material they
are currently reading or watching. This suggests that
the toread tag could function like a colleague's e-mail
suggesting that the article is interesting and worthy of
a little of your time. As a tag, it functions as an indica-
tor of interest.

Tags that appeared to be related to specific
projects, such as acronyms or tags which looked very
much like university course codes, were present in
both previous studies, however these tags were not
included in this study as finding them is often a hit or
miss proposition. Similar to the toread tags, though,
these tags could be highly useful for finding specific
information about specific projects or locating materi-
lial that other students or professors found useful for a
course.

5.3 Affective Tags

The affective tags were noted in the previous two
studies as an oddity in what appeared to be a rudimen-
tary distributed classification effort. Tags such as
cool or fun do not appear to add anything to the sub-
ject classification of an item and would also not seem
to be good candidates for search terms for informa-
tion retrieval.

Kipp and Campbell (2006) suggested that affect-
tive tags could represent an attempt by users to add
an additional personal aspect to classification. These
terms presumably indicate the user's emotional reac-
tion to the document, or perhaps the emotional reac-
tion the user expects to have after putting information
in the document into practice. These terms are obvi-
ously subjective and have thus far been excluded
from classification systems for this reason alone.
However, the use of such terms in social bookmark-
ning tools suggests that they are meaningful for users.

5.4 Non Subject Tags With Subject Tags

An analysis of subject tags in combination with
non subject tags shows that users of CiteULike and
Connotea do indeed find some scholarly articles to
be 'fun' or 'cool'. Especially fun were articles in the
realm of mathematics, physics and computation. In-
terestingly, the tag 'fun' was most commonly linked to
articles in the realm of physics, while 'toread' was
most commonly linked to articles in biology.

Non Subject Tags: Citeulike
Title: Symmetry and Self-Organization in Complex
Systems
URL: http://arxiv.org/pdf/cond-mat/0609274
taglist: automata, fun, graphs, mathematics, net-
works, statistical-mechanics, symmetry

Non Subject Tags: Connotea
Title: Foundations for engineering biology
URL:http://www.nature.com/nature/journal/v438/n7067/full/nature04342.html
While the latter tags express an emotional connection to the document, the former show evidence of a desire to attach personal information management information to documents. This desire to combine personal information management and document classification echoes findings in document use research at Xerox in which users categorised items in order to better understand their relationship to other items and to tasks the users wished to perform. (Sellen and Harper 2002) Use of these non subject tags suggests an active engagement with the text and show that users perceive the subject matter of their tagged documents as being contextually related to: a specific task, a specific set of interests or specific emotional reactions. Non subject tags express a dynamic relationship between users and documents, suggesting possible new ways of modelling information access.

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?

Libraries have begun to include social tagging systems either directly in their online public access catalogues (OPACs) or as addons to organise material that may be of interest to users. The PennTags project\(^1\) at the University of Pennsylvania (Allen and Winkler 2007), the Steve Museum project\(^2\) (Trant 2006) and LibraryThing for Libraries\(^3\) are specific examples of systems which combine traditional classification with social tagging. Examination of these systems as they develop will provide valuable insight in how users combine traditional classification and social tagging.

The examination of how users seek and use information is an important aspect of library and information science. Another important aspect of this is how they relate to information. (Bates 1998, 1048) Findings from this study suggest that users consider information within a contextual web of their own personal tasks, projects and emotional responses to everyday life.

**Acknowledgements**

Aspects of this study have been previously presented at the 2007 Information Architecture Summit in Las Vegas, Nevada and at the 2008 Visual Re-

---

1 http://tags.library.upenn.edu/
2 http://www.steve.museum/
3 http://www.librarything.com/forlibraries/

Proceedings of the 36th annual conference of the Canadian Association for Information Science (CAIS), University of British Columbia, Vancouver, June 5-7, 2008
References and bibliography


APPENDIX: Non Subject Tags Collected

Affective Tags
@cool
amusing
awesome
bastards
boring
cool
curious
exciting
favorite
favourite