

## Do Tags Really Provide More Semantic Concepts than LCSH Does?

Jihee Beak

School of Information Studies  
University of Wisconsin-Milwaukee  
jbeak@uwm.edu

### Abstract

The purpose of this study is to investigate the patterns that appear in LCSH and tags in fiction genres. After identifying the patterns, this study examines whether tags truly provide more subjects or not through semantic equivalence controlling. In library practice, subject access to fiction by such authority controlled vocabularies and/or other cataloging standard tools is more restrictive than subject access by social tagging. On the other hand, social tagging is created by larger number of different users. Taggers are more likely to create tags after reading books. There is no guideline or manual to create tags.

Many studies about social tagging have been identifying the characteristics of social tagging. The significant advantages of social tagging are; 1) to be user-friendly terms, 2) to provide larger number of terms, 3) to show latent subjects, and 4) to include non-subject related terms like personal information. As the attention about social tagging is greater, recently researchers concerned with whether social tagging can be combined with the library system. In terms of incorporating social tagging into the library system, social tagging has been criticized for its quality issues. Social tagging tends to be ambiguous, uncontrolled, and inconsistent. These natures of social tagging impede the implementation of social tagging in the library system.

Nevertheless of social tagging's quality issues, some libraries have already started to expand tagging services. In order to explore more plausible possibility of the implementation, recent studies focus on the comparison or linkage between controlled vocabularies and social tagging (Lu, Park, & Hu, 2010; Yi & Chan, 2009; Smith, 2007; Rolla, 2009; Heyman & Garcia-Molina, 2009). Most of these studies focus on tags created for academic resources or image resources rather than fiction. The problematic phenomenon in library system is that fiction has less subject access points than non-fiction like academic resources. Given that social tagging provides more subject access points than libraries do, the library environment may take advantage of social tagging to enhance the subject access to fiction. Therefore, the study about comparison between social tagging and controlled vocabularies for fiction is required.

The data collection was done during April 2011. The study selected 120 fiction recommended by RUSA (Reference & User Services Association)

from 2008-2011. The RUSA suggests 8 fiction genres: Adrenaline, Fantasy, Historical fiction, Horror, Mystery, Romance, Science fiction, and Women's fiction. 15 fiction books were selected from each genre. By using books' ISBNs, LCSH was collected from Los Angeles Public Library (LAPL). Given that the collection of Library of Congress (LC) is likely to have non-fiction rather than fiction, the study decided to collect the LCSH from a public library. LAPA provides MARC records. MARC records enabled the study to distinguish whether subject headings are from LC or locally devised topical subjects. LCSH of 6XX level was collected. The ISBNs of the selected fiction was also used to collect tags from LibraryThing website. LibraryThing website shows 30 popular tags in a main webpage of each book. The collected tags come from these popular tags rather than all tags assigned by users. In total, the 120 fiction books have 3,600 tags by users and 600 Library of Congress subject terms.

This study consists of two parts: 1) explores the patterns of terms that appear in fiction from a social tagging website and LCSH, and 2) examine impact of semantic equivalent control to the number of overlapped terms between tags and LCSH. If individual tags are considered as access points, the patterns of tags will mean facets of information as metadata elements. Therefore, the study tries to contribute to suggest user-friendly metadata elements for fiction by identifying the patterns of tagging. Furthermore, the study investigates to what extent social tagging is overlapped with LCSH in fiction. While social tagging gets praised for a greater number of subject access points, it also incurs blame for being uncontrolled and inconsistent. It illustrates that there are many terms syntactically and semantically equivalent. Therefore, this study statistically shows whether there is significant difference in overlapped terms between before and after controlling terms by semantic equivalence.

### References

Heymann, Paul & Garcia-Molina, Hector. (2009). Contrasting controlled vocabulary and tagging: *Do Experts Choose the Right Names to Label the Wrong Things?*. In *Second ACM International Conference on Web Search and Data Mining (WSDM 2009), Late Breaking Results Session*, February 9-13, 2009, Barcelona, Spain.

Lu, Caimei, Park, Jung-ran, & Hu, Xiaohua. (2010). User tags versus expert-assigned subject terms: A comparison of LibraryThing tags and Library of Congress Subject Headings. *Journal of Information Science*, 36(6). 763-779.

Rolla, Peter J. (2009). User tags versus subject headings: Can user-supplied data improve subject access to library collection? *Library Resources and Technical Services*, 53(3). 174-184.

Smith, Tiffany. (2007). Cataloging and you: Measuring the efficacy of a folksonomy for subject analysis. *In Proceedings 18th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research*, Milwaukee, Wisconsin.

Yi, Kwan & Chan, Lois M. (2009) Linking folksonomy to Library of Congress subject headings: An exploratory study. *Journal of Documentation*, 65(6).