**Samuel Kai Wah Chu**
**Division of Information & Technology Studies, Faculty of Education, The University of Hong Kong**
**Dinesh Rathi**
**School of Library and Information Studies, University of Alberta**
**Helen S. Du**
**Department of Computing, The Hong Kong Polytechnic University**

# Users' Behaviour in Collaborative Tagging Systems

**Abstract:** This study draws upon the work of Golder and Huberman (2005) and examines the users' behaviour of two social bookmarking websites, Delicious and Connotea, by analyzing over 200,000 bookmarks from each website. The study also explores the cause(s) for similarity and difference in the behaviour of users of both websites.

**Résumé :** Basée sur les travaux de Golder et Huberman (2005), cette étude examine le comportement des usagers de deux sites web d'étiquetage social : Delicious et Connotea par l'analyse de plus de 200 000 signets dans chaque site. L'étude explore également les origines des similitudes et des différences dans le comportement des usagers des deux sites.

Social bookmarking services have attracted a large number of users. For example, Delicious (http://del.icio.us) has over two million users since its launch in 2003 (Schachter, 2007). These users are adding a large amount of data (Fichter, 2006) including the trail of their activities and the users' behaviour has not yet been studied extensively. The majority of articles written on social bookmarking have focused predominantly on social bookmarking tools (Hammond et al., 2005; Lund et al., 2005; Gordon-Murnane, 2006), the use of social bookmarking in aiding search queries (Heymann et al., 2008; Yanbe et al., 2007), and in development of ontologies (Zacharias and Braun, 2007), etc.. However, there is a limited work on user behaviour on folksonomy-based web collaborative system(s). Hotho et al. (2006) claimed that "there are currently virtually no scientific publications about folksonomy-based web collaboration systems. The main discussion on folksonomies and related topics is currently taking place on mailing lists only" (p. 3).

There is a proliferation of resources on the Web. For example, Google's index grew from 26 million pages in 1998 to one billion pages in 2000 and by 2008, the number of unique URLs grew to 1 trillion (Alpert and Hajaj, 2008). Thus, there is an information overload for the users and they might experience difficultly staying up-to-date on changes and updates on the web resources related to their area of interest. The use of social bookmarking tools indicates that web-and-folksonomy-based approaches seems to be the solution for overcoming this difficulty (Hotho et al., 2006). Social bookmarking services, such as Delicious, are basically server-based systems with simple-to-use interfaces that allows users to organize and share bookmarks on the Internet. These bookmarks are not saved in a particular browser that the users use on their personal computers but are in the cloud. Thus, these bookmarks are machine as well as browser independent. Most social bookmarking tools allow people to assign a number of tags (keywords) to a bookmark. The tags assigned to the bookmarks help the users organize their information resources using natural language as opposed to controlled vocabulary. Mathes (2004) examined user-generated metadata (i.e., tags created by users of a document) and the findings of the study suggest that the tagging function of bookmarking tools encourages users to organize information in their own way. In addition, these tools provide opportunities for other users to contribute in

tagging (Godwin-Jones, 2006). This is a significant change when compared to the traditional information structure, where subject specialists will classify and organize the information.

Gordon-Murnane (2006) explores the use of the social bookmarking tools and argues that these tools have begun to serve the academic and scientific community as well. Examples of such tools include CiteULike [http://www.citeulike.org/] and Connotea [http://www.connotea.org/]. CiteULike is a free service designed for academics to "share, store, and organize the academic papers they are reading" and the software "automatically extracts the citation details" ("CiteULike FAQ", n.d.). Hammond et al. (2005) and Lund et al. (2005) present the advantage of social bookmarking tools. For example, Connotea allows the placement of tagged material hierarchically in different folders, facilitates the organization of the material on the web, and at the same time eases the sharing of personal collections of resources with other users (by being on the Web). Lund et al., (2005) argues that tagging makes the organization of bookmarks more flexible, multi-faceted and spacious. The openness benefits "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" (p. 4).

The research work of Golder and Huberman (2005) is one of the most cited works in this area and has over 500 citations (as of April 10, 2010) as reflected in Google Scholar. They investigate the usage patterns as well as the structure of a collaborative tagging system, which included regularities in the user activity, tag frequencies, kinds of tags used, bursts of popularity in bookmarking, and stability in the relative proportions of tags within a given URL. The findings of the study reveal that when users discover new interests, they may add new tags to categorize and describe the bookmarks. This leads to the expansion of users' tags over time and the tag expansion may have different growth rates. The authors also argue that the tags could map the development and change in users' interests over time. Golder and Huberman (2005, p. 207) state that "while this research was focused on Delicious, we expect that these findings will apply to other, similar tagging system". Golder and Huberman's (2005) pioneer article was one of the first in understanding the users' patterns of the tagging system. However, the data size used in the study was a relatively small dataset (i.e., 229 users with a total of 68,668 bookmarks) which was collected from a single social bookmarking website, i.e., Delicious. While Golder and Huberman's (2005) study presents many interesting findings, this work needs to be explored (as well as expanded) further by using a much larger dataset and evaluating more than one social bookmarking website.

This study examines the behaviour of users on the two social bookmarking websites i.e., Delicious and Connotea. The proposed study draws upon the work of Golder and Huberman (2005). This study is conducted on a much larger sample size that was collected from these two social bookmarking services as compared to the data set used by Golder and Huberman (2005) in their study. In addition, the study also examines whether there is any surprising deviation in users' behaviours since 2005, as identified by Golder and Huberman (2005). The proposed work also explores the cause(s) for similarity and difference in the behaviour of users of both websites.

The data for the proposed study was collected from the two social bookmarking websites viz. Connotea and Delicious by using data crawler which was coded using Java 1.4.2. The collected data included different types of information. For example, the users' account names, the URL of the web pages that the users bookmarked, date stamp of their bookmark, and the tags that the users used to classify the content of the web pages that they bookmarked. The websites were crawled and the bookmarks (and

related data such as tags) were downloaded from both the website. Over 200,000 bookmarks were downloaded from Connotea and over 600,000 bookmarks were downloaded. Since there was difference in the data size (nearly 1:3 for Connotea to Delicious) in terms of total number of bookmarks downloaded from both the website, hence we decided to randomly select around 200,000 bookmarks from each website's downloaded data. Our earlier aim was to select equal number of bookmarks (i.e., 200,000) for both the websites from the downloaded data but in doing so, the profile of the (randomly selected) last user (i.e., the 'last user' in the randomly selected list of users) was represented partially. Thus to avoid the partial representation of the last user's profile, we decided to have all the bookmarks of that particular user (i.e., the last user). This led to a minor difference in the total count (i.e., exactly not the same count) of bookmarks selected from each website.

The preliminary analysis indicates that the bookmarking behaviors of Connotea and Delicious users differ significantly. For example, Delicious users in the sample pool created bookmarks at a much faster rate than many of the Connotea users' who took much longer (average) time to tag a bookmark. The possible reason for the difference in user behaviours could be due to the difference in the target audience of the two bookmarking websites. For example, Delicious is generally patronized by users who use this tool for general purposes. Connotea, on the other hand, is a social bookmarking service that caters specifically to the management of scientific reference sources found on the Web. Thus a large number of Connotea users might be from a scientific community who use this tool for conducting research for any other academic purposes. Such users are more likely to stop using Connotea upon the completion of the research projects and/or assignment and would return when they will start another project. Since all the data collection is complete and the analysis is under progress, the detailed findings will be completed prior to the CAIS/ACSI conference and will be presented at the conference.

To the best of our knowledge, there has been no follow-up to Golder and Huberman's (2005) study, hence the findings of this study will contribute towards the advancement of understanding users' behaviour including comparative evaluation of user behaviour on the different social bookmarking websites and provide insight into changes in users' behaviour over a period of 4 years as discovered by Golder and Huberman in 2005.

The proposed research work fits well with the conference theme on Social Networking and User Participation in Knowledge Structure.

**References**

Alpert, J. and Hajaj, N. (2008) We Knew the Web was Big, As Accessed on January 15, 2010 from http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html.

CiteULike FAQ. (n.d.). CiteULike FAQ. As Accessed on July 27, 2007 from http://www.citeulike.org/faq/all.adp.

Fichter, D. (2006). Intranet applications for tagging and folksonomies. Online, 30(3), 43-45.

Godwin-Jones, R. (2006). Tag clouds in the blogosphere: Electronic literacy and social networking. Language Learning & Technology, 10(2), 8-15.

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. Red Orbit. As Accessed on 27 July, 2007 from http://www.redorbit.com/news/technology/ 541192/social_bookmarking_folksonomies_ and_web_20_tools/

Hammond, T., Hannay, T., Lund, B., & Scott, J. (2005). Social bookmarking tools (I): A general review. D-Lib Magazine, 11(4). As Accessed on July 27, 2007 from http://www.dlib.org/ dlib/april05/hammond/ 04hammond.html.

Heymann, P., Kourtika, G. and Garcia-Molina, H. (2008) Can Social Bookmarking Improve Web Search?, Proceedings of the International Conference on Web Search and Web Data Mining, 195-206

Hotho, A., Jäschke, R., Schmitz, C., & Stumme, G. (2006, Dec). Trend detection in folksonomies. Paper presented at the 1st International Conference on Semantics and Digital Media Technology, Athens, Greece.

Lund, B., Hammond, T., Flack, M., & Hannay, T. (2005). Social bookmarking tools (II): A Case Study - Connotea. D-Lib Magazine, 11(4). As Accessed on July 27, 2007 from http://www.dlib.org/dlib/april05/lund /04lund.html.

Mathes, A. (2004). Folksonomies - Cooperative classification and communication through shared metadata. Computer Mediated Communication, LIS590CMC, 1-13.

Schachter, J. (2007). del.icio.us / blog. As Accessed from 27 July, 2007 from http://blog.del.icio.us/blog/2007/03/index.html.

Yanbe, Y., Jatowt, A., Nakamura, S. and Tanaka, K. (2007) Can Social Bookmarking Enhance Search in the Web?, Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, 107 – 116.