Automated Generation of Social Tag Values for Popular Content – Critical Assessment of Two Semantic Technologies with Five Embedded Knowledge Organization Systems

Dr. Denise A. D. Bedford, Goodyear Professor of Knowledge Management
Information Architecture and Knowledge Management
Kent State University
Kent OH 44242-0001
Dbedfor3@kent.edu

Relevance to Workshop: Role of Social Tagging and Relevance of Social Tagging to NKOS Applications, NKOS-Fueled Semantic Technologies

Proposal Abstract: Research has demonstrated that there is a clear role for tagging in the semantic future. Tagging can both augment access to knowledge and add value to Knowledge Organization Systems (KOS). While there is value to be derived from social tags, there are significant challenges associated with managing and integrating tag values into Knowledge Organization Systems. Tags are often of questionable quality (Golder and Huberman 2011) (Guy and Tonkin 2006) (Mathes 2004) (Merholz 2004). Tag values are often confounded across facets or attributes (Spiteri 2010) (Yi 2010) (Fernandez-Tobias 2006). Tag values may represent different levels of specificity (Golder and Huberman 2011).   Management and integration often requires human manual judgment and intervention. Manual approaches to integration are labor-intensive and time-consuming, and raise questions of cost versus benefit. Researchers have already demonstrated the feasibility of generating user tags through clustering and semantic similarity measures (Shiri 2009) (Razikin et al 2011) (Fu et al 2010). However, to date no automated solution has demonstrated the feasibility of generated well formed and managed tag values for integration into a Knowledge Organization System. A possible solution to these challenges involves using semantic technologies with embedded Knowledge Organization Systems to automatically generate tags from which users can select. Such a solution could mitigate the management and integration challenges. This approach might produce tags which are by definition well formed because they derive from a managed KOS. Slight variations in terms which result in redundancies or scattered tag values may also be managed from the KOS source. For this solution to have value, though, the tags suggested by the semantic technologies must coincide with what the user would have suggested.

This research expands upon research that was presented to the NKOS 2011 Workshop. The preliminary research reported in 2011 suggested that:

1. 90% of the time, the semantic engine when powered by a KOS identified the core topic terms suggested by end users.
2. The semantic engine, when powered only by a topic-focused KOS, identified 45% of all of the topic-focused tag values suggested by end users. In this case, the remaining tag values pertained to other metadata attributes such as People, Geographic Locations, Products, or Events.
3. Semantic technologies will generate between 1.5 to 10 times as many topical tags as are suggested by single users. This result may provide a richer set of tag values from which end users may choose.

The original research was conducted on a data set of tagged items, including books, journal articles and press stories in five disciplines, e.g. agriculture, environment, health, education and transport. The tagged items were derived from the open web and from other sites such as Amazon.com, CiteULike, del.icio.us, and Digg. The current research builds upon the original work in three ways. First, it tests our ability to increase the rate of coverage of tag values by expanding the Knowledge Organization Systems used. Four Knowledge Organization Systems are leveraged, covering People Referenced, Geographic Locations, Products and Events. Second, the original research is further expanded to include the use of two different semantic technologies – the SAS/Content Categorization Suite and ClearForest. Different semantic approaches may be better suited to different types of metadata. Finally, the data set is expanded to include popular content which provides a more rigorous test of the new metadata attributes.

References

Fernandez-Tobias, I., Cantador, I. and Bellogin, A. Tag: Semantic Contextualisation of Social Tags, Accessed online on June 15, 2012 at http://ir.ii.uam.es/rim3/publications/sasweb11-fernandez-tobias.pdf

Fu, W.-T., Kannampallil, T., Ruogu, K. and He, J. (2010). Semantic Imitation in Social Tagging, *ACM Transactions on Computer-Human Interaction*, Vol. 17, No. 3,

Golder, S. and Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science Vol.* 32, No. 2, 198--208.

Guy, M., and Tonkin, E. (2006). Folksonomies: Tidying up tags? *D-LibMagazine Vol.* 12, No. 1. Accessed online on June 15, 2012 at http://www.dlib.org/dlib/january06/guy/01guy.html.

Mathes, A. (2004). Folksonomies: Cooperative classification and communication through shared metadata. Accessed online on June 15, 2012 from http://adammathes.com/academic/computer-mediatedcommunication/folksonomies.html

Merholz, P. (2004). Metadata for the masses. Accessed online on June 15, 2012 from http://www.adaptivepath.com/ideas/essays/archives/000361.php

Razikin, K., Goh, D. H., Chua, A. Y. K. and Lee, C.S. (2011). Social tags for resource discovery: a comparison between machine learning and user-centric approaches, Journal *of Information Science* Vol. 37 No. 4, 391–404

Shiri, A. (2009). An examination of social tagging interface features and functionalities: An analytical comparison of social tagging, Online *Information Review* Vol. 33 No. 5, 901-919

Spiteri, L. (2010). Incorporating Facets into Social Tagging Applications: An Analysis of Current Trends, *Cataloging & Classification Quarterly* Vol. 48:94–109

Yi, K. (2010). A Semantic Similarity Approach to Predicting Library of Congress Subject Headings for Social Tags, *Journal of the American Society for Information Science and Technology*, Vol. 61, No. 8, 1658–1672,