

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

E-JASL 1999-2009 (volumes 1-10)

E-JASL: The Electronic Journal of Academic  
and Special Librarianship

---

Summer 2008

## Library 2.0 and the Problem of Hate Speech

Margaret Brown-Sica  
*Auraria Library*

Jeffrey Beall  
*Auraria Library*

Follow this and additional works at: <https://digitalcommons.unl.edu/ejasljournal>



Part of the [Communication Technology and New Media Commons](#), [Scholarly Communication Commons](#), [Scholarly Publishing Commons](#), and the [Social Media Commons](#)

---

Brown-Sica, Margaret and Beall, Jeffrey, "Library 2.0 and the Problem of Hate Speech" (2008). *E-JASL 1999-2009 (volumes 1-10)*. 99.  
<https://digitalcommons.unl.edu/ejasljournal/99>

This Article is brought to you for free and open access by the E-JASL: The Electronic Journal of Academic and Special Librarianship at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in E-JASL 1999-2009 (volumes 1-10) by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

# Electronic Journal of Academic and Special Librarianship



v. 9 no. 2 (Summer 2008)

## Library 2.0 and the Problem of Hate Speech

Margaret Brown-Sica  
Auraria Library, Denver, Colorado, USA

Jeffrey Beall  
Auraria Library, Denver, Colorado, USA

Library 2.0 applications benefit library users by providing rich, peer-generated content that adds value to online library databases and systems. However, not all of this shared content is beneficial, for it's possible for library users to abuse library 2.0 applications by uploading words, pictures, or other content that constitutes hate speech. Internet lawyer Christopher Wolf warns of, " ... the sudden and rapidly increasing deployment of Web 2.0 technologies to spread messages, sounds and images of hate across the Internet and around the world" [1]. As academic libraries make available Web 2.0 systems that allow user-generated content, they must incorporate into these systems quick, effective, and consistent means of dealing with user-generated hate speech.

Hate speech is " ... usually thought to include communications of animosity or disparagement of an individual or a group on account of a group characteristic such as race, color, national origin, sex, disability, religion, or sexual orientation" [2]. To promote research, learning, and the generation of new ideas, universities and colleges have opposed limiting speech on their campuses to support of the long-standing traditions of academic freedom, opposition to censorship, and freedom of speech. However, universities and colleges today view hate speech as outside the realm of protected speech. Hate speech violates the terms of most institutions' code of conduct and merits decisive action. Also, many college libraries are big players in their universities' overall mission to value and promote diversity. Perhaps nothing can poison this mission more than a library web site filled with racist, homophobic, or other defamatory speech.

## **Identifying Hate Speech in Library 2.0**

Library 2.0 applications allow several different types of user-generated content. Some, with features that include adding a numerical or “number of stars” rating to an online resource, are immune to the problem of hate speech. “Systems that allow user-generated content, however, provide ample opportunities for users to upload their hate speech, which can range from photos to text to URLs” [3]. Even systems that only allow social tagging can be abused by using unwarranted, hateful terms in the tags.

Libraries need to define hate speech in the content guidelines that are part of their Library 2.0 applications and use these definitions to identify user-generated hate speech. Context and user intent play a large role in identifying hate speech, and automated systems, like word matching, may not effectively carry out the task. Word matching fails to distinguish a word that’s a slur in one context but acceptable in another. Also, in some academic contexts, it may be perfectly acceptable to discuss ethnic slurs as words, and the slurs may appear in literature, including books and online resources in a library’s collection. Literary criticism may legitimately include such terms.

Malicious users of Web 2.0 applications are becoming adept at creating hate speech that isn’t overtly hateful or defamatory. They may imply their statements without actually stating them. They may use synonymous terms like “those south-of-the-border people” to refer to a group. Identifying hate speech on the Internet often needs to rely on the contributor’s intent than the actual word-for-word text. Libraries need to address malevolent intent in their definitions of unacceptable content.

## **Options for Dealing with Hate Speech**

### **Human Moderation**

The most complete, labor intensive solution is to have a human moderate or edit the user input to screen for contributions that violate content guidelines. With this method, added content must be approved by a human before it is posted in an application. A human being can catch the subtleties of hate-speech and remove it before it’s publicly viewable. However, humans are not always consistent and the question of what constitutes hate speech and what the library finds acceptable may be difficult to incorporate into policy and practice. Human moderation is also slow, a problem that affects the immediacy of the library 2.0 participation experience. It may cause many who did not see their comments added immediately to not want to post a comment again. Also, some libraries may find that their already over-worked staff lacks the time to perform human moderation of library 2.0 contributions.

## **Automated Moderation and Filtering**

This method uses an automated filter to screen all user-contributed content for predetermined undesirable phrases or words, and either rejects the posting or sends questionable items to a human moderator. This is more time-effective than human moderation and is accomplished fast enough to provide immediate posting of most entries and provides a reason for rejecting others. However, as outlined above, language is often hard to filter effectively because some words are acceptable in some contexts and not in others.

## **Ranking-Based or Demotion**

An interesting option outlined in the article “Fighting Spam on Social Web Sites” is “to design the system to reduce the prominence of content likely to be spam.” [\[4\]](#) For example, in a tag system, tags constituting hate-speech are probably not as numerous as tags which accurately describe a resource, and when sorted by a ranking system, the more frequent (and accurate) tags display first, thus deemphasizing the offensive tags, which display last. However, this method is more difficult to carry out when real-time ranking is important, and it is more difficult to do for textual comments.

## **Reporting Abuse**

In this method, the library depends on its user community to report abuse through “Report abuse” buttons in the application or to the library directly, such as through email. This method ensures that there is a way to have items examined and possibly removed when hate-speech is identified. It relies on the library community rather than library staff to identify the problem. This method also has the advantage of user input not having to wait to be approved before being posted. Still, the library may face the problem of judging what constitutes hate speech and what does not. The library must also accept that hate-speech may go unreported or may be posted for a period of time before someone reports it and it is removed. However, if it is not reported, the library may be able to assume that it is not sufficiently offensive to justify removal.

## **Disallowing Comments**

Another way to prevent hate-speech is to limit input to options other than comments. Typically, this includes ranking systems, pre-selected controlled vocabulary and/or tags, the submission of topic-appropriate internet links to be shared, etc., instead of allowing prose. However, this method reduces much of the rich content users provide in a more open environment.

## Requiring Logins

This method may be the best option for academic libraries. In this method, users are required to log in using their library card or university ID number in order to add content to a library 2.0 application. Requiring logins makes it easy to identify and ban problem users and also discourages users from using hate-speech in the first place. However, the downside of this method is that comments are limited to those in the library or institution's immediate community, perhaps decreasing the number of valuable contributions. Also, some users may feel discouraged from making legitimate comments because they do not want to be publicly identified, so allowing logged-in users to post anonymously may work to increase participation.

## Conclusion

The problem of hate speech in library 2.0 applications is likely to increase and will require academic libraries to establish policies and procedures to prevent it. Libraries will need content guidelines that address hate speech, and they will need systems able to identify and eliminate it when it occurs in library 2.0 applications. By taking measures to deal with hate speech, libraries will be able to ensure that user contributions enrich library databases without poisoning them.

## Notes

1. Christopher Wolf, "The Dangers Inherent in Web 2.0," Anti-Defamation League, [http://www.adl.org/main\\_internet/Dangers\\_Web20.htm](http://www.adl.org/main_internet/Dangers_Web20.htm).
2. John T. Nockleby, "Hate Speech," in Encyclopedia of the American Constitution. Ed. Leonard W. Levy and Kenneth L. Karst. Vol. 3. 2<sup>nd</sup> ed. Detroit: Macmillan Reference USA, 2000. p. 1,277-1,279.
3. Paul Heymann, Georgia Koutrika, Hector Garcia-Molina, "Fighting Spam on Social Web Sites: A Survey of Approaches and Future Challenges," IEEE Internet Computing, vol. 11, no. 6, pp. 36-45, November/December, 2007.
4. Ibid., p. 40.

## [Back to Contents](#)

[http://southernlibrarianship.icaap.org/content/v09n02/brown-sica\\_m01.html](http://southernlibrarianship.icaap.org/content/v09n02/brown-sica_m01.html).