

Enhancing social tagging with a knowledge organization system

Koraljka Golub
UKOLN - The University of
Bath, UK
k.golub@ukoln.ac.uk
Brian Matthews, Science
and Technology Facilities
Council, UK
B.M.Matthews@rl.ac.uk
Marianne Lykke Nielsen
Royal School of Library and
Information Science,
Denmark
mln@db.dk

Catherine Jones,
Science and Technology
Facilities Council, UK
C.M.Jones@rl.ac.uk
Jim Moon
University of Glamorgan,
UK
jnjmoon@glam.ac.uk
Douglas Tudhope
University of Glamorgan,
UK
dstudhope@glam.ac.uk

Abstract

Traditional subject indexing and classification are considered infeasible in many digital collections. Automated means and social tagging are often suggested as the two possible solutions. Both, however, have disadvantages and, depending on the purpose of use or context, require additional manual input. This study investigates ways of enhancing social tagging via knowledge organization systems, with a view to improving the quality of tags for increased information discovery and retrieval performance. Benefits of using both social tags and controlled terms are also explored, including enriching knowledge organization systems with new concepts.

Keywords: folksonomy; social tagging; knowledge organization system; controlled vocabulary

1. Introduction

Knowledge organization systems have been used as tools for information discovery and retrieval in libraries and abstracting and indexing services, some for more than a century. Their benefits for improved information retrieval in the digital environment have been well acknowledged and recognized. They have devices to reduce the ambiguity of natural language when describing and retrieving items, and to allow access via browsing and navigation. However, there are costs associated with use of knowledge organization systems – manual indexing or classification are a significant resource, especially when performed by trained indexers.

Social tagging applications, such as Flickr (2008) and Del.icio.us (2008) with their community-based user interfaces encouraging social tagging activity, currently attract much attention and are seen as key elements of new Web 2.0 services. They hold the promise of reducing indexing costs by drawing end-users into contributing, adding value as part of their interaction with information services. However, social tagging is less concerned with consistency than with making it easier for end-users to describe information items and to have access to other users' descriptions. Existing social tagging applications have not been designed with information discovery and retrieval in mind. The resulting folksonomies (collections of tags) are completely uncontrolled, lacking even basic control of word forms such as spelling variants, synonyms and disambiguation of homonyms (cf. Spiteri 2007; Guy & Tonkin 2006). Many users use tags only to organize own documents, and not to help the community (cf. Tonkin et al. 2008). On the other hand, natural language tags could cover aspects that are not available in a knowledge organization system, especially when it comes to new concepts; as such, they could help update the knowledge organization system.

The EnTag project explores the combination and comparison of controlled and folksonomy approaches to semantic interoperability in the context of repositories and digital collections. The aim is to investigate the effect on both indexing and retrieval when using only social tagging versus when using social tagging in combination with a knowledge organization system. The remainder of this paper discusses related work (section 2), the overall approach (section 3), and concluding remarks with future steps (section 4).

2. Related work

The need for knowledge organization systems in relation to folksonomies has been reported in the literature. Weller (2007) compares ontologies and folksonomies, suggesting that they are not to be seen as rivals but complement each other. Noruzi (2007) provides seven arguments for why a folksonomy-based system should use a thesaurus, emphasizing that there is no way to maintain consistency over time or across folksonomy users without a thesaurus. For Connotea (2008), a service for organizing references, recently an add-on tool has been developed that allows taggers to select terms from a knowledge organization system (Entity Describer 2007).

Smith (2007) explores the connection between folksonomies and Library of Congress Subject Headings (LCSH) and describes advantages and disadvantages of each. She suggests that their product (called LibraryThing for Libraries) may “provide a compromise between the constraints of controlled vocabulary and the relative wilderness of the folksonomy”. Hayman (2007) argues for combining the best of the two worlds and describes its application on Australian collection of education resources (education.au). In their collection, knowledge organization systems are used for metadata creation and searching, and in order to keep pace with user needs, folksonomies are being explored. Users can tag resources by choosing from an established taxonomy or by entering their own terms. Users’ own terms will be used later to feed back into the taxonomy to improve its quality. The Library of Congress is collaborating with Flickr (2008), in order to enhance bibliographic records for its images by end-user tags (Raymond 2008).

3. Overall approach

The main focus of investigation will be the effect of an enhanced tagging system. The enhanced system, with the capability of offering suggestions via a knowledge organization system, will be compared against free social tagging. Two different contexts will be explored: tagging by readers (Intute) and tagging by authors (Science and Technology Facilities Council (STFC)). For each of these a separate demonstrator will be developed, one operating on data extracted from Intute (Intute 2008), and the other operating over STFC’s repository (STFC ePublication Archive 2008) in which tagging will be conducted by authors submitting papers to the repository. A user study will be conducted for each demonstrator, which will allow a general comparison of a repository versus digital collection context, a different knowledge organization system, interface and user community. In the EnTag Intute demonstrator, the Dewey Decimal Classification (DDC) forms the knowledge organization system.

Two major methods planned to collect user data are a questionnaire and log analysis. The evaluation of the Intute demonstrator will compare basic and advanced system for indexing and retrieval implications. Indexing aspects included are perspective, specificity, exhaustivity; language (e.g., word class, single word/compound, spelling); consistency; efficiency (time used, user satisfaction); and, use (tags selected, clouds consulted, order of consultation). Retrieval evaluation will be based on examining the degree of match between user and system terminology as to user tags, DDC tags, controlled Intute keywords, title terms, and text terms. Evaluation will also consider user interface issues. The test setting is planned to comprise some several dozen of students in political science and documents covering three or four topics of relevance for the students.

4. Concluding remarks

The next steps include pilot testing of the demonstrators, the user studies, analysis and reports. All developments and results are regularly updated at the project's Web site, <http://www.ukoln.ac.uk/projects/enhanced-tagging/>. Final results are expected to be obtained by the end of 2008.

In future work, it is hoped to generalize the knowledge organization elements considered useful as terminology services. We also intend to conduct studies with more consideration of longitudinal effects and a wider range of repositories and user groups.

Acknowledgements

This work is funded by UK's Joint Information Systems Committee (JISC) Capital Programme. Thanks are due to Intute for supplying an extract of their social science collection and to OCLC Research for providing the Dewey Decimal Classification.

References

- Connotea. (2008). Retrieved March 19, 2008, from <http://www.connotea.org/>.
- Del.icio.us. (2008). Retrieved March 19, 2008, from <http://del.icio.us/>.
- Entity Describer. (2007). Retrieved March 19, 2008, from <http://www.connotea.org/wiki/EntityDescriber>.
- Flickr. (2008). Retrieved March 19, 2008, from <http://www.flickr.com/>.
- Guy, Marieke; Tonkin, Emma. (2006). Folksonomies: Tidying up tags? *D-Lib Magazine, January 2006*. Retrieved March 19, 2008, <http://www.dlib.org/dlib/january06/guy/01guy.html>.
- Hayman, Sarah. (2007). Folksonomies and tagging: New developments in social bookmarking. *Ark Group Conference: Developing and Improving Classification Schemes 27-29 June, Rydges World Square, Sydney*. Retrieved March 19, 2008, from <http://www.educationau.edu.au/jahia/webdav/site/myjahiasite/shared/papers/arkhayman.pdf>.
- LibraryThing. (2008). Retrieved March 19, 2008, from <http://www.librarything.com/>.
- Noruzi, Alireza. (2007). Folksonomies: Why do we need controlled vocabulary? Editorial. *Webology, Vol. 4, No. 2*. Retrieved March 19, 2008, from <http://www.webology.ir/2007/v4n2/editorial12.html>.
- Raymond, Matt. (2008). My friend Flickr: A match made in photo heaven. *Library of Congress Blog*. Retrieved March 19, 2008, from <http://www.loc.gov/blog/?p=233>.
- Smith, Tiffany. (2007). Cataloging and you: Measuring the efficacy of a folksonomy for subject analysis . In Lussky, Joan, Eds. *Proceedings 18th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research*, Milwaukee, Wisconsin. Retrieved March 19, 2008, from <http://dlist.sir.arizona.edu/2061/>.
- Spiteri, Louise F. (2007). Structure and form of folksonomy tags: The road to the public library catalogue. *Webology, Volume 4, Number 2, June, 2007*. <http://www.webology.ir/2007/v4n2/a41.html>
- STFC ePublication Archive. (2008). Retrieved March 19, 2008, from <http://epubs.cclrc.ac.uk/>.
- Tonkin, Emma; Corrado, Edward M.; Moulasion, Heather Lea; Kipp, Margaret E. I.; Resmini, Andrea; Pfeiffer, Heather D.; Zhang, Quiping. (2008). Collaborative and social tagging networks. *Ariadne 54*. Retrieved March 19, 2008, from <http://www.ariadne.ac.uk/issue54/tonkin-et-al/>.
- Weller, Katrin. (2007). Folksonomies and ontologies: two new players in indexing and knowledge organization. *Online Information 2007*, 108-115.