The Structure and Form of Folksonomy Tags: The Road to the Public Library Catalog

Louise F. Spiteri

This article examines the linguistic structure of folksonomy tags collected over a thirty-day period from the daily tag logs of Del.icio.us, Furl, and Technorati. The tags were evaluated against the National Information Standards Organization (NISO) guidelines for the construction of controlled vocabularies. The results indicate that the tags correspond closely to the NISO guidelines pertaining to types of concepts expressed, the predominance of single terms and nouns, and the use of recognized spelling. Problem areas pertain to the inconsistent use of count nouns and the incidence of ambiguous tags in the form of homographs, abbreviations, and acronyms. With the addition of guidelines to the construction of unambiguous tags and links to useful external reference sources, folksonomies could serve as a powerful, flexible tool for increasing the user-friendliness and interactivity of public library catalogs, and also may be useful for encouraging other activities, such as informal online communities of readers and user-driven readers' advisory services.

One of the most daunting challenges of information management in the digital world is the ability to keep, or refind, relevant information; bookmarking is one of the most popular methods for storing relevant Web information for reaccess and reuse (Bruce, Jones, and Dumais 2004). The rising popularity of social bookmark managers, such as Del.icio.us, addresses these concerns by allowing users to organize their bookmarks by assigning tags that reflect directly their own vocabulary and needs. The collection of user-assigned tags is referred to commonly as a folksonomy. In recent years, significant developments have occurred in the creation of customizable user features in public library catalogs. These features offer clients the opportunity to customize their own library Web pages and to store items of interest to them, such as book lists. Client participation in these interfaces, however, is largely reactive; clients can select items from the catalog, but they have little ability to organize and categorize these items in a way that reflects their own needs and language.

Digital document repositories, such as library catalogs, normally index the subject of their contents via keywords or subject headings. Traditionally, such indexing is performed either by an authority, such as a librarian or a professional indexer, or is derived from the authors of the documents; in contrast, collaborative tagging, or folksonomy, allows anyone to freely attach keywords or tags to content. Demspey (2003) and Ketchell (2000) recommend that clients be allowed to annotate resources of interest and to share these annotations with other clients with similar interests. Folksonomies can thus make significant contributions to public library catalogs by enabling clients to organize personal information spaces; namely, to create and organize their own personal information space in the catalog. Clients find items of interest (items in the library catalog, citations from external databases, external Web pages, and so on) and store, maintain, and organize them in the catalog using their own tags.

In order to more fully understand these applications, it is important to examine how folksonomies are structured and used, and the extent to which they reflect user needs not found in existing lists of subject headings. The purpose of this proposed research is thus to examine the structure and scope of folksonomies. How are the tags that constitute the folksonomies structured? To what extent does this structure reflect and differ from the norms used in the construction of controlled vocabularies ,such as Library of Congress Subject Headings? What are the strengths and weaknesses of folksonomies (for example, reflect user need, ambiguous headings, redundant headings, and so forth)?

This article will examine a selection of tags obtained from three folksonomy sites, Del.icio.us (referred to henceforth as Delicious), Furl, and Technorati, over a thirty-day period. The structure of these tags will be examined and evaluated against section 6 of the NISO guidelines for the construction of controlled vocabularies (NISO 2005), which looks specifically at the choice and form of terms.

Definitions of folksonomies

Folksonomies have been described as "user-created metadata . . . grassroots community classification of digital assets" (Mathes 2004). Wikipedia (2006) describes a folksonomy as "an Internet-based information retrieval methodology consisting of collaboratively generated, open-ended labels that categorize content such as Web pages, online photographs, and Web links." The concept of collaboration is attributed commonly to folksonomies (Bateman, Brooks, and McCalla 2006; Cattuto, Loreto, and Pietronero 2006; Fichter 2006; Golder and Huberman

Louise F. Spiteri (Louise.Spiteri@dal.ca) is Associate Professor at the School of Information Management, Dalhousie University, Halifax, Nova Scotia, Canada.

This research was funded by the OCLC/ALISE Library and Information Science Research Grant Program.

2006; Mathes 2004; Quintarelli 2005; Udell 2004). Thomas Vander Wal, who coined the term *folksonomy*, argues, however, that:

> the definition of Folksonomy has become completely unglued from anything I recognize. . . . It is not collaborative . . . it is the result of personal free tagging of information and objects (anything with a URL) for one's own retrieval. The tagging is done in a social environment (shared and open to others). The act of tagging is done by the person consuming the information" (Vanderwal.net 2005).

It may be more accurate, therefore, to say that folksonomies are created in an environment where, although people may not actively collaborate in their creation and assignation of tags, they may certainly access and use tags assigned by others. Folksonomies thus enable the use of shared tags.

Folksonomies are used primarily in social bookmarking sites, such as Delicious (http://del.icio.us/) and Furl (http://www.furl.net/), which allow users to add sites they like to their personal collections of links, to organize and categorize these sites by adding their own terms, or tags, and to share this collection with other people with the same interests. The tags are used to collocate bookmarks within a user's collection and bookmarks across the entire system, so, for example, the page http://del.icio.us/tag/blogging will show all bookmarks that are tagged with *blogging* by any member of the Delicious site.

Benefits of folksonomies

Quintarelli (2005) and Fichter (2006) suggest that folksonomies reflect the movement of people away from authoritative, hierarchical taxonomic schemes that reflect an external viewpoint and order that may not necessarily reflect users' ways of thinking. "In a social distributed environment, sharing one's own tags makes for innovative ways to map meaning and let relationships naturally emerge" (Quintarelli 2005). Vander Wal (2006) adds that "the value in this external tagging is derived from people using their own vocabulary and adding explicit meaning, which may come from inferred understanding of the information/object."

An attractive feature of folksonomies is their inclusiveness; they reflect the vocabulary of the users, regardless of viewpoint, background, bias, and so forth. Folksonomies may thus be perceived to be a democratic system where everyone has the opportunity to contribute and share tags (Kroski 2006). The development of folksonomies may reflect also the difficulty and expense of applying controlled taxonomies to the Web: building, maintaining, and enforcing a sound, controlled vocabulary is often simply too expensive in terms of development time and of the steep learning curve needed by the user of the system to learn the classification scheme (Fichter 2006; Kroski 2006; Quintarelli 2005; Shirky 2004). A further limitation of taxonomies is that they may become outdated easily. New concepts or products may emerge that are not yet included in the taxonomy; in comparison, folksonomies easily accommodate such new concepts (Fichter 2006; Mitchell 2005; Wu, Zubair, and Maly, 2006). Shirky (2004) points out that the advantage of folksonomies is not that they are better than controlled vocabularies, but that they are better than nothing.

Folksonomies follow desire lines, which are expressions of the direct information needs of the user (Kroski 2006; Mathes 2004; Merholz 2004). These desire lines also may reflect the needs of communities of interest: taggers who use same set of tags have formed a group and can seek each other out using simple search techniques. "Tagging provides users an easy, yet powerful method to express themselves within a community" (Szekely and Torres 2005).

Weaknesses of folksonomies

Folksonomies share the problems inherent to all uncontrolled vocabularies, such as ambiguity, polysemy, synonymy, and basic level variation (Fichter 2006; Golder and Huberman 2006; Guy and Tomkin 2006; Mathes 2004). The terms in a folksonomy may have inherent ambiguity as different users apply terms to documents in different ways. The polysemous tag *port* could refer to a sweet fortified wine, a porthole, a place for loading and unloading ships, the left-hand side of a ship or aircraft, or a channel endpoint in a communications system. Folksonomies do not include guidelines for use or scope notes. Folksonomies provide for no synonym control; the terms *mac, macintosh,* and *apple,* for example, are all used to describe Apple Macintosh computers. Similarly, both singular and plural forms of terms appear (for example, flower and flowers), thus creating a number of redundant headings. The problem with basic level variation is that related terms that describe an item vary along a continuum of specificity ranging from very general to very specific, so, for example, documents tagged *perl* and *javascript* may be too specific for some users, while a document tagged *programming* may be too general for others. Folksonomies provide no formal guidelines for the choice and form of tags, such as the use of compound headings, punctuation, word order, and so forth; for example, should one use the tag *vegan cooking* or *cooking, vegan*? Guy and Tomkin (2006) provide some general suggestions for tag selection best practices, such as the use of plural rather than singular forms, the use of underscore to join terms in a multiterm concept (for example, open_source), following conventions established by others, and adding synonyms. These suggestions are rather too vague to be of much use, however; for example, under what circumstances should singular forms be used (such as noncount nouns), and how should synonyms be linked?

Applications of folksonomies

Other than social bookmarking sites, folksonomies are used in commercial shopping sites, such as Amazon (http://www.amazon.com/), where clients tag items of interest; these tags can be accessed by people with similar interests. Platial (http://www.platial.com/splash) is used to tag personal collections of maps. Examples of the use of folksonomies for intranets include IBM's social bookmarking application Dogear, which allows people to bookmark pages within their Intranet (http://domino.watson.ibm.com/cambridge/research.nsf/99751d8eb5a20c1f852568db004efc90/1c181ee5fbcf59fb852570fc0052ad75?OpenDocument), and Scuttle (http://sourceforge.net/projects/scuttle/), an open-source bookmarking project that can be hosted on Web servers for free. PennTags (http://tags.library.upenn.edu/) is a social bookmarking service offered by the University of Pennsylvania Library to its community members. Steve Museum is a project that is investigating the incorporation of folksonomies into museum catalogs (Trant and Wyman 2006). Another potential application of folksonomies is to public library catalogs, where users can organize and tag items of interest in user-specific folders; users could then decide whether or not to post the tags publicly (Spiteri 2006).

Analyses of folksonomies

Analysis of the structure, or composition, of tags has thus far been limited; there has been more emphasis placed upon the co-occurrence of tags and their frequency of use. Cattuto, Loreto, and Pietronero (2006) applied a stochastic model of user behavior to investigate the statistical properties of tag co-occurrence; their results suggest that users of collaborative tagging systems share universal behaviors. Michlmayr (2005) compared tags assigned to a set of Delicious bookmarks to the DMOZ (http://www.dmoz.org/) taxonomy, which is designed by a community of volunteers. The study concluded that there were few instances of overlap between the two sets of terms.

Mathes (2004) provides an interesting analysis of the strengths and limitations of the structure of Delicious and Flickr, but does not provide an explanation of the methodology used to derive his observations; it is not clear, for example, for how long he studied these two sites, how many tags he examined, what elements he was looking for, or what evaluative criteria he applied.

Golder and Huberman (2006) conducted an analysis of the structure of collaborative tagging systems, looking at user activity and kinds and frequencies of tags. Specifically, Golder and Huberman looked at what tags Delicious members assigned and how many bookmarks they assigned to each tag. This study identified a number of functions tags perform for bookmarks, including identifying the:

- subject of the item;
- format of the item (for example, blog);
- ownership of the item; and
- characteristics of the item (for example, funny).

While the Golder and Huberman study provides an important look at tag use, their study is limited in that they examined only one site for a period of four days; their results are an excellent first step in the analysis of tag use, but the narrow focus of their population and sample size means that their observations are not easily generalized. Furthermore, this study focuses more on how bookmarks are associated with tags (for example, how many bookmarks are assigned per tag and by whom) rather than at the structural composition of the tags themselves.

Guy and Tonkin (2006) collected a random sampling of tags from Delicious and Flickr to see whether "popular objections to folksonomic tagging are based on fact." The authors do not explain, however, over what period the tags were acquired (for example, over a one-day period, over a month), nor to they provide any evaluative criteria. The tags were entered into Aspell, an open source spell checker, from which the authors concluded that 40 percent of Flickr and 28 percent of Delicious tags were either misspelled, encoded in a manner not understood by Aspell, or consisted of compound words of two or more words. Tags did not follow convention in such areas as the use of case or singular versus plural forms. While this study certainly focuses upon the structure of the tags, the bases for the authors' conclusions are problematic. It is not clear that the use of a spell checker is a sufficient measure of quality. Does the spell checker allow for cultural variations in spelling (for example, labor or labour)? How well-recognized and comprehensive is the source vocabulary for this spell checker? Furthermore, if a tag does not exist in the spell checker, does this necessarily mean that the tag is incorrect? Tags may include several neologisms, such as *podcasting,* that may not yet exist in conventional dictionaries but are well-recognized in a particular domain. The authors do not mention whether they took into account the correct use of the singular form of such tags as noncountable nouns (for example, air) or tags that describe disciplines or emotions (for example, history and love). If a named entity (person or organization) was not recognized by Aspell, does this mean that the tag was classified as incorrect? Lastly, the authors seem to imply that compound words of two or more words are necessarily incorrect, which may not be the case (for example, open source software).

The pitfalls of folksonomies have been well-documented; what is missing is an in-depth analysis of the linguistic structure of tags against an established benchmark. While popular opinion suggests that folksonomies suffer from ambiguous and inconsistent structure, the actual extent of these problems is not yet clear; furthermore, analyses conducted so far have not established clear benchmarks of quality pertaining to good tag structure. Although there are no guidelines for the construction of tags, recognized guidelines do exist for the construction of terms that are used in taxonomies. Although these guidelines discuss the elucidation of inter-term relationships (hierarchical, associative, and equivalent), which does not apply to the flat space of folksonomies, they contain sections pertaining to the choice and formation of concept terms that may, in fact, have relevance for the construction of tags.

Methodology

Selection of folksonomy sites

Tags were chosen from three popular folksonomy sites: Delicious, Furl, and Technorati (http://www.technorati.com/). Delicious and Furl function as bookmarking sites, while Technorati enables people to search for and organize blogs. These sites were chosen because they provide daily logs of the most popular tags that have been assigned by their members on a given day. The daily tag logs from each of the sites were acquired over a thirty-day period (February 1–March 2, 2006). The daily tags for each site were entered into an Excel spreadsheet. A list of unique tags for each site was compiled after the thirty-day period; *unique* refers to the single instance of a tag. Some of the tags were used only once during the thirty-day period, while others, such as *travel,* occurred several times, so travel appears only once in the list of unique tags. Variations of the same tag—for example, car or cars, Cheney or Dick Cheney—were considered to constitute two unique tags. Only English-language tags were accumulated.

The analysis of the tag structure in the three lists was conducted by applying the NISO guidelines for thesaurus construction, which are the most current set of recognized guidelines for the:

> contents, display, construction . . . of controlled vocabularies. This Standard focuses on controlled vocabularies that are used for the representation of content objects in knowledge organization systems including lists, synonym rings, taxonomies, and thesauri (NISO 2005, 1).

While folksonomies are not controlled vocabularies, they are lists of terms used to describe content, which means that the NISO guidelines could work well as a benchmark against which to examine how folksonomy tags are structured as well as the extent to which this structure reflects the widely accepted norm for controlled vocabularies. Section 6 of the guidelines (term choice, scope, and form) was applied to the tags, specifically the following elements (see appendix A for the expanded list):

6.3 Term choice
6.4 Grammatical form of terms
6.5 Nouns
6.6 Selecting the preferred form

Only those elements in section 6 that were found to apply to the lists of unique tags are included in appendix A. For each site, the section 6 elements were applied to each unique tag; for example, it was noted whether a tag consists of one or more terms, whether the tag is a noun, adjective, or adverb, and so on. The frequency of occurrence of the section 6 elements was noted for each site and then compared across the three sites in order to determine the existence of any patterns in tag structure and the extent to which these patterns reflect current practice in the design of controlled vocabularies.

Definition and disambiguation of tags

The meanings of the tags were determined based upon (1) the context of their use; and (2) their definition in three external sources, namely Merriam Webster online dictionary (http://www.m-w.com/); Google (http://www.google.com/); and Wikipedia (http://www.wikipedia.org/). Merriam-Webster was used specifically to define all tags other than those that constitute unique entities (for example, named people, places, organizations, or products) and to determine the various meanings of tags that are homographs (for example, *art* or *web*). The actual concept represented by homographs was determined by examining the sites or blogs to which the tag was assigned.

Merriam-Webster also was used to determine the grammatical form of a tag; for example, noun, verbal noun, adjective, or adverb. Determining verbal nouns proved to be complicated, especially given that NISO relies only on examples to illustrate such nouns. Some tags could serve as both verbal and simple nouns; for example, the tag *clipping* could describe the activity *to clip* or an item that has been clipped, such as a newspaper clipping. Similarly, does *skiing* refer to an activity, or the sport? If the dictionary defined a tag as an *activity,* the tag was classified as a verbal noun. In the case of tags that were defined as both verbal nouns and simple nouns, the context in which the tag was used determined the final classification.

The dictionary also was used to determine the type of concept represented by a tag. The NISO guidelines do not define any of these seven types of concepts outlined in section 6.3.2; they provide only a short list of examples for each type. If the term represented by the tag was defined as an activity, property, material, event, discipline or field of study, or unit of measurement, it was classified as such unless the context of the tag suggested otherwise. If none of these six types was defined in the dictionary, the default value of *thing* was assigned to the tag. These definitions were then compared to the context in which the tag was used. In the case of the tag *art,* for example, an examination of the sites associated with this tag indicated that it refers to art objects, rather than the discipline, so it was classified as a *thing*.

Merriam-Webster was used to determine whether a tag constitutes a recognized term in standard English (both United States and United Kingdom variants); for example, the tag *blogs* is a recognized term in the dictionary, while *podcasting* is not. NISO does not provide a clear definition of slang, neologism, or jargon, other than to say that they are nonstandard terms not generally found in dictionaries. Is the term *podcasting,* for example, an instance of slang, jargon, or neologism? At what point does jargon become a neologism? Because of the difficulty of distinguishing among these three categories, it was decided to use the broader category *nonstandard terms* to cover tags that (1) could not be found in the dictionary; or (2) are designated as vulgar or slang in the dictionary.

Google and Wikipedia were used to define the meanings of tags that constitute unique entities. Wikipedia also was used to distinguish the various meanings of tags that constitute abbreviations or acronyms via its disambiguation pages; for example, the tag *NFL* is given eight possible meanings. In this case, the tag *NFL* is used to refer specifically to the *National Football League,* so the tag is a homograph, noun, and unique entry.

Tagging conventions and guidelines of the folksonomy sites

Delicious

Delicious defines tags as:

> one-word descriptors that you can assign to your bookmarks. . . . They're a little bit like keywords but non-hierarchical. You can assign as many tags to a bookmark as you like and easily rename or delete them later. Tagging can be a lot easier and more flexible than fitting your information into preconceived categories or folders" (Del.icio.us 2006a).

The Delicious help page for tags encourages people to "enter as many tags as you would like, each separated by a space" in the tag field. This paragraph explains briefly that two lists of tags may appear under the entry form used to enter a bookmark. The first list consists of popular tags assigned by other people to the bookmark in question, while the second consists of recommended tags, which contains a combination of tags that have been assigned by the client in question as well as other users (Del.icio.us 2006b). It is not clear how the two lists differ in that they both contain tags assigned by other people to the bookmark at hand.

The only tangible guideline provided about how tags should be structured is the sentence "your only limitation on tags is that they must not include spaces." Delicious thus addresses only indirectly the fact that it does not allow multiterm tags; the examples provided suggest ways in which compound terms can be expressed; for example, San-francisco, SanFrancisco, San.franciso (Del.ico.us 2006b). Punctuation thus appears to be allowed in the construction of tags, which is confirmed by the suggestion that asterisks may be used to rate bookmarks: "a tag of * might mean an OK link, *** is pretty good, and a bookmark tagged ***** is awesome" (Del.icio.us 2006b). It is thus possible that tags may not consist of recognizable terms, even though asterisks are neither searchable nor indicative of content.

Furl

The Furl Web site uses the term *topics* rather than tags, but provides no guidelines or instructions for how to construct these topics. Furl mentions only that when entering a bookmark, "a small window will pop up. It should have the title and URL of the page you are looking at. Enter any additional details (i.e., topic, rating, comments) and click Save" (Furl 2006). Furl provides all users with a list of default topics to which one can add at will. Furl provides no guidelines as to whether single or multiword topics may be used; it is only by trial and error that the user discovers that the latter are, in fact, allowed.

Technorati

In its tags help page, Technorati encourages users to "think of a tag as a simple category name. People can categorize their posts, photos, and links with any tag that makes sense" (Technorati 2006). A tag may be "anything, but it should be descriptive. Please only use tags that are relevant to the post" (Technorati 2006). Technorati tags are embedded into individual blogs via the link rel="tag"; for example: <a href="http://technorati.com/tag/global+warming" rel="tag">global warming</a>. The tag will appear as simply global warming. No other guidelines are provided about how tags should be constructed.

As can be seen, the three folksonomy sites provide very few guidelines or conventions for how tags should be constructed. Users are not pointed to the common problems that exist in uncontrolled vocabulary, such as ambiguous headings, homographs, synonyms, spelling variations, and so forth, nor are suggestions made as to the preferred form of tags, such as nouns, plural forms, or the distinction between count nouns (for example, dogs) and mass nouns (for example, air). Given this lack of guidance, it is not unreasonable to assume that the tags acquired from these sites will vary considerably in form and structure.

Findings

Unless stated otherwise, the number of tags per folksonomy site is 76 for Delicious, 208 for Furl, and 229 for Technorati.

Homographs

The NISO guidelines recommend that homographs—terms with identical spellings but different meanings—should be avoided as far as possible in the selection of terms. Homographs constitute 22 percent of Delicious tags, 12 percent of Furl tags, and 20 percent of Technorati tags. Unique entities constitute a significant proportion of the homographs in all three sites, with 71 percent in Delicious, 43 percent in Furl, and 55 percent in Technorati. The most frequently occurring homographs across the three sites consist predominantly of computer-related terms, such as Ajax and CSS.

Single-word versus multiword terms

The NISO guidelines recommend that terms should represent a single concept expressed by a single or multiword term, as needed. Single-term tags constitute 93 percent of Delicious tags, 76 percent of Furl tags, and 80 percent of Technorati tags. The preponderance of single tags in Delicious may reflect the fact that it does not allow for the use of spaces between the different elements of the same tag; for example, *open source*.

Types of concepts

NISO provides a list of seven types of concepts that may be represented by terms; while this list is not exhaustive, it represents the most frequently occurring types of concept. Table 1 shows the percentage of tags that correspond to each of the seven types of concepts.

Tags that represent *things* are clearly predominant in the three sites, with activities and properties forming a distant second and third in importance. None of the tags represent events or measures, and only a fraction of the Technorati tags represent materials. The NISO guidelines provide no indication of the expected distribution of the types of concepts, so it is difficult to determine to what extent the three folksonomy sites are consistent with other lists of descriptors. None of the tags fell outside the scope of the seven types of concepts.

Unique Entities

Unique entities may represent the names of people, places, organizations, products, and specific events (NISO 2005). Unique entities constitute 22 percent of Delicious tags, 14 percent of Furl tags, and 49 percent of Technorati tags. There is no consistency in the percentage of unique entities: Technorati has nearly twice the percentage of tags than Delicious has, and nearly triple the percentage of tags than Furl has. Computer-related products constitute 100 percent of the unique entities in Delicious, 63 percent in Furl, and 38 percent in Technorati. The remainder of the unique entities in Furl and Technorati represent places, people, and corporate bodies. The unique entities in Technorati are closely related to developments in current news events, an occurrence that is likely due to the site's focus on blogs rather than Web sites. As will be discussed in a subsequent section, the unique entries constitute a significant proportion of the tags that represent ambiguous acronyms or abbreviated terms, such as *Ajax* or *PSP*.

Table 1. Concepts represented by the tags

	Delicious (%)	Furl (%)	Technorati (%)
Things	76	82	90.0
Materials	0	0	0.4
Activities	12	10	4.0
Events	0	0	0.0
Properties	8	6	4.0
Disciplines	4	3	1.0
Measures	0	0	0.0

Grammatical forms of terms

The NISO standards recommend the use of the following grammatical forms of terms:

- Nouns and noun phrases
  - verbal nouns
  - noun phrases
  - premodified noun phrases
  - postmodified noun phrases
- Adjectives
- Adverbs

Table 2 shows the distribution of the grammatical forms of tags.

If all the types of nouns are combined, then 95 percent of Delicious tags, 94 percent of Furl tags, and 97 percent of Technorati tags constitute types of nouns. The grammatical structure of the tags in the three folksonomy sites thus reflects very closely the NISO recommendations that tags consist of mainly nouns, with the added proviso that adjectives and adverbs be kept to a minimum. None of the folksonomy sites used adverbs as tags, and the number of adjectives was very small, forming an average total of 5 percent of the tags.

Nouns (plural and singular forms)

NISO divides nouns into two categories: Count nouns (how many?), and noncount, or mass nouns (how much?). NISO recommends that count nouns appear in the plural form and mass nouns in the singular form. NISO specifies other types of nouns that appear typically in the singular form:

- Abstract concepts
  - beliefs; for example, *Judaism, Taoism*
  - activities; for example, *digestion, distribution*
  - emotions; for example, *anger, envy, love, pity*
  - properties; for example, *conductivity, silence*
  - disciplines; for example, *chemistry, astronomy*
- Unique entities

Table 3 shows the distribution of the singular and plural forms of noun tags. The term *singular nouns* was used to collocate all the types of non-plural nouns.

Table 3 represents the number of tags that constitute count nouns; this does not mean, however, that the tags appeared correctly in the plural form. Of the count nouns, 36 percent of Delicious tags, 62 percent of Furl tags, and 34 percent of Technorati tags appeared correctly in the plural form. It should be noted that although table 3 indicates that properties constitute 8 percent of Delicious, 6 percent of Furl, and 4 percent of Technorati tags, most of these tags are adjectives, and thus are not counted in the table. The NISO guidelines do not suggest the typical distribution of count versus singular nouns, but table 3 indicates that at least among the three folksonomy sites, singular nouns form the bulk of the tags.

Table 2. Grammatical form of tags

	Delicious (%)	Furl (%)	Technorati (%)
Nouns	88	71	86
Verbal Nouns	5	6	4
Noun Phrases—Premodified	1	15	4
Noun Phrases—Postmodified	0	2	3
Adjectives	6	6	3
Adverbs	0	0	0

Table 3. Count and noncount noun tags

	Delicious (%)	Furl (%)	Technorati (%)
Count nouns	18	35	23
Noncount nouns	77	59	74
Mass nouns	36	32	19
Activities	12	10	4
Properties	3	0	1
Disciplines	4	3	1
Unique	22	14	49
Total	**95**	**94**	**97**

Spelling

The NISO guidelines divide the spelling of terms into two sections: warrant and authority. With respect to warrant, NISO recommends that "the most widely accepted spelling of words, based on warrant, should be adopted," with cross-references made between variant spellings of terms. As far as authority is concerned, spelling should follow the practice of well-established dictionaries or glossaries.

While spelling refers normally to whole words, I included in this analysis acronyms and abbreviations used to denote unique entities, such as countries or product names, as there are recognized spellings of such acronyms and abbreviations. Table 4 shows the tags from the three sites that do not conform to recognized spelling; the terms in italics show the accepted spelling.

The number of tags that do not conform to spelling warrant is clearly very few, constituting a total of 4 percent of the Delicious tags, 3 percent of the Furl tags, and 2 percent of the Technorati tags. Two of the nonrecognized spellings in Delicious are likely due to the difficulty of creating compound tags in this site, as was discussed earlier. The remainder of the tags conformed to recognized spellings as found in the three reference sources consulted. The findings suggest that tags are spelled consistently and in keeping with recognized warrant across the three folksonomy sites. Because of the international nature of the three folksonomy sites, no default English spelling was assumed. Table 5 shows those tags whose spellings reflect regional variations.

None of the three folksonomy sites featured lexical variants of any one tag. As the three sites are United States–based, the preponderance of American spelling is not surprising. What is surprising, however, is that Technorati features only the British variants in the total of tags examined in this study. It should be pointed out that the two lexical variants of these terms do appear in the three folksonomy sites; the two variants simply did not appear in the daily logs examined. No system to enable cross-referencing (for example, *Humour* USE or SEE *Humor*) exists in any of the three folksonomy sites, nor is cross-referencing discussed in the help logs of the sites.

Table 4. Tags that do not conform to spelling warrant

Delicious (N=76)	Furl (N=208)	Technorati (N=229)
Howto (*How to*)	Hollywood b-day (*Hollywood birthday*)	Met-art pics (*Metropolitan art pictures*)
Opensource (*Open source*)	Med-books (*Medical books*)	Superbowl (*Super Bowl*)
Toread (*To read*)	Oralsex (*Oral sex*)	Web-20 (Web2.0)

Abbreviations, initialisms, and acronyms

NISO recommends that the full form of terms should be used. Abbreviations or acronyms should be used only when they are so well-established that the full form of the term is rarely used. Cross-references should be made between the full and abbreviated forms of the terms. Abbreviations and acronyms constitute 22 percent of Delicious tags, 16 percent of Furl tags, and 19 percent of Technorati tags. The majority of these abbreviations and acronyms pertain to unique entities, such as product names (for example, *Flash, Mac,* and *NFL*). In the case of Delicious and Furl, none of the abbreviated tags is referred to also by its full form. Four of the abbreviated Technorati tags have full-form equivalents:

- Cheney/Dick Cheney
- IE/Internet Explorer
- Sheehan/Cindy Sheehan
- UAE/United Arab Emirates

Abbreviations and acronyms play a significant role in the ambiguity of the tags from the three sites; they represent 71 percent of the abbreviated Delicious tags, 45 percent of the abbreviated Furl tags, and 73 percent of the abbreviated Technorati tags. Furl and Technorati are very similar in the proportion of abbreviated tags used, but Delicious is significantly higher. The Delicious tags are focused more heavily upon computer-related products, which may explain why there are so many more abbreviated tags, as many of these products are often referred to by these shorter terms; for example, CSS, Flash, Apple, and so on.

Table 5. Tags that reflect regional spelling variations

Delicious (N=76)	Furl (N=208)	Technorati (N=229)
Humor (U.S. spelling)	Humor (U.S. spelling)	Favourite (British spelling)
	Jewelry (U.S. spelling)	Humour (British spelling)

Neologisms, slang, and jargon

The NISO guidelines explain that neologisms, slang, and jargon terms are generally not included in standard dictionaries and should be used only when there is no other widely accepted alternative. Nonstandard tags do not constitute a particularly relevant proportion of the total number of tags per site; they account for 3 percent of the Delicious tags, 10 percent of the Furl tags, and 6 percent of the Technorati tags. The nonstandard tags refer almost exclusively to either computer- or sex-related concepts, such as *Podcast, Wiki,* and *Camsex*.

Nonalphabetic characters

This section of the NISO guidelines deals with the use of capital letters and nonalphabetic characters. Capitalization was not examined in the three folksonomy sites, as none of them are case sensitive; Delicious and Furl, for example, post tags in lower case, regardless of whether the user has assigned upper or lower case, while Technorati shows capital letters only if they are assigned by the users themselves. The NISO guidelines state that nonalphabetic characters, such as hyphens, apostrophes (unless used for the possessive case), symbols, and punctuation marks, should not be used because they cause filing and searching problems. Table 6 shows the occurrence of nonalphabetic characters in the three folksonomy sites.

A very small proportion of the tags in the three folksonomy sites contains non-alphabetic characters, namely 1 percent of the Delicious tags, and 3 percent of the Furl and Technorati tags. As was discussed previously, the Delicious help screens may encourage people to use nonalphabetic characters to construct compound tags; in spite of this, however, such characters are not, in fact, used very frequently. It should be noted that the terms above were all searched, with punctuation intact, in their respective sites; in all three cases, the search engines retrieved the tags and their associated blogs or Web sites, which suggests that nonalphabetic characters may not negatively impact searching.

Table 6. Nonalphabetic characters

	Delicious (N=76)	Furl (N=208)	Technorati (N=229)
Hyphens	—	Hollywood b-day; URL-Project	Consumer-Credit; Web-2.0
Apostrophes	—	Mom's medical (possessive)	Valentine's Day (possessive)
Underscore	Safari_export	Blogger_life	—
Full stop	—	Web 2.0 (part of product name)	Web-2.0 (part of product name)
Forward slash	—	—	/Africa
+ sign	—	JCR+	—

Discussion and Recommendations

The tags examined from the three folksonomy sites correspond closely to a number of the NISO guidelines pertaining to the structure of terms, namely in the types of concepts expressed by the tags, the predominance of single tags, the predominance of nouns, the use of recognized spelling, and the use of primarily alphabetic characters.

Potential problem areas in the structure of the tags pertain to the inconsistent use of the singular and plural form of count nouns, the difficulty with creating multiterm tags in Delicious, and the incidence of ambiguous tags in the form of homographs and unqualified abbreviations or acronyms. As has been seen, a significant proportion of tags that represent count nouns appears incorrectly in the singular form. Because many search engines do not deploy default truncation, the use of the singular or plural form could affect retrieval; a search for the tag *computer* in Delicious, for example, retrieved 208,409 hits, while one for *computers* retrieved 91,205 hits. Some of the results from the two searches overlapped, but only if both the singular and plural forms of the tags coexist. It would thus be useful for the help features of the folksonomy sites to explain the difference between count and noncount nouns and to discuss the impact of the form of the noun upon retrieval.

While all three sites conform to the NISO recommendation that single terms be used whenever possible, some concepts cannot be expressed in this fashion, and thus folksonomy sites should accommodate the use of multiterm tags.

Furl and Technorati allow for their use, but make no mention of this feature in their help screens, which means that such tags may be constructed inconsistently—for example, by the insertion of punctuation—where a simple space between the tags will suffice. As has been seen, Delicious does not allow directly for the construction of multiterm tags, and in its instructions it actually promotes inconsistency in how various punctuation devices may be used to conflate two or three separate tags, once again at the detriment of retrieval, as is shown below:

Opensource: 103,476 hits
Open_source: 91, 205 hits
Open.source: 26,494 hits

Delicious should consider allowing for the insertion of spaces between the composite words of a compound tag; without this facility, users may be unaware of how to create compound tags. Alternatively, Delicious should recommend the use of only one punctuation symbol to conflate terms, such as the underscore. Furl and Technorati should explain clearly that compound tags may be formed by the simple convention of placing a space between the terms.

Ambiguous headings constitute the most problematic area in the construction of the tags; these headings take the form of homographs and abbreviations or acronyms. In the case of computer-related product names, it may be safe to assume that in the context of an online environment it is likely that the meaning of these product names is relatively self-evident. In the case of the tag *Yahoo,* for example, none of the sites or blogs associated with this tag pertained to "a member of a race of brutes in Swift's Gulliver's Travels who have the form and all the vices of humans, or a boorish, crass, or stupid person" (Merriam-Webster 2007), but referred consistently to the Internet service provider and search engine. On the other hand, the tag *Ajax* was used to refer to Asynchronous JavaScript and XML technology as well as to a number of mainly European soccer teams. Given the international audience of these folksonomy sites, it may be unwise to assume that the meanings of these homographs are self-evident.

Library of Congress Subject Headings often uses parenthetical qualifiers to clarify the meaning of terms—for example, *Python (Computer program language)*—even though this goes against NISO recommendations. It is unlikely, however, that such use of parentheses will be effective in the folksonomy sites. A search for *Opera (browser)*, for example, will likely imply an underlying AND Boolean operator, which detracts from the purpose and value of the parenthetical qualifier; this was confirmed in a Furl search, where the terms *Opera* and *Browser* appeared either immediately adjacent to each other or within the same document.

The application of the section of the NISO guidelines pertaining to abbreviations and acronyms is particularly difficult, as it is important to balance between using abbreviated forms of concepts that are so well-known that the full version is hardly used versus creating ambiguous tags. The fact that abbreviated forms appear so prominently in the daily logs of the three folksonomy sites suggests that the full forms of these tags are, in fact, very well-established. At face value, therefore, many of the abbreviated tags are ambiguous because they can refer to different concepts, but it is questionable whether such tags as *CSS, Flash, Apple,* and *RSS,* for example are, in fact, ambiguous to the users of the sites. The use of the full forms for these tags seems cumbersome, as these concepts are hardly ever referred to in their full form. It could possibly be argued, in fact, that in some cases, the full forms may not be familiar; I may know to what concept *RSS* refers, for example, without knowing the specific words represented by the letters R, S, S.

The possible ambiguity of abbreviated forms is compounded by the fact that none of the three folksonomy sites allows for cross-references between equivalent terms, which is a standard feature of most controlled vocabularies, for example:

NFL/National Football League
USE National Football League/Used For NFL

The help screens of the three sites do not address the notion of ambiguity in the construction of tags: They do not draw people's attention to the inherent ambiguity of abbreviated forms that may represent more than one concept. The sites also fail to address the fact that abbreviated forms (or any tag, for that matter) may be culturally based, so that while the meaning of *NFL* may be obvious to North American users, this may not be the case for people who live in other geographic areas. It may be useful for the folksonomy sites to add direct links to an online dictionary and to Wikipedia, and to encourage people to use these sites to determine whether their chosen tags may have more than one application or meaning; I had not realized, for example, that *RSS* could represent twenty-three different concepts until I used Wikipedia and was led to a disambiguation page. Access to these external sources may help users decide which full version of the abbreviation to use in the case of ambiguity.

The examination of the structure of the tags pointed to some deficiencies in section 6 of the NISO guidelines, specifically its occasional lack of sufficient definition or explanation of some of its recommendations. The guidelines list seven types of concepts that are typically represented by controlled vocabulary terms, but rely only upon a few examples to define the meaning and scope of these concepts. The guidelines thus provide no consistent mechanism by which the creators of terms can assess consistently the types of concepts represented. How, for example, is a *discipline* to be determined? Does the term *business* represent a discipline if it is a subject area that is taught formally in a post-secondary institute, for example? Is it necessary for a discipline to be recognized as such among a majority of educational institutions? In its examples for *events,* NISO lists *holidays* and *revolutions*. It is unclear, however, what level of specificity applies to this concept; would *Christmas,* for example, be considered an *event* or a *unique entity/proper noun* (which is listed separately from types of concepts)? It is only later in the guidelines, under the examples provided for unique entities (for example, *Fourth of July*), that one may assume that a named event should be considered a *unique entity*. Verbal nouns also are difficult to determine based only upon the NISO examples, and once again no guidelines are provided to determine whether a noun represents an activity or a thing, or possibly both; for example, *skiing* or *clipping*.

The lack of clear definitions in NISO also appeared in the section pertaining to slang, neologisms, and jargon, which are considered to be nonstandard terms that do not generally appear in dictionaries. As was discussed previously, it is not clear at what point a jargon term or a slang term becomes a neologism. All of the slang tags found in the three sites (for example, *babe*) appeared in Merriam-Webster, which may serve to make this NISO section even more ambiguous.

Conclusion

The most notable suggested weaknesses of folksonomies are their potential for ambiguity, polysemy, synonymy, and basic level variation as well as the lack of consistent guidelines for the choice and form of tags. The examination of the tags of the three folksonomy sites in light of the NISO guidelines suggests that ambiguity and polysemy (such as homographs) are indeed problems in the structure of the folksonomy tags, although the actual proportion of homographs and ambiguous tags each constitutes fewer than one-quarter of the tags in each of the three folksonony sites. In other words, although ambiguity and polysemy are certainly problematic areas, most of the tags in each of the three sites are unambiguous in their meaning and thus conform to NISO recommendations.

The help sites of the three folksonomy provide few tangible guidelines for (1) the construction of tags, which affects the construction of multiterm tags; and (2) the clear distinction between the singular and plural forms of count versus noncount nouns. As has been shown, the use of the singular or plural forms of terms, as well as the use of punctuation to form multiterm tags, affects search results. A large proportion of the tags in all three sites consists of single terms, which mitigates the impact on retrieval, but the inconsistent use of the singular and plural forms of nouns is indeed significant and thus may have marked effect upon retrieval. Synonymy and basic level variation were not examined in this study, but are certainly worthy of further exploration.

In other areas, the tags conform closely to the NISO guidelines for the choice and form of controlled vocabularies. The tags represent mostly nouns, with very few unqualified adjectives or adverbs. The tags represent the types of concepts recommended by NISO and conform well to recognized standards of spelling. Most of the tags conform to standard usage; there are few instances of nonstandard usage, such as slang or jargon. In short, the structure of the tags in all three sites is well within the standards established and recognized for the construction of controlled vocabularies.

Should library catalogs decide to incorporate folksonomies, they should consider creating clearly written recommendations for the choice and form of tags that could include the following areas:

- The difference between count and noncount nouns, as well as an explanation of how the use of the singular and plural forms affects retrieval.
- One standard way in which to construct multiterm tags; for example, the insertion of a space between the component terms, or the use of an underscore between the terms.
- A link to a recognized online dictionary and to Wikipedia to enable users to determine the meanings of terms, to disambiguate amongst homographs, and to determine if the full form would be preferable to the abbreviated form. An explanation of the impact of ambiguous tags and homographs upon retrieval would be useful.
- An acceptable use policy that would cover areas of potential concern, such as the use of potentially offensive tags, overly graphic tags, and so forth. Although such terms were not the focus of this study, their presence was certainly evident in some cases, and would need to be considered in an environment that includes clients of all ages.

With the use of such expanded guidelines and links to useful external reference sources, folksonomies could serve as a very powerful and flexible tool for increasing the user-friendliness and interactivity of public library catalogs, and also may be useful for encouraging other activities, such as informal online communities of readers and user-driven readers' advisory services.

Works Cited

Bateman, S., C. Brooks, and G. McCalla. 2006. *Collaborative tagging approaches for ontological metadata in adaptive e-learning systems.* http://www.win.tue.nl/SW-EL/2006/camera-ready/02-bateman_brooks_mccalla_SWEL2006_final.pdf (accessed Jan. 11, 2007).

Bruce, H., W. Jones, and S. Dumais. 2004. *Keeping and re-finding information on the web: What do people do and what do they need?* Seattle: Information School. http://kftf.ischool.washington.edu/re-finding_information_on_the_web3.pdf (accessed Jan. 11, 2007).

Cattuto, C., V. Loreto, and L. Pietronero. 2006. *Collaborative tagging and semiotic dynamics.* http://arxiv.org/PS_cache/cs/pdf/0605/0605015.pdf (accessed Jan. 11, 2007).

Del.icio.us. 2006a. *Del.ico.us/about*. http://del.icio.us/about/ (accessed Jan. 11, 2007).

Del.icio.us. 2006b. *Del.ico.us/help/tags*. http://del.icio.us/help/tags (accessed Jan. 11, 2007).

Dempsey, L. 2003. The recombinant library: portals and people. *Journal of Library Administration* 39, no. 4: 103–36.

Fichter, D. 2006. Intranet applications for tagging and folksonomies. *Online* 30, no. 3: 43–45.

Furl. 2006. *How to save a page in Furl*. http://www.furl.net/howToSave.jsp (accessed Jan. 11, 2007).

Golder, S. A., and B. A. Huberman. 2006. Usage patterns of collaborative tagging systems. *Journal of Information Science* 32, no. 2: 198–208.

Guy, M., and E. Tonkin. 2006. Tidying up tags? *D-Lib Magazine* 12, no. 1. http://www.dlib.org/dlib/Jan.06/guy/01guy.html (accessed Jan. 11, 2007).

Ketchell, D. S. 2000. Too many channels: making sense out of portals and personalization. *Information Technology and Libraries* 19, no. 4: 175–79.

Kroski, E. 2006. *The hive mind: folksonomies and user-based tagging.* http://infotangle.blogsome.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging/ (accessed Jan. 11, 2007).

Mathes, A. 2004. *Folksonomies—Ccooperative classification and communication through shared metadata.* http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html (accessed Jan. 11, 2007).

Merholz, P. 2004. *Ethnoclassification and vernacular vocabularies.* http://www.peterme.com/archives/000387.html (accessed Jan. 11, 2007).

Merriam-Webster. (2007). *Yahoo*. http://www.m-w.com/ (accessed Jan. 11, 2007).

Michlmayr, E. 2005. A case study on emergent semantics in communities. http://wit.tuwien.ac.at/people/michlmayr/publications/michlmayr_casestudy_on_emergentsemantics_final.pdf (accessed Jan. 11, 2007).

Mitchell, R. L. 2005. Tag teams wrestle with Web content. *Computerworld 38*, no. 16: 31.

NISO. 2005. *Guidelines for the construction, format, and management of monolingual controlled vocabularies*. ANSI/NISO Z39.19-2005. Bethesda, Md.: National Information Standards Organization. http://www.niso.org/standards/resources/Z39-19-2005.pdf (accessed Jan. 11, 2007).

Quintarelli, E. 2005. *Folksonomies: Power to the people.* http://www.iskoi.org/doc/folksonomies.htm (accessed Jan. 11, 2007).

Shirky, C. 2004. *Folksonomy.* http://www.corante.com/many/archives/2004/08/25/folksonomy.php (accessed Jan. 11, 2007).

Spiteri, L. F. 2006. The use of folksonomies in public library catalogues. *The Serials Librarian* 51, no. 2: 75–89.

Szekely, B., and E. Torres. 2005. *Ranking bookmarks and bistros: Intelligent community and folksonomy development.* http://torrez.us/archives/2005/07/13/tagrank.pdf. (accessed Jan. 11, 2007).

Technorati. 2006. *Technorati help:Tags*. http://www.technorati.com/help/tags.html (accessed Jan. 11, 2007).

Trant, J., and B. Wyman. (2006). *Investigating social tagging and folksonomy in art museums with steve.museum.* http://www.archimuse.com/research/www2006-tagging-steve.pdf (accessed Jan. 11, 2007).

Udell, J. 2004. *Collaborative knowledge gardening.* http://www.infoworld.com/article/04/08/20/34OPstrategic_1.html (accessed Jan. 11, 2007).

Vander Wal, T. 2006. *Understanding folksonomy: Tagging that works.* http://s3.amazonaws.com/2006presentations/dconstruct/Tagging_in_RW.pdf (accessed Jan. 11, 2007).

Vanderwal.net. 2005. *Folksonomy definition and Wikipedia.* http://www.vanderwal.net/random/entrysel.php?blog=1750 (accessed Jan. 11, 2007).

Wikipedia. 2006. *Folksonomy.* http://en.wikipedia.org/wiki/Folksonomy (accessed Jan. 11, 2007).

Wu, H., M. Zubair, and K. Maly. 2006. *Harvesting social knowledge from folksonomies.* http://delivery.acm.org/10.1145/1150000/1149962/p111-wu.pdf (accessed Jan. 11, 2007).

Appendix A: List of NISO elements

6.3 Term Form
6.3.1 Single Word vs. Multiword Terms
6.3.2 Types of Concepts
Terms for things and their physical parts
Terms for materials
Terms for activities or processes
Terms for events or occurrences
Terms for properties or states
Terms for disciplines or subject fields
Terms for units of measurement
6.3.3 Unique Entities
6.4 Grammatical Forms of Terms
6.4.1 Nouns and Noun Phrases
6.4.1.1 Verbal Nouns
6.4.1.2 Noun Phrases
6.4.1.2.1 Premodified Noun Phrases
6.4.1.2.2 Postmodified Noun Phrases
6.4.2 Adjectives
6.4.3 Adverbs
6.5 Nouns
6.5.1 Count Nouns
6.5.2 Mass Nouns
6.5.3 Other Types of Singular Nouns
6.5.3.1 Abstract Concepts
6.5.3.2 Unique Entities
6.6.2 Spelling
6.6.2.1 Spelling—Warrant
6.6.2.2 Spelling—Authorities
6.6.3 Abbreviations, Initialisms, and Acronyms
6.6.3.1 Preference for Abbreviation
6.6.3.2 Preference for Full Form
6.6.3.2.1 General Use
6.6.3.2.2 Ambiguity
6.6.4 Neologisms, Slang, and Jargon
6.7.1 Capitalization and Nonalphabetic Characters