

Exploring the Use of Category Archives in Library and Information Science Topical Blogs

Kristina M. Spurgin¹

SUMMARY. This paper examines category archive usage in library and information science (LIS) topical weblogs, or blogs. The research answers the following questions: 1) Are category archives used in LIS topical blogs?; 2) If so, how are they used?; and 3) Do some category labels appear frequently across LIS topical blogs? The study finds that category archives are used in a small percentage of the blogs in the sample. The number of categories in individual blogs varies widely. Finally, some category labels do appear frequently across blogs. Conceptual categories frequently used in archives also emerge.

KEYWORDS. Blogs, weblogs, categories, World Wide Web

The weblog, or blog, is a relatively recently developed type of website. The exact history of the development of blogs has been written about and debated extensively both in print and online. (Blood 2000; Winer 2002; Blood 2002; Rodzvilla 2002; Bausch, Haughey, and Hourihan 2002). Before weblogs and the tools used to create, publish, and host them on the web emerged, creating and maintaining a web presence required that a would-be web author knew how to create web pages using HTML code or HTML authoring applications, had access to web server space, and knew how to transfer files to and manage files on that server space. The availability of blogging software and services lowered the threshold of technical savvy required for publishing online.

¹ The author is a doctoral student at the School of Information and Library Science, University of North Carolina at Chapel Hill, CB #3360, 303 Manning Hall, Chapel Hill, NC 27599-3360. She holds an M.L.S. from SUNY Albany. E-mail: kspurgin@email.unc.edu

Now any Web user can start and maintain a blog merely by filling out a form or two, typing in content, and clicking “publish” or “post.” This has resulted in an explosion of information being added to the web in blog format. Understanding the usage and characteristics of this new information format is imperative for advancing knowledge about today’s rich information environment.

Before proceeding, it is necessary to define the terms used to describe this research: blogs, topical blogs, and category archives. The co-creators of the popular blog publishing tool Blogger define the blog as a format, explaining that “at their core, weblogs are pages consisting of several posts or distinct chunks of information per page, usually arranged in reverse chronology from the most recent post at the top of the page to the oldest post at the bottom” (Bausch, Haughey, and Hourihan 2002). Likewise, the definition of weblog in the forthcoming *Encyclopedia of Narrative Theory* (Walker 2005) focuses on format before describing the variations of blogs based on content and use characteristics. Blog content varies widely and may consist of intimate personal confessions, lists of one’s daily activities, political rants, current events and news on a particular topic, photos or multimedia, or research ideas. A blog may be authored by one person, or by any number of contributors. It may be a personal venture or an official publication of a company or institution. It may consist only of links—with or without commentary—or it may contain no links at all.

There is no generally accepted taxonomy of types of blogs. Blood (2002) places blogs into three extremely broad categories: short form journal blogs, notebooks, and filters. Krishnamurthy (2002) developed a matrix for classifying blogs on two axes: personal/topical and individual/community. Bar-Ilan (2004) suggests categorization of weblogs along three dimensions: content, format, and authorship. As of this writing, the Wikipedia entry for weblogs outlines seven types of blogs: personal (includes topical, thoughtful, and friendblogs),

collaborative, political, directory, corporate, advice, and format-based (ie, photo blog or audio blog) blogs. (Weblog 2001) Other sources have identified research blogs (Mortensen and Walker 2002) and knowledge management blogs, or k-logs, (Röll 2003) as distinct types of blogs.

Due to the variation widely noted in both the content and use of blogs, it is of limited usefulness to attempt to generalize findings across all types of blogs. Because there is currently no clear agreement on how to define the categories of blogs, researchers must be careful in constructing samples and generalizing and comparing their findings across blog types and blog studies.

Topical blogs

This paper examines the use of one blog feature in library and information science (LIS) topical blogs. A topical blog is a specific type of blog, defined by its content. The content is, not surprisingly, focused on a particular topic. A topical blog may be about anything—a professional field, a subject discipline, a hobby, a favorite musical group, or anything an author is interested in—however, the prototypical topical blog will keep off-topic and personally focused content to a minimum. Along with LIS blogs, some other professional groups noted to be creating topical blogs are legal professionals (blawgs) and medical professionals (medblogs). The topical blog may be written by a single author, or it may be made up of contributions from a group of authors or the community at large. Cohen refers to topical blogs as informational blogs and discusses how reading and writing them can help information professionals keep current in their field (Cohen 2003).

It should be noted that even if there did exist an agreed-upon taxonomy of blog types, it would be difficult to categorize individual blogs into such a scheme. The content and scope of a blog tends to shift over time as the author's interests, activities, or priorities change. One often mentioned example of this is that many blogs with previously different foci began to focus on

politics and news for a period after the attacks of September 11, 2001. Also, prototypical examples of the recognized types of blogs are difficult to find. Rare is the topical blog that does not, from time to time, contain personal news or information. Some topical blogs may cover more than one topic. Because a limited number of the most recent posts appear on the front page of a blog, what a reader sees when visiting that page at any one time may not be representative of the overall topic or scope of the blog.

Organizing Information in Blogs: Chronology and Category

The predominant method of organizing information in blogs has always been chronologically. An author-specified number of entries are ordered in reverse chronological order on the front page of a blog site. Usually the entries are displayed in a wide column which is flanked by a narrower column, referred to as a sidebar, containing various links and other information. Some blogs may have a sidebar on either side of the column of entries. The addition of new entries pushes older content off of the front page, and access to older content is provided via chronological archives, which are automatically generated by all blog publishing programs.

The sidebar almost always contains one or more links providing access to chronological archives. These links take three main forms. The first is a list of hyperlinked date ranges, usually months. Clicking on a date range takes the reader to a new page containing all entries made in that date range. Second, the sidebar may contain a small calendar with hyperlinked dates on days that entries were made. Clicking on a date takes the reader to all entries made on that day. Finally, some sidebars contain a link to a separate archives page that provides further access to the archives. Any blog may feature one or more of these types of access to chronological archives. Further, blog publishing software gives blog authors the option to turn off the archiving features altogether, though this seems to be rarely done.

This chronological ordering of information makes it easy for readers to get only the newest information from a blog, which enables keeping current whether the topic is one's profession or what one's friends did yesterday. However, purely chronological organization is linear, flat, and semantically limited, making it difficult to access older information when desired, unless the approximate or exact date of the publication of that information is known. Some blog publishing applications include a blog search function which can be used for locating specific information in a blog. However, searching is not a standard feature in all blogs, search functionality is simplistic, and the activity of searching a blog is quite different from how people typically interact with blogs through browsing. The use and usefulness of search in blogs has been discussed within the blogging community (Baumgart 2004), but to the researcher's knowledge there exists no empirical research on the presence or use of the feature in blogs. As such, it is an open area for further research. In the presence of flat and simplistic methods of organizing and accessing blog content, some blog authors desired a way to organize their blog entries conceptually. Around the middle of 2001, the introduction of category archive features in blog publishing software (such as MovableType) made this type of content organization possible.

In order to archive entries by category, a blog author must first create categories by entering category labels into the blog software package. When an entry is created, the entry composition interface will allow the author to choose one or more of those categories into which the entry will be archived. This may also be conceptualized as the development of a set of keywords, and the application of one or more keywords to each entry. For each category in a blog containing entries, an archive page containing all entries in that category is generated. Categories labels may be edited after they are created. The changes in the labeling will then update throughout the blog. Entries may also be re-categorized at any time. Finally, there is no requirement that any entry in a blog with a defined set of categories actually be placed into those categories.

Blogs with category archives indicate this using two main conventions. First, a hyperlinked list of the names of all categories in the blog may appear in a sidebar. Clicking on a category name will load a page containing all entries that have been placed in that category. Second, the header or footer of each entry may indicate that the entry has been “filed under” a category, or placed into a “department.” The category name here will also be hyperlinked, and clicking on it takes a reader to the archive page for that category. Some blogs use both methods of indicating category archives.

Affordances of category archives

Affordances are the action possibilities offered by objects (McGrenere and Ho 2000). Affordances may be examined at several different levels. Sellen and Harper (2002) examined the affordances of both paper (a medium) and digital content management systems (software applications) for managing information in the workplace. Gaver (1991) explored the affordances of small graphical components of interface design such as buttons and scrollbars.

As features in blogs, category archives also have affordances. The presence of a list of categories in a blog gives the reader an overview of the topics covered in a blog in order to make decisions about whether the blog is of interest. If each blog entry indicates the category into which it has been placed with a link to that category archive, a reader is able to easily access more content on the same topic. In blogs with both categories and site summary feeds, a user can subscribe to receive only the content from specific categories.²

Personal information management is commonly cited by blog authors as one motivation for starting or continuing to maintain a blog. One blog author refers to his blog as his “outboard brain” (Doctorow 2002). In her exploration of how blogs fit into knowledge work, Efimova

² Explaining what site feeds are and how they work is beyond the scope of this paper. Commonly referred to using such terms as syndication, aggregation, RSS, and ATOM, this technology concept and how it works in practice are nicely explicated in print by Cohen (Cohen 2003) and online by many, including Fagan (Fagan 2003).

(2004) characterizes blogs as “one-user content management systems,” and states that, “although far from providing full functionality, weblogs are used as personal information management tools.” By facilitating user-defined grouping of entries, category archives may enhance this personal information management function.

Finally, if the categorization of an entry is an indication of the aboutness of the entry, then category archives can be used as one “hook” on which to begin hanging applications to link or aggregate like content automatically across blogs (Cayzer). For example, the Internet Topic Exchange project is a promising tool for aggregating topical content across blogs, relying heavily on blog category archives. The blog author locates a Topic Exchange category that matches a topic that he writes about in his own blog. The author then manually configures the blog category that matches the Topic Exchange category so that the Topic Exchange is pinged when a post is published in that blog category. In this way, the author makes his concept of the intended content of the category explicit, and blog posts from various authors are automatically re-published in topic channels that readers may to subscribe to (Paquet and Pearson 2004).

The current body of published research literature on blogs is small, but growing. One group has published a series of papers exploring people’s motivations for starting and keeping blogs, their use of blogs in general, and the nature of social interactions in blogs in order to make design recommendations for blog software (Nardi et al. ; Schiano et al. ; Nardi, Schiano, and Gumbrecht). Another group has examined the characteristics of blogs as a genre, and as social networking and linking tools (Herring et al. 2004, 2005). None of these studies, however, has addressed the usage of the category archives feature of blogs.

The first two questions of the present research—are category archives used, and if so, how are they used?—are small steps toward beginning to determine whether the actual usage of category archives in blogs is such that it enables these affordances.

The third question asked in the current research—do some categories appear frequently across LIS topical blogs?—is based on the idea that examination of blog category archives, as information organization structures in an informal method of professional and scholarly communication, could constitute a form of domain analysis. The topics discussed in the blogging community of a discipline could indicate what topics participants in that discipline are focused on, and how those topics are conceptualized and referred to, before those topics make their way into such formats as the formal scholarly literature, indexes, or classification schemes (Hjørland 2002). The answer to this research question will begin to shed light on whether category archives are useful for this purpose.

This research has been carried out in topical blogs within the LIS discipline for two reasons. First, LIS is one discipline among several that have embraced blogs and are known to be using them widely in various ways. Second, it made sense to begin examining the use of category archives among a group of professionals that, on the whole, knows more about organizing and categorizing information than nearly anyone else. It would be interesting in the future to compare the LIS blogging community's use of category archives with that in other topical blogging communities.

Methodology

Previous blog studies sampled in various ways: random sampling from websites that host blogs or list recently updated blogs, purposive selection of one or two blogs of interest for careful analysis, and location-based convenience sampling with further snowballing of the sample. None of these methods would work for the current study since it examines blogs in only one particular discipline. Several valid methods for establishing sample frames for general web research have been proposed, including the use of “collector sites.” This method is recommended for gathering sites on a particular topic (Weare and Lin 2000). Many listings of LIS blogs exist, but two of the

largest, which are also highly regarded in the LIS blogging community, were used to populate the sample for this study. The first is the Open Directory Project's listings for LIS weblogs and LIS collaborative weblogs (http://dmoz.org/Reference/Libraries/Library_and_Information_Science/Weblogs/). Each listing in the Open Directory Project is maintained by an individual editor. There is also a listing for LIS personal blogs which was not consulted. The second listing is Libdex.com's list of library weblogs (<http://www.libdex.com/weblogs.html>). Libdex is also maintained by an individual, and contains many kinds of blogs related to libraries and LIS. Together, these two sources contained 355 listings. All LIS topical blogs in these listings were identified for inclusion in the sample through the process described below.

Each of the 355 listings was briefly examined to determine whether it should be added to a list of blogs to be more closely examined. At this stage, 180 listings were discarded for various reasons. Some of the blogs appeared in both sources, so one of the listings was discarded. Some of the blogs were identified as institutional or organizational blogs—official blogs of libraries or information centers which tend to serve as public relations tools. These were removed because they serve a different function than topical blogs. Some of the listings were direct links to LIS category archive pages within blogs, and were also removed. Finally, others were removed due to file not found errors, content in languages other than English and French that could not be read by the researcher, absence of entries, or presence of test entries only. After removing these, 175 unique LIS blogs remained for closer analysis.

The content of each of these 175 blogs was examined to remove personal blogs from the list. As mentioned before, not only does the topical focus of a blog tend to change over time, but some blogs have a mix of both personal and topical content by design. However, for the purpose of this study, an LIS topical blog was defined as one in which fewer than half of the entries on the front

page at the time of examination were of a personal nature. Thirty-one blogs were identified as personal blogs using this criteria, leaving a total of 144 topical blogs.

For each of these 144 topical blogs, data on three variables were collected. First, the presence or absence of category archives was recorded. Second, each blog was coded as either an individual blog or a collaborative blog. An individual blog is a blog written by one person, while a collaborative blog may have two or more contributors. Because collaborative blogs have more than one author—and some have many authors—they may have greater numbers of entries than individual blogs, seemingly increasing the value of category archives to organize the entries. Also, a category scheme in a collaborative blog is used and possibly designed by more than one person, with all the implications of conceptual misalignment that could generate. Finally, the blog software used to publish and maintain the blog was recorded. It is conventional for a blog to display a graphical or textual link to the software with which it is published. When such a link was absent, the source code of the blog was examined for identifying code. When the software used could not be verified using the source code, an email of inquiry was sent to the blog author if an email address was included on the blog. Further, the online documentation for each blog software tool used by blogs in the sample was examined to determine whether or not the software supports category archiving.

Of the 144 topical blogs, 43 were found to use category archives. Since weblogs change frequently, the content of the blogs with categories was downloaded to the researcher's computer in order to freeze the blogs for analysis. This was accomplished in most cases by using the HTTrack and WebReaper website mirroring software tools, though the content of a few of the blogs had to be manually downloaded using the Save As function of the web browser.

Next, the category archive labels from each of the blogs were entered into an Excel spreadsheet. Upon closer examination of the category archive labels, it became apparent that two

of the blogs were not LIS topical blogs. One is maintained by a librarian in an academic library, but the categories and content of the blog focus on issues of higher education and pedagogy. The second appeared to be an LIS topical blog because, on initial examination, the front page entries concerned search engines and information literacy. However, the vast majority of this blog's content is comprised of links to non-LIS websites that may be of use to internet researchers or reference librarians. This is an example of how the presence of a categories in a blog can help a user determine the scope of the blog's content. Since these blogs did not focus on LIS content and topics, they were removed from the sample, leaving a total of 142 LIS topical blogs, 41 of which contain category archives. The data collected previously were adjusted to reflect the removal of these blogs from the sample.

The number of categories per blog was counted and SPSS was used to generate descriptive statistics on the number of categories per blog overall, and number of categories per individual and collaborative blog. Then, a frequency count of the author-assigned category labels was conducted to identify categories that are used most often in LIS topical blogs.

Finally, the content of categories with conceptually similar category labels was examined. Categories that covered primarily the same topics and links, but had differing category labels were brought together under an iteratively devised conceptual category scheme. In this case, for example, category labels such as "blogs," "blogging," "weblogs," and "weblogging" were brought together. Conversely, some categories with identical category labels were found to contain divergent topics. For example, in some cases the category of "books" was used to collect book reviews, while in others it was used to collocate entries on news stories regarding books. A frequency count was run on these conceptual categories to identify category concepts used most often in LIS topical blogs, to account for basic differences in terminology usage by blog authors.

Limitations

There are several limitations in this study and its methodology. The first is that it is a basic, descriptive, and exploratory examination of a particular type of blog in one discipline. Due to the sampling method used, there is no guarantee that the sample includes, or is representative of all LIS topical blogs. The findings cannot be generalized outside the data gathered. However, given the current lack of empirical research about category archive usage in blogs, beginning to understand how this feature is used, even in a limited study, is valuable.

Second, classifying blogs as topical or personal based on their front page entries at a given point in time may result in misclassification. Defining criteria for inclusion in a sample always runs this risk. Due to the relatively small number of blogs excluded due to their personal content, given that the Libdex.com listing includes personal blogs, it is assumed that the number of blogs misclassified is very small.

Finally, conceptual coding based on semantic similarity of author-assigned category labels does not truly gather all conceptually similar categories under one term. The inconsistent level of granularity at which authors have assigned categories, and the extent of overlap in the content contained within and across author-assigned categories make it impossible to draw hard boundaries for what author-assigned categories belong in a conceptual category. For example, an author-assigned category named “legislation” may contain entries about the Freedom of Information Act, the USA Patriot Act, the Children’s Internet Protection Act, and other topics. Meanwhile the same blog (to say nothing of all blogs in the sample) may contain other author-assigned categories such as “Patriot Act” or “Filtering” which contain the same or similar entries. In the absence of scope notes for all author-assigned categories, it is impossible for the researcher to know the blog authors’ criteria for what belongs in a particular category, so the researcher’s judgment of semantic and content similarity is necessarily somewhat subjective. In the interest of exploration of these research questions, with no expectation of generalization, conceptual

categorization based on author-assigned category labels and examination of content within each category is deemed sufficient to gain an idea of the concepts discussed in these blogs. A more objective analysis of the content of blogs would require statistical analysis of the text and links contained in the blogs, which is beyond the scope or aim of this research, but a promising direction for future research.

Findings and Interpretation

Question 1: Are category archives used?

Forty-one (29%) of the 142 LIS topical blogs in the sample have category archives. This low number is misleading on its own, however. The left-most three columns of Table 1 show the frequency of software tools used to publish all the blogs in the sample. Only 71 (50%) of the blogs are published using software that supports category archives. Software support for category archives is unclear for 9 (6%) of the blogs. In the case of 8 of the blogs, the software used is unknown, either because there was no way to contact the author, or because there was no response to the inquiry email. In these cases, support for category archives cannot be determined. Bloxom does not include category archives by default, but there is a plugin to enable category archiving. Hand coded blogs could conceivably include category archives, but none of the hand coded blogs examined here did, likely because of the complexity of maintaining another set of archives without the support of a database-driven blogging software package.

Several factors could influence the choice of software used to maintain a blog. The first is the level of technical sophistication of the blog author. Some software tools such as Blogger and LiveJournal allow a user to create a blog simply by filling out a web form. Others, such as php Weblog and Slash are open source projects requiring installation on a server and offering little documentation. Access to or ownership of server space on the web is another factor. Blogger's Blog*Spot service hosts blogs for free, while RadioUserland and TypePad host more advanced

blogs for a fee. Packages such as MovableType require that a blog author install the software and host the blog on his own web server space. Finally, the date a blog was started is a factor. A blog started before mid-2001 would not have been able to implement category archives from its inception, because software supporting the feature had not yet been released. Migration of blog content from one software package to another is possible, but requires a certain level of technical knowledge and confidence to attempt and accomplish.

Of the 71 blogs known to be produced with software that supports category archiving by default, 38 (54% of blogs with categorization features available) make use of the category archives. Support for category archives and other features increases the complexity of blog software. Some blog authors may not realize that category archiving is available to them. Also, setting up a category archive scheme requires an investment of time and cognitive effort that some blog authors may not be willing to spend. The questions of what motivates the choice of blog software and the use of features in blog software remain to be explored in future research.

Question 2: How are category archives used?

Determination of the number of categories used in blogs is a first step toward answering this question. In total, there were 676 categories in the 41 LIS topical blogs with category archives. Table 2 presents basic descriptive statistics on the number of categories in blogs overall, and the number of categories in individually-authored and collaborative blogs. These blogs are examined separately to explore whether collaborative blogs in the sample tend to have more categories than the individually-authored blogs. Since collaborative blogs have more authors, they are likely to have more entries, and perhaps a need for more categories. Figure 1 visually represents the number of categories per blog. Each column indicates a blog, with the height of the column relative to the number of categories in that blog. The black bars indicate collaborative blogs.

The mean number of categories per blog is 16.5. This is skewed high due to the extreme case of a collaborative blog with 79 categories. This blog has over twice as many categories as the blog with the next highest number of categories. The sample does not meet assumptions for comparison of means using an independent samples t-test, but examination of Figure 1 shows that though several collaborative blogs are clustered in the half of the chart indicating more categories, the two blogs with the most categories (excluding the extreme value) are individual blogs. Likewise, one of the blogs with the fewest categories is a collaborative blog. In this sample, there is no clear indication that collaborative blogs have more categories.

Question 3: Do some categories appear frequently across LIS topical blogs?

Table 3 presents the author-assigned category archive labels that appear most frequently in the LIS topical blogs in this sample. This table may be read as follows: “books” appears as an author-assigned category in 8 blogs, or in 19.5% of all blogs with categories archives.

Table 4 presents the conceptual categories that are most frequently represented in the blogs. It should be noted that the fact that a conceptual category appears 20 times does not indicate that it appeared in 20 blogs. For example, a single blog may have contained one category called “Google,” and another called “search engines.” Both of these categories would have been placed into the “searching” conceptual category. This table may be read as follows: 19 author-assigned categories fall into the conceptual category “searching.”

Some of the conceptual category labels are self explanatory, but others may not be. “Resources” groups categories populated with links to and information about sites, tools, and materials that other information professionals might find helpful in their professional work, or that users might find helpful in their own information seeking. “Searching” represents categories containing search tips and news about databases and search engines. “About the Blog” groups categories that hold entries that describe changes and updates to the blog, and information about

the authors' credentials and why they maintain the blog. "Open source" categories contained both links and reviews of open source software, along with musings about the open source philosophy. "Librarians" groups categories containing entries about librarians, the librarian image, and the culture of librarianship. "Links" are distinguished from "resources" by the fact that entries in these categories did not indicate that the linked sites were of professional interest for use in information seeking. "Project status and reports" categories were used to report on professional activities and accomplishments. Finally, "LIS research" categories contained entries about conducting research in the discipline of LIS.

The author-assigned categories that appear frequently reflect, for the most part, traditional library interests and concerns. Grouping the categories conceptually introduces a number of concepts that are not as traditionally library-bound, as well as others that are very important in libraries. Perhaps one explanation of the difference is that library science (LS) has a much longer history as a discipline than does information science (IS). Perhaps the phrases and terms used to refer to the issues in the former are more established and stable than those still forming for the latter, and therefore more likely to be uniformly applied spontaneously by blog authors.

Even with the introduction of non-library specific concepts in the latter table, the conceptual categories are still very much library- and librarian-related. This is likely an artifact of the sample, in that the directories used to populate the sample listed mostly library-focused blogs, despite the fact that dmoz.org includes IS in its directory's category labels. Perhaps this is because library blogs are easy to recognize and classify as such because the word "library" is usually quite prominent in the site. IS as a discipline is more diffuse than LS. Blog authors who write about IS topics may identify as information architects, user interface designers, database professionals, or a wide variety of other information-related roles. Because of this, IS blogs are scattered and more difficult to define and identify.

Because of this, the interpretation of whether blog category labels as information structures would be useful for domain analysis work is not clear-cut. Within LS blogs, the analysis of category labels may be useful for tracing the topics, trends, and vocabulary used to refer to concepts in the field. LS is relatively cohesive as a discipline or domain. LS topical blogs are easily identified as such, and often appear to be interconnected with each other in online communities of practice.

Because IS is a relatively new discipline and meta-field that cuts across traditional topically-oriented disciplines and is applied in them, the boundaries of IS as a content domain are difficult to limn (Bates 1999). To illustrate this with an extreme example, consider that both information seeking by children and bioinformatics fall within the scope of information science, however blogs on these topics would be likely to have very little in common. The boundaries of a domain must be identified before one can begin to analyze it through its products and artifacts.

At least in this sample, the blogs that the researcher identified as IS topical blogs were self identified as being about sub-topics within, or interdisciplinary with, information science such as information architecture, information design, or computer science. Due to the previously mentioned sample bias toward library science blogs, and because of this tendency of IS blogs within the sample to cluster into topical subgroups, the results of this study are inconclusive about the usefulness of IS topical blog categories as content for domain analysis.

Conclusions and further questions

Category archives are indeed used in LIS topical blogs, but blogs with categories are still the minority. The current stage of this research does not indicate much about how category archives are actually used in these blogs. The number of categories in blogs varies widely, but collaborative blogs do not necessarily have more categories than individual blogs. Finally, a number of author-assigned and conceptual categories do appear frequently across the blogs in the

sample. The frequently appearing categories do appear to be consistent with what is expected in this domain. However, the difficulty of devising conceptual categories as discussed above is indicative of how inconsistently category labels are applied across these blogs.

In light of the small percentage of blogs that use category archives, and the heterogeneity of terminology used to label categories even within a field such as library science, other approaches, such as text co-occurrence and link co-citation analysis appear to be more promising for automatic clustering of related blog content on a large scale than will any category archive-based approaches. However, it is important to understand how the new communication format of weblog is being used, in its various applications. In order to further that goal, more investigation of the use of the category archive feature is necessary.

To understand the motivations in choosing blog software and whether or not to use category archive features, further research must be conducted with blog authors. The use, or non-use, of category archives by both authors and readers of blogs should be investigated. A number of further questions present themselves. How do blog authors devise category archive schemes? How are entries placed into these categories? Are they evenly distributed across categories, or do entries in a blog tend to cluster in a few core categories? What functions do category archives in blogs serve for both blog authors, for organizing and accessing their content, and readers, for finding and accessing new information?

As the number of blogs continues to grow, and the format is used to organize and communicate increasing amounts and differing types of information, the body of research on the format, its uses, and features should also grow so that we may understand this new information phenomenon.

REFERENCES

Bar-Ilan, Judit. 2004. An outsider's view on "topic-oriented" blogging. In *International World*

Wide Web Conference: Proceedings of the 13th international World Wide Web conference, 28-34. New York: ACM Press.

Bates, Marcia J. 1999. The invisible substrate of information science. *Journal of the American Society for Information Science* 50, no. 12: 1043-50.

Baumgart, Jessica. "The failings of search engines and why calendars on blogs are handy," *j's scratchpad*, 3 March 2004, <<http://blogs.law.harvard.edu/jkbaumga/2004/03/03#a850>> (6 February 2005).

Bausch, Paul, Matt Haughey, and Meg Hourihan. 2002. *We blog : Publishing online with weblogs*. Indianapolis: Wiley.

Blood, Rebecca. "Weblogs: A history and perspective," *Rebecca's Pocket*, 7 September 2000, <http://www.rebeccablood.net/essays/weblog_history.html> (6 February 2005).

———. 2002. *The weblog handbook: Practical advice on creating and maintaining your blog*. Cambridge, MA: Perseus Publishing.

Cayzer, Steve. 2004. Semantic Blogging: Spreading the Semantic Web Meme. In *Proceedings of XML Europe 2004*. <http://www.idealliance.org/papers/dx_xmle04/papers/03-05-03/03-05-03.pdf>

Cohen, Steven M. 2003. *Keeping current : advanced Internet strategies to meet librarian and patron needs*. Chicago: American Library Association.

Doctorow, Cory. "My blog, my outboard brain," *O'Reilly Network*, 31 May 2002, <<http://www.oreillynet.com/pub/a/javascript/2002/01/01/cory.html>> (6 February 2005).

Efimova, Lilia. 2004. Discovering the iceberg of knowledge work: a weblog case. White paper. <<https://doc.telin.nl/dscgi/ds.py/Get/File-34786>>.

Fagan, Michael. "Explanation of RSS, how you can use it, and finding RSS feeds," *Fagan Finder*, 19 February 2004, <<http://www.faganfinder.com/search/rss.shtml>> (6 February

2005).

Gaver, William W. 1991. Technology affordances. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology*, 79-84. New York: ACM Press.

Herring, Susan C., Lois Ann Scheidt, Sabrina Bonus, and Elijah Wright. 2004. Bridging the gap: A genre analysis of weblogs. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04)*. Los Alamitos, CA: IEEE Computer Society.

Herring, Susan C., Inna Kouper, John C. Paolillo, Lois Ann Scheidt, Michael Tyworth, Peter Welsch, Elijah Wright, and Ning Yu. 2005. Conversations in the blogosphere: An analysis "from the bottom up". In *Proceedings of the thirty-eighth Hawai'i international conference on system sciences (HICSS-38)*, 107b. Los Alamitos, CA: IEEE Press.

Hjørland, Birger. 2002. Domain Analysis in Information Science - Eleven Approaches - Traditional as Well as Innovative. *Journal of Documentation* 58, no. 4: 422-62.

Krishnamurthy, Sandeep. 2002. The multidimensionality of blog conversations: The virtual enactment of September 11. Presented at *Internet Research 3.0: Net / Work / Theory*.

McGrenere, Joanna, and Wayne Ho. 2000. Affordances: Clarifying and evolving a concept. In *Proceedings of graphics interface 2000*.

Mortensen, Torill, and Jill Walker. 2002. Blogging thoughts: Personal publication as an online research tool. In *Researching ICTs in context*. Edited by Andrew Morrison, 249-78. Oslo: University of Oslo Press.

Nardi, Bonnie A., Diane J. Schiano, and Michelle Gumbrecht. 2004. Blogging as social activity, or, would you let 900 million people read your diary? In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 222-31. New York: ACM Press.

Nardi, Bonnie A., Diane J. Schiano, Michelle Gumbrecht, and Luke Swartz. 2004. The

blogosphere: Why we blog. *Communications of the ACM* 47, no. 12: 41-46.

Paquet, Sebastien, and Phillip Pearson. 2004. A topic sharing infrastructure for weblog networks.

In *Proceedings of the Communication Networks and Services Research (CNSR) 2004*. Edited by Ali A. Ghorbani, 301-4. Los Alamitos, CA: IEEE Computer Society Press.

Pikas, Christina. "BloggerCon II: Librariansesque session with Jessica Baumgart," *Christina's LIS Rant*, 18 April 2004,

<http://christinaslibraryrant.blogspot.com/2004_04_18_christinaslibraryrant_archive.html#108233010104782384> (6 February 2005).

Robinson, D. Keith. "Blog searching -- Your take," *Asterisk: Where web design lives*, 21

December 2003, <http://www.7nights.com/asterisk/archives/blog_searching_your_take.php> (6 February 2005).

Rodzvilla, John. 2002. *We've got blog : how weblogs are changing our culture*. Cambridge, MA: Perseus Publishers.

Röll, Martin. 2003. Business weblogs -- A pragmatic approach to introducing weblogs in medium and large enterprises. Presented at *BlogTalk: A European conference on weblogs*.

Sellen, Abigail J., and Richard H. R. Harper. 2002. *The myth of the paperless office*. Cambridge, MA: The MIT Press.

Walker, Jill. 2005. Weblogs. To be published in *The Routledge Encyclopedia of Narrative Theory*. Edited by David Herman, Manfred Jahn, and Marie-Laure RyanRoutledge.

Weare, C., and W. Y. Lin. 2000. Content Analysis of the World Wide Web - Opportunities and Challenges. *Social Science Computer Review* 18, no. 3: 272-92.

"Weblog." *Wikipedia, the free encyclopedia*. 6 February 2005,

<<http://en.wikipedia.org/wiki/Weblog>> (6 February 2005).

Winer, Dave. "The history of weblogs," *Weblogs.com news*, 17 May 2002,

<<http://newhome.weblogs.com/historyOfWeblogs>> (6 February 2005).

Software Used	# of blogs	% of sample	# with categories	% with categories
Software has Category Archive feature				
MovableType	46	32%	24	52%
RadioUserland	8	6%	4	50%
Manila	6	4%	3	50%
b2	2	1%	1	50%
Conversant	1	1%	1	100%
IBlog	1	1%	1	100%
php Weblog	1	1%	1	100%
Scoop	1	1%	1	100%
Slash	1	1%	1	100%
Typepad	1	1%	0	0%
Zope	1	1%	1	100%
LiveJournal	2	1%	0	0%
Subtotal	71	50%	38	54%
Presence of Category Archive feature unclear				
Cannot determine	8	6%	2	25%
Blosxom	1	1%	1	100%
Subtotal	9	6%	3	33%
Software has no Category Archive feature				
Blogger/Blogspot	54	38%		
Hand Code	6	4%		
Blogcity	1	1%		
Crimsonblog	1	1%		
Subtotal	62	44%		
Total	142	100%	41	

Table 1: Blog Software Used in LIS Topical Blogs and Category Archive Usage in Blogs Using Software Supporting the Feature

	All	Individual Blogs	Collaborative Blogs
n=	41	31	10
Mean	16.5	13.7	25
Std. Dev.	13.4	8.8	20.9
Median	12	11	18
Range	75	34	74

Table 2: Descriptive Statistics on Number of Categories per Blog

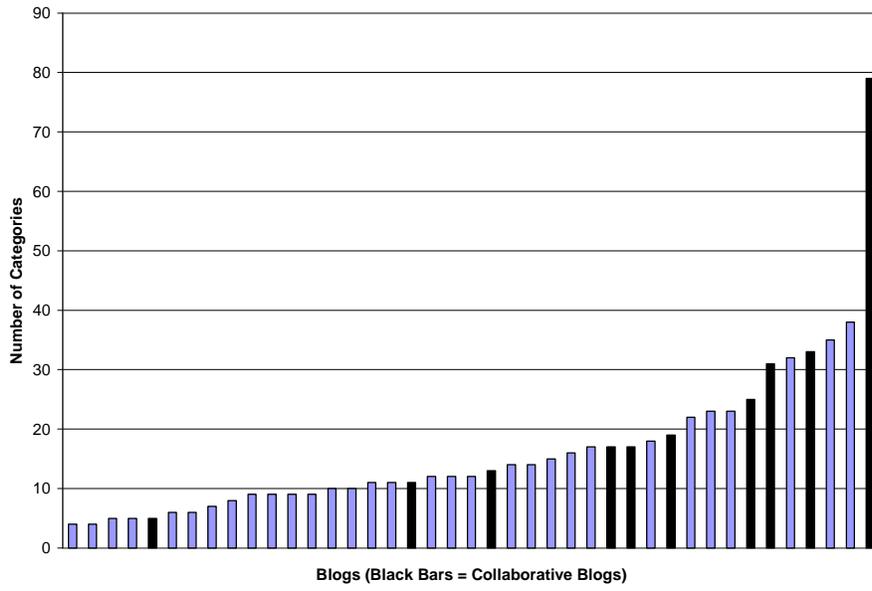


Figure 1: Number of Categories per Blog

# of blogs (% of total blogs)	Author Assigned Category Labels
8 (19.5%)	Books
7 (17%)	News
6 (14.6%)	Events; Knowledge management; Libraries; Technology
5 (12.2%)	Censorship; Information literacy; Intellectual property; Librarianship; Patriot Act; Search engines
4 (10%)	Copyright; Digital libraries; Education; Librarians; Research; Resources; RSS; Software

Table 3: Author-Assigned Categories Appearing in 4 or More Blogs

# of Appearances	Conceptual Categories
31	General library news
23	Technology
22	Miscellaneous
21	Resources
19	Searching
15	Blogs
14	About the blog; Copyright
13	Conferences; Digital Library
12	Book news; Metadata: RSS
10	Accessibility/open access; Knowledge management; Personal
9	Open source
7	Email; Legislation; Education of librarians; Librarians; Links; LIS literature; Patriot Act; Project status and reports; LIS research; Work issues

Table 4: Most Frequently Occurring Conceptual Categories