
Contentious Categories: Creating the Category System of Wikipedia

Katherine Thornton

The Information School
Box 352840
Mary Gates Hall, Ste 370
Seattle, WA 98195
thornt@uw.edu

Dissertation Overview

The amount of digital content that is accessible via the Internet is so large that many people and many different systems have control over how it is organized and presented. Diverse communities face similar challenges of how to organize the content they wish to present in order to make it more accessible, and more meaningfully located in relationship to similar, or pertinent, content. These communities range from corporations, to libraries, to communities of users contributing content to a website. As the members of each of these communities work through the challenges of designing and implementing systems of information organization, many articulate the need for models of how information structures should be designed to best meet their needs. As an approach to creating such models I propose to undertake an analysis of information structures that have been discussed, implemented, and tested in the field of knowledge organization as examples of structures designed by experts. I will also discuss models that have been created by large-scale collaborative groups in response to the current needs of communities that seek to organize digital content online.

In order to consider models for information structures for the purpose of information organization specific to digital content, I have identified two aspects which I propose to focus my discussion around: collaborative work and relationship types. I chose these two aspects because of the fact that existing systems for information organization were not designed with these issues in mind, thus they will require the most extensive consideration when proposing new models. The new models I will propose will be informed by the study of existing structures of information organization.

In this dissertation I will examine how web-based category systems that are collaboratively created articulate relationships between concepts. I will take Wikipedia as a case study. I will compare the Wikipedia category structure to structural design requirements in the extant literature of knowledge organization. I have three research questions: a) What does the Wikipedia community see as the purpose of the category system? b) How does the Wikipedia community use the category system? c) Does the category system meet the needs of the community?

The study I propose will make three contributions to interdisciplinary literature. First, by examining the Wikipedia community I will elucidate how one collaboratively-created system of information organization differs from well-understood systems from the library tradition, which were created by smaller committees of executives with editorial power. Second, I will elaborate a framework for the analysis of relationships expressed between concepts in Wikipedia. Third, I will make design recommendations for what to consider when evaluating models of future systems for the organization of digital content online. These

recommendations will emphasize how to manage input from a large number of contributors of different levels of expertise, how to build a model that is robust to the concern of how the system will change over time, and how to label and display information about relationship types to contributors and users so that the system will be easier to build in a consistent manner and will provide more rich contextual mapping of content in relationship to other content.

Biographical Sketch

I am a third year PhD student at the University of Washington. I have advanced to candidacy, and I plan to defend my dissertation proposal in October, 2012. I earned my MSIS at the University of Washington in 2011. I also earned my MA in Chinese Linguistics at the University of Washington in 2008.

Advisor Information

My advisor is Dr. Allyson Carlyle. Dr. Carlyle's primary research areas are evaluation and use of online catalogs, and foundations of descriptive cataloging. She has also written about conceptual modeling of documents, and plans future work in that area. Her contact information is: Allyson Carlyle (acarlyle@uw.edu).

CV

EDUCATION

PhD, Information Science (expected 2013)

University of Washington

Advisor: Allyson Carlyle

Research Interests: Classification Theory,
Computer-Supported Cooperative Work

MS, Information Science December, 2011
University of Washington

MA, Asian Languages and Literature (Chinese Linguistics) August, 2008
University of Washington
Thesis: Early Chinese Loans in Vietnamese
Advisor: Zev Handel

BA, summa cum laude, International Studies (China) December, 2002
University of North Carolina at Chapel Hill

ACADEMIC PAPERS

Huang, Jeff, Thornton, Katherine & Efthimiadis, Efthimis. Conversational Tagging in Twitter. In: *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia HT'10*, June 13–16, 2010, Toronto, Ontario, Canada, 5 pages.

Thornton, Katherine. "Contentious Categories: Designing the Category System in Wikipedia". In: *Proceedings of the North American Symposium of Knowledge Organization*, June 16-17, Toronto, Ontario, Canada, 10 pages.

Tennis, Joseph T. , Thornton, Katherine and Andrew Filer. "Some Temporal Aspects of Indexing and Classification: Toward a Metrics for Measuring Scheme Change".

PROFESSIONAL PRESENTATIONS

June 2004 "Tonogenesis As Seen Through Rhyme: Tonal Irregularity in the *Jiu Zhang*"

paper presented at 12th Annual Conference of the International Association for Chinese Linguistics, Tianjin, P.R.C. **Nominated for the IACL Young Scholar Award.**

2007 "Etymological Doublets in Sino-Vietnamese" paper presented at the 9th Annual Southeast Asian Studies Graduate Student Conference, Cornell University.

January 2009 "The Contact Relationship Between Chinese and Vietnamese" paper presented at the UCLA – UC Berkeley Joint Conference on Southeast Asian Studies.

June 2011 "Contentious Categories: Designing the Category System in Wikipedia". In: *Proceedings of the North American Symposium of Knowledge Organization*, Ryerson University, Toronto, Canada.

TEACHING EXPERIENCE

Co-Teaching, University of Washington
Archive and Manuscript Services 1 quarter, 30 graduate students
Applied Research Seminar (Category System of Wikipedia) 1 quarter, 10 students
Research Methods 1 quarter, 30 students

Teaching Assistant, University of Washington
Information Services and Resources 1 quarter, 90 graduate students
Information Behavior 3 quarters, 80 graduate students per quarter
Management for Information Organization, 1 quarter, 40 graduate students
Information Organization, 2 quarters, 80 graduate students per quarter

Co-Director, University of Washington (Netherlands)
Exploration Seminar: Dutch Designs, 1 month, 30 students

SERVICE ACTIVITIES

2011-2012 President of the Information School
Doctoral Student Association, UW Seattle

2011-2012 Graduate and Professional Student Library
Advisory Committee, UW Seattle

2006- 2010 graduate student representative to the
Faculty Council on University
Libraries, UW Seattle

2008- 2009 graduate student representative to the
Scholarly Communications
Committee, UW Seattle

2006-2007 graduate student representative to the
President's Advisory Council on
Women, UW Seattle

HONORS AND AWARDS

Vietnamese Advanced Studies Institute Fellowship 2006

Foreign Languages Area Studies Fellowship 2005-6

UW Library Student Assistant Scholarship 2005

Chester Fritz Scholarship Jackson School of
International Studies 2004-2005

Nominated for the Young Scholar Award by the
International Association for Chinese
Linguistics 2003

Awarded Highest Honors on undergraduate thesis, 2002

