

Meaningful Concept Displays: The First Step

Xia Lin

College of Information
Science and Technology
Drexel University
Philadelphia, PA
U.S.A.
linx@drexel.edu

Dagobert Soergel

Department of Library
and Information Studies
University at Buffalo
Buffalo, NY
U.S.A.
dsoergel@buffalo.edu

Murtha Baca

The Getty Research
Institute
Los Angeles, CA
U.S.A.
MBaca@gatty.edu

In this presentation, we will discuss our on-going research project on designing, implementing, and evaluating ways to improve users' experience with libraries and museums through Meaningful Concept Displays (MCD). The project is aimed at supporting "high-end" searches for topics and problems that require thorough exploration and high-recall results.

With the recent developments of Terminology Services and Linked Open Data that make available many high-quality Knowledge Organization Systems (KOS), there is a great potential for significant increase in the adoption and practical use of KOS in the digital environment. What is still missing is how these KOS can be connected to digital resources and how users can interact with the rich knowledge embedded in KOS. We believe that an essential component to promote and spread the use of KOS in the digital environment is innovative KOS interfaces. The goal of our project thus is to develop a new framework of "Meaningful Concept Displays" (MCD) on which new KOS interfaces can be developed for specific KOS and digital resources. The MCD interfaces will take advantage of the underlying KOS structures and present meaningful views to the user.

We will build a pilot MCD appliance that will sit between the user and Web search engines, digital library resources, museum catalogs, etc. to support enhanced searching by users; it will also be usable as an API (application programming interface) for search engines to access data from multiple KOS. It will provide integrated access to multiple KOS, both from its own database and from the Web at large (including linked open data). For the user, the appliance will function as a meta search engine that accepts user queries and a list of systems to be searched, sends improved queries to the various systems (including improved free-text queries to Google, Bing, and other Web search engines), and return results. The user can either choose automatic query enhancement (either fully transparent or with information about the query sent) or fully engage with KOS conceptual structure (through concept maps, hierarchical displays, etc.) to explore the topic and formulate one or more queries.

The MCD appliance aims to significantly enhance users' searching, browsing, and learning experience. To achieve this the appliance will:

- Map users' queries to KOS terms and allow users to modify their queries interactively and visually. This will help to clarify users' information needs and transform the search experience and success of a wide range of users.
- Visualize users' query terms in the context of KOS structures and digital resource mapping. The emphasis of this is to create visual concept displays that are meaningful to users and easy to interact with, and that will evolve and improve as more people interact with them.
- Send enhanced queries to the search engines and other digital resources of the users' choice and return results, at the users' option arranged by ad-hoc clusters or a classification of the users' choice.
- Support integration of KOS and digital resources and make it easy for users to interact with both the KOS and the resources even when the resources are not fully indexed by the KOS. This will be achieved through content mapping both at the resource level and at the subject domain level.

This is a three-year research project funded by IMLS. Partners of the project include the Getty Research Institute (GRI), ARTstor, and the Indianapolis Museum of Art (IMA), The University at Buffalo, and Drexel University. The first sets of KOS we have included in our database are the *Getty's Art & Architecture Thesaurus®* (AAT), *Union List of Artist Names®* (ULAN), and *Getty Thesaurus of Geographic Names®* (TGN). Sample datasets are also collected from both ARTstor and IMA. What we will present at this workshop is our first year's activities and some preliminary MCD designs, prototypes, and results. Some of the main discussion topics include:

- The idea and concept of MCD and its relevant literature and related projects.
- The implementation of a unified database for multiple KOS.
- The use of Wikipedia as a knowledge source to map free-text queries to KOS terms.
- The initial design and implementation of the KOS-based visualization interfaces.

Through the presentation, we hope to interest workshop participants in the concept of MCD and extend the discussion of how to create effective KOS interfaces and enhanced searching services for digital resources.