Margaret E.I. Kipp (mkipp@uwo.ca)
Faculty of Information and Media Studies, University of Western Ontario, London, Ontario

Tagging Practices on Research Oriented Social Bookmarking Sites

Abstract: This paper examines the tagging practices evident on CiteULike, a research oriented social bookmarking site for journal articles. Tagging practices were examined using standard informetric measures for analysis of bibliographic information and term use. Additionally, tags were compared to author keywords and descriptors assigned to the same article.

Résumé : Cette communication examine les pratiques d'étiquetage par mots-clés qui sont utilisés sur CiteULike, un service d'étiquetage social, pour les articles de périodiques. Ces pratiques de marquage ont été examinées en utilisant les mesures informétriques habituellement utilisées pour l'analyse d'information bibliographique et d'utilisation de mots-clés. En outre, les étiquettes ont été comparées aux mots-clés utilisés par les auteurs et aux descripteurs attribués à ces mêmes articles.

1. Introduction

The ability to quickly locate relevant information is becoming increasingly important as more information becomes available digitally. Much of this information is unsorted and retrieval relies on free text search, user created hyperlinks and a large dose of serendipity.

Information organisation is a core area of library and information science dealing directly with the ability to increase the relevance of information retrieval by increasing the ability to at once collocate and distinguish material. In a digital world, one of the important tasks of library and information science is to reduce the difficulty inherent in searching large document spaces for information. A classification system using terms and keywords, appropriate to the context of the intended user, can help make the difference between a usable document space and a space which is difficult to navigate and find the information sought.

Universal hierarchical classification systems and subject specific taxonomies have a long history, but the design and application of these systems has largely been left to professional intermediaries such as librarians. As the amount of information available for user search increases and users begin to demand increasingly specialised information in search, these systems are often found to be at once too generic and too specific for user needs. Full text search, which can provide fine grained access to information has, however, the fault of doing so at the expense of precision resulting from the use of differing terminology.

User tagging and folksonomies created in a distributed fashion through social bookmarking sites have been suggested as a potential solution to these problems (Mathes 2004; Hammond et al 2005) since user tagging could provide the additional access points at less cost. However, this relies on many assumptions, such as the assumption that user tagging provides a similar or better search context to free text searching or intermediary assigned index terms.

1

This study builds on a previous study (Kipp 2006) examining the emerging phenomenon of social bookmarking or tagging in comparison to existing classificatory structures from traditional cataloguing and classification research. A sample of articles from the field of library and information science was examined for contextual differences in keyword usage between users of social bookmarking sites and authors and intermediaries (cataloguers or indexers). This study found many similarities and some intriguing differences in context, specifically in the realm of personal information management. Users tagging articles on social bookmarking tools tend to use terms such as 'toread' and 'todo' to indicate their interest in further use or study of an item. (Kipp 2006) A study of del.icio.us found that approximately 16% of tags in the sample were time and task related tags having a personal information management edge. (Kipp and Campbell 2006) Additional differences included the fact that "intermediaries considered geographic location to be an important part of the description of the aboutness of an article, authors and users tended to assume it was somewhat less important than the other contexts of the articles." (Kipp 2006) Many tags were related to terms in the formal thesaurus from which the descriptors were located, but were not formally in the thesaurus. In some cases this was due to new or emerging terminology, in others to material being used in related but different areas of a field (e.g. information seeking versus information retrieval). (Kipp 2006)

The current study expands upon the findings from this earlier study using a larger collection of articles from the field of biology tagged by users of CiteULike (http://CiteULike.org/), social bookmarking site which is specialised for academic articles. The chosen journals were restricted to journals known to request author assigned keywords and to journals indexed in Pubmed, which provides intermediary assigned controlled vocabulary for searchers. Thus, each article in the study has three sets of keywords assigned by three different classes of metadata creators. As in the previous study, the data will be analysed using thesaural comparisons for depth of specificity at various levels as well as statistically for term usage and frequency.

Analysis of this new data set from a different field will help to strengthen the conclusions of the earlier study by showing that users in different fields also provide useful sets of tags. This study has implications for the design of systems for accessing, indexing and searching document spaces.

2. Social Bookmarking Tools
Social Bookmarking sites have become increasingly popular since their inception. Sites such as del.icio.us report over a million users with additional users signing up every day. (http://blog.del.icio.us/blog/2006/09/million.html) Interest is increasing in academic circles. In particular, researchers from library science and computer science examine the growth of an Internet phenomenon with potential applications to both fields. (Voss 2007; Kipp 2006; Kipp and Campbell 2006; Hammond et al. 2005). One of the most interesting aspects of social bookmarking sites is the phenomenon of social tagging that has grown along with them as users are encouraged to provide a few key terms they consider most useful in categorising the item they are bookmarking.

Tagging, which began on social bookmarking sites like del.icio.us, allowed users to store their bookmarks (favourite URLs) in a publicly accessible fashion and associate these bookmarks with a series of descriptive tags the user thought might be helpful in aiding

the process of finding the URL again. Early adopters found that the automatic clustering of bookmarked URLs by their associated tags led to the discovery of other useful URLs on similar topics. (Shirky 2005) The number of sites utilising user tagging as a form of information organisation is increasing and tagging is beginning to be integrated into web sites with more traditional hierarchical organisational systems such as on-line book stores (e.g. Amazon.com).

CiteULike (http://CiteULike.org/) is a social bookmarking service specialised for use by academics who wish to bookmark academic articles for later retrieval. CiteULike was created by Richard Cameron in November 2004. (http://www.CiteULike.org/faq/all.adp)

Figure 1: Screenshot of CiteULike

Similar to the more commonly known del.icio.us, CiteULike allows users to assign an arbitrary number of tags to the articles in their library. Users may search by tag to relocate articles in their own library, as well as in the libraries of other users. User and overall tag clouds allow users to see commonly used or popular tags for an article or for the entire tool.

Since CiteULike tags are often associated with journal articles, it is possible to collect author keywords and descriptors for many of the articles. Thus, a comparison can be made between user tags, author keywords and intermediary descriptors attached to a single article.

3. Related Studies

Bowker and Star (1999) suggest that classification is a basic practice of all humans. (Bowker and Star 1999) Traditional classification methods have tended to rely on trained indexers, cataloguers or taxonomists to organise and describe information. While other groups have been involved in creating keywords or index terms (for example, journal article authors who are asked to provide a certain number of keywords with their submitted articles), these keywords generally have a small circulation and are not widely used. Such small scale indexing is common but generally covers a narrow range of topics and is specific to the article. Additionally, such keywords are often derived from the work itself and may or may not have wide circulation outside a small subset of the field. Collaborative tagging systems such as CiteULike allow users to publicly participate in the classification of journal articles.

To discover if tags can truly provide a useful replacement or enhancement for controlled vocabularies, it is important to examine whether or not they provide a similar contextual dimension to the existing classification systems. While it seems unlikely that untrained users will produce a full featured classification system similar to the traditional library systems, it is possible to examine the tags they do assign to see how they compare to the descriptors assigned by a trained indexer and to keywords assigned by authors.

Adam Mathes (2004) notes that there are three major groups that are commonly involved in the classification of documents. These groups are authors, intermediaries and users. (Mathes 2004) While intermediary index terms (often subject headings) have been widely promulgated, author keywords and user terminology have tended to be relatively local. In fact, author keywords have received relatively little attention in the literature. (Kipp 2006; Ansari 2005; Voorbij 1998) While intermediaries have been indexing documents for some time, the development of large scale user created collections of tagged documents is new.

This leads one to ask if user categories are indeed different from subject headings or author keywords and if so, how they differ? Are there differences in context, type, or some other semantic relationship? If so, it could be quite important to examine the differences between these categories and the reasons that they do not appear in traditional classification systems. Perhaps these categories are considered to be too short term, too user centric or too subjective to be included? Terms such as @toread and cool after all, do not describe the aboutness of a document and would seem to be of little use in the organisation and retrieval of information. Yet, they are an important part of the phenomenon of tagging. (Kipp 2007) These short term and highly specific tags suggest important differences between user classification systems and author or intermediary classification systems.

Descriptive statistics can be used to make a basic comparison of the indexing practices of each of the three groups involved in the classification of journal articles (users of a document, authors of a document, and intermediaries or indexers of a document). Additionally, a comparison can be made at the level of the assigned metadata itself. Tags can be examined to see how well they fit the aboutness of the document and to see how closely they match the existing descriptors and author keywords already assigned to the documents.

A few studies have made comparisons of different types of keywords. Voorbij (1998) studied the correspondence between words in the titles of monographs in the humanities and social sciences and librarian assigned descriptors existing in the online public access catalogue of the National Library of the Netherlands. His study used the different relationships in a thesaurus as an indication of closeness of match, beginning with an exact (or almost exact) match, continuing to synonyms, narrower terms, broader terms, related terms, relationships not formally in the thesaurus, and terms which did not appear in the title at all. (Voorbij 1998, 468) A similar study by Ansari (2005) examined the degree of exact and partial match between title keywords and the assigned descriptors of medical theses in Farsi. She found that the degree of match was greater than 70 per cent. (Ansari 2005, 414) Both studies suggest that title keyword searching alone and controlled vocabulary searching alone lead to failure to find some articles. However, there is very little research in this area. Consequently, this study continues to examine the question of convergence between tags, keywords and descriptors by exploring the tagging phenomenon as it is growing at CiteULike.

This study posed the following research question:

- To what extent do term usage patterns of user tags, author keywords and intermediary descriptors suggest a similar context between users, authors and intermediaries?

4. Methodology

This study builds on previous work (Kipp 2006) which examined three forms of index term creation originating from three different groups: users of a document, authors of a document and intermediaries or indexers of a document. In Kipp (2006) it was found that while users often did use terms which were directly from the thesaurus used to assign descriptors to the articles, terms were also often similar or related terms which were not formally linked in the thesaurus. The most prominent example was the use of information retrieval versus information seeking (related but distinct areas of research). Additionally, users tended to include personal information management terminology such as 'toread' in their tag sets, but were less likely to include geographic information. (Kipp 2006) While the findings from the preliminary study showed that there were differences in the way users, authors and intermediaries classified documents, the size of the data set--165 articles--made it difficult to generalise these findings to larger data sets from other fields. A larger data set, from a different field, which showed similar patterns of term usage and thesaural matches would strengthen conclusions from the earlier study.

Tag data for the current study was collected from CiteULike between January 12, 2007 and January 24, 2007 via a python script (CiteULike.py). Author keywords and descriptors were collected from on-line journal databases and Pubmed respectively using additional python scripts.

Journals selected for this study were chosen because they are: a) biology related, b) require authors to submit keywords for their articles and c) are indexed in Pubmed using Medical Subject Headings (MeSH). Two journals were selected for this study: Proteins and Journal of Molecular Biology. All articles from these selected journals, which have been tagged on CiteULike by at least one user, were collected. To ensure that all articles from these journals were collected, the python script was designed to collect under all common variants of their names (e.g. J. Mol. Biol. for Journal of Molecular Biology). (These results were parsed to exclude currently untagged articles. To aid in the location of new articles, CiteULike also provides listings for articles from selected journals that have not yet been tagged.)

Data collected included title, journal name, volume, issue, page numbers, author names, abstract where available, and URLs providing access to the article or its abstract. URLs were collected for each article and automatically separated into categories as potential sources of keywords or descriptors. Digital Object Identifiers (DOIs - http://www.doi.org/) were selected by preference as a source of author keywords for journal articles and Pubmed URLs were used to locate descriptors (in this case MeSH indexing terms).

All articles were then located in Pubmed and on publicly available abstract pages from on-line journal database sites using the URLs collected from CiteULike. Where possible, pubmed URLs and DOI URLs were used directly, otherwise a series of scripts was used

to locate pubmed URLs given the DOI, the DOI given the pubmed ID or, in extreme cases, Google Scholar was used to locate articles using the article title and other bibliographic information. A total of 19 items could not be located on Pubmed, via a DOI (all had at least a DOI or a Pubmed ID) or on Google Scholar. These 19 were excluded from the following study.

This resulted in a total of 1083 articles for analysis. Since many articles were tagged by more than one user, this resulted in a total of 1588 posts with tag lists for analysis.

Journal Name	Number of Articles	Number of Posts
Journal of Molecular Biology	649	931
Proteins	434	657
Total	1083	1588

Table 1: Journals with author assigned keywords

In the end, each article selected for this study had 3 sets of keywords assigned by three different classes of metadata creators. The data was stored in a MySQL database and preliminary informetrics analysis was done using SQL scripts as suggested by Wolfram (2005). Descriptive statistics and basic informetric data were collected to provide a good picture of the scope of the collected data. Additionally, a sample of highly tagged articles was selected to have its tags, keywords and descriptors examined for term usage.

5. Results
5.1 Authors, Users and Journals
Bibliographic data for a total of 1083 articles was collected from CiteULike. This data set included all articles tagged by at least one user from the journals: Proteins and Journal of Molecular Biology. The data set thus contained a total of 1588 posts.

Unique user names present in the sample totalled 239. Due to the use of user selected user names and the fact that it is possible to sign up for an account under different e-mail addresses, it is not possible to ensure that these are indeed 239 distinct persons.

Each user name was associated with at least one post in the data set. One user had posted 94 of the 1588 collected posts. Many other users had posted significantly fewer posts. A total of 94 users (39%) had posted only one post in the data set. Of the users who posted more frequently in this data set, 42 (18%) posted 10 or more times.

Username	Number of Articles Posted
ana	94
barry	65
marcius	64
bicko	44
lna	43

Table 2: Top 5 Taggers

A similar drop off can be seen in the data set when examined based on the number of users who have posted a link to a specific article. In this case, the maximum number of users per article was 14, the minimum 1, and the median 2.

Number of Users/Posts	Article Title
14	Principles of docking: An overview of search algorithms and a guide to scoring functions.
7	Comparing protein-ligand docking programs is difficult.
6	Protein flexibility predictions using graph theory.
6	Binding MOAD (Mother Of All Databases).
6	The Relationship between the Flexibility of Proteins and their Conformational States on Forming Protein-Protein Complexes with an Application to Protein-Protein Docking

Table 3: Number of users who posted a link to a specific article

In fact, the number of users who posted more than one article dropped off quite quickly (799 articles were posted only once, median was 1 post per article). This matches findings from citation analysis which show that a few articles tend to be highly cited while many others are infrequently cited.

The number of authors per article collected ranged from a maximum of 48 authors to a minimum of 1. One article had 48 authors while 61 articles had 1 author. Over 80% of articles had between 2 and 5 authors. This is to be expected since scientific articles tend to have more authors.

5.2 Tags, Keywords and Descriptors
The total number of descriptors in the sample was found to be extremely high. This is due to the fact that Pubmed articles tend to have many descriptors assigned to increase recall, precision and relevance when searching pubmed.

	Tags	Keywords	Descriptors
Unique	1136	3181	2746
Total	3788	4866	12473

Table 4: Number of indexing terms of each type

Additionally, Pubmed descriptors include both major and minor descriptors covering as many aspects of the work as possible. This finding suggests that Pubmed's descriptors are likely to provide a very thorough description of the article in question. The ratio of unique terms to total terms is highest for author keywords. This supports findings from the previous study in which author keywords were found to be more diverse than tags or descriptors. Author keywords were also less likely to match tags or descriptors. (Kipp 2006)

Many tags, keywords and descriptors occurred frequently in the collected data.

The most popular tag was 'protein_structure', used 140 times; the most popular keyword was 'protein folding', used 58 times; and, the most popular descriptor was 'Models, Molecular', used 649 times in the data set.

Frequency	Tag
140	protein_structure
114	no-tag
114	protein
103	structure
97	docking

Table 5: Most commonly used tags

A total of 645 tags were used only once in the data set and 185 tags were only used twice. The median number of times a tag was used in the data set was 1.

In comparison, author keywords were much more diverse with 2548 of the keywords being used only once once in the data set. The maximum number of times a keyword was used was 58, minimum 1 and median 1. As previously noted in Kipp (2006) the author keywords were less likely to match descriptors or tags suggesting that there is a distinct difference between the context of the user of the article and the author of the article. (Kipp 2006)

Frequency	Author Keywords
58	protein folding
49	protein structure
46	molecular dynamics
38	protein structure prediction
31	docking

Table 6: Most commonly used author keywords

Descriptors were heavily reused in the data set, with some descriptors being used hundreds of times. The maximum number of times a descriptor was used in the data set was 649, minimum 1 and median 2.

Frequency	Descriptors
649	Models, Molecular
511	Protein Conformation
388	Proteins
306	Amino Acid Sequence
280	Binding Sites

Table 7: Most commonly used descriptors

Out of a total of 2746 unique descriptors, 731 descriptors were used only once and 249 were only used twice. This is a higher reuse rate than that for author keywords.

When examined at the article level, there are similar patterns of usage of tags, keywords and descriptors. While some articles were highly tagged, the majority had only a few tags. The maximum number of tags assigned to an article was 29, minimum 1 and median 2. The article with 29 tags was tagged by 14 users, suggesting that this is still an example of users assigning some 1-3 tags to an article.

Frequency	Article Title
29	Principles of docking: An overview of search algorithms and a guide to scoring functions.
20	Binding MOAD (Mother Of All Databases).
19	Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function.
18	How different amino acid sequences determine similar protein structures: The structure and evolutionary dynamics of the globins
18	Using a neural network and spatial clustering to predict the location of active sites in enzymes.

Table 8: Number of Tags per Article (top 5)

An examination of the number of tags per post (an article may be posted multiple times thus generating multiple posts per article) shows smaller numbers of tags. The maximum number of tags per post was 15, minimum 1 and median 2.

Similarly, the maximum number of keywords found for an article in the data set was 13, minimum 1, median 5. One reason why the median number of keywords is higher than for tags is due to the fact that many journals have a set number of author keywords they request, often 5 or 6.

Frequency	Article Title
13	Automated prediction of domain boundaries in CASP6 targets using Ginzu and RosettaDOM.
13	Automated prediction of CASP-5 structures using the Robetta server.
11	Structure modeling, ligand binding, and binding affinity calculation (LR-MM-PBSA) of human heparanase for inhibition and drug design.
11	Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model
10	Minimizing false positives in kinase virtual screens.

Table 9: Number of Keywords per Article (top 5)

The total number of descriptors users in the data set was 12743, but the number of unique descriptors was only 2746. An examination of the number of descriptors per article shows that many articles had a much larger number of assigned descriptors than either tags or

keywords. The maximum number of descriptors assigned was 36, minimum 2, median 11. This high median suggests that Pubmed indexers attempt to provide as broad a list of relevant descriptors as possible to aid in information retrieval.

Frequency	Article Title
36	Crystal structure of cone arrestin at 2.3A: evolution of receptor specificity.
30	G-protein-coupled receptor domain overexpression in Halobacterium salinarum: Long-range transmembrane interactions in heptahelical membrane proteins.
29	A Snapshot of Viral Evolution from Genome Analysis of the Tectiviridae Family.
28	Computer-assisted identification of cell cycle-related genes: new targets for E2F transcription factors,
27	Catalytic Independent Functions of a Protein Kinase as Revealed by a Kinase-dead Mutant: Study of the Lys72His Mutant of cAMP-dependent Kinase

Table 10: Number of Descriptors per Article (top 5)

An interesting measure for examining term usage in tagging is the measure of user vocabulary length, most often used to analyse search query logs. (Wolfram 2005) This data represents all the tags used by a specific user in the data set. The largest user vocabulary length in the data set was 62, the smallest 1 and the median 2. This suggests that most users tend to use a small number of tags (as noted in previous studies), while a small number of users will use more tags.

When the user vocabulary length is broken down at the individual article level, the largest length was 15 tags for one article.

User	Max tag list length	Min tag list length	Number of articles posted
3109	7	2	15
3063	6	1	73
4068	15	2	9

Table 11: User Vocabulary Length by Article

5.3 Term Usage

Examining the tags from a specific article (788), "Computer modeling 16 S ribosomal RNA", it was noted that 9 tags were applied to the article. Two of the tags came directly from the title, namely 'rna' and '16s'. It is interesting that taggers chose to use the term 'algorithms' rather than a term like 'computer modeling', which was used for other items in the data set, despite the fact that computer modelling is a term from the title. In fact 'computer modeling' is one of the author keywords for this article and the term 'computer simulation' occurs in the descriptor list.

Additional terms that do not come directly from the title were 3d, prediction, distance_geometry, bioninformatics, structure and structure_prediction. The term bioinformatics is an excellent example of an extremely generic term for computer modelling and analysis as related to biology, which one would not necessarily expect in

10

the descriptor list since it would likely be a broader term. Seen across all three sets of indexing terms are variants on '16s rna'.

Tags	Keywords	Descriptors
3d	16 S RNA	Base Sequence
algorithms	ribosome	Computer Simulation
prediction	computer modeling	Cross-Linking Reagents
rna	distance geometry	Escherichia coli
16s		Models, Molecular
distance_geometry		Molecular Sequence Data
bioinformatics		Nucleic Acid Conformation
structure		RNA, Ribosomal, 16S
structure_prediction		

Table 12: Tags, Keywords and Descriptors for Article 788

6. Discussion and Conclusions

This preliminary examination of a larger data set from CiteULike shows that the results from the previous study (Kipp 2006) using a smaller data set from library science are relevant to other fields and to larger data sets.

As previously discussed in Kipp (2006), Kipp and Campbell (2006) and Hammond et al (2005), users tend to use some terminology which is rare or completely absent from author keyword lists or descriptor lists. Time and task related terminology were present in the current study as well as earlier studies. Terms such as 'to_be_tagged', 'toread' and a number of calendar dates (e.g. 31/03/06, 1998) were found as tags assigned to articles in this data set.

Many user terms were found to be related to the author and intermediary terms but were not part of the formal thesaurus used by the intermediaries and, thus, not formally linked to the intermediary terms. In some cases this was due to faceting of terms for example 'protein' and 'structures' used separately in the tag lists where they were linked in the thesaurus or the use of abbreviations such as 'PDB' for 'Databases, Protein'.

Terms such as 'human', 'animal', and 'family-studies' showed that users tagging biology related articles are extremely interested in methodology and user groups associated with articles. This is distinct from the previous study where such terms were more common in the descriptors unless they described extremely specific kinds of methodologies, such as 'pubmed-mining' for data-mining of Pubmed. (Kipp 2006)

This study has implications for the design of systems for accessing, indexing and searching document spaces. The popularity of Google has demonstrated that users prefer to be able to search for items in a more natural way using one interface to locate items of a varied nature. However, users also express frustration at being unable to locate items or narrow their search results from a huge search set (e.g. 300 000 hits on Google). (Campbell and Fast 2004) Controlled vocabularies help to narrow a search set to a manageable size, but controlled vocabulary usage can be expensive.

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: by association rather than by categorisation. (Bush 1945) 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 and task relationships as well as other associations which are new and novel.

References

Ansari, Mariam. 2005. Matching Between Assigned Descriptors and Title Keywords in Medical Theses. *Library Review* 54(7), 410-4.

Bowker, Geoffrey C.; Star, Susan Leigh. 1999. *Sorting Things Out: Classification and its Consequences*. Cambridge, MA: MIT Press.

Bush, Vannevar. 1945. As We May Think. *The Atlantic Monthly*, July 1945. Available from http://ccat.sas.upenn.edu/~jod/texts/vannevar.bush.html

Campbell, D. Grant, and Karl V. Fast. 2004. Panizzi, Lubetzky, and Google: How the Modern Web Environment is Reinventing the Theory of Cataloguing. *Canadian Journal of Information and Library Science* 28(3), 25-38.

Hammond, Tony, et al. 2005. Social Bookmarking Tools (I): A General Review. *D-Lib Magazine* 11(4). Available from http://www.dlib.org/dlib/april05/hammond/04hammond.html

Kipp, Margaret E.I. 2006. Complementary or Discrete Contexts in Online Indexing: A Comparison of User, Creator, and Intermediary Keywords. *Canadian Journal of Information and Library Science*. Available from http://eprints.rclis.org/archive/00008379/ (forthcoming)

Kipp, Margaret E.I. and Campbell, D. Grant. 2006. Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices. *Proceedings of the 2006 Annual Meeting of the American Society for Information Science and Technology, Austin, November 3-8, 2006*. Available from http://eprints.rclis.org/archive/00008315/

Kipp, Margaret E.I. 2007. @toread and Cool: Tagging for Time, Task and Emotion. *8th Information Architecture Summit, March 22-27, Las Vegas, Nevada*.

Mathes, Adam. 2004. Folksonomies - Cooperative Classification and Communication Through Shared Metadata. Adammathes.com. Available from http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html

Shirky, Clay. 2005. Ontology is Overrated: Categories, Links, and Tags. Clay Shirky's writings about the internet. Available from http://shirky.com/writings/ontology_overrated.html

Voorbij, Henk J. 1998. Title Keywords and Subject Descriptors: A Comparison of Subject Search Entries of Books in the Humanities and Social Sciences. *Journal of Documentation* 54(4), 466-76.

Wolfram, Dietmar. 2005. Applications of SQL for Informetric Data Processing. *Proceedings of the 33rd conference of the Canadian Association for Information Science, London, June 2-4, 2005*. Available from http://cais-acsi.ca/proceedings/2005/wolfram_2005.pdf

Voss, Jakob. 2007. Tagging, Folksonomy & Co-Renaissance of Manual Indexing? Available from http://arxiv.org/abs/cs.IR/0701072v2