

LiveJournal: Content, Genre, and Personal Security

INF 397C: Introduction to Research in Information Studies
December 5, 2006

Allison King
Jesse Martinez

I. Abstract

Weblogs ("blogs") are a relatively new form of internet communication used for online publishing. Online security is a problem that has recently received considerable attention. Personal journal blogs, such as those found on LiveJournal.com, allow the user various levels of privacy control. In this study we will to conduct an unobtrusive content analysis on 200 random LiveJournal user's public blogs in order to analyze and compare the frequency of use of specific LiveJournal privacy controls to the types of personal information that the user makes public. By searching through each user's 50 most recent public posts we will count the number of private posts made within the selection while also identifying information that could lead to the unambiguous identification of the user. Furthermore, we will be looking to discover how many and what types of blog genres are used by the typical LiveJournal user. Our study will further previous research by bringing together the areas of blog genre and blog security through a comparison of the personal information revealed publicly to that of blog genres used. By doing so, we hope to be able to make recommendations to companies who provide blogging software in order to help them increase the personal security of their users.

Table of Contents

I. Abstract	p. 1
II. Introduction	p. 3
1. What are Blogs	
2. Special Features of LiveJournal	
III. Problem Statement	p. 5
IV. Review of Literature	p. 6
1. Content and Genre	
2. Blogs, Privacy and Security	
V. Objectives & Research Questions	p. 9
VI. Research Design	p. 9
1. Methodology	
2. Sampling	
3. Body of Research	
4. Timeline	
5. Limitations	
6. Resources needed	
VII. Anticipated Results	p. 14
VIII. Conclusion	p. 15
IX. Bibliography	p. 16
X. Appendices	p. 17
1. Appendix A	
2. Appendix B	

II. Introduction

1. What are Blogs?

In the late 1990s, a new form of website, the weblog began to emerge. According to the Webster's New Millennium™ Dictionary of English (n.d.), a weblog is "a personal chronological log of thoughts published on a Web page." The term weblog was created in 1997 by Jorn Barger and was soon shortened to just "blog" (Herring, Scheidt, Bonus, & Wright, 2004). Blog entries, or posts, usually appear in reverse chronological order, with the newest entry at the top of the web page (Herring et al., 2004). A blog post is generally textual in nature and can contain images and or hyperlinks to other websites. In more recent years, bloggers (people who write blogs) have started to include other multimedia files, such as movies, in their posts (Nardi, Schiano, & Gumbrecht, 2004).

As soon as blogging started to become popular, people began to come up with websites that allowed people to create and maintain their own blogs without having to have any knowledge of how to make a blog or even a website. Some of these sites include LiveJournal (www.livejournal.com), Open Diary (www.opendiary.com), Pitas (www.pitas.com), and Blogger (www.blogger.com), which was bought by Google in 2003 (Blog – Wikipedia, n.d.). Each of these sites have a different features and options that their user can customize. More recently social networking sites, such as MySpace (www.myspace.com) and Facebook (www.facebook.com), have also added a way for their users to create blogs within their sites. Thanks to these sites, the popularity of blogging has grown even more. According to Riley (2006), as of July 19, 2006, there are currently over 70 million blogs with 63 million of them hosted on blog sites. However, it should be noted that not all of these blogs are active; there are many blogs on the internet that were started and then abandoned by their creator (Riley, 2006).

2. Special features of LiveJournal

LiveJournal (www.livejournal.com) is an incredibly popular "online journal service" that was started by Brad Fitzpatrick in 1999 as a way to keep in touch with his friends by giving them access to the same database program that he was using to update his own journal (LiveJournal FAQ, 2006). According to the status page there are now over 11 million accounts (both individual user blogs and community blogs); over 1 million of these accounts are "active in some way" and, as of November 29, 2006, almost 650 thousand have been updated in the past 7 days.

There are a number of features of LiveJournal that seem to set it apart from other blog hosting sites. The features that are important for this research project are the ability for a user to "friend" other accounts, the specialized <lj-cut> tag for posts, and the ability for a user to set security levels on individual posts.

Once a user creates an account, she or he can then "friend" other bloggers who also have a LiveJournal account. The primary function of the friends feature is to help create an online community on the site. Each user has a unique "Friends Page" where he or she can view the blogs of those whom the user has added as Friends. A user may then create Custom filter groups, in which she or he may place as few or as many of his or her friends as she or he desires. (LiveJournal FAQ, 2006). However, there are other advantages to "friending" other users, as talked about in the discussion of LiveJournal security below. There has been some controversy over the strength of the term "Friend" as well as online drama being caused if a user friends another user and that user does not friend the first user back (Bandlow & Jensen, 2002).

When creating a LiveJournal entry, the user has the option of hiding all or part of an entry behind a link. This is done while the user is writing an entry by adding in the code <lj-cut text="[Read More]">[content]</lj-cut> into the body of the post. This will then appear in the post as ([Read More]) and when the reader clicks on the link, the [content] will be revealed. Although the lj-cut feature does not technically provide any security benefits, some users claim that they feel more secure placing personal information "behind a cut tag" (Bandlow & Jensen, 2002).

The one feature of LiveJournal that seems to set it apart from other online blogging sites is that it gives the user the option of setting security levels for his or her individual posts at the time of entry creation. There are four levels of posting: Public, Friends-only, Custom, and Private. Anyone who is browsing the internet and comes across a user's LiveJournal may read posts marked as Public. A Friends-only post is only visible to other users whom the blogger has added as Friends and only when these other users are logged into their accounts. The user may also choose to Custom "lock" the post so that only friends who have been added to a specific filter or filters may read the post. Lastly, by designating an entry as Private, only the user may read the post he or she created and only when logged into the system. (LiveJournal FAQ, 2006). If a post is locked in a way that does not allow someone to read it, then the post just will not show up for them, either on a Friends Page or when viewing the user's blog (Bandlow & Jensen,

2002). However, the fact that a post was made on a certain day will show up on the entry count of a user's archive or calendar page.

III. Problem statement

The more we develop into an information based society, the more ordinary people are starting to use the public arena of the internet to disclose personal thoughts and information. One of the developments in this society has been that of the web log (more commonly called "blog"). A blog allows a user to instantly publish online their thoughts, ideas, daily life and anything else that they wish for anyone to read. LiveJournal is one of the more popular sites where users can create and maintain their own blogs. An interesting feature of LiveJournal's services is that it can allow the user to set a level of privacy for their blog entries: public, friends, custom friends, or private. Another growing trend is the concern towards users' online security, especially in terms of stalking and identity theft. Much research has been done on the genre of blogs and their creators' knowledge and feelings about online security. Viegas (2005) mentions that some LiveJournal users feel safer placing personal information in posts that are set to a higher level of security. We propose to do a content analysis of blog posts made by LiveJournal users who post both publicly and with set levels of security to find out what sorts of personal information they still decide to make publicly available as well as how they convey this information. We will be looking at the nature of the posts, such as personal diary, social invites, and filters, as well as what sorts of private information these public posts contain, such as location of user, schedule, contact information, or other information that could lead to unambiguous identification of the user. In doing so, we hope to find out if in practice LiveJournal users do in fact hide this information as the ones in Viegas's survey claimed to do. Also, we hope to be able to make a link between the type of genre a blogger uses and the type of information they reveal. Finally, we would like to be able to help improve blogging websites by making suggestions on how they can increase the security of their users, both in terms of improvements and additions to current features as well as increasing their users' awareness of the importance of privacy control.

IV. Review of Literature

With the advent of the weblog came the research into blogs and bloggers. Research has been done by bloggers themselves, such as Rebecca Blood (www.rebeccablood.net) and even the authors of this paper, and also by non-bloggers who are trying to understand the blogging phenomenon. There have been myriad topics that these researchers have looked into, such as why people blog, the social networking that happens when blogs and bloggers are connected over the internet, how a blog is different than a paper journal, classifying the content and genres of blogs, and blogs and security. For the purpose of this study, we will examine previous research into the latter two areas.

1. Content and Genre

As soon as blogging started to become popular, researchers have tried to find ways to classify them not only in terms of how blogs differ from other online websites, but to try to figure out what are the types and characteristics of the blogs themselves. Several studies have analyzed both the different types of blogs that exist on the internet and also what sort of content users place into these different types of blogs.

Herring et al. (2004) placed the blogs in their study into three main types: filters, k-logs, and personal journals. In a filter blog, the blogger finds interesting websites and then links to them in his or her posts, often with personal comments about the websites' contents. Filter blogs were once thought to be the most popular form of blogs, but according to Herring et al. (2004), only 12.6% of the blogs they analyzed fell into this category. K-log is short for knowledge-log and is used by organizations, institutions, or groups to share information, either internally or externally. K-logs are fairly uncommon, as Herring et al. (2004) only came across 6 out of the 199 blogs that they examined. Herring et al. (2004) reported that the most common type of blog is the personal journal, in which the user posts personal content, much like a paper diary. Their research showed 70.4% of the blogs they analyzed to be of this type. They state that, "This result is all the more notable in that we excluded journal sites such as LiveJournal.com and Diaryland.com from our data collection, so that their popularity would not overshadow the other blogs in the sample (p. 6)." To further concur with this, 83% of the 486 bloggers who responded to Viegas' survey described their blogs as "personal musings" (2005). Herring et al (2004) also

identified some blogs as mixed, containing posts of two or more of the three primary types, or "other", in which the posts did not fit well into any of the other groups.

Nardi et al. (2004) took a deeper look into the content of the personal journal as both a personal space and social activity. They found that there were several different types of posts that personal journal bloggers made, three which involved strong considerations towards readers and two which were more personal, but where the blogger was still aware of an audience. The first of the three where known or unknown readers were strongly recognized were blogs that updated others to activities or whereabouts. These posts tended to be written either in a more narrative style, "Yesterday I went to the store," or as announcements, "Hey everyone, I'm going to Club X tomorrow night, you should come too." This type is reminiscent of the reason why Fitzpatrick first started LiveJournal, as stated above. The purpose of the second type is to express opinions or influence others, either through the use of filter-type posts with comments about it or through personal motivation. The third type is to seek other's opinions and feedback about anything from poems the blogger wrote to what to name the new kitten. In terms of the more personal posts, Nardi et al. (2004) found that some people used their blogs as a space to "Think by Writing," where the blog was used as a way to improve writing and communication skills. The final type they found was when people used their blog as a way to release emotion that they may have had trouble expressing to others in person or as a way to just "let off steam."

2. Blogs, Privacy and Security

Blogging may be a relatively new form of Internet communication, but issues of privacy and information control are not new. By virtue of being posted online, blogs are both easy to make and yet easy to lose control over. Due to the widespread nature of the Internet, what is posted online is quickly disseminated, resulting in "personal information being broadcasted over the Internet (Viegas, 2005, p. 3)." Although much attention has been given to the realm of identity theft and unwanted surveillance, public opinions on privacy are often at odds with demonstrated actions, including online activities. According to Spiekermann, Grossklags and Berendt (2001), most users in their study who claimed to be privacy-conscious demonstrated a tendency to reveal personal information at a rate comparable to others who had much less concern for privacy. The notion of privacy is a subjective one, and can change according to context.

As mentioned above, of the three primary types of blogs identified by Herring, et al. (2004), the personal journal was the most common type. The study by Nardi et al. (2004) states that in this more personalized diary-like journal, people seem less aware or even interested in the security of their journal posts. They found the diary-like nature of the personal journal motivated the bloggers to create avenues for communication, seek feedback, and vent emotional tension, all the while becoming more forthright in their digital compositions. When asked about the prospect of having, in theory, 900 million people read their blog, most of the interviewees responded ambivalently. In fact, many people in the sample rarely made any of their blog posts private. Viegas (2005) reinforces this concept when she states how 76% of her respondents said they do not limit who gets to read what they post. However, she goes on to mention that some of the respondent who are LiveJournal users claim that they do, in fact, filter who can read their posts by utilizing the site's security features.

Other research has shown that a remarkable amount of private information is available on personal blogs. Herring et al. (2004) showed that 92% of their sample revealed some sort of self-identifying name. Huffaker (2006) in his content analysis of teenage blogs, including those on LiveJournal, showed 72% of teenagers revealed their first name, 67% revealed their age, and 61% revealed their contact information. Viegas (2005) showed very similar results when her research found 81% of her respondents had used some sort of accurate self-identification, including 55% posting their first name on their blogs. Similarly, Nardi et al. (2004) claim that many of the blogs in their sample also provided accurate identification. However, Bandlow & Jensen (2002) note that many LiveJournal users say that they use the site's filters and locks to give out personal information to their "Friends" only.

Previous research has also raised questions about the motivations to not conceal one's identity. Huffaker (2006), in his unobtrusive content analysis of blogs, questions whether his teenage sample, ages 13 to 17, is mature enough to understand the consequences of placing private information on their public journals. Nardi et al. (2004), on the other hand, argues that some of the bloggers in her sample, ages 18 to 60, take comfort in the notion that their blog is one of millions therefore there is probably little public interest in their specific one. Their results suggest that maturity is not the offending parameter in the lack of proper privacy control, but rather a willingness to share and gain welcome attention is the motivation for bloggers to post personal information.

V. Objectives & Research Questions

The purpose of this exploratory study is to determine if, as in Viegas's study, LiveJournal users do in fact post less personal information publicly since they have the option to "lock" some of their posts. In order to fulfill this objective, we have a number of questions that we will use this research to try to answer.

Which of the genres that researchers have discovered – non-diary types, such as filters, k-logs, opinions or influencing posts, reviews of other of media (such as TV shows or movies), and the more personal journal types such as narrative life updates, announcements, surveys and quizzes that the users has done, and emotional posts, and creative writing – do LiveJournal users writing in the context of the personal journal use? Are there any other genres that have not yet been discovered?

What types of personal information - part or all of the user's real name, age or birthday, contact information (ie: e-mail, chat id, phone number), location, and current school or place of employment – does the typical LiveJournal user reveal about him or herself? How much of this personal information is publicly disclosed through posting, as opposed to on the profile page? Can the information given throughout blog entries be enough to unambiguously identify the real-life identity of a LiveJournal user?

Which genres of blog posts tend to include more personal information than others? Is there a noticeable reason for this?

Does the lj-cut feature give LiveJournal users a false sense of security and encourage more personal information to be posted publicly than when lj-cut is not used in public entries?

VI. Research Design

1. Methodology

For this study we wish to conduct an unobtrusive content analysis on approximately 200 different English-speaking LiveJournal users. For this content analysis we wish to observe and codify four aspects of each user: 1) within the 50 most recent public posts, how many, if any, posts are deemed private, 2) within the 50 most recent posts what types of genres are observed, 3) what kinds of personal or private information can be found within the posts that would lead to an unambiguous identification of the user, and 4) the frequency of cut tags used to hide personal

information from the front page. For sampling purposes, we will not make a distinction between paid and free user accounts.

It is important to remain unobtrusive in our content analysis as to not deter any user from posting as he or she normally does. Having the user aware of the study may affect their use of private posts, editing or deleting past public posts, or in any other way negatively alter the conditions for the user to make posts.

2. Sampling

To accomplish the near-random sampling of LiveJournal users we will utilize the “latest posts” feature on LiveJournal (<http://www.livejournal.com/stats/latest.bml>) to select from the 150 most recent posts at that moment, the first 50 posts written in English. We will repeat this step 10 times at random intervals throughout a one-week period. We want to avoid taking a sample at regular intervals as to prevent the exclusion of posters who may not post during particular parts of the day, due to factors such as being at work or sleeping.

If the 10 samples taken from the “latest posts” feature on LiveJournal does not sum up to 200 users who fit the criteria discussed below, then another 10 iterations of random samplings will be conducted as needed and added to the original set of users. Ideally, there should be more than 200 valid and unique users within the larger sample at the end of the selection process. From this point, the larger set of users will be randomly arranged into an array, where N is the length of the array. From here, every $N/200^{\text{th}}$ user (rounding down when necessary) will be selected in a systematic sampling.

3. Body of research

The body of research will take place in five stages: sample collection, sample elimination, data collection, data coding, and statistical analysis.

Stage 1 - Sample collection

To offset the tedious nature of sampling and of human error, we will utilize computer scripts to automate the sample collection of users. The first script will to fetch the most current listing of the “latest posts” page at ten random intervals throughout one weeklong period. The script will save the page layout in html format onto a dedicated hard drive for later retrieval.

Stage 2 - Sample elimination

At the end of the one-week period, the research assistants will parse through the saved html pages and eliminate users based on these conditions:

- The user does not post in English as a primary language
- The user has not made 50 public posts within a one-year period. (This one-year period is designated as, for example, from November 2006 to October 2005.)
- The user already exists as a valid sample member

Each valid user will be entered into a database with a unique name. Generally, the user's handle, or online name, will be unique enough to be sufficient. If there does not exist at least 200 valid users, then the process will repeat once again at stage 1. We do not suspect that this should happen at all, less more than once.

Stage 3 - Data collection

From this point on a second script will take the database listing and randomly select 200 users by the process listed in stage 2. Next, a third script will do the bulk of retrieval by fetching each user's 50 most recent posts (including all parts placed behind the lj-cut tag), determining if there exist any posts that were made private, and determining valid information from the user's profile page. Since private posts are by their nature not readable to the general public, we can only note their presence in this study. This script will use the calendar feature of each user's profile to indirectly count the number of posts made private and save this number in a separate file. This script's details are explained in Appendix A. After this third script runs its course it will have saved each individual post in html format for each user in chronological order beginning from most recent to least recent. All information posted behind a LiveJournal cut tag will be saved in full and designated as being "hidden" behind the cut tag as well. These posts will also be separated by user's unique identification. To prevent overwhelming LiveJournal's web servers there will be a slight pause between each post being fetched. Overall there should be 10,000 individual posts saved in total.

Stage 4 - Data coding

This step in the research design will utilize four research assistants to manually analyze each post and code it based upon a set list of genres. To gauge interrater reliability between the research assistants twenty posts from five users, who are not part of the final sample set, will be taken from the larger sample. These four research assistants will each independently code the five users and their 100 posts according to the coding specification. Once it is known that there is a statistically high level of agreement then the 200 users will be evenly distributed between all the research assistants so that each user's complete set of posts are kept together. Otherwise, if there is not a statistically high level of agreement or if there is a consensus that there needs to be changes to the coding scheme based on the test posts, changes will be made to our coding scheme and this stage will be repeated. Each research assistant will enter in the coded information of each post into a database for further analysis. Appendix B contains the proposed coding categories as well as a sample coding chart to be used in this study.

Stage 5 - Statistical analysis

The results from the database will be statistically analyzed using appropriate statistical software such as SPSS. We will use descriptive statistics to analyze the frequency and averages of the different types of genres and types of personal information given both individually and when mixed together. From these statistics we will attempt to find correlations between the percentage of users that use private posts versus those who do not, and if this relates to the levels of personal information being made public. We will also search for patterns of users using cut tags and the amount of private posts made. Should the situation warrant it, further analysis will include inferential statistics, such as a Chi-squared test, to infer additional conclusions based on our data.

4. Expected Timeline

The timeline for entire body of research is as follows:

Stage 1 – Sample collection:	1 to 2 weeks
Stage 2 – Sample elimination:	1 week
Stage 3 – Data collection:	1 to 2 days
Stage 4 – Data coding:	4 to 6 weeks
Stage 5 – Statistical analysis:	2 weeks
Overall estimated time:	12 weeks

5. Limitations

The largest limitation is the lack of a truly random sample. We wish to take as close to a random sample of LiveJournal users as possible. Unfortunately, there is no feasible way to randomly select users from LiveJournal's vast database of 11 million users. The LiveJournal website does not offer a function to randomly select an active account without first being an administrator. We have tried to overcome this by randomly sampling the “latest posts” feature. By sampling this “latest posts” site multiple times at random intervals we hope to gather a large set of people who are actively posting onto their blogs. While this is not a true random sample we are sampling a live base of people at random times.

We have made our coding scheme distinct enough to accommodate a wide range of genres, but there may be some ambiguity within the scheme that may lead to some difficulty. While we will try to encapsulate all the major types of genres in the user’s posts there is bound to be some exceptions. We may also find it hard to make a distinction between two genres intermixed within a post. The interarater reliability test is designed to overcome and test our coding scheme.

By using individual LiveJournal user names, we hope to maintain a level of anonymity. However, due to our study trying to discover if unambiguous identification of users can be made, some of the users in our sample may become identifiable to our researchers. Should this happen, none of this information will be revealed in our results.

6. Resources Needed

a. Institutional resources

Computers: Four Apple Macintosh computers running OS X with internet access.

Disk space: On average the text of an average post is about 6kb. With images, this could be up to 1Mb. 1Mb x 10k posts is approximately 10Gb. Factoring in memory block sizes and other overhead, this will require about 20Gb of storage space.

Software: SPSS (statistical software), OpenSQL database management software, PERL scripting language

b. Personnel

Four research assistants will be required to complete this study. Each assistant will take up to one-quarter of the task loads, which include eliminating users from our large sample, providing statistically high level of interrater reliability, and encoding 2500 different posts. Each assistant will need to dedicate from 10 to 20 hours a week for the duration of the user elimination and data coding stages of the study. The proposed timeline estimates this to be about 5 weeks.

VII. Anticipated Results

We hope to find what sorts of blog genres people who post in an online personal journal use. We will be looking at the number of posts that fall into the individual categories as well as on average how many genres are used by an individual LiveJournal user.

We want to see if LiveJournal users use levels of posting security, as will be discovered by finding out how many hidden posts were made within the 50 most recent posts that were analyzed. While we will not be able to read their posts and thus not be able to find out what they posted or why the post was locked, we will be able to get a sense of how often the average LiveJournal user uses the security features of the site.

We also will look at the personal information revealed to try to determine if an unambiguous identification of a user is generally possible or not possible. We will also be finding out what sort of personal informational items tend to be revealed and through what types of posts. We will also look at the averages of these types to find out which are divulged with greater frequency.

We also hope to discover if LiveJournal users are utilizing the lj-cut feature in their postings to try to block people from finding out personal information about them.

VIII. Conclusion

The creation and the naming of blogs have only existed since the late 1990s. Not surprising, previous research on the nature of blogs as well as the privacy content of blogs have only been recently published. Our study has focused on two major results of this prior research. First, the personal journal type is the most prevalent of all types of blogs, and second, bloggers who have personal journal type blogs tend to reveal personal information in high percentages. These statistics raise grave concerns for the role of privacy in this fledgling publishing field. Our focus of research will focus on the content analysis for publicly posted information as well as analysis for genre. Our study will take 200 randomly selected LiveJournal users and analyze the 50 most recent posts of each user. Through the content analysis of the posts, we hope to gain an understanding of what types of personal information bloggers make available to anyone with an internet connection and web browser when they have the option of posting to only a selected group of people. Through the genre analysis we hope to categorize each user's posts and then compare this information to the levels of private information publicly posted and the number of private posts, from which we deduce through various LiveJournal mechanisms. By doing so, we wish to statistically analyze any connections and correlations between privacy and genre. We would also like to analyze the role of LiveJournal specific privacy features, such as lj-cuts, in the "hiding" of sensitive or personal information.

Bibliography

- Bandlow, A., & Jensen, P. (2002). LiveJournal: Personal thoughts in a public forum. Unpublished midterm paper for "The Design of Online Communities" class (CS6470). Georgia Institute of Technology.
- Blog. (n.d.). Blog – Wikipedia, the free encyclopedia. *Wikipedia*. Retrieved November 29, 2006, from <http://en.wikipedia.org/wiki/Blog>
- blog. (n.d.). Webster's New Millennium™ Dictionary of English, Preview Edition (v 0.9.6). Retrieved November 29, 2006, from Dictionary.com website: <http://dictionary.reference.com/search?q=blog>
- Herring, S. C., Scheidt, L. A., Bonus, S., & Wright, E. (2004). Bridging the gap: A genre analysis of weblogs. *Proceedings of the 37th Hawai'i International Conference on System Sciences (HICSS-37)*. Los Alamitos: IEEE Computer Society Press. Retrieved January 5, 2005 from <http://www.blogninja.com/DDGDD04.doc>
- Herring, S. C., Scheidt, L. A., Bonus, S., & Wright, E. (2005). Weblogs as a bridging genre. *Information, Technology & People*, 18(2), 142-171.
- Huffaker, D., (2006). Teen blogs Exposed: The Private Lives of Teens Made Public. Retrieved October 14, 2006 from http://www.soc.northwestern.edu/gradstudents/huffaker/papers/Huffaker-2006-AAAS-Teen_Blogs.pdf
- LiveJournal FAQ. (November, 2006). Retrieved November 29, 2006 from <http://www.livejournal.com/support/faq.bml>
- Nardi, B. A., Schiano, D. J., and Gumbrecht, M. (2004). Blogging as Social Activity, or, Would You Let 900 Million People Read Your Diary? *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, 222-231.
- Riley, D. (July 19, 2006). Blog count for July: 70 million blogs. *The Blog Herald*. Retrieved November 29, 2006 from <http://www.blogherald.com/2005/07/19/blog-count-for-july-70-million-blogs/>
- Spiekermann, S., Grossklags, J., and Berendt, B. (2001). E-Privacy in 2nd Generation E-Commerce: Privacy Preferences versus actual Behavior. *Proceedings of the ACM Conference on Electronic Commerce (EC-01)*, 38-47.
- Viegas, F.B., (2005). Bloggers' Expectations of Privacy and Accountability: An Initial Survey. *Journal of Computer-Mediated Communication*, 10(3), article 12, <http://jcmc.indiana.edu/vol10/issue3/viegas.html>.

Appendix A

The planned PERL script will work as follows:

For LiveJournal user XYZ

- 1) Fetch <http://XYZ.livejournal.com/calendar> and then parse the html code for strings that match ` View Subjects` where yyyy is the year and mm is the month in numerical format beginning with the current year and month.

- 2) Next, parse through the same page and collect all links that match the string ` % ` where the % symbol represents the number of posts that day. Make sure to only collect the postings that have the exact same yyyy and mm values as in step 1. For instance, <http://XYZ.livejournal.com/2004/09> would be the calendar link while for step 1 while <http://XYZ.livejournal.com/2004/09/23> would be the link for that day's postings.

- 2) Carefully collect which days of the selected month showed a post of some sort. These links will show a combined posting history including public and private postings.

- 3) Follow each link that was found in the step 1 which then shows a listing of all the subjects for each public post made that month.

- 4) Compare the listings that are displayed to the collected values from step 2. Any discrepancies will be due to private posts not being displayed.

- 5) Fetch all available links and save the resulting html files, including any lj-cuts, in the appropriate folder hierarchy.

6. Go to step 1, decrease the month and year if necessary until 50 posts are found.

Appendix B

Proposed Coding Scheme:

Genre:

To be used only for Posts analysis.
Multiple codes may be used for the same post.

Ann	Announcement
Emo	Emotional or very personal in nature
Flt	Filter
Img	Embedded images
Klg	Knowledge-log
Mov	Embedded movie
Opn	Expressing opinions or Influencing others
Pln	Making plans
Pol	Poll to have readers answer
Quz	Quiz results (may also be called "a meme")
Rvw	Movie, TV show, or book review
Srv	Survey filled out by the user (often referred to as "a meme" by users)
Tdo	To do lists
Tnk	Thinking by writing
Upd	Narrative life update
Wrt	Creative Writing (poem, short story)

Personal Information:

To be used in both the Profile Page and Posts analysis.

Since there can be multiple facets, coding will be put together as follows:

Subject Title: subsection/subsection,
so that if, for example a user lists their full name, it will read RN:fn/lm

In the Profile Page section, only the lowercase coding will be used.

In the Posts section, if one of sections is not used, it will simply not be coded (this is to prevent an over abundance of "XX:no" codes.

RN	Real Name
fn	first name
fi	first initial
ln	last name
li	last initial
ot	other
no	none

- WB Website (for Profile Page section only)
- pr personal
 - ot other
 - no none
- LN Location
- co country
 - st state
 - ct city
 - zp zip code
 - fc fictional (ie: another planet, fantasy world)
 - ot other
 - no none
- BD Birthday
- yr year
 - mo month
 - dt date
 - ag age
 - ot other (ie: "in five days", "next month")
 - no none
- CI Contact Information
- em e-mail
 - ch chat (ie: AIM, Yahoo! ID)
 - ph phone number
 - sa street address
 - ot other
 - no none
- SW School and/or Work
- sc current school
 - wk current work place
 - ot other
 - no none

Sample Coding Page:

User Name:		
Profile Page:		
Real Name		
Website		
Location		
Birthday		
Contact Info		
School/ Work		
Posts:		
Post #	Genre	Personal Information
1		
2		
3		
Etc...		