Terminology Services

Diane Vizine-Goetz
Senior Research Scientist
OCLC Research
A version of this presentation was given at:

- **New Dimensions in Knowledge Organization Systems: A Joint NKOS/CENDI Workshop**
  The World Bank
  Washington, DC
  11 September 2008

- **OCLC / ISKO-NA Preconference**
  Université de Montréal
  Montréal, Canada
  5 August 2008
Moving Vocabularies to the Network Level

Requirements:

- Expressive data structures
- Vocabularies encoded for the Web
- Access mechanisms for search and retrieval
- URI accessible content
- Use of open protocols and standards
Success will be measured by the appearance of applications that use or combine vocabulary data to create new derivative works or tools.
OCLC Terminology Services Prototype

Employs library and Web standards to make the terms and relationships in controlled vocabularies available as Web resources.
Top-level site intended for machines
Human interface for everyone else

Terminology Services
Experimental Services for Controlled Vocabulary

Highlights

- Search descriptions of controlled vocabularies
- Search for concepts/headings in a controlled vocabulary
- Retrieve a single concept/heading by its identifier
- View relationships for a concept/heading including equivalence, hierarchical, and associative
- Retrieve concepts/headings in multiple representations including HTML, MARC XML, SKOS, and Zthes.
- Search using SRU CQL syntax

Technical Resources

- Overview
- URI-patterns
- Indexing
- Standards
- Issues
- Roadmap

Vocabulary Resources

Jump to: meta fast gsafo icshac icshn mean ictgn gmnpcl

Name: Controlled Vocabulary Metadata
Description: Descriptions of controlled vocabularies
Date: 2006-03
Identifier: meta
Links: About | SRU Interface | Examples | MARC Statistics
Applications

Query Expansion

Searching Heterogeneous Collections

Metadata Creation

Vocabularies

fast

gsafd

lcsh

mesh

lctgm & gmgpc

Content

Web Services
Content

Vocabularies

- fast
- gsafd
- lcsh
- mesh
- lctgm & gmgpc

Web Services

Applications

- Query Expansion
- Searching Heterogeneous Collections
- Metadata Creation
Types of Controlled Vocabularies (Hodge 2000)
* indicates availability in Terminology Services Prototype

Term Lists
- Authority Files*
- Glossaries
- Dictionaries
- Gazetteers

Classifications and Categories
- Subject Headings*
- Classification Schemes*
- Taxonomies
- Categorization Schemes

Relationship Lists
- Thesauri*
- Semantic Networks
- Ontologies
1. Faceted Application of Subject Terminology (fast)
2. Form and Genre Terms for Fiction and Drama (gsafd)
3. Library of Congress Subject Headings (lcsh)
4. Medical Subject Headings (mesh)
5. Thesaurus for Graphic Materials: TGM I, Subject Terms (lctgm)
6. Thesaurus for Graphic Materials: TGM II, Genre and Physical Characteristics (gmgpc)
Data Structures for Controlled Vocabularies

MARC 21 Format for Authority Data

- A format for the use and exchange of information about the authorized forms of names and subjects used as access points in MARC bibliographic records.

Simple Knowledge Organization System (SKOS)

- SKOS Core is a model and an RDF vocabulary for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, 'folksonomies', [etc.]

Zthes

- A model for representing thesauri* and a specification for expressing them in XML. Zthes also provides specifications for searching Zthes compliant data using SRU/SRW or Z39.50.
Encoding Mechanisms

XML (Extensible Markup Language)
- A data-interchange format for custom markup languages.

RDF (Resource Description Framework)
- A data-interchange format for the representation of graph models.

JSON (Javascript Object Notation)
- A data-interchange format based on a subset of the JavaScript Programming Language defined by the ECMA-262 3rd Edition standard.
Access Mechanisms

REST (Representational State Transfer)
• A software architecture style used for building distributed systems that retrieve Web resources.

SRU & SRW (Search/Retrieve via URL)
• A standard search protocol that utilizes the Contextual Query Language (CQL) syntax to retrieve Web resources.
Applications

- Query Expansion
- Searching Heterogeneous Collections
- Metadata Creation

Vocabularies
- fast
- gsafd
- lcsh
- mesh
- lctgm & gmgpc
{URL for the service}/{vocabulary}/{identifier}.{format}

- http://tspilot.oclc.org/lcsh/sh95000541.html
- http://tspilot.oclc.org/lcsh/sh95000541.json
- http://tspilot.oclc.org/lcsh/sh95000541.marcxml
- http://tspilot.oclc.org/lcsh/sh95000541.skos
- http://tspilot.oclc.org/lcsh/sh95000541.zthes
The template shows how to link to vocabulary data in the prototype

{URL for the service}/{vocabulary}/{identifier}.{format}

1. Base URL for the service - http://tspilot.oclc.org

2. Vocabulary - the code for the controlled vocabulary in the MARC code list for terms
   - http://www.loc.gov/marc/relators/relasour.html#rela6xx
   - http://www.loc.gov/marc/relators/relasour.html#rela655

3. Identifier - a control number associated with a concept or term (e.g., an LCCN - Library of Congress Control Number)

4. Format - the representation of the vocabulary data
In the Queue

- More vocabularies
- Access to complete term hierarchies
- MADS profile
- Additional mappings
Terminology Services Prototype

- [http://tspilot.oclc.org/resources/](http://tspilot.oclc.org/resources/)
- [http://tspilot.oclc.org](http://tspilot.oclc.org) (machine interface)

Project page