Maximizing the Usage of Value Vocabularies in the Linked Data Ecosystem:

Meaningful Concept Displays for KOS-based Searching and Browsing

Xia Lin
The College of Computing and Informatics
Drexel University
Philadelphia, Pennsylvania, USA
Use of KOS for Searching and Browsing

• Challenges in the Linked Data environment
  – More KOS vocabularies are available; but they are all in different formats through different APIs.
  – More resources/documents are searchable, but they are not indexed by the specified KOS.
  – There are few practical solutions to map concepts from one KOS to another or from KOS to natural languages.
  – Users demand easy-to-use interfaces and simple search logic.
Use of KOS for Searching and Browsing

• A Proposed solution – MCD Appliance
  – Meaningful Concept Displays (MCD) that can be adapted to different KOS and different collections/digital libraries
    • A unified vocabulary data structures for multiple KOS
    • KOS Web Services and APIs with concept mapping
    • KOS-based query expansion algorithms and interfaces
    • Visual and interactive concept displays
The Original Design

The MCD Appliance Architecture

Web Services/API

Extendable SKOS-XL Data Model

Meaningful Concept Displays

KOS-based Concept Map Interface

- Drag-and-drop editing
- Concept maps

Unified KOS concept displays

- Detail text Displays
- Semantic Link Displays
- Semantic Map displays
- Hierarchical Displays

KOS-enhanced Search Interface

- Multi-displayed search interface
- KOS-enhanced tagging interface

Search engines and resources

Other KOS/Terminology Services

KOS Data Acquisition

User term spaces and KOS vocabulary spaces mapping

Search log terms
Tagging Terms
Text of documents
KOS Mapping data set
Statistic al term Associations
Published KOS
Concept Maps
Linked Data

The Original Design
The Challenges

• Data structures: Converting KOS relational databases to unified data triples
• Concept mapping: Improving matching rates between two KOS terms.
• Displays: Applying visualization techniques for concept displays
Data Structures

• From relational databases to data triples
  – Three different versions of databases for Getty vocabularies were implemented and compared
    • A modified ISO 25964-1 database schema
      – Relatively “flat”; easy to query
    • A unified KOS (UMLS-style) database
      – Emphasize the difference of concepts, terms, and strings
    • A triple-based database
      – All the relationships are defined in a triple form
        » Entity 1 – has-relationship – entity 2
      – Easy to convert to RDF
Data Structures

• Everything becomes either an entity or a relationship:
  – concept1 → has-preferred-label – term1
  – Concept1 → has-parent – concept2
  – Concept1 → has-concept-type – term2
  – Concept1 → has-descriptive-note – longString1
  – Concept1 → is-derived-from – term3
  – Concept1 → is-in-facet – term4
  – Concept1 → is-a-component-of – term5
Concept Mapping

• **Goal:** To map from one KOS to another KOS
  – Exact matches (based on string mapping after normalization) success only about 15%
  – Partial matches increase the matching rate to 60 to 80%, but false drop is increased significantly as well.
  – Additional information is needed to improve the matching rate
    • Semantic structures
    • Syntactic components
Concept Mapping

• Experiment:
  – to map Getty AAT terms in the Materials facet to (free text) indexing terms for “Materials” of an ARTStor images collection.
  – Procedures
    • Normalize string for exact matches
    • Identify sub components if any
      – (i.e., “oil on canvas” → “oil AND canvas”)
    • Identify Material sub-facets:
      – “color” (so they can be temporarily removed for the mapping purpose, i.e., “blue pastel on white canvas” -> “pastel on canvas”)
      – “surface” (canvas, wood, glass, etc.)
      – “technique” (drawing, engraving, sketching, etc.)
      – “coating” (paint, ink, pencil, etc.)
Concept Mapping

AAT Material Terms
- Search
- Find in AAT

ARTStor Material Terms
- Search
- Find in ARTStor

Click an AAT concept to search ARTStor
- balk
- poles
- bondstone
- perpend
- damp course
- construction joints
- scarf joints
- interlocking joints
- butterfly joints
- masonry joints
- bed joints
- blind joints
- concave joints
- rough-cut joints
- raked joints
- shoved joints
- stripped joints
- struck joints
- metal joints
- rigid joints
- macadamized roads
- inorganic material

Search result: raked joints. Click an ARTStor material term to search AAT using
- Sørl
- Jaccard

No result found: raked joints
Meaningful Concept Displays

• Meaningful concept displays (MCD)
  – Mapping and simplifying concept structures or relationships based on KOS and content collections.
  – Conveying the concept structures visually to help users understand and learn
  – Providing useful functions to help users complete their information tasks
    • Searching, browsing, exploring, tagging, and learning
MCD and Visualization

• Meaningful
  – The picture conveys semantic structures or relationships that the viewer can understand

• Trustful
  – The structures and relationships on the picture match the semantic structures of the underlying data.

• Useful
  – What users get from the picture will help them do something useful.
Three approaches:

– Visualizing existing structures or relationships
  • in KOS or link structures

– Visualizing learned structures or relationships
  • through machine learning
  • through linked data

– Visualizing structures through visual metaphors
  • to serve a purpose better
  • to make it easy to understand
MCD for Visualizing Existing Structures
MCD for Visualizing Learned Structures

--Interactive Concept Map of “aging”
MCD by Metaphors

Visual Query Expansion

-- A ripple style query expansion
MCD by metaphors
Summary

• To creative effective and useful MCD, we have made progress on three major areas:
  – Developing a triple-based database structure for integrating multiple KOS.
  – Developing new procedures and algorithms for concept mapping.
  – Designing and implementing effective styles of displays and interaction for concept-based searching, browsing, and learning.
Acknowledgments

• Thanks to IMLS for funding this project.
• The work presented here is done by the whole project team:
  – Dagobert Soergel, Univ. of Buffalo
  – William Ying, ARTstor
  – Murtha Baca, Getty Research
  – Jae-wook Ahn, Data Scientist, Drexel
  – Mi Zhang, Mike Zarro, Haozhen Zhao, Drexel Ph.D. students
DDC Visualization

-- Visualizing DDC classes based on document relationships

Dewey Digital Universe