

## Introduction

Although other social media tools and websites have begun to gain recognition, blogging still remains one of the most popular ways for libraries to promote their resources, facilities, and upcoming events to their users. Libraries can establish a flow of information to reach out to their patrons by posting to their blogs that allow patrons to engage in interactive communication and open discussion online through the use of commenting (Stephens, 2006). Commenting is one of the important features in blogs as it allows users not only to interact and share their opinions with the library but also to engage in discussions with other users at the same time (Stuart, 2009).

However, not all comments submitted to a blog are appropriate. Some of these comments may be unwanted and considered as spam due to their malicious content and intent. E-mail spam is the most commonly known spamming technique where unsolicited bulk e-mail messages are sent to a number of recipients indiscriminately. In blogs, spamming occurs in the form of comment spam. Mishne *et al.* (2005) defined comment spam as a “link spam originating from comments and responses added to web pages which support dynamic user editing.” Comment spam usually provides links to websites or some text that are unrelated to the blog post. Sometimes, such comments may be considered offensive because of their inappropriate language and the inclusion of links to websites with graphic content. The comments include these links because blogs are given a high level of prestige in search rankings as they are frequently updated and share a density of links (Mishne *et al.*, 2005). This kind of spamming can either be done by a person submitting comments to a blog one at a time or, often times, by computer programs designed to automatically submit large numbers of spam comments in a short period of time. Automated spam is considered a much greater problem than human generated spam because of the sheer volume of comments it produces (Six Apart, n.d.). According to Akismet, a content filtering service for blogs, eighty-four percent of all comments are spam (Akismet, n.d.).

Comment spam is a problem for blogs since it has the potential to interfere with genuine interactions and discussions between libraries and their users (Crosby, 2010, p.85). It also violates the current context of a blog, as comment spam pertains to completely different issues and topics than the original post and adds no value to the discussion (Six Apart, n.d.). Comment spam is also an issue for library blogs, because it can be enough of a nuisance to lead some administrators to turn off comments on their blogs (Rutherford, 2008). Turning off comments clearly poses a problem for the library blogger, who is deprived of a valuable chance for interactive communication with their users. If libraries do not wish to cut off this communication line it is important to understand there are a number of effective comment spam prevention and detection techniques and tools available. Often times, these techniques and tools are freely available or could be implemented with minimal effort. As current library literature on the subject (Crosby, 2010; Rutherford, 2008) is very limited and does not address the problem at length, results of this study provide empirical evidence regarding level of commenting and the impact of comment spam on library blogs. The results and findings of the study can offer guidance to libraries that are reconsidering whether or not to allow commenting in their blogs and to those that are planning to establish a blog to reach out to their users while keeping this online environment engaging and interactive. In the context of this study, blogs affiliated with libraries where the library is responsible for content published on the blog are defined as library blogs (Bar-Ilan, 2007). This study focuses on the impact of comment spam in library blogs. The following section begins with a review of the relevant literature on blogs in the context of

libraries, and continues with review of the literature on anti-spam strategies and their use in social websites, specifically in blogs.

### **Literature Review**

The level of commenting is an important indicator used to determine success of a blog. However, both Clyde (2004) and Rutherford (2008) argued that the level of commenting in library blogs is low and blogs are not updated frequently. This level of commenting may be attributed to a number of factors including purpose of the blog, salience of blog posts, and spam prevention techniques used (Rutherford, 2008; Clyde, 2004; Six Apart, n.d.). Professional bloggers also rely on additional measures including the number of unique visitors, number of subscribers to the blog's feed, and number of links from other sites (trackbacks) to assess the success of their blogs (Crosby, 2010; Sussman, 2009).

Clyde (2004) and Bar-Ilan (2007) found that some library blogs are established for providing news and updates for library users, and as such, these blogs only facilitate one way communication from library to the user. Crosby (2010) identified additional measures that apply to library blogs that do not attract a high volume of traffic or comments. These measures included circulation counts of materials featured on the blog. Bar-Ilan (2007) and Rutherford (2008) found that some blog posts (e.g., gaming-related) attracted more comments than others as they seem more appealing to a younger segment of the population. A recent study on generational differences in online activities found that teens (ages of 12 to 17) and Gen Y (ages of 18 to 32) spend more time in reading blogs than other age groups (Johns and Fox, 2009). Therefore salient blog posts about needs or interests of users in these age groups are more likely to motivate users to engage with the blog and generate more comments. Models based on theories of motivation for participation (Kraut *et al.*, 2010) may be used to explain the low level of, or lack of, commenting in library blogs. Maher (2010) proposed a set of motivation categories which may be used to explore user participation in library blogs. A few of these motivations included ideology (purpose of being part of a larger cause), fun (purpose of enjoyment or excitement), reward (purpose of earning tangible rewards), and recognition (purpose of receiving public or private acknowledgment).

As blogs are designed to foster an interactive communication between the blog owner and readers through commenting, it is in a libraries' best interest to make commenting easy and less time consuming for readers so that they may be more inclined to share their thoughts, ideas, and get involved with the library. Of course, making it easy for readers to add comments may also allow spammers to easily take advantage of this medium and abuse it (Six Apart, n.d.). There are a number of ways to combat comment spam in blogs. However, the problem of spam prevention and detection is about a problem of convenience as blog administrators want to block as much comment spam as possible with minimal effort and let all legitimate comments go through (Six Apart, n.d.).

Heymann *et al.* (2007) identifies three main anti-spam strategies for social websites: prevention-based, detection-based, and demotion-based. The prevention-based strategy aims at making adding spam content more difficult by limiting automated interaction (e.g., limiting number of comments submitted to a post, challenge-response tests) or by changing system details (e.g., changing the name of the comment script). This strategy attempts to prevent spammers from submitting malicious or unwanted content. The detection-based strategy attempts to combat spam after it is successfully submitted. This strategy works in two stages. In the first stage, likely

spam content may be identified by human moderators or by automated tools (e.g., text classification, link analysis). In the next stage, the system may automatically delete likely spam content or flag them as likely spam. The demotion-based strategy aims at reducing the prominence of content which is more likely be spam. However, this strategy is not applicable to blogs as comments are listed in a reverse-chronological order and the recency of comments is important. Therefore, relevant literature on prevention- and detection-based methods is reviewed below.

### Prevention-based Methods

One of the most commonly used spam prevention technique is the Turing Test. Turing Tests were originally designed to test machines for intelligence, but their use as a standard security measure on websites has been well established (Heymann *et al.*, 2007, Ahn *et al.*, 2008). The most popular Turing Test used in blogs is the Completely Automated Public Turing Test to Tell Computers and Humans Apart (CAPTCHA), which is a test that is solvable by humans, but challenges the current capabilities of computers (Heymann *et al.*, 2007). Usually, a CAPTCHA involves a distorted word presented as an image that a human can type in a text box to show they are human and not a computer program. CAPTCHAs tend to be very effective at blocking automated processes from submitting spam, but they may pose important accessibility issues (May, 2005). Some CAPTCHAs take a different approach by posing simple questions to users. These questions may be random mathematics questions (e.g., What is  $3 + 5$  ?) or simple questions (e.g., What is the opposite of hot?) that could easily be answered. These questions are also displayed as a distorted image to prevent automated systems from recognizing its content. CAPTCHAs are also successfully used to support book digitization initiatives while preventing spam. This application of CAPTCHA is called reCAPTCHA. The reCAPTCHA displays a distorted version of word which was not recognized by optical character recognition (OCR) applications during book digitization along with a known word which is also distorted (Ahn *et al.*, 2008).

A World Wide Web Consortium (W3C) Working Group report (May, 2005) noted that CAPTCHAs can pose major accessibility problems to those who are blind, dyslexic, or have low vision. There are currently solutions in place to provide audio alternative of the CAPTCHA words, but added background noise and issues due to localization (i.e., English versus other languages) may serve as additional accessibility barriers for visually disabled users (Yan and Ahmad, 2008). Although CAPTCHA-based solutions do not completely prevent or eliminate spamming, they are very effective in slowing down spammers (Timmer, 2008; Prasad, 2008).

Because spam comments are usually generated in large numbers by an automated program, they can come in waves (Six Apart, n.d.). One spam prevention technique that uses this aspect of spamming to block comment spam is called “comment throttling.” Comment throttling enables blog moderators to put a limit on the number of blog comments that can be submitted to a given post or in a given amount of time (Ringalda, 2005). However, this technique applies an “all or nothing” approach that might keep real users from commenting. Because all comments are cut off once the throttle is reached, it’s possible that legitimate comments could be blocked by the throttle (Six Apart, n.d.).

Another prevention approach is aimed at making it difficult for automated spam programs to locate the comment script file name by changing its name or adding additional required fields. Changing the script name or adding additional fields does not impact the

functionality of the blog but it helps to slow down the spam. In addition, a JavaScript may be incorporated to generate comment fields. Since fields on a comment serve as variables when a comment is submitted and JavaScript must be rendered to generate the form, spammers must take an extra step to circumvent this measure. Although this method does not eliminate spam, it is an effective method in preventing spam.

No matter how many spam comments come in, they are usually submitted anonymously. Therefore, another useful spam prevention method requires some kind of consistent authentication or identification information from users to add their comments. Hosted blogs and content management systems provide tools to authenticate users if the blog administrator decides to use them. This method is very effective at preventing spam in that it requires the commenter to have an account and lets the administrator moderate comments to prevent any spam that does get through. However, the signup process may discourage casual commentators from joining in conversations (Six Apart, n.d.).

Though most spam prevention methods are automated, blog administrators can still choose to manually moderate some of the comments or all of them that are submitted to their blogs. Moderating only old entries or closing commenting on old entries is another prevention approach as spammers are more likely to target their efforts on these entries because of their perceived lower visibility. However, this does nothing to fight spam in recent posts. Moderating is very effective at blocking comment spam, because comments are presented to a human moderator for approval before they are published on the blog. However, comment moderation can increase the time spent on spam prevention if the level of commenting including comment spam is very high in a blog.

Since spammers add linked comments to blogs to increase their websites' ranking by manipulating search engine ranking algorithms, Google introduced a new tag for hyperlinks (`rel="nofollow"`) as an automated spam prevention method on blogs in early 2005 (Cutts and Shellen, 2005). Basically this tag would be automatically added any link submitted as an entry on a blog and these links will not get any credit in website rankings when search engines index these blog websites. Although major search engines and social software providers agreed to support this tag, this method has not proven to be useful in preventing spam as it did not prevent unsuspecting users from clicking such links and ignoring the value of legitimate links (Çelik and Marks, 2009). Google is now revisiting the way this tag linking is handled (Newcomb, 2009).

### Detection-based Methods

Content Filtering has been best known for its use in fighting email spam, but it is also a popular technique for preventing comment spam. This technique involves parsing comments and their headers to check for differences between legitimate comments and spam (Six Apart, n.d.). One model that seems to have success in filtering spam comments in the context of blogs is language model disagreement. This filtering technique checks how much the comment resembled the original post's language to detect spam (Mishne *et al.*, 2005). Another model that is suggested in the literature is collaborative spam filtering. In this approach, a blog searches to check if a comment has been identified as spam by other blogs in a peer to peer network (Han et al. 2006). There are also statistical content filtering techniques such as Bayesian spam filtering, Controllable Regex Mutilator (CRM114), and DSPAM. These methods calculate probabilities of certain words or phrases occurring in a text to distinguish a legitimate comment from spam (Thomason, 2007). Though such content filtering methods show promise in their ability to detect

and eliminate comment spam, it should be noted that they may also block legitimate comments from getting through (Thomason, 2007; Mishne *et al.*, 2005).

A promising content filtering tool for blogs is the Automatic Kismet which is offered as a Web service (Akismet, n.d.). This tool is set up to filter and detect spam comments submitted to blogs. When a comment is added, it is automatically transmitted to Akismet, which runs a number of tests on the comment to determine if it is spam and flags the comment for the blog administrator to approve. There is not much information available regarding the specific tests Akismet runs, but it appears that it relies not only on statistical and collaborative content filtering techniques mentioned above, as well as other email spam classification techniques.

## Methodology

The goal of this research was to identify and describe the impact of comment spam in library blogs. Three research questions guided the study to achieve this goal:

- What is the current level of commenting in library blogs?
- How do library blog administrators perceive problem of comment spam?
- How do library blog administrators address problem of comment spam?

The researchers used a quantitative approach to investigate these research questions. Since the literature in the area of comment spam in libraries is very limited, four academic and three public libraries that ran active blogs were contacted for informal interviews to develop a better understanding of the problem and then to develop an appropriate data collection instrument. These libraries were sent emails asking what blog software they used and how many spam comments they received. They were also asked what spam detection and prevention methods they used and what they did if any spam comments made it through the filter. If the libraries had detailed comment policies, they were also asked how they came up with those policies. Based on the feedback received from these blog administrators, a survey questionnaire was developed. The questionnaire was pilot tested with the same group of blog administrators and several questions were refined as suggested.

The final questionnaire consisted of eleven questions including demographic questions to determine the size of the population served and type of library. There were also questions about how many entries the blog generates and how many comments the blog receives. Survey participants were also asked about which comment spam prevention and detection methods they use, how much time per week they spend on spam prevention, and how much of a problem they think comment spam is for the blog. The questionnaire was distributed online using SurveyMonkey.com, an online survey service. In addition, two twenty-five dollar Amazon.com gift certificates were offered to two randomly selected respondents as an incentive offer and token of appreciation for participation.

The researchers identified four mailing lists (see Appendix A for more information) that primarily focused on marketing and technology in libraries. The questionnaire was sent out to these mailing lists. In addition, a number of library blogs were identified from the Blogging Libraries Wiki (2010) and blog administrators were invited to take the survey. The survey was left open for two months. During this period, the survey received 108 responses. One response

was incomplete and was not included in the analysis. Descriptive statistics were used to analyze the survey data.

## Results

The libraries that responded to the survey were relatively evenly distributed among public and academic libraries. Slightly more public libraries responded to the survey than academic libraries as shown in Table 1. The great majority of responses were from public libraries and four-year academic libraries. Two two-year academic libraries also responded to survey, but no special libraries. Though a library that labeled itself a “statewide multi-type consortium” and another as a “library system” and chose the “other” category, both libraries are categorized as public libraries. Two libraries chose “other” because they offered a mix of two and four year academic programs or they offered graduate programs in addition to their four year programs. An academic medical library and a digital library consortium covering ten academic institutions also chose “other”. These responses were categorized as academic libraries.

Table 1. Types of libraries (n=107)

	Percentage
Academic Library	48.6 %
Public Library	51.4 %

The community size served by library blogs appears to be very diverse as shown in Table 2. About thirty percent of the blogs served a community size of fewer than 10,000 and a majority of them belonged to academic libraries. A little over forty-five percent of the blogs served a community size larger than 10,000 but less than 100,000, and a majority of these blogs were affiliated to public libraries. About twenty-three percent of the blogs served a community size of larger than 100,000 and only one academic library reported that its blog serves a community of 100, 000 or more.

Table 2. Size of community the blog serves by library type (n=106)

		Library Type		
		Academic Library	Public Library	Cumulative
Community Size	Under 10,000	94%	6%	31%
	10,000 – 19,999	57%	43%	13%
	20,000 – 49,000	42%	58%	18%
	50,000 – 99,999	27%	73%	14%
	Above 100,000	4%	96%	24%

The libraries did not seem to have a large volume of comments. Regardless of the library type and the size of the community these blogs served, about eighty-five percent of respondents

reported they received fewer than five comments per week. About the same number of blog administrators also indicated that they put out less than five blog posts per week as shown in Table 3. It appears that there were not a lot of blog entries to comment on. However, data show that an increase in the number of blog entries did not necessarily increase the level of commenting. Half of the blogs that generated between six and fifteen posts in a week received more than five comments on average each week. On the other hand, one respondent who posted between sixteen to twenty entries a week received five or fewer comments a week and one blog received more than twenty comments for five or fewer blog posts per week. In addition fourteen respondents reported that commenting is disabled in their library blogs.

Table 3. The Number of Blog Posts Generated and Level of Commenting (n=104)

		Blog Posts (per week)					
		5 or fewer	6 – 10	11 – 15	16 – 20	21 or more	Cumulative
Comments (per week)	5 or fewer	91.0% (n=81)	66.7% (n=6)	20.0% (n=1)	100% (n=1)	-	85.6%
	6 – 10	4.5% (n=4)	33.3% (n=3)	40.0% (n=2)	-	-	8.7%
	11 – 15	-	-	20.0% (n=1)	-	-	0.95%
	16 – 20	-	-	20.0% (n=1)	-	-	0.95%
	21 or more	4.5% (n=4)	-	-	-	-	3.8%

Nearly half of the respondents (n=53) indicated they take some sort of preventive measure against comment spamming as shown in Table 4. Of the respondents who reported using a spam prevention strategy, about twenty-eight percent replied that commenting is disabled in their blogs. Overall, this figure corresponds to a little over fourteen percent of all respondents. The rest indicated that they use a combination of various techniques including keeping the blog software updated, Turing tests (CAPTCHAs), and requiring registration for commenting as shown in Table 4. Comments left by respondents to this question indicate that blog administrators are more inclined to adopt built-in spam prevention tools in the blog software they use, as keeping the software updated keeps spam prevention tools up-to-date. A few blog administrators reported the use of more technical measures including comment throttling, adding new fields to the comment form, and obfuscating comment script in preventing spam.

Table 4. Spam Prevention Techniques (n=53)

	Response Totals (n=53) Percentage	Cumulative (n=107) Percentage
--	--------------------------------------	----------------------------------

Keep Blog Software Updated	43.4%	21.5%
Turing Tests	30.2%	14.9%
Disable Comments	28.3%	14.1%
Require Registration	20.8%	10.3%
Comment Throttling	5.6%	2.8%
Add Additional Fields to the Comment Form	3.8%	1.9%
Obfuscate the Comment Script	1.9%	0.9%

About one third of those who reported commenting is disabled on their blogs indicated they use at least one form of filtering technique to detect spam. A few of those indicated they ran multiple blogs with commenting enabled in one of them and disabled in another or commenting is enabled for certain blog posts but not for all.

The use of filtering and comment moderation is also used as a strategy to identify and eliminate comment spam in library blogs which appears to be more popular than spam prevention techniques. Four respondents reported using Akismet service or WordPress spam filter but they did not select the “content filtering” option in the survey. Since Akismet is the default comment spam detection module in WordPress and is a content filtering tool, the researchers took these responses into consideration when analyzing survey data (WordPress Codex, n.d.). Three respondents noted they moderate comments after they are published on the blog and did not select the “moderating all comments” option. Since this is a form of moderating all comments added to blogs, these responses were also taken into consideration in data analysis. Of the seventy percent of respondents who reported using filtering techniques on their blogs, most indicated they moderated comments on all entries. About twenty six percent chose content filtering programs and about thirteen percent said they monitored or closed comments on old entries. It should be noted that respondents used a combination of these filtering techniques as shown in Table 5. Overall, about fifty-nine percent of libraries responding to this survey chose to moderate the comments added to their blogs while nearly thirty percent of blog administrators chose not to use a filtering technique to detect comment spam. Survey data also indicated that use of spam prevention and filtering techniques does not appear to be related to the library type or community size the blog served. Almost all academic libraries used at least one form of filtering technique while about twenty-five percent of public libraries chose not to use any filtering technique.

Table 5. Types of Filtering Techniques Used\* (n=75)

	Response Totals (n=75) Percentage	Cumulative (n=107) Percentage
Moderating all Comments	84.0% (n=63)	58.9% (n=63)
Content Filtering	26.7% (n=20)	14.9% (n=20)



Moderating or Closing Comments on Old Posts	13.3% (n=10)	9.3% (n=10)
No Filtering	-	29.9% (n=32)

\*Multiple responses allowed.

One of the most important factors in combating comment spam is how much time a blog administrator has to spend to address the problem. The survey asked respondents to indicate amount of time they spent on spam prevention and detection, on average, per week. About sixty-five percent of respondents spent 30 minutes or less in dealing with comment spam. A little over twenty-five percent of respondents indicated they spent no time on spam prevention and about half these respondents chose to disable commenting; therefore, they did not have to contribute any time for spam prevention. Less than eight percent of respondents spent between 31 and 60 minutes and nearly two percent of respondents spent between 61 and 120 minutes. No one reported spending over 120 minutes per week on combating comment spam as shown in Table 6. Those who rated comment spam prevention on their blog as somewhat important or higher still primarily (about seventy percent) spent about 30 minutes or less on spam prevention and detection. However, nearly all of the respondents (ninety percent) who spent over 30 minutes on spam prevention rated the problem as somewhat important or higher. Twenty percent of them rated it as very important or essential.

Table 6. Time Spent on Spam Prevention per week (n=107)

	Percentage
No Time	25.2%
1-30 Minutes	65.4%
31-60 Minutes	7.5%
61-120 Minutes	1.9%

These figures did not change significantly based on how respondents rated the overall problem of comment spam on the library blog. No matter how many comments the blog was receiving, the majority of survey respondents did not consider comment spam to be an important problem for their blog. About fifty-seven percent of respondents rated the problem of comment spam as not important while a little over thirty-one percent indicated it was somewhat important and about ten percent felt the problem was important or essential as shown in Table 7.

Table 7. Library Blog Administrators' Perception of Comment Spam and Offensive Comment Problems

	Perception of Comment Spam problem (n=106)	Perception of Offensive Comment problem (n=105)
Not Important	57.5%	60.0%

Somewhat Important	31.1%	29.5%
Very Important	6.6%	8.6%
Essential	4.7%	1.9%

A majority of public librarians (over fifty-seven percent) felt comment spam was not important while about fifty-six percent of academic librarians did not think comment spam was an important problem as shown Table 8. The figures for those who felt the problem was very important or essential were different for public and academic libraries. Nearly fifteen percent of public libraries regarded the problem as very important or essential while over seven percent of academic libraries viewed the comment spam problem as very important or essential. In addition, data showed that perception of comment spam as a problem is not necessarily a function of number of comments a blog receives, as ten respondents whose library blogs received ten or less comments per week rated comment spam as very important or essential problem. On the other hand, three respondents whose library blogs received more than sixteen comments in a given week indicated that comment spam is a somewhat important problem.

Table 8. Perception of comment spam problem by library type (n=106)

	Academic Library (n=52)	Public Library (n=54)
Not Important	55.8%	57.4%
Somewhat Important	36.6%	27.8%
Very Important	3.8%	9.3%
Essential	3.8%	5.5%

## Discussion and Conclusions

Three research questions guided this study to identify and describe the comment spam problem in library blogs. The first research question attempted to understand the level of commenting in library blogs so that the threat posed by comment spam could be better understood in terms of its ability to interrupt communication process in blogs. Over eighty percent of libraries that responded to the survey reported receiving five or fewer comments for their blog entries in a given week. Interpreting this number objectively is challenging as there are no standards to measure this against; however, the level of commenting appears fairly light in library blogs as seventy-seven percent of library blogs served communities larger than 10,000 people. This finding is consistent with Clyde (2004) and Rutherford's (2008) assessments on low level of commenting in library blogs. As Clyde (2004) and Bar-Ilan (2007) argued that the way communication flows in the blog may be an important factor impacting user participation. A few comments left in the survey appear to support this assessment as respondents argued that their blogs may be seen "boring" since they are oriented towards sharing updates about collections and news.

About ten percent of respondents reported requiring users to register to be able to comment on blog posts, and nearly all of these respondents indicated that they receive five or fewer comments per week, as did those who do not require users to register to be able to comment. Only one of those respondents reported receiving about six to ten comments each week. Both those who reported receiving five or fewer comments and those received six or more comments per week utilized a combination of various spam prevention and detection techniques. Data were not conclusive that use of such strategies plays an important role in influencing user participation in library blogs.

The second research question is aimed at understanding how library blog administrators perceive the problem of comment spam, including offensive spam. A majority of blog administrators (about ninety percent) reported that they spend thirty minutes or less in a given week to deal with comment spam and about one-third of these respondents indicated they spend no time at all. These blog administrators reported that comment spam does not pose a major threat nor does it take too much of their time to manage. A little over thirty percent of respondents (n=33) considered comment spam somewhat important while about fifty-seven percent of respondents (n=61) indicated that comment spam does not pose a threat. Nearly all of these respondents (n=92) who considered comment spam not important or somewhat important reported spending thirty minutes or less to deal with comment spam. Although it appears there is a positive relationship between time spent on combating spam and perception of blog administrators, a majority of respondents (n=10) who considered comment spam as a very important or essential problem also reported spending thirty minutes or less to combat spam. The survey data does not provide evidence to explain this viewpoint. However, a recent experience with a failing spam prevention or detection tool may have influenced blog administrators' perceptions. Two representative comments from respondents included “[o]ur touring test app doesn't work, so a lot of time is devoted to mass deleting spam comments,” and “[i]t [content filter] can't seem to tell junk from real comments. Granted the comments have gotten smarter and smarter.”

The third research question is aimed at understanding measures taken by blog administrators to address potential problems posed by comment spam. Data suggest that regardless of the library type and community size a blog serves, both spam prevention and detection techniques were used in library blogs where commenting is allowed. Detection-based approaches, specifically comment moderation, were more popular than prevention-based approaches. Prevention-based techniques used by blog administrators suggest they have a tendency to adopt spam prevention tools that are readily available in the blog software. Though the survey questionnaire did not include a question regarding the blog software used, a review of the library blogs listed at “the blogging libraries wiki” indicated that Blogger and WordPress are the most frequently used platforms for library blogs as shown in Table 9 (“The Blogging Libraries Wiki,” 2010; Blogger, n.d.; WordPress, n.d.).

Table 9. Blog Platforms by Library Type (n=825)

	Academic Library	Public Library	Total
Blogger/Blogspot	152	285	437
WordPress	116	108	224

Movable Type	12	11	23
Type Pad	12	11	23
B2Evolution	3	14	17
Drupal	9	6	15
Engaged Patrons	0	6	6
Other	14	9	23
Undetermined	23	34	57

The Blogger, the most common blogging platform used by libraries, is a free and hosted blog publishing platform. Since blogs created using Blogger reside on Blogger's servers, the spam prevention and detection tools offered are limited to Turing tests (i.e., CAPTCHA), user registration, and comment moderation or disabling commenting (Blogger Help, 2010). The WordPress as an open source content management system is also popular as a blogging platform among libraries, and it is offered as a hosted service similar Blogger or can be hosted by libraries themselves. About half these libraries using WordPress use the hosted WordPress service in which the libraries' ability to prevent and detect spam is limited to Akismet and comment moderation or disabled commenting. Libraries that choose to host their own WordPress, however, have a number of solutions offered not only by WordPress but also by products created by third party developers. Keeping blog software updated is rated as the most commonly used spam prevention technique as it updates the software as a whole including third party products should there be an update available for them (this update process may work differently for different blogging software).

Although spam prevention and detection techniques do not seem to impact user behavior on library blogs, libraries that require users to register themselves with the blog to add their comments should consider alternative methods to simplify this process. For example, about one in two Americans had Facebook accounts at the time of writing this article; allowing users to add their comments using their Facebook credentials will not only reduce the burden on blog administrators in managing user accounts but also lower the access barrier for users and save them time (Facebook, 2010). Alternatively, OpenID is a promising open source single sign-on initiative that blog administrators should consider adopting as major social software providers support it ("OpenID Foundation," n.d.). The OpenID allows users to use an existing account (e.g., Facebook) to log on to a number of websites without needing to create a new account.

Although a small number of libraries chose not to allow commenting on their blogs, the research has provided empirical evidence that comment spam is not regarded as an important problem and doesn't interfere with the communication process in library blogs. Library blog administrators are making a conscious effort to prevent and identify spam and offensive comments by successfully using a combination of various methods and tools. However, these tools and methods were often limited to features available in blogging platforms or software used to create the blog. Although creating a blog is a relatively inexpensive investment as there are a number of free platforms and software available. Maintaining an up-to-date and spam-free blog that reaches out to the community and keeps them engaged requires an investment in library resources.

This study demonstrates that there is a need for developing metrics to assess success of a library blog so that the libraries can make informed decisions about their blogs when allocating resources, making marketing plans, or assessing online library services. Further study is needed to investigate how libraries are evaluating their blogs, what kind of user tracking systems are being utilized, how they are being promoted to the community, and how effectiveness of various methods of communicating with library users is assessed. A review of blogging guidelines and policies in libraries may also shed a light to better understand the communication behavior of users in library blogs.

## Appendix A

### Descriptions of Mailing Lists Used to Solicit Participation.

#### 1. Academic PR

Listserv for topics in library marketing. Dedicated to library professionals who are involved in marketing their libraries. The list currently has 760 subscribers.

<http://lists.ala.org/wws/info/academicpr>

#### 2. Lita-L

This mailing list is focused on all aspects of information technology in libraries and is directed at those who work in positions that work with IT in libraries.

<http://www.ala.org/ala/mgrps/divs/lita/litamembership/litaldisclists/litalotherdiscussion.cfm>

#### 3. PUBLIB

This mailing list is focused on library-based web functions. It is aimed towards librarians who work with web applications.

<http://lists.webjunction.org/publib/>

#### 4. Web4Lib

This mailing list is focused on library-based web functions. It is aimed towards librarians who work with web applications. The list currently has about 4,700 subscribers worldwide.

<http://lists.webjunction.org/web4lib/>

## References

- Ahn, L. v., Maurer, B., McMillen, C., Abraham, D. and Blum, M. (2008), "reCAPTCHA: Human-Based Character Recognition via Web Security Measures", *SCIENCE*, Vol. 321, No. 12, pp. 1465-1468.
- Akismet (n.d.), "Stop comment and trackback spam", available at: <http://www.akismet.com/> (accessed 10 September 2010).
- Bar-Ilan, J. (2007), "The Use of Weblogs (blogs) by Librarians and Libraries to Disseminate Information", *Information Research*, Vol. 12, No. 4, available at: <http://informationr.net/ir/12-4/paper323.html> (accessed 10 September 2010).
- Blogger (n.d.), "Create your free blog", available at: <http://www.blogger.com> (accessed 10 September 2010).
- Blogger Help (n.d.). "Preventing Unwanted Comments and Comment Spam", available at: <http://www.google.com/support/blogger/bin/answer.py?answer=42064> (accessed 10 September 2010).
- Clyde, L. A. (2004), "Library Weblogs", *Library Management*, Vol. 25, No. 4/5, pp. 183-189.
- Crosby, C. (2010), *Effective Blogging for Libraries*, Neal-Schuman Publishers, Inc., New York, NY.
- Cutts, M. and Shellen, J. (2005), "Preventing comment spam", *The Official Google Blog*, available at: <http://googleblog.blogspot.com/2005/01/preventing-comment-spam.html> (accessed 10 September 2010).
- Çelik, T. and Marks, K. (2009), "rel="nofollow"", available at: <http://microformats.org/wiki/rel-nofollow> (accessed 10 September 2010).
- Facebook (2010), "Statistics", available at: <http://www.facebook.com/press/info.php?statistics> (accessed 10 September 2010).
- Han, S., Moon, S., Ahn, Y. y. and Jeong, H. (2006), "Collaborative blog spam filtering using adaptive percolation search", in *WWW 2006 Workshop on Weblogging Ecosystem: Aggregation, Analysis and Dynamics*, Edinburgh, UK,
- Heymann, P., Koutrika, G. and Garcia-Molina, H. (2007), "Fighting Spam on Social Websites: A Survey of Approaches and Future Challenges", *IEEE Internet Computing*, Vol. 11, No. 6, pp. 36-45.
- Jones, S. and Fox, S. (2009), "Generational differences in online activities", The Pew Research Center available at: <http://www.pewinternet.org/Infographics/Generational-differences-in-online-activities.aspx> (accessed 10 September 2010).
- Kraut, R., Maher, M. L., Olson, J., Malone, T. W., Pirolli, P. and Thomas, J. C. (2010), Scientific Foundations: A Case for Technology Mediated Social Participation Theory, *Technology*

- Mediated Social Participation Workshop*, available at [http://www.tmsp.umd.edu/TMSPreports\\_files/1.IEEE-Computer-TMSP-Theory-Kraut-100812.pdf](http://www.tmsp.umd.edu/TMSPreports_files/1.IEEE-Computer-TMSP-Theory-Kraut-100812.pdf) (accessed 11 September 2010).
- Maher, M. L. (2010), Motivation and Collective Intelligence: Design Lessons, *Technology Mediated Social Participation Workshop*, available at [http://www.tmsp.umd.edu/position%20papers/Maher\\_TMSP-workshop.pdf](http://www.tmsp.umd.edu/position%20papers/Maher_TMSP-workshop.pdf) (accessed 11 September 2010).
- May, M. (2005), "Inaccessibility of CAPTCHA - Alternatives to Visual Turing Tests on the Web", *W3C Working Group*, available at: <http://www.w3.org/TR/turingtest/> (accessed 10 September 2010).
- Mishne, G., Carmel, D. and Lempel, R. (2005), "*Blocking blog spam with language model disagreement*", in proceedings of the First International Workshop on Adversarial Information Retrieval on the Web, Chiba, Japan, available at: <http://airweb.cse.lehigh.edu/2005/mishne.pdf> (accessed 10 September 2010).
- Newcomb, K. (2009), "Google Changes Course on Nofollow", available at: <http://searchenginewatch.com/3633972> (accessed 10 September 2010).
- OpenID Foundation Website (n.d.), available at: <http://openid.net> (accessed 10 September 2010).
- Prasad, S. (2008), "Google's CAPTCHA busted in recent spammer tactics", *websense:essential information protection*, available at: <http://securitylabs.websense.com/content/Blogs/2919.aspx> (accessed 10 September 2010).
- Ringalda, P. (2005), "Real comment throttle", *Phil Ringalda: A Digital Magpie*, available at: <http://weblog.philringalda.com/mt-plugins/real-comment-throttle/> (accessed 10 September 2010).
- Rutherford, L. L. (2008), "Implementing Social Software in Public Libraries: An Exploration of the Issues Confronting Public Library Adopters of Social Software", *Library Hi Tech*, Vol. 26, No. 2, pp. 184-200.
- Six Apart (n.d.), "Six Apart guide to comment spam", available at: [http://www.sixapart.com/pronet/comment\\_spam/](http://www.sixapart.com/pronet/comment_spam/) (accessed 10 September 2010).
- Stephens, M. (2006), "Blogs", *Library Technology Reports*, Vol. 42, No. 4, pp. 184-200.
- Stuart, D. (2009), "Social Media Metrics", *ONLINE*, Vol. 33, No. 6, available at: <http://www.onlinemag.net/nov09/Stuart.shtml> (accessed 10 September 2010).
- Sussman, M. (2009), "The What and Why of Blogging - The State of the Blogosphere 2009", available at: <http://technorati.com/blogging/article/day-2-the-what-and-why2/> (accessed 10 September 2010).



The Blogging Libraries Wiki (2010), available at: <http://www.blogwithoutalibrary.net/links/> (accessed 10 September 2010).

Thomason, A. (2007), "Blog Spam: A Review", in proceedings of the *Fourth Conference on Email and Anti-Spam (CEAS)*, Mountain View, California, available at: <http://www.ceas.cc/2007/papers/paper-85.pdf> (accessed 10 September 2010).

Timmer, J. (2008), "Computer scientists find audio CAPTCHAs easy to crack", available at: <http://arstechnica.com/security/news/2008/12/computer-scientists-find-audio-captchas-easy-to-crack.ars> (accessed 10 September 2010).

WordPress (n.d.), "Blog Tool and Publishing Platform", available at: <http://wordpress.org/> (accessed 10 September 2010).

WordPress Codex (n.d.), "Combating Comment Spam", available at: [http://codex.wordpress.org/Combating\\_Comment\\_Spam](http://codex.wordpress.org/Combating_Comment_Spam) (accessed 10 September 2010).

Yan, J. and Ahmad, A. S. E. (2008) *Usability of CAPTCHAs or usability issues in CAPTCHA design*, in proceedings of the 4th symposium on Usable privacy and security, Pittsburgh, PA, available at: <http://cups.cs.cmu.edu/soups/2008/proceedings/p44Yan.pdf> (accessed 10 September 2010).