

Du, H., Chu, S., & Lam, F. (2009). Social Bookmarking and Tagging Behavior: An Empirical Analysis on Delicious and Connotea. *Proceedings of the 2009 International Conference on Knowledge Management* [CD-ROM]. Hong Kong, Dec 3-4, 2009.

SOCIAL BOOKMARKING AND TAGGING BEHAVIOR: AN EMPIRICAL ANALYSIS ON DELICIOUS AND CONNOTEA

HELEN S. DU

*Department of Computing, The Hong Kong Polytechnic University,
Hung Hom, Kowloon, Hong Kong
E-mail: cshelen@inet.polyu.edu.hk*

SAMUEL K.W. CHU¹ and FLORENCE T.Y. LAM²

*Division of Information Technologies, Faculty of Education, The University of Hong Kong¹
Department of Statistics and Actuarial Science, The University of Hong Kong²
E-mail: samchu@hkucc.hku.hk¹ and florentia.lam@gmail.com²*

Social bookmarking services have shown themselves as common and popular Internet tools by successfully acquiring millions of users, with Delicious being one of the most popular social bookmarking services to the public. While Delicious is used mainly for general purposes, Connotea, another social bookmarking site that primarily serves academic and scientific interests, has become equally popular among researcher groups. This paper attempts to analyze and compare users' bookmarking and tagging behavior in Connotea and Delicious. The results show that there is a distinctive difference in usage behavior among these two groups of users. Delicious users create bookmarks more frequently than Connotea users, but Connotea users tend to use more distinctive tags for their bookmarks than Delicious users. Moreover, our result from the analysis indicates that the number of bookmarks created is a significant predictor of the quantity of tags used. This study is a starting point from which to explore the reasons behind the difference in social bookmarking and tagging behavior among different usage orientation groups.

1. Introduction

Social bookmarking sites have successfully acquired millions of online users in the recent years. A social bookmarking site provides the service for storing, sharing and discovering a collection of web bookmarks with the help of *user-generated taxonomies* (folksonomies). The term folksonomy came from the words *taxonomy* and *folk*, which is used to name the growing phenomenon of users *collaboratively* creating and managing metadata by *tagging* pieces of digital information with their own searchable keywords (Dye, 2006). Folksonomy is also known as *social tagging*. For simplicity, this paper uses the word *tagging* to refer to *social tagging*. In recent years, research has been done on tagging and folksonomy to understand folksonomic patterns (Al-Khalifa, & Davis, 2007) and their trends (Hotho, Jäschke, Schmitz, & Stumme 2006). Very few articles have been written on social

bookmarking. The ones that have been published mainly give definitions on social bookmarking and its related concepts (Golder & Huberman, 2006) and provide general discussions on social bookmarking tools (Hammond, Hannay, Lund and Scott, 2005; Lund, Hammond, Flack and Hannay, 2005).

Delicious, created by Joshua Schachter in 2003, is one of the most popular social bookmarking service sites. Until now, it has acquired over 5 million users and 150 million bookmarked URLs [http://en.wikipedia.org/wiki/Delicious, viewed June 8, 2009]. Diigo is another similar social bookmarking service like Delicious, which allows users to bookmark and tag web pages. It also contains the functions to highlight and paste sticky notes, so that users can create their personal notes on the content of WebPages they visit. While Delicious and Diigo are social bookmarking tools that serve general purposes, some other social bookmarking tools such as CiteULike and Connotea are being targeted at academic and scientific purposes (Gordon-Murnane, 2006). These two social bookmarking service sites operate in the similar style as of Delicious, in addition to capturing the bibliographic information from scientific articles and journals.

From a preliminary analysis using a data mining technique (association rule mining), we find a notable difference in the usage pattern of the top 100 frequently used tags¹ between Connotea and Delicious. As shown in Figure 1, Delicious has more diversified relationships of tags than Connotea.

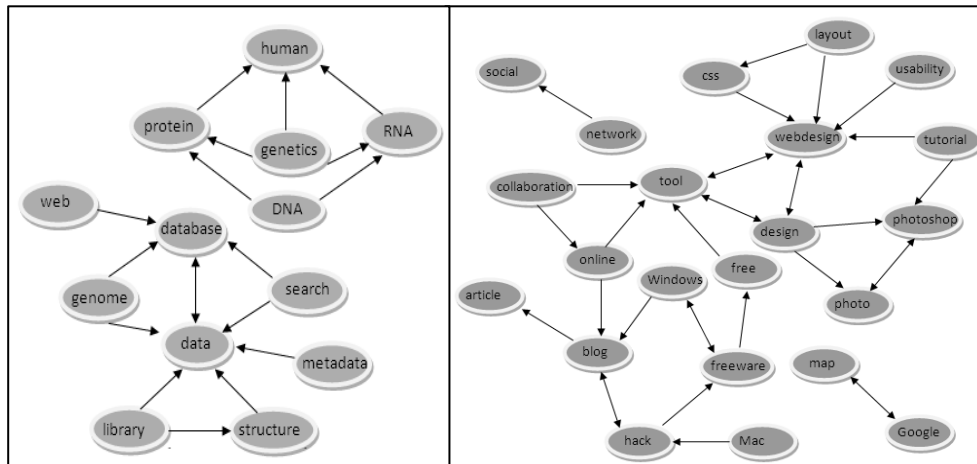


Fig 1a. Connotea

Fig 1b. Delicious

Fig. 1. Association rules found in Connotea and Delicious with .1% minimum support and 50% confidence.

Note. Each element in the graph represents a tag. The arrow connecting any tags implies their association

¹ The 100 frequently used tags are extracted from the same data set used in the analysis described in session 3.

relationship. For example, one user uses the tag “protein” may also use “human” to bookmark the same URL (“unidirectional relationship”); one user uses the tag “protein” will also use “RNA” and vice versa (“bidirectional relationship”).

Based on the above preliminary analysis, we suspect that there exists different tagging behavior among the users of Delicious and Connotea that requires further investigation. This investigation is worthwhile as Delicious and Connotea are two very popular social bookmarking sites that serve different target user groups (general-purpose users and researchers respectively). So far, no published articles on social bookmarking have yet attempted to examine and compare social bookmarking services with different orientations. Therefore, this study investigates both Delicious and Connotea, and examines their similarities and differences in users’ bookmarking and tagging behavior.

2. Literature review

2.1 *What is social bookmarking?*

When people start to rely on the use of Internet, one of the greatest challenges they face is to remember and retrieve items that they have previously found to be useful. The common approach to arrange information on the Web is through the use of personal bookmarks (Millen, Feinberg, & Kerr, 2005). The desire to share information among communities has led to the development of shared bookmarking systems. Social bookmarking is a way to locate, classify and share Internet resources through the use of shared lists of user-created Internet bookmarks. Social bookmarking tools allow users to create tags for bookmarks they saved, and organize all users’ tags so that users can search and browse the tags to find out not only their own bookmarks but also other users’ bookmarks.

2.2 *Social bookmarking services*

Delicious [<http://delicious.com>], being the most commonly and widely used social bookmarking service among online users, is a server-based system with a simple-to-use interface that allows users to organize and share bookmarks on the Internet. Hotho et al. (2006) saw such service yields benefits for each individual user (e.g. organizing one’s bookmarks in a browser-independent, persistent fashion) without too much overhead, while there is a proliferation of resources in the Web that makes it difficult to remain up-to-date and to keep track on documents that are related to one’s own area of interest. The usage of social bookmarking services indicates that folksonomy-based approaches seem to be the solution to overcome this difficulty.

A review on social bookmarking services, particularly Connotea [www.connotea.org], and their advantages were summarized by Hammond et al. (2005) and in its companion paper by Lund et al. (2005). Connotea makes sharing among personal collections of resources much easier than before. Instead of placing materials hierarchically in folders, Connotea allows users to create simple tags to the bookmarks. Tagging allows the organization of bookmarks

to be more flexible, multi-faceted and spacious. Furthermore, all bookmarks posted by these tools are visible to registered users, which take the concept of sharing to a higher level and benefit “not just from the ease with which it allows explicit sharing with friends and colleagues, but from many users storing their bookmarks in the same space” (Lund et al. 2005, p. 4).

2.3 Advantages/ Disadvantages of social bookmarking

Several researchers (Hotho et al, 2006; Laura, 2006; Menchen, 2005; Millen, 2006) have identified various advantages of social bookmarking. Social bookmarking services provide online storage that facilitates a single repository (Menchen, 2005). They allow an individual to create personal collections of bookmarks and facilitate sharing of bookmarks (Menchen, 2005; Millen, 2006). Social bookmarking services also allow users to create tags, which help organize and categorize users’ collection of information (Gordon-Murnane, 2006; Millen, 2006). By searching or browsing specific tags, users can retrieve all the items bookmarked by other users, which are tagged with specific keywords (Gordon-Murnane, 2006). Besides, bookmarks become portable since they are web-based. Users can access to the links and sites from computers, in contrast to the traditional way of accessing bookmarks from dedicated computers (Gordon-Murnane, 2006).

However, tags lack hierarchy so that a search of a specific term will only yield results on that term and not provide the full body of related terms that might be relevant to the user’s information needs and goals (Gordon-Murnane, 2006). Tags can be considered uncontrolled vocabulary shared across the entire social bookmarking system, and they have inherent ambiguity as different users apply terms to documents in different ways. Tags have no synonym control in the system, words with similar intended meanings, plural and singular forms will also appear in the system (Mathes, 2004). Al-Khalifa and Davis (2007) conducted a study to analyze tags in Delicious and classify them into three groups. They found that the non-standardized forms of tags make it difficult to capture them in a general form. For example, there are spelling variations, compound tags with different combination etc.

2.4 Users’ tagging behavior

Hotho et al. (2006) showed how topic-specific trends can be discovered in Delicious, one of the folksonomy-based systems. They collected users’ profiles and tags in Delicious within a time frame, and analyzed the tags to discover topic-specific trends. They argued that their analysis can be done regardless of the types of the underlying resources which would make folksonomies interesting for multimedia applications.

Kipp’s (2007) study found that a surprising number of tags used in the three social bookmarking tools examined (Delicious, Connotea and Citeulike) were not subject-related. These non subject-related tags can be classified into two groups: affective tags and time and tasks related tags. These behaviors have suggested that “users appear to want to store more than just the subject of the documents they are bookmarking” (Kipp, 2007, p. 3). They revealed specifically the users’ desire to “express an emotional connection to the document”

and to “attach personal information management information to documents” (Kipp, 2008, p. 5).

Golder and Huberman (2005, p. 198) investigated the usage patterns of the tagging system in order to “identify regularities in user activity, tag frequencies, kinds of tags used, bursts of popularity in bookmarking and the stability in the relative proportions of tags within a given URL”. It is found that the frequencies of tagging vary much among users. In particular, there is no strong connection between the length of the user account’s existence and the number of days the user takes to create one or more bookmarks. Similarly, the number of bookmarks created by users has very little association with the number of tags used in each bookmark as well. However, a user’s tagging behavior could possibly be used to reflect his or her development of interests. For instance, as a tag grows steadily over time, it might indicate the user’s continual interest in that particular subject. On the other hand, if one tag suddenly grows rapidly, it might reveal the user’s newfound interest (Golder & Huberman, 2005, p. 202).

Other trends in social bookmarking are also observed, such as the time taken for an URL to reach its peak popularity. Although it is found that a majority of the URLs would indeed reach their peak of popularity as soon as they are introduced in Delicious, others have actually taken longer time before they are “rediscovered” and experience a sudden jump (Golder & Huberman, 2005, p. 204-205). In addition, it is often believed that social bookmarking is chaotic, unstructured, and imprecise because the collection of tags depends on each individual’s personal preference and level of knowledge. However, Golder and Huberman’s study claims that each tag’s frequency for a particular URL is a nearly fixed proportion of the total frequency of all tags used. More interestingly, this stability becomes apparent after fewer than 100 bookmarks (Golder & Huberman, 2005, p. 205-206).

2.5 Research Gap

Golder and Huberman (2006) analyzed the structure of collaborative tagging systems of Delicious and discovered regular patterns in user’s activity and tag frequencies. He expected that such findings could be applied to other similar tagging systems. However, little has been done regarding his claim. Therefore, it is worthwhile to conduct a research using similar methodology to that of Golder and Huberman (2006), but with a social bookmarking service designed for scientific purposes, such as Connotea.

The preliminary investigation using data mining also suggests that there is different usage pattern between the two sites – users of Delicious create tags for all kinds of topics while Connotea users create tags more for research purposes. Therefore, it is meaningful to study the users’ tagging behavior in Delicious alongside Connotea to find out if there are any significant similarities and differences between social bookmarking services of different orientations.

3. The Study

3.1 Data Collection and Sampling

Analysis is conducted on two sets of data (one for Delicious and one for Connotea) which contain the activity profiles of some of its users. A program is written using Java 1.4.2 to retrieve data that query both databases. The ways to decide which user should be included are slightly different between Delicious and Connotea because Connotea has a smaller user base than Delicious. For Delicious, the user's history of activity profile is collected and analyzed if one or more of his/her bookmarks reached the top 500 most popular bookmarks on Feb 16, 2009 (Midnight GMT). This information is found through the public RSS feeds of the 'popular' page and then a portion of the website from the user's activity profile is *crawled*. For Connotea, the list of users is compiled by examining at the bookmarks posted between April 1, 2008 (Midnight GMT) and May 31, 2008 (23:59 GMT).

This time period is chosen because of two reasons: 1) Most of the researchers and lecturers would have their summer break in June, it would be best to set the time frame before summer to represent the use of Connotea among researchers; 2) the user base for Connotea is smaller than Delicious and a longer period is needed to have a similar amount of bookmarks tagged. Based on the list of users who bookmarked in the aforementioned period, all the bookmarks they have posted since the activation of their accounts till Feb 17, 2009 (0:00:00 GMT) were retrieved. This procedure stopped when over 200,000 bookmarks from both databases were collected. As a result, we collected 5454 users' data from Connotea and 440 users' data from Delicious. Note here that the sample size of Connotea is over twelve times more than that of Delicious because Delicious users on average create much more bookmarks than Connotea users (see Table 1).

Table 1. Descriptive Statistics of user profile

	Delicious	Connotea
Mean	1404.175	49.724
Median	711.500	4
Min	3	1
Max	19100	15067

Note. The values are presented in number of bookmarks

3.2 Data Analysis

This analysis studies the bookmarking and tagging behavior of Connotea and Delicious respectively. The result between Connotea and Delicious is also compared to examine any significant differences.

3.2.1 The descriptive statistics of users' bookmarking behavior

From the statistics (see Table 2), it is found that the bookmarking behaviors of Connotea and Delicious users differ significantly. It is shown from the standard deviation value and range value that users of Delicious do not deviate a lot in their bookmarking behavior. All users in

the sample pool create at least one bookmark every month. However, some Connotea users may only create one bookmark every two years.

Table 2. Descriptive Statistics of users' bookmarking behavior of Connotea and Delicious

	Delicious	Connotea
Mean	1.485	8.394
Standard Deviation	2.356	23.614
Minimum	0	0
Maximum	24.221	676.863

Note. The values are present in days.

3.2.2 Relationship between the length of the account since activation and the total number of bookmarks they created.

In this regression analysis, both Connotea and Delicious' data are best fitted into the exponential model. See Table 3 for the exponential regression results.

Delicious. As a result, the length of the user account since activation is a significant predictor of users' bookmarking behavior variance ($F = 74.323, p < .05$). In other words, the length of the user account since activation accounts for 14.3% of users' bookmarking behavior variance. Therefore, it was likely for Delicious users who have created an account for a longer time to create more bookmarks ($\beta = .381, p < .05$).

Connotea. The length of the user account's existence since activation is a significant predictor of users' bookmarking behavior variance ($F = 5806.223, p < .05$). In other words, the length of the user account's since activation accounts for 51.6% of users' bookmarking behavior variance. Connotea users who have created an account for a longer time were therefore more likely to create more bookmarks ($\beta = .718, p < .05$).

Table 3. Univariate Regressions: Delicious versus Connotea
(Dependent Variable: number of bookmarks created)

	R ²	df	F	Sig.	β	t
Delicious ^a	.145	1	74.323	.000	.381	8.621
Connotea ^a	.516	1	5806.223	.000	.718	76.199

^apredictor: the length of the user account's existence in days

In addition, a two-way between-subjects analysis of variance is conducted to compare the effect of the length of account since activation on number of bookmarks created in Delicious and Connotea. The between-subjects factors are length of the existence of the account and

nominal value with two levels (1 or 0 representing Connotea or Delicious, respectively). A significant interaction effect was found, $F(1, 5890) = 1071.98, p < .000$. The positive coefficient of interaction term (0.0039) suggests the length of account's existence has a greater effect on number of bookmarks created in Connotea than Delicious.

3.3.3 Relationship between the number of bookmarks a user creates and the number of tags they use in those bookmarks.

In this analysis, the overall relationship between the number of bookmarks users created and the number of tags they used in those bookmarks will be analyzed. The lower end of the scale (users keeping fewer than 30 bookmarks) and the upper end (users keeping more than 500 bookmarks) will also be examined. Both Connotea and Delicious' data are best fitted into the exponential model for regression analysis. Table 4 gives the exponential regression results.

Delicious. As a result, the number of bookmarks a Delicious user creates is a significant predictor of users' tagging behavior variance ($F = 1207.6, p < .05$). In other words, the number of bookmarks a user creates is able to account for 73.7% of users' tagging behavior variance. Users who create more bookmarks are also likely to use more tags for those bookmarks ($\beta = .858, p < .05$). The relationship is found to be weaker at the lower end of the scale with users having fewer than 30 bookmarks ($R = .652, R^2 = .425, p < .05$), but stronger at the upper end with users having more than 500 bookmarks ($R = .716, R^2 = .512, p < .05$).

Connotea. The number of bookmarks a Connotea user creates is a significant predictor of users' tagging behavior variance ($F = 38941.46, p < .05$). The results suggest that the number of bookmarks a user creates is able to account for 87.7% of users' tagging behavior variance. Users who create more bookmarks will also likely to use more tags for those bookmarks ($\beta = .937, p < .05$). It is stronger at the lower end of the scale, with users having fewer than 30 bookmarks ($R = .845, R^2 = .715, p < .05$) but comparatively weaker at the upper end, with users having more than 500 bookmarks ($R = .825, R^2 = .681, p < .05$).

Table 4. Univariate Regressions: Delicious Versus Connotea
(Dependent Variable: number of tags they use in bookmarks)

	N	R ²	df	F	β	t
Delicious^a						
Overall	442	.737	1	1207.6**	.858	34.751
Lower end (< 30 bookmarks)	14	.425	1	9.604**	.652	3.099
Upper end (>500 bookmarks)	277	.512	1	289.76**	.716	17.022
Connotea^a						
Overall	5455	.877	1	38941.466**	.937	197.336
Lower end (< 30 bookmarks)	4334	.715	1	10848.069**	.845	104.154
Upper end (>500 bookmarks)	92	.681	1	194.527**	.825	13.947

^a predictor: number of bookmarks a user creates

** $p < .01$.

In addition, a two-way between-subjects analysis of variance is conducted to compare the effect of the number of bookmarks created on number of tags used in Delicious and Connotea. The between-subjects factors are number of bookmarks created and nominal value with two levels (1 and 0 represents Connotea and Delicious respectively). A significant interaction effect was found, $F(1, 5884) = 5.79, p < .05$. The negative coefficient of interaction term (-0.054) suggests the number of bookmarks created has a greater effect on number of tags used in Delicious than Connotea.

3.3.3 Proportion of unique tags over all tags used per user account

Figure 3 shows a significant difference in tagging behavior between Connotea and Delicious users. Almost half of Connotea users use mainly distinctive tags to organize or classify their bookmarks, while most of Delicious users use comparatively less unique tags for organization and classification.

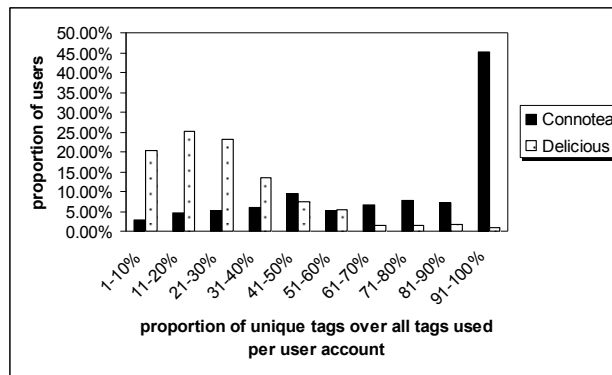


Fig. 3. Proportion of unique tags over all tags used per user account in Connotea and Delicious
Note. 6 users from Delicious are excluded since they do not have any tags.

4. Discussion and Implications

Golder and Huberman (2006) expected that their findings on Delicious can be applied to other similar tagging systems. However, our findings suggest that Delicious and Connotea are quite different from one another. Sections 3.2.1 and 3.2.2 analyze the bookmarking behavior of Delicious and Connotea users, and show a distinctive difference between them. The reason behind this finding can be explained by the different user orientations of these two social bookmarking services. Since Delicious is used for general purposes, its users may use Delicious to bookmark websites related to personal interests or entertainment etc., which may occur frequently. Connotea, on the other hand, is a social bookmarking service that caters specifically to the management of scientific references. Users of Connotea mainly use its

tools for doing research or for other academic purposes. Once they complete their assignments or research projects, they usually stop using Connotea until they start another project.

Sections 3.2.3 and 3.2.4 examine the tagging behavior of Delicious and Connotea users. Section 3.2.3 shows that both groups of users will use more tags if they have more bookmarks. It is noted that Connotea requires users to use at least one tag per bookmark while Delicious has no such restriction. These findings may suggest that tagging is a useful tool for users to manage, classify and organize bookmarks so that users are willing to create tags on bookmarks even if they are not required to do so. However, users of Delicious and Connotea have different tagging behavior on the bookmarks they created. Section 3.2.4 shows that Connotea users are more likely to use distinctive tags on their bookmarks than Delicious users. Connotea users, when conducting their research or projects, may classify their online resources into specific topics for future retrieval or to facilitate division of labors, so that more unique tags are used to classify different topics within a project. Delicious users, in contrast, are likely to use more general tags to classify their bookmarks, such as *music*, *sport*, etc., in order to present a general topic.

Since no publication is found on studying the similarities and differences of users' behavior in using two different bookmarking services that target at different user groups, this study helps fill the research gap by comparing the bookmarking and tagging behavior of Delicious and Connotea users.

5. Limitation and further research

This study examines the usage pattern in Connotea and Delicious, by analyzing the log data crawled from these two social bookmarking websites. The analysis can only identify the usage pattern and the difference in behavior of these two social bookmarking users, and the reason behind such pattern or difference is yet to be investigated. Further study can be done from the users' perspective to examine in what ways social bookmarking users utilize the services and how effective the services are in helping them manage or organize their online resources.

6. Conclusion

Although social bookmarking has received growing attention and interest in the general public as well as in academia, the users' bookmarking and tagging behavior have not been distinguished out of social bookmarking services which target at different user groups. The reasons behind such difference are yet to be discovered. The results of this study suggest that the bookmarking behaviors of Connotea and Delicious users have distinctive difference. Most of the Delicious users (in the sample pool) create bookmarks frequently, while Connotea users deviate a lot in their bookmarking behavior. The discrepancy in their bookmarking behavior can be partly explained by the existence period of user accounts, which has a greater effect on Connotea users than Delicious. The study also finds that there is a strong link between bookmarks created and tags used. In particular, Connotea users tend to use more

unique tags for their bookmarks than Delicious users. Further investigation is needed to probe deeper into the reasons behind the social bookmarking users' behavioral differences and assess the effectiveness of different bookmark management strategies employed by these users.

Acknowledgement

The research team would like to thank Mr. Wenfeng Han for his contribution in the preliminary analysis.

References

- Al-Khalifa HS, Davis HC (2007): Towards Better Understanding of Folksonomic Patterns. In: Conference on Hypertext and Hypermedia, Proceedings of the 18th Conference on Hypertext and Hypermedia, Manchester, UK: ACM Press, pp.163–166.
- Dye, J. (2006). Folksonomy: A game of high-tech (and high-stakes) tag. *E-Content*, 29(3), 38-43.
- Golder, S.A., Huberman, B.A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2), 198-208.
- Gordon-Murnane, L. (2006). Social bookmarking, folksonomies and web 2.0 tools. *Searcher*, 14(6):26-38.
- Hammond, T., Hannay, T., Lund, B., & Scott, J. (2005). Social bookmarking tools. A general review. Part 1. *DLib Magazine*, 12(1).
- Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: Search and ranking. In: Sure, Y., Domingue, J. (eds.) ESWC 2006. LNCS, vol. 4011, pp. 411–426. Springer, Heidelberg (2006)
- Kipp, M.E.I. (2006a). @toread and cool: Tagging for time, task and emotion. 17th ASIS&T SIG/CR
- Lund, B., Hammond, T., Flack, M., & Hannay, T. (2005). Social bookmarking tools (II). A case study - Connotea. *D-Lib Magazine*, 11(4).
- Mathes, A. (2004). Folksonomies - Cooperative classification and communication through shared metadata. *Computer Mediated Communication, LIS590CMC*, 1-13.
- Menchen, E. Feedback, Motivation and Collectivity in a Social Bookmarking System. In *Kairosnews Computers and Writing Online Conference*. 2005.
- Millen, D., Feinberg, J., Kerr, B.: "Social Bookmarking in the Enterprise", *ACM Queue*, 3, 9 (2005), 28-35.