

Selecting and Weighting Semantically Discovered Concepts as Social Tags

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, and NKOS-Fueled Semantic Technologies

Proposal Abstract: Research has demonstrated the value of social tags for accessing and marking content on the web. The value of social tags has been described in the literature as both added metadata value and as a location marker or memory aid for the end user. This presentation investigates whether NKOS-fueled semantic technologies can provide as good or better solution than end user suggestions. While there is significant value from social tagging, there are also challenges. Social tags must be integrated into search systems with other values derived from NKOS. This introduces a burden for information managers. The tags are often of questionable quality, and often cannot be integrated into an NKOS application without editing. There is considerable redundancy across social tags suggested by end users. This introduces quality control issues without adding value.

A possible solution to these challenges involves using NKOS-fueled semantic technologies to generate tag clouds of concepts from which users can select and assign social tags. This solution addresses each of the concerns raised above without a loss of functionality for end users. In fact, it provides an opportunity to leverage NKOS-generated and managed concepts for ranking and promoting social tags.

The goal of this research is to propose a new and sustainable approach to social tagging. End users can select tags from a semantically generated cloud of concepts. Selections can be used to rank concepts. Rankings can be added to the search index as they are created by end users. Rankings can be used an important indicators for the importance of concepts for indexing. Also, the burden of integrating social tags into existing NKOS applications is reduced since users are selecting from among acceptable tags.

This presentation describes ongoing research into four questions:

1. whether most social tagging values suggested by end users already exist as high frequency concepts in the referenced item;
2. whether there is redundancy across social tagging values for the same item;
3. whether NKOS-fueled semantic engines can generate concept tags which are as good as or better than users' suggestions;
4. Whether semantically-generated concept tag clouds can be used to facilitate user identification and selection of social tags.

The research is conducted on a data set of 1,500 items. The data set represents 300 socially tagged items, including books, journal articles and press stories in five disciplines, e.g., agriculture, environment, health, education, and transport. The tagged items are derived from Amazon.com, CiteULike, del.icio.us, and Digg.

The research methodology includes: (1) constructing an NKOS-fueled semantic engine for the five disciplines of interest; (2) generating concepts for each of the 1,500 items; (3) evaluating the redundancy and closeness of concepts semantically generated and social tags assigned; (4) generating a tag cloud of semantically generated concepts; and (5) end user testing for suggesting vs. selecting social tag values. The SAS/Teragram Content Categorization Suite semantic technologies are used to generate the NKOS structures, and to also process the individual content items. Redundancy and closeness comparisons are conducted manually by the researcher and a graduate assistant. End user testing of concept selection from tag clouds is scheduled to begin on September 1 and conclude by September 15, 2011.