LCSH is to Thesaurus as Doorbell is to Mammal: Visualizing Structural Problems in the Library of Congress Subject Headings

Simon Spero
UNC Chapel Hill, USA
ses@unc.edu

Keywords: LCSH; Controlled vocabularies; Hierarchical relationships; visualization

The Library of Congress Subject Headings (LCSH) has been developed over the course of more than a century, predating the semantic web by some time. Until the 1986, the only concept-to-concept relationship available was an undifferentiated "See Also" reference, which was used for both associative (RT) and hierarchical (BT/NT) connections. In that year, in preparation for the first release of the headings in machine readable MARC Authorities form, an attempt was made to automatically convert these "See Also" links into the standardized thesaural relations.

Unfortunately, the rule used to determine the type of reference to generate relied on the presence of symmetric links to detect associatively related terms; "See Also" references that were only present in one of the related terms were assumed to be hierarchical. This left the process vulnerable to inconsistent use of references in the pre-conversion data, with a marked bias towards promoting relationships to hierarchical status.

The Library of Congress was aware that the results of the conversion contained many inconsistencies, and intended to validate and correct the results over the course of time. Unfortunately, twenty years later, less than 40% of the converted records have been evaluated.

The converted records, being the earliest encountered during the Library's cataloging activities, represent the most basic concepts within LCSH; errors in the syndetic structure for these records affect far more subordinate concepts than those nearer the periphery. Worse, a policy of patterning new headings after pre-existing ones leads to structural errors arising from the conversion process being replicated in these newer headings, perpetuating and exacerbating the errors.

As the LCSH prepares for its second great conversion, from MARC to SKOS, it is critical to address these structural problems. As part of the work on converting the headings into SKOS, I have experimented with different visualizations of the tangled web of broader terms embedded in LCSH. This poster illustrates several of these renderings, shows how they can help users to judge which relationships might not be correct, and shows just exactly how Doorbells and Mammals are related.

References:

Kiczales, Gregor, Jim Des Rivi`eres, and Daniel Gureasko. (1991). *Bobrow. The art of the metaobject protocol*. Cambridge, Mass.: MIT Press.

Tarjan, Robert. Depth-First Search and Linear Graph Algorithms. (1972) *SIAM Journal on Computing 1*(2), 146-160.