

Agatha S. Donkar. I'm In Ur Bookmarks, Stealin' Ur Tags!: Closed Communities and Their Influence On Consistent Vocabularies In User Developed Folksonomies. A Master's Paper for the M.S. in L.S. April, 2007. 35 pages. Advisor: Jane Greenberg.

Metadata technology allowing users to create and modify their own personal descriptive metadata for World Wide Web pages has also given rise to similarly-interested communities of web users registered at sites such as Delicious, who are refining their own content vocabularies. This research examined these vocabularies to determine if trends, patterns and unspoken vocabulary policy exists amongst the users. This study extracted data from Delicious' URL history pages and analyzed the data via content analysis. The research found that vocabulary consistency exists within the community, despite individually and independently generated data. The analysis was based specifically on type and content descriptor identifiers for the Stargate: Atlantis fandom, which is a community of fans of the Sci Fi television show.

#### Headings:

- Cataloging -- Electronic Data Archives
- Standardization of bibliographic records -- Metadata
- Standardization of bibliographic records -- Internet -- Cataloging
- Information systems -- Cataloging
- Bookmarks (Internet) -- Tagging (Internet)

I'M IN UR BOOKMARKS, STEALIN' UR TAGS:  
CLOSED COMMUNITIES AND THEIR INFLUENCE ON CONSISTENT  
VOCABULARIES IN USER DEVELOPED FOLKSONOMIES

by  
Agatha S. Donkar

A Master's paper submitted to the faculty  
of the School of Information and Library Science  
of the University of North Carolina at Chapel Hill  
in partial fulfillment of the requirements  
for the degree of Master of Science in  
Library Science.

Chapel Hill, North Carolina

April 2007

Approved by

---

Jane Greenberg

## I. Introduction

With the advent in the early 21<sup>st</sup> century of "tagging"<sup>1</sup> supported by social software, and of the sudden ability of the generic World Wide Web (web) user has become involved in preserving, cataloging and classifying information resources (in the form of hypertext websites, digital photographs, and electronic audio files). Despite this development, the volatile nature of the World Wide Web remains a problem. Any Internet resource could disappear, be removed from the World Wide Web by the resources' original creator, and given these hurdles, cataloging attempts in the style of the Library of Congress cataloging for monographs will never be viable for WWW resources .

However, given the new technology available in such social bookmarking hypertext websites such as Del.icio.us [<http://del.icio.us>; henceforth Delicious], the base user of the World Wide Web is able to create his or her own catalog; a catalog using personal vocabulary, allowing the user to save and locate websites of importance to him or her. Social bookmarking<sup>2</sup> and the tags that form the baseline search structure of these websites are in the process of revolutionizing the cataloging and classification genre, making classification an easily understood, easily used file-saving structure, and this

---

<sup>1</sup> **Tagging:** the action of associating identifying words or phrases with information (hyperlinks, photographs, etc) in a public, searchable database.

<sup>2</sup> **Social bookmarking:** World Wide Web sites that allow users to save ("bookmark") World Wide Web pages to a publicly viewable database; social bookmarks are generally identified with "tags" and are available on any computer which is connected to the Internet.

development will continue to change the way information on the internet is stored and searched in the future.

Delicious, the primary focus in this proposal, is defined by its creator as:

[A] social bookmarking service that allows users to tag, save, manage and share Web pages from a centralized source. With emphasis on the power of the community, del.icio.us greatly improves how people discover, remember and share on the Internet. (Schacter, 2004)

This resource currently has more than 1 million registered users, and more than 50 million individual links saved by the user collective. With one of the fastest growing user bases on social software websites, Delicious was the first social bookmarking site of its kind (other sites, such as furl.net and digg.com, have since been launched and developed their own individual user bases), and remains the benchmark for "collective information classification" World Wide Web sites. It is the registered user base at Del.icio.us on which this study will be focusing its investigation. The research reported on in this paper addresses the question of about the relationship between the user generated tags on Delicious and the community to which the user belongs, drawing a sample from Delicious. The following section includes a literature review that discusses metadata generation, classification theory, and social software and folksonomies. This is followed by a presentation of the research questions, results and a discussion, and the final section includes a conclusion with recommendations for further study.

## **II. Literature Review**

Since the World Wide Web became readily available to the general public in the early 1990s, librarians and information managers alike have asked questions about cataloging and classifying the information available via hypertext web pages, FTP sites, and public domain electronic resources.

### **Author Generated Metadata**

The Dublin Core Metadata Initiative [<http://www.dublincore.org>] exists to monitor and control specific metadata standards. Metadata can be placed in World Wide Web pages, embedded in the HTML headers or in the document body itself, via XML [Extensible Markup Language]. Metadata that is created for World Wide Web pages seeks to provide indexing assistance to search engines such as Google. World Wide Web search engines index World Wide Web sites for easier retrieval by users when searching. The indexes created by search engines are not viewable or even remotely accessible by users; the indexes are computer files that hold data about web site contents, but they are only machine readable and are machine created, so unlike human-created text indices, search engine indexing is not necessarily accurate.

In the last three years, web sites that purport to be part of the Web 2.0 movement - a movement inviting the web's enormous user base to participate in content creation, cataloging and classification of web resources - have introduced a practice that is known as "tagging", adding searchable, grouped keywords to World Wide Web content. Tagging is the natural extension of author-generated metadata, and is a new format of metadata: user-generated metadata.

The research that has been done on the subject of author-generated metadata dates back to the late 1990s, and informs and directs the research that is being done in the field of user-generated metadata; early adopters of the user-generated metadata tools such as Del.icio.us [<http://del.icio.us>], a social bookmarking site that allows users to share World Wide Web site bookmarks with the general public, were frequently World Wide Web authors who had *used* and *created* author-generated metadata before the advent of user-

generated metadata.

User-generated metadata is the process of creating metadata about web sites originally authored by users who are not the metadata creators; web sites like Del.icio.us that contain millions of saved links, each marked with individual tags that describe the web sites' content, are coming closer and closer to creating an actual index of the World Wide Web. The process is accessible and keyword-searchable indices are beginning to form via user-generated metadata.

In the 2001 paper "Author-generated Dublin Core Metadata for Web Resources: A Baseline Study in an Organization", Greenberg et al examined the ability of resource authors to create acceptable metadata for their own resources in specific organizational contexts, specifically with authors applying Dublin Core Metadata standards to the resources. Greenberg et al state that "the design of a simple form, with selective use of features, may be the best means for author-generated metadata" (2001). Author-generated metadata may be created by the authors, who are not the "metadata professionals" studying the subject with understanding of both the procedure and the tools of creation, but the authors can in fact create metadata "equivalent to that of a metadata professional" if provided with an appropriate framework within which to create the metadata.

This paper is interestingly applicable to the ideas behind user-generated metadata, though it was written three years before the advent of Del.icio.us [henceforth Delicious]. Greenberg et al focus heavily on the tools available to metadata creators, specifically the authors, and the user-generated metadata arenas that exist are specifically free-form arenas. The quality of metadata in Greenberg et al is high, because of control; the lack of

consistency in user-generated metadata could be construed, via this paper, to be a distinct flaw in the user-generated metadata system.

Zhang and Jastram (2006) sought to examine the metadata publishing behavior and response of professional publishing companies on the World Wide Web. Their results showed that the use of metadata in any domain has not yet standardized, and authors are inclined to either over or under utilize the ability to create easily searchable tags for their web pages with metadata. Zhang and Jastram's findings are interesting in comparison to Greenberg et al because of the definite and opposite findings; unlike Greenberg et al, Zhang and Jastram studied the metadata created "in the wild" - author-generated metadata that was not created in a controlled environment.

Their findings state that "there is no centralized control over the quality or content of either the visible text or the embedded metadata". This finding is directly applicable to the generated of user metadata; control is the key element in these studies of author-generated metadata, and the lack of control in user-generated metadata is a key problem in the creation of accurate metadata. Spammers who use Delicious, trying to promote their own sites on the "popular" or "recent" pages of Delicious, may tag web sites with highly irrelevant metadata that simply is used frequently for relevant tagging - this only leads to the creation of irrelevant or incorrect metadata as outlined in Zhang and Jastram.

In Hunter (2001), the author outlines a specific plan for creating a metadata thesaurus that would allow for metadata access across platforms. Hunter suggests that the system would lead to "efficient, scalable machine-translations across heterogeneous metadata descriptions" and discusses possible merging of projects to create a single "super" ontology. Hunter's paper specifically digs into problems with large-scale metadata

systems, discussing the "inherently ambiguous nature of language" - applied to a system such as Delicious, where vocabularies are not controlled or even partially agreed upon, Hunter's listed problems are even more applicable. Not only are the differences in American and British or Canadian English an issue in creating standard metadata, but they are an even greater problem in uncontrolled user-generated metadata.

The difficulty of determining the usefulness or necessity of a specific metadata term comes into play when an attempt to apply the Harmony project's ideals to Delicious; the more important conclusion that can be drawn from Hunter's paper and applied to user-generated metadata are the problems inherent in user-generated metadata - the knowledge of these problems, as identified specifically by Hunter, can be used to police and structure Delicious and the user-generated metadata from a revisionist standpoint.

### **Controlled Vocabularies**

Controlled vocabularies<sup>3</sup> are predominantly found in indexing efforts in the library world. Standardization of vocabulary allows professionals to determine the language that is used in retrieval situations, and with "use" and "use for" terminology, vocabulary is structured and presented to users in a conceptual framework that improves retrieval efforts by limiting the choices by which searches can be completed. Research has not studied consistency of the vocabularies used by generic web users, in large part because of the newness of both the web, generally, and the technology that employs folksonomies, specifically. Work in the area of controlled vocabularies, however, may

---

<sup>3</sup> **Controlled vocabulary:** An established list of preferred terms from which a cataloger or indexer must select when assigning subject headings or descriptors in a bibliographic record, to indicate the content of the work in a library catalog, index, or bibliographic database <sup>(Reitz, 2006)</sup>. In the context of this research question, refers to previously used "tags" in the user-developed folksonomy, which can be referred to when choosing identifying vocabulary for future information management.



help researchers understand vocabulary consistency, and its benefits to users and to the web's future.

In "The Vocabulary Problem in Human-System Communication" (1987), Furnas et al concentrate on the issue of language, a pressing issue in metadata creation. Furnas et al focus specifically on language used in search and retrieval instances, rather than instances of metadata creation, but their agreement with Hunter about the "vocabulary problem" of their title is obvious; Furnas et al write that "many functions of most large systems depend on users typing in the right words," and they note that "people often disagree on the words they use for things" (964). A user's chosen search vocabulary may accurately describe the resources they are attempting to access, but if the user's vocabulary does not match the system's assigned keyword vocabulary for the resources, their attempts at retrieval will fail.

In their study, which allowed both professionals and users to assign keywords to resources and then attempt to retrieve them, Furnas et al found that the vocabulary problem was not as simple as they had hoped: with the assignment of only an applicable "name" to a resource, retrieval success stood only at 10-20%. Contrary to this, an empirically based, "frequency weighted" table of possible synonyms provided a success rate anywhere from 50 to 100%, far above the success rate of the single name. Furnas et al draw several conclusions from the outcomes of their study, the most notable being that full-text indexing or extensive vocabulary alias creation is far more likely to lead to success in resource retrieval and discovery.

Individual vocabularies, Furnas et al outlined, will always allow for the creator to retrieve their resources, but not necessarily for anyone else to retrieve the resources.

Because the vocabularies of Delicious users are not created in a vacuum of single users viewing only their own resources, the individual vocabularies begin to alter -- as Furnas et al note, users can often be trained to use different language, and this principle is readily at hand in the vocabularies of Delicious users. Their vocabularies are initially built for themselves, but the social aspect of the Delicious interface begins to alter this; a vocabulary is chosen for themselves and for the other users who might attempt to access their resources. Users develop unrefined "tables" of vocabulary aliases, and through these tables, refining occurs, synonyms are developed and adopted, and the vocabulary problem does not disappear, but is taken into the hands of the users and turned into an advantage.

L.C. Howarth, in "Enabling Metadata: Creating Core Records for Resource Discovery" (2005), raises more questions than are answered, but one of these questions is an important one: what metadata is essential, or 'core', to providing access to a digital object? Howarth focuses on Dublin Core, and the elements within Dublin Core's already defined "essential elements" which should be considered even more essential -- the "core of the core" elements.

In folksonomies and user-generated metadata, there exist no pre-defined core elements for creation, but users themselves will slowly begin to determine the most important vocabulary choices for retrieving resources. These vocabulary choices might slant toward subject tags, or toward tagging resources with identifying information about author, location or even source from which the resource was received originally, but core elements, the most important vocabulary choices for users' own retrieval, will begin to emerge.

In "Moving towards shareable metadata" (2006), Shreeves, Riley & Milewicz outline the academic uses of shareable metadata. They suggest that "sharing metadata and the resultant aggregations benefit users, particularly those users whose subject interest cuts across disciplinary boundaries." The authors also outline the six principles of excellent shared metadata: content, consistency, coherence, context, communication and coherence to standards. Their focus on the *sharing* of metadata is important because individually created metadata can, as outlined in Zhang and Jastram, spread so far across so standards and content that the metadata becomes useless to users other than the creators.

Shreeves, Riley & Milewicz note that "the ability to think critically about the shareability of ones' own metadata and the commitment to make the necessary changes will be key for the next stage of effective digital library services." In user-generated metadata specifically, standards cannot be enforced, but in the Delicious community, strong users - influential users who tag frequently, thoroughly or well - can influence. If, in a user-generated metadata community, a strong user could adhere themselves to Shreeves, Riley & Milewicz's principles, the overall community could be swayed towards those same principles. The standards will never be adhered to 100%, but the acknowledgement of the existence of standards, as outlined in "Moving toward shareable metadata", is a step toward the use of these standards and principles in user-generated metadata.

"Moving toward shared metadata" is a recent paper, published in mid-2006 and influenced by the changing technology; in "Who Will Create Metadata For The Internet?" (1998), Thomas & Griffin set the stage for Shreeves, Riley & Milewicz's later research. Thomas & Griffin provide seminal ideas, new when they were written and still applicable

to the subject of author- and user-generated metadata today.

Thomas & Griffin address the problem of descriptive language as metadata format, and about "the case for metadata as a potential remedy to the problem of finding relevant information on the Internet". The authors go on to state that "most of the printed and electronic literature on this topic reaches the same conclusion, that an overarching standard will solve the problem of a chaotic Internet," but "careful consideration of this solution indicates a very significant likelihood that the popular approach will be inadequate to tame the Internet."

The commercial indexing services that Thomas & Griffin were thinking of are not necessarily the commercial indexing services that have developed in the wake of the dot com boom and the advent of Web 2.0, but that does not reduce the power of their words. Delicious is, in fact, a commercial indexing service - a different variety of indexing than Google's search algorithm provides, certainly, but an indexing service nonetheless - and Delicious's users are making their best attempt to impose order on every piece of electronic information that has been developed. Written in 1998, Thomas & Griffin show a foresight in their language and their ideas for the purposes of metadata, even without knowing how the development of the actual technology would progress.

Each user's individual bookmarks on Delicious are a collection of information resources, as is the entire database of bookmarks that belong to each of Delicious' more than one million users. Using someone else's bookmarks to find information that you are seeking absolutely depends on representation and organization in the collection. Poorly organized bookmarks, with metadata that does not meet Shreeves, Riley & Milewicz's outlined standards, will be unusable by the general public, if not also by the user himself;

by improving representation (metadata) and organization (structure of metadata and adherence to Shreeves, Riley & Milewicz's standards) of bookmarks stored on Delicious, a user can adhere to Jacob's principles of classification: that is, that each individual item is unique, and carries "its own unique set of characteristics."

In "Classification and categorization: a difference that makes a difference", Jacob discusses the idea that classification is the notion of dividing information into groups or categories based on subject or experience; by applying a single tag to multiple World Wide Web links on Delicious, users are perhaps classifying these bookmarks as similar in topic, author or scope.

### **Folksonomies and Bookmarks**

In her 2006 article, Jessica Dye notes that information architects believe that the World Wide Web will always need a clear taxonomical structure to make digital surfing manageable. Dye goes on to state the difference between author-generated and user-generated metadata is that user-generated metadata - the tags applied to information by users who sought the information out and wished to save it - allows more applicable information to "float to the top" in a simple search.

"Collaboration through collective tagging gives [users] the chance to build their own search systems from the ground up, based on their vocabularies," Dye writes, and this is the center of the user-generated metadata question: the users can create their own search systems with their own vocabularies, but there is no guarantee to the inherent usefulness of these vocabularies or search systems (Dye 40). "Most folksonomists pattern their tags after others" (42), Dye writes, and the social aspect of folksonomies is the important issue when studying the user-generated metadata of "flat" (non-hierarchical)

folksonomies; there is no hierarchy, as there are in structured taxonomies, but the policing by the user community improves the metadata that they are creating.

The seminal paper on, specifically, the user-generated metadata community of Delicious is Adam Mathes' "Folksonomies - Cooperative Classification and Communication Through Shared Metadata" (2004). Mathes notes that a study of Delicious "reveals that the most popular tags are primarily subject descriptor keywords at various levels of specificity." Mathes goes on to discuss the limitations of these flat folksonomies, and like Hunter, outlines problems that are similar to those found in synthesizing and collecting structured metadata: ambiguity of language and the tagging of pages with multiple synonyms or even "lack of synonym control can lead to different tags being used for the same concept, precluding collocation".

These problems are not limited to user-generated or author-generated metadata; language of descriptors will always be a problem. Mathes, however, also notes the pros to a system like Delicious's folksonomy<sup>4</sup>: "This tight feedback loop leads to a form of asymmetrical communication between users through metadata. The users of a system are negotiating the meaning of the terms in the folksonomy, whether purposefully or not, through their individual choices of tags to describe documents for themselves." Mathes has discovered that the problems of language in folksonomies can also be the positive aspects of the system, because the constant development - folksonomies are never static, always dynamic - of the vocabularies used by the creators allow for revision, refinement and adjustment. Unlike author-generated metadata, which is static and structured, folksonomies and their user-generated metadata can change as necessary. This constant

---

<sup>4</sup> **Folksonomy:** First coined by Thomas Van Der Wal, "folksonomy" is a hybrid of "folk taxonomy", a taxonomy created by the general public.

flow may eventually solidify into a structure, but for the moment, it is fluid and that is its power: the ability to change, to grow, and to become better.

Users of Delicious bookmark WWW resources on their own accounts for many reasons: later personal retrieval, desire to share a resource with friends or community members, simply noting an interesting resource that may or may not be later retrieved. By creating their own controlled vocabularies, they ease their own retrieval; by sharing vocabulary choices with community members, they ease the retrieval efforts of the full community. Bookmarking resources in a social setting moves past the late 20<sup>th</sup> bookmarking efforts, where resources were stored locally on a user's computer, in single folders. The ability to apply controlled vocabularies, with hierarchical tags and multiple vocabulary keywords, to bookmarks in a social setting raises the storage, and consequently the retrieval, bar. A controlled vocabulary improves precision and recall of resources; a user-generated controlled vocabulary improves precision and recall of specific resources important to a specific community.

The literature does not yet exist to provide a full overview of user-generated metadata; the concept of user-generated metadata and folksonomies are still too new, technologically, to have provoked much study. However, the literature that exists on author-generated metadata and standard classification informs the future study of user-generated metadata; while the overarching principles of user-generated and author-generated metadata are not the same, they are similar, and the historical study of author-generated metadata can inform and guide both the development of user-generated metadata and the study of it. Although research has examined author-generated metadata, research has not begun to examine the next-generation folksonomies,

containing metadata generated by users rather than authors or professionals. This study intends to study a cross-section of user-generated metadata in small communities with shared interests, and determine if there is any structured agreement in vocabulary use, contradicting Furnas et al's assertion that users cannot agree on vocabulary choices, much less the "aboutness" of a resources.

### **III. Research Questions**

This study examines if shared vocabularies underlying social indexing communities exhibit elements of a controlled vocabulary. The specific research questions guiding the research is:

1. Does the vocabulary developed by shared communities, via social software, exhibit elements of a rudimentary structured consistent vocabulary?

By examining if patterns exist in a selected subject/interest-specific community within Delicious, and examining the bookmarking of resources relevant to their interest, this research may be able to draw conclusions about the influence of community on vocabulary in user-generated folksonomies, and further research can be recommended.

In isolating small communities with individual specialized interests, the study will examine whether these communities or their individual users, in saving and sharing links that are related to their specialized interests, exhibit any controlled vocabulary behavior in their tagging language. If controlled vocabulary behavior can be isolated in small communities and their users, better information retrieval and storage systems can be refined, based on the needs of individuals and communities.

### **IV. Method**



The method underlying this study was a content analysis, which was performed on the data contained in Delicious' individual URL history pages. Content analysis was chosen because Delicious contains a great deal of pre-generated data; rather than apply content analysis to data gathered via surveys or interviews, the study analyzed the existing data for content. Content analysis was necessary in this study because the data that exists has not been vetted or monitored in the study, and content analysis was the only method that can be scientifically applied to the data. Content analysis allowed the study to reach objective conclusions without the free-form creation affecting the study.

As seen in Figure A below, Delicious maintains a publicly viewable World Wide Web page for each URL that is saved in the Delicious system by a user; each page contains a record of the tags that each individual user assigned to the particular URL, as well as a numerical list of the most popular tags for the URL and a listing of any notes that the individual users assigned to the URL.

The screenshot shows a web browser window displaying a Delicious URL history page. The address bar shows the URL: <http://del.icio.us/url/05cacc45a07d44a9c680f2840b8a6773>. The page title is "del.icio.us / url". Below the title, there is a search bar and a "check url" button. The main content area displays the URL history for <http://www.kekkai.org/synechdochic/sg>. The page includes a "common tags" section with a list of tags and their counts, a "posting history" section with a list of users and their posting dates, and a "related items" section with a list of related URLs. The page also features a "user notes" section with several notes and their authors.

**common tags**

count	tag
133	sga
85	mckay/sheppard
52	synechdochic
51	fic
32	slash
26	gen
22	rodneymckay
21	angst
21	mckay
19	futurefic
14	long
13	au
13	earthside
13	sheppard/mckay
12	fanfic
11	rodney
10	fanfiction
10	moshap
9	deathfic
7	john/rodney
7	reos
6	atlantis
5	future
4	earth

**posting history**

user	date
first posted by wylolur to sga rodneymckay mcshep	Nov '06
by leyrin to sga fic	
by thefakeprince to slash sga mckay/sheppard	
synechdochic john mckay deathfic earthside future-fic	
angst au to-read	
by khraiken to sga mckay/sheppard fanfic synechdochic	
fic slash atlantis stargate	
by photojenny to sga fic fandom sci-fi	
by zor0 to mckay/sheppard sga	
by KaKira to fic sga fandom long	
by makasnewmanadie to sga	
by vardoger to atlantis sheppard/mckay	

**user notes**

- Beautiful, bittersweet  
- leyrin
- SGC finally stops calling him by December. Rodney celebrates by writing a final exam for his relatively class so difficult that it reduces four students to tears in the exam hall. Upon reflection, he decides to be meriful and offer partial credit.  
- thefakeprince
- One of the best fanfics I have ever read in the Stargate Atlantis fandom and one of the few I've reread countless times without it losing the impact. There are John/Rodney overtones, but please don't let that keep you from reading it if you aren't fond of  
- khraiken
- The one with Rodney after the mission is over.  
- zor0
- the one where john is dead  
- kaKira
- Rodney's life post-Atlantis as a University professor.  
- sasa808
- Futurefic in which Rodney gets a job as a physics professor--it's much more complex than that, but the backstory unfolds so beautifully that it would be a crime to spoil it. A believable, powerful character study with non-explicit John/Rodney.  
- watersofthompson
- Five and a half years of loving someone, of being loved, as much as you can love in the city on the edge of forever, which is sometimes too much and sometimes not enough at all. They'd both known how it could end at the very beginning.  
- kaymanah
- Rodney McKay, R. "SGC finally stops calling him by December. Rodney celebrates by writing a final exam for his relatively class so difficult that it reduces four students to tears in the exam hall."  
- watersword
- There is a gravestone in Arlington National Cemetery. It sits watch over an empty grave, and there is a vine with small blue flowers, one the gardeners can't identify from any book, twining up its side.  
- sevier
- Every year, a few more flowers bloom.  
- pasterstar
- A bitter Rodney hides away in second rate university and almost accidentally trains the next generation of geeks to save the universe and all it entails.  
- korlian
- Rodney can't find a textbook that doesn't make him want to spit nails, so he gets by with photocopying all the least-egregiously-wrong journal articles he can find.  
- sink
- There is a gravestone in Arlington National Cemetery. It sits watch over an empty grave, and there is a vine with small blue flowers, one the gardeners can't identify from any book, twining up its side. It is an utterly unremarkable memorial, name and  
- antihoneyai
- bring tissue  
- versani

Figure 1. Screenshot of Delicious URL History page.

### **Sample Group**

This study focused on a small community of users within the overall Delicious user base, and group selection will be based on similarity of user interests. The small communities within Delicious are identified by their primary interests: recipes; fanfiction (stories about television or literary characters, written by someone other than the medium's primary, original authors); sports news; technology advances; Web 2.0 social networking sites. As outlined, the overall user base of Delicious does not conform to any standardization of vocabulary, in large part because the user base is both so large and so diverse in their personal applications of the tagging technology, but smaller, individual communities that share interests and consequently share links between users with great regularity may exhibit more control over their share vocabularies, because they also share interests or personal connections with some or all community members, and seek (without necessarily knowing that they seek) a standardization that leads them to easier resource sharing and discovery.

This study identified these groups, the highly representative users within the chosen groups, and the primary tag applied to each of their bookmarks, and from a union of these identifications, select individual URL history pages on which the tags chosen will be analyzed for content. The analysis was applied to both the pre-identified primary tag as well as the secondary tags, and hierarchies, patterns and diversity will all be examined within the tagging schemes that emerge from the URL history page.

The sample group for this study is the Stargate: Atlantis fandom. This group includes readers and writers of fanfiction about the Sci Fi Channel's original series Stargate: Atlantis. The target study group may also bookmark non-fanfiction resources relevant to Stargate: Atlantis's fictional canon, the show's

actors and actresses, and future speculation regarding the plot arc, but this study focuses only on fanfiction bookmarked by this user group. The primary tag will be "sga" (a shorthand notation for the show used by fans), but the intersection of "sga+fic" (fic: a shorthand notation for "fanfiction" used by fans) will also be considered.

This group was chosen for their forefront positions as Delicious users; tracked to Delicious from personal websites or blogs about the subjects by which they were grouped, the groups' Delicious use was clearly and easily slotted in to a primary small community. Either by their chosen bookmarks (primarily fanfiction) or by their public identifying information, the group was identified as just that: groups. This group was selected for their large-scale use of Delicious to share, save and retrieve information on their given topics of interest. Groups that were considered and discarded included political bloggers (writers of weblogs devoted to political topics), food bloggers (writers of weblogs devoted to recipes, restaurants, cooking and eating), registered users of the website Slashdot.org, and knitters. These groups were rejected because their use of Delicious and Delicious' network feature did not project a cohesive user group on Delicious, despite an off-site user community.

Highly representative users within these groups will be identified by one of three criteria, and noted in the analyses for each group's chosen URLs:

- A user who bookmarks a large numerical value of links per day.
- A user who is always among the first bookmarkers on a link (so that their tags and their comment-style is seen early and often by network readers).
- A user who brings their own audience from an outside source (such as a blogger with a large following whose readers follow the blogger's bookmarks on Delicious via the Delicious network or by directly viewing the blogger's Delicious page).

Highly representative users who can be identified before the URLs for analysis have been chosen will act as guides; bookmarks that have been saved by these strong users in their

small communities will be examined first in choosing the URLs for analysis, as the strong users' bookmarking may lend influence to the bookmarking of other users.

Within these three groups, once highly representative users have been identified and noted on the coding instrument, 15 individual URL history pages will be selected for analysis. These pages were selected via purposive sampling; because Delicious contains thousands of URLs, examining every piece of fanfiction bookmarked would be unfeasible. Based on primary tag and strong user knowledge of the pre-selected small communities, a sample of URLs bookmarked by these user groups will be sufficient to study the vocabulary of the small communities.

The criteria under which the pages were selected are as follows:

**Occurrence of the primary tag:** the URLs will be selected from the overall Delicious tag pages for each primary tag; these pages list every instance of a URL tagged with the chosen tag. The format of the URL from which the pages will be chosen is [http://del.icio.us/tag/\[primary tag here\]](http://del.icio.us/tag/[primary tag here]).

**Number of bookmarkers:** because small groups provide a less wide vocabulary, URLs that have been bookmarked by fewer than 75 users will not be considered. A maximum number of bookmarkers will not be considered, but preference will be given to URLs that have been bookmarked by more than 75 users, for breadth of vocabulary.

**Date of bookmarking:** because the study examines the development, pages with a substantial history will be used, but not so substantial that the development has become static. Pages that were first bookmarked no more than 36 months and that have been bookmarked at least once within 2 months prior to the sample date will be considered; any pages that do not meet these criteria of date will not be considered.

Because this study was reliant on users choosing to register and save World Wide Web URLs at Delicious, the chosen option of non-probability purposive sampling is the best applicable to the study.

The data recorded for content analysis will be as follows:

**Number of overall bookmarkers:** numerical value, provided by Delicious' database. The URL history page states clearly at its beginning: "this url has been saved by X people."

**Primary tag:** the most common tag for the URL.

**Common tags:** in the right hand corner of any URL history page on Delicious, a list (or cloud) of common tags applied to the URL is easily visible. This study will view these common tags in the "list" format, with a numeric value provided for each common tag, and no more than the top five common tags will be recorded and displayed both in a standard numeric value and as a percentage as determined by their ratio of use compared to overall bookmarkers.

**Unique tags:** underneath the "common tags" interface on Delicious is a list of all users who have bookmarked this URL, as well as their choice of tags for the URL. These tags will be reviewed and any "noteworthy" (author, genre, subject material) tags will be noted as well, in numeric value occurrence as well as percentage to the overall bookmarking number.

The coding instrument that was used is included in Appendix A.

After the data was collected from the 15 pre-selected URL history pages on Delicious, it was analyzed within the small user group in an effort to determine if there is a majority tagging scheme being applied by the group. Relevance, to the URL's actual content, of the tags applied was considered, and the tags were analyzed, stand-alone, via percentage of overall bookmarkers, and common tags will be noted and described in the findings as hierarchical, consistent and "agreed upon" (where "agreed upon" can be defined as being in use by more than 50%, a majority, of the users who have bookmarked a particular URL).

## V. Results and Discussion

Analysis of 15 individual URL history pages was completed, and data defined by types of tags used by members of the small community (members of the Stargate: Atlantis fandom, bookmarking pieces of Stargate: Atlantis fanfiction on Delicious), as well as number of users who applied various vocabulary choices and tag types, is

discussed in the following section. Primary tag occurrence, followed by tag type identifiers and vocabulary choices, is identified and explained.

The analysis clearly indicated that highly representative users could not be identified via the URL pages. It remains possible that "strong users" are influencing their communities via the Delicious network feature, which allows users to subscribe to other users' bookmarks, but individual users made no significant impact on the data derived from URL history pages. Of the 15 URL history pages analyzed, no single user had been the "first tagger" identified by the Delicious interface; all 15 URLs were introduced to the system by different users, and no single user exhibited any signs of influencing the trends more than others.

The lack of influential user, however, does not appear to have affected the Stargate: Atlantis community's *vocabulary*. The most-used tag on all 15 stories was "sga", a shorthand notation for the name of the television show for which the fanfiction has been written (see figure 2). On average, 77.76% of all users who bookmarked a URL chose to identify it with the tag "sga". Comparatively, 0.48% of users chose to use "atlantis", 0.17% chose "stargate:atlantis", 0.41% used "stargate\_atlantis", and 0.23% used "stargateatlantis" (see figure 3). This is a community language differential: while there are outliers using vocabulary that does not match the majority, over the course of a 24 month period, more users choose the vocabulary term "sga" over any of the alternate term choices.

	<b>sga</b>	<b>mckay/sheppard</b>	<b>mckay/zelenka</b>
<b>Secret Lives Of Scientists</b>	70.00%		66.00%
<b>A Beautiful Lifetime Event</b>	72.07%	59.91%	
<b>About A Lamp</b>	79.83%	63.03%	
<b>20,000 Leagues Under The Sea</b>	81.25%	62.50%	
<b>Abstain</b>	76.84%	63.16%	
<b>Intersections</b>	78.38%	56.76%	
<b>Scenes From A Lesser War</b>	74.51%	64.71%	
<b>Boom Boom Room</b>	78.68%	61.93%	
<b>24 Hours With A Rodneysaur</b>	82.11%	24.21%	
<b>Freedom's Just Another Word</b>	74.04%	38.30%	
<b>Ratio Of Burning</b>	83.00%	65.00%	
<b>Lost In Waiting</b>	76.19%	66.67%	
<b>Second Skin</b>	77.39%	60.87%	
<b>Take Clothes Off As Directed</b>	76.92%	61.54%	
<b>Weapons Of Some Distraction</b>	85.14%	63.51%	
	<b>77.76%</b>	<b>58.01%</b>	<b>66.00%</b>

Figure 2. Stories identified by name with tag percentages for primary tag and pairing tag based on total sample.

	<b>atlantis</b>	<b>stargate:atlantis</b>	<b>stargate_atlantis</b>	<b>stargateatlantis</b>
<b>Secret Lives Of Scientists</b>		3.00%	3.00%	
<b>A Beautiful Lifetime Event</b>				
<b>About A Lamp</b>				
<b>20,000 Leagues Under The Sea</b>		2.50%		
<b>Abstain</b>			3.16%	
<b>Intersections</b>	2.03%			
<b>Scenes From A Lesser War</b>				
<b>Boom Boom Room</b>				
<b>24 Hours With A Rodneysaur</b>				
<b>Freedom's Just Another Word</b>	2.98%			
<b>Ratio Of Burning</b>				
<b>Lost In Waiting</b>				
<b>Second Skin</b>				3.48%
<b>Take Clothes Off As Directed</b>	2.20%			
<b>Weapons Of Some Distraction</b>				
	<b>0.48%</b>	<b>0.17%</b>	<b>0.41%</b>	<b>0.23%</b>

Figure 3. Stories identified by name with tag percentages for alternate vocabulary tag outliers based on total sample.

The tag "sga" is the primary tag on each of these URL history pages; it serves as the identifier tag, the commonality between all the users that helps them seek each other out in the general population. In Figure 4, below, a secondary tag is outlined: the "type" tag. In 13 of 15 URL pages, the resource was tagged with the vocabulary "fic", short for fiction or fanfiction, thereby identifying to the community what the resource actually *is*. "fic" is the clear choice of the community, as other outliers are presented below, but the low percentage of users who employ a tag called "fic" is notable. While the tag appears

enough times to occur in the top five tags of 13 of 15 resources, its low occurrence and absence suggests a community outlook on both vocabulary and necessity of identification. The vocabulary was chosen by the community, which made a clear choice of "fic" over any of the other options used by outliers, but the community at large, in the majority, did not feel it necessary to *use* the prescribed vocabulary. Possible reasons for this include being previously aware of the type of resource they were identifying with tags and not feeling the need to include a piece of, to them, redundant information, and not feeling that type was a necessary identifier for the community at large. The former reasoning suggests a individual-oriented mindset of the user; the latter suggests a community mindset, where assumptions are made of the community.

These community assumptions are based in vocabulary, as well -- because the vocabulary is shared, knowledge of the resource's content is also shared. This suggests that users who choose to exclude the "type" tags on their resources make assumptions of common knowledge beyond vocabulary. These assumptions may be conscious or unconscious, but they exist, when users choose to exclude a vocabulary choice that is preferred by the community. Likewise, users who choose to *include* the type tag make the assumption that resource-seekers may *not* know the content of resources, and the inclusion of the "type" tag alerts seekers to instances of a certain type of resource that they may seek, or may seek to avoid. With a consensus on vocabulary, but disagreement on inclusion, the "type" tag is worthy of further, deeper study.



	fic	fiction	fanfiction	fanfic	
<b>Secret Lives Of Scientists</b>	27.00%		2.00%		13
<b>A Beautiful Lifetime Event</b>	25.68%		2.25%		
<b>About A Lamp</b>	25.21%				24.44%
<b>20,000 Leagues Under The Sea</b>	21.25%				
<b>Abstain</b>	25.26%		2.11%	2.11%	
<b>Intersections</b>	29.05%				
<b>Scenes From A Lesser War</b>	23.53%				
<b>Boom Boom Room</b>	26.90%				
<b>24 Hours With A Rodneysaur</b>	24.21%				
<b>Freedom's Just Another Word</b>	31.49%				
<b>Ratio Of Burning</b>	29.00%				
<b>Lost In Waiting</b>	22.86%				
<b>Second Skin</b>					
<b>Take Clothes Off As Directed</b>			3.30%		
<b>Weapons Of Some Distraction</b>	27.03%				
	<b>20.76%</b>	<b>0.14%</b>	<b>0.51%</b>	<b>1.63%</b>	

**Figure 4. Stories identified by name with tag percentages for "type" tags, with primary and outliers based on total sample.**

In 13 of the 15 selected URL history pages (see Figure 2 above), the second most popular tag was the same *part* of the vocabulary, designating the story's primary romantic pairing, if not exact tag ("mckay/sheppard", used to designate a romantic pairing, is replaced by "mckay/zelenka", a different romantic pairing, on one story). In all 15 of the URL history pages analyzed, the primary pairing tag appeared in the top five most common tags. Another vocabulary choice by the community is demonstrated in this tag choice; by choosing to represent a story about characters named McKay and Sheppard as "mckay/sheppard", the community raises the identifiers that it has attached to the URL to a higher level, allowing themselves to locate resources not only by title or URL, but also by the persons involved in the resource.

In Figure 5, below, alternate pairing tags are identified. These tags demonstrate that, even when a story is tagged by less than 50% of the population as "mckay/sheppard", it is not because the community is using an alternate vocabulary; as with the development of the community vocabulary to a general consensus on "sga" versus "stargate\_atlantis" or "atlantis", the data below demonstrates that in dealing with resources featuring a romantic

pairing of McKay and Sheppard, "mckay/sheppard" is preferred to the reverse, or to stories that use a shortened version of the pairing or first names as opposed to last.

	<b>zelenka/mckay</b>	<b>john/rodney</b>	<b>sheppard/mckay</b>	<b>rodney/john</b>	<b>mcshep</b>
<b>Secret Lives Of Scientists</b>	3.00%				
<b>A Beautiful Lifetime Event</b>		2.70%			
<b>About A Lamp</b>					
<b>20,000 Leagues Under The Sea</b>			12.50%	2.50%	
<b>Abstain</b>					
<b>Intersections</b>					
<b>Scenes From A Lesser War</b>					
<b>Boom Boom Room</b>					
<b>24 Hours With A Rodneysaur</b>					
<b>Freedom's Just Another Word</b>					
<b>Ratio Of Burning</b>					
<b>Lost In Waiting</b>		2.86%			3.00%
<b>Second Skin</b>					
<b>Take Clothes Off As Directed</b>		2.70%			
<b>Weapons Of Some Distraction</b>	0.20%	0.37%	0.83%	0.17%	0.20%

**Figure 5: Outlier pairing usages on each of 15 stories based on total sample.**

In the instance of Freedom's Just Another Word, the instance of "mckay/sheppard" appearing as the second most used tag but at a percentage under 50% can be attributed to a reader's analysis of the content (see below, Figure 6):

	<b>au</b>	<b>slash</b>	<b>gen</b>	<b>angst</b>
<b>Secret Lives Of Scientists</b>		28.00%		
<b>A Beautiful Lifetime Event</b>		32.88%		
<b>About A Lamp</b>		26.89%		
<b>20,000 Leagues Under The Sea</b>		28.75%		
<b>Abstain</b>		30.53%		
<b>Intersections</b>	24.32%	24.32%		
<b>Scenes From A Lesser War</b>		30.39%		
<b>Boom Boom Room</b>	66.50%	32.49%		
<b>24 Hours With A Rodneysaur</b>			33.68%	
<b>Freedom's Just Another Word</b>		22.55%	16.17%	
<b>Ratio Of Burning</b>		27.00%		40.00%
<b>Lost In Waiting</b>		29.52%		20.95%
<b>Second Skin</b>		25.22%		
<b>Take Clothes Off As Directed</b>	56.04%	35.16%		
<b>Weapons Of Some Distraction</b>		32.43%		
	<b>9.79%</b>	<b>23.05%</b>	<b>3.32%</b>	<b>4.06%</b>

**Figure 6: Tags used as content descriptors, where descriptors appeared attached to more than one URL history page based on total sample.**

Tags appear as content descriptors when the community identifies a quality or a topic within the resource that is notable. The tag "au" designates resources in which the story

deviates from the canon outlaid by the source material, creating an alternate universe, and of the three resources for which the tag "au" appears, two are distinctly alternate settings of already established characters. The appearance of a low-percentage "au" tag on Intersections may be attributed to readers' analysis of the content as not matching what has occurred later in the source text.

The tags of "slash" and "gen" are vocabulary terms specific to the fandom community at large, with the former being defined as a resource containing a non-canon homosexual relationship, and the latter being defined as a resource with no outright romantic relationships, shortened from "general" or "generic". Rarely do these tags occur together, as the content of resources does not generally allow it. The secondary placement of "mckay/sheppard" on the list of tags for Freedom's Just Another Word while still maintaining a less-than-majority percentage of use, however, can be linked to the remainder of the top five most used tags on that resource, as of the 15 URL histories chosen, Freedom is the only history which shows a high use of both "slash" *and* "gen".

The dichotomy of language in this tagging suggests not that the community could not agree on the correct vocabulary, but rather that they could not agree on the *content* of the resource. The "slash" tag links strongly to the "mckay/sheppard" pairing tag, but "gen" indicates that many readers did not see or did not have interest in the romantic relationship that underpins the resources. The community agrees on the language used to describe various content, but a consensus of language does not necessarily equate a consensus of "aboutness". Consensus on vocabulary emerges from the necessity of categorizing resources, but if resources span several categories, vocabulary broadens, changes, or overlaps in order to cover the content categorization.

Content descriptor tags occur when the content of a resource is highly specialized, as well. The community identifies the content that they consider relevant to future search and retrieval efforts, and chooses a vocabulary and assignment scheme accordingly. Beautiful Lifetime Event was tagged with the content descriptor "kidfic" by 35.59% of users; this vocabulary choice, within the community, indicates a resource in which children, usually belonging to the main characters, play a major role. 15.79% of users chose the tag "aliensmadethemdoit" for Abstain, indicating a resource in which aliens, in fact, make the main characters have sex.

Second Skin and Take Clothes Off As Directed are tagged with, respectively, "genderfuck" and "kink", and "bdsm", indicating subjects that might be out of a seeker's comfort zone; descriptor tags can, in this fashion, function as warnings to the information seeker in the community that some resources may be distasteful to them. The vocabulary of the community in describing content not only describes but also guides, steering users from things they may not wish to see, read, or experience.

Scenes From A Lesser War exhibits both the primary vocabulary behavior, with 19.61% of users tagging the story "dadt", community shorthand for a story that deals with the implications of Don't Ask, Don't Tell, and homosexual relationships in the American military, but also features outlier tags "outed" (2.94%) and "comingout" (2.94%), indicating that the content is clearly of interest to the community, but the vocabulary, while preferential to the tag "dadt", has not solidified into a community standard as it has in other instances, such as the primary "sga" tag or the type or pairing tags.

Content descriptor tags function in defining the "aboutness" of a resource, but can also function as descriptors of *genre*. Two of 15 URL history pages show "angst" as a

top five primary tag, which indicates not content but rather mood of the resource. 40% of users used a content descriptor tag to identify Ratio Of Burning as a story with a overall mood of "angst", and 20.95% used the same content descriptor for Lost In Waiting. The absence of this content descriptor on any of the 13 other resources is likely linked directly to content: none of the remaining 13 resources had content that could be accurately described by a content descriptor of "angst". In addition, the absence of a tag for "humor" or "funny" -- the lighter side of a content descriptor of "angst" -- indicates that either none of the resources is a particularly funny story, in such a way that a community would need to identify it as such, or the community at large does not feel the need to warn other resource seekers for stories that are funny.

It appears that the community is exercising judgment via both their vocabulary and their tagging choices; by choosing not to have a vocabulary term that describes a resource opposite to a resource tagged "angst", the community suggests to its information seekers that identifying those stories is less important to them than identifying the stories that are emotionally manipulating toward the sad end of the scale.

The results show that there is no unanimous vocabulary choice, even within the community, but they also show that, given a choice, the community will choose one vocabulary choice over others in a predominant fashion. All users may not employ the type of tag (content descriptor, type identifier), but those users that do use the same vocabulary. A move toward standardization is found, though the hypothesis cannot be proven without further study, outlined below.

## **VI. Conclusions**

This study was limited by several factors. The primary factor is the non-standardization of URLs used on the Internet; because many of the resources retrieved from Delicious are available at several different URLs, the URL pages retrieved and analyzed may not represent the full scope of tags applied by users to said resources. The same resources may have been saved on different URL history pages on Delicious, simply because of the way the users initially retrieved and recorded the URL. This is both a problem of Delicious, which does not recognize similar URLs (<http://google.com> and <http://www.google.com>) as the same resource, and of the study, which retrieved only single URL history pages for each resource chosen and did not seek out any further, independent URL history pages for the same resource. In recreating the study, it would be advised to retrieve as many URL history pages for each resource as possibly, either via a title search on Delicious, or simply application of alternate possible URLs into the search via URL on Delicious.

In addition, the subject group was limited by time and analysis restrictions; in a further study, expanding the study to include resources bookmarked by a smaller number of users from the chosen community, as well as selecting a much larger number of resources to analyze. A broader selection of resources -- 50 as opposed to 15, for example -- would give a better sense of whether the community vocabulary choices spread across all resources, rather than the study's selection.

The community, while all sharing an interest, do not necessarily all interact with each other. There is a divide into smaller social groups within the already small community, and these divides may have affected language choices in smaller areas, such as the "type" tag and the content descriptors. In a further study, the identification of small groups

within the pre-identified community would be recommended; an analysis of the influence of internal groups' language on both other members of their group and the small community as a whole would be a variable worth pursuing in a larger study.

Regardless, the study concludes that a small community, creating metadata for their own purposes on a set of resources specific to that community, will find a vocabulary that allows them to describe their resources in as in-depth a fashion as they would like. Because the community of users on Delicious is not policed by a governing body, linguistic outliers will continue to exist, ignoring or twisting the community standards for their own purposes, but the overall community will seek to standardize tags and language for ease of retrieval, description, and sharing. The shared vocabulary brings the community together, suggesting that even in uncontrolled metadata creation situations, some consensus will be demonstrated simply because a community exercises its own need for agreement on vocabulary choice.

The fallibility of an uncontrolled group continues to exist, but a group can exist that controls itself to the extent that any group not governed by an authorizing body can succeed in organizing data and presenting a unified front. While there is not a consensus -- nor should one be expected from a large community of people who, while they share interests, may not share friendships or even casual relationships -- this study shows that within a closed community, when saving and sharing information related to their subject around which the community is created, a distinct trending toward a formalized vocabulary occurs. While Delicious, as a whole, remains a morass of uncontrolled language with users having no connection, vocabulary-wise, to their nearest neighbors,

the communities *within* Delicious are beginning to forge forward in finding the shared language, in the form of consistent vocabularies, of their interests and populations.



### Works Cited

- Dye, J. (2006). Folksonomy: A game of high-tech (and high-stakes) tag. *E-Content Magazine*, 29(3), 38-45.
- Furnas, G.W., Landauer, T.K., Gomez, L.M., & DuMais, S.T. (1987). The vocabulary problem in human-system communication. *Communications of the ACM*, 30(11).
- Greenberg, J., Pattuelli, M. C., Parsia, B., & Robertson, W. D. (2001). Author-generated dublin core metadata for web resources: A baseline study in an organization. *Journal of Digital Information*, 2(2).
- Howarth, L. C. (2005). Enabling Metadata: Creating Core Records for Resource Discovery; 2004 Update on Activities of the IFLA Cataloguing Section Working Group on the Use of Metadata Schemas. *International Cataloguing and Bibliographic Control*, 34(1) p. 14-17.
- Hunter, J (2001). MetaNet - a metadata term thesaurus to enable semantic interoperability between metadata domains. *Journal of Digital Information*, 1(8).
- Jacob, E. K. (2004). Classification and categorization: A difference that makes a difference. *Library Trends*, 52(3) p. 514-540.

- Mathes, A. (2004). *Folksonomies - cooperative classification and communication through shared metadata*. Retrieved October 19, 2006 from <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
- Reitz, J. M. *Online Dictionary for Library and Information Science*. Retrieved November 4, 2006 from <http://www.lu.com/odlis/>
- Schacter, J. *Del.icio.us About Page*. Retrieved November 4, 2006 from <http://del.icio.us/about/info/>
- Shreeves, S. L., Riley, J., & Milewicz, L. (2006). Moving towards shareable metadata. *First Monday*, 11(8), 10/19/2006. [http://www.firstmonday.org/issues/issue11\\_8/shreeves/index.html](http://www.firstmonday.org/issues/issue11_8/shreeves/index.html)
- Thomas, C.F. & Griffin, L.S. (1998). Who will create metadata for the internet? *First Monday*, 3(12), 10/20/2006. [http://www.firstmonday.org/issues/issue3\\_12/thomas/index.html](http://www.firstmonday.org/issues/issue3_12/thomas/index.html)
- Zhang, J., & Jastram, I. (2006). A study of the metadata creation behavior of different user groups on the internet. *Information Processing and Management*, 42(4), 1099-1122.

## Appendices

### A. Coding

**URL:** [url here]

**Date of first bookmarking:** [Month/Year]

**# of Taggers:**

**Title:**

	Numerical Value	Percentage of Overall Taggers	Agreed Upon?
Primary tag:			
Common tag #1:			
Common tag #2:			
Common tag #3:			
Common tag #4:			
Common tag #5:			

	Numerical Value	Percentage of Overall Taggers	Agreed Upon?
Unique tag #1:			
Unique tag #2:			
Unique tag #3:			
Unique tag #4:			
Unique tag #5:			