Steps towards a Culture Web

Tumbling Walls
&
Building Bridges
Interoperability: tearing down the walls between collections

- Musea have increasingly nice websites
- But: most of them are driven by stand-alone collection databases
- Data is isolated, both syntactically and semantically
- If users can do cross-collection search, the individual collections become more valuable!
The Web: “open” documents and links
The Semantic Web: “open” data and links

Painting
“Green Stripe (Mme Matisse)”
Royal Museum of Fine Arts, Copenhagen

Painter
“Henri Matisse”
Getty ULAN

URL
Web link
creator
Dublin Core

URL
Matisse, Henri (French)

Roles:
- artist (preferred)
- painter
- printmaker
- sculptor
- designer
- writer

Gender: male

Birth and Death Places:
- Born: Le Cateau-Cambrésis (Nord, Nord-Pas-de-Calais, France) (inhabited place)
- Died: Nice (Alpes-Maritimes, Provence-Alpes-Côte d'Azur, France) (inhabited place)

Related People or Corporate Bodies:
- apprentice was: Jolin, Einar 1911-1913
- parent of: Duthuit, Marguerite Matisse
- patron was: Barnes, Dr. Albert C.
- student of: Corom, Fernand
- student of: Moreau, Gustave

Nationality: French
Levels of interoperability

• Syntactic interoperability
  – using data formats that you can share
  – XML family is the preferred option

• Semantic interoperability
  – How to share meaning / concepts
  – Technology for finding and representing semantic links
Simple Knowledge Organisation System (SKOS)

SKOS Core | SKOS Mapping | SKOS Extensions

This page: Specifications | RDF Vocabularies | Development


SKOS is an area of work developing specifications and standards to support the use of knowledge organisation systems (KOS) such as thesauri, classification schemes, subject heading lists, taxonomies, other types of controlled vocabulary, and perhaps also terminologies and glossaries, within the framework of the Semantic Web.

There are three RDF vocabularies under active development: SKOS Core | SKOS Mapping | SKOS Extensions. There is also the SKOS API, a web service API for interacting with a KOS datasource.

SKOS Specification Development

The following specifications are under development within the W3C Semantic Web Best Practices and Deployment Working Group:

- **SKOS Core Guide**

  This document is a guide using the SKOS Core Vocabulary, for readers who already have a basic understanding of RDF concepts. It is the authoritative guide to recommended usage of the SKOS Core Vocabulary at the time of publication.

- **SKOS Core Vocabulary Specification**

  This document gives a reference-style overview of the SKOS Core Vocabulary as it stands at the time of publication. It is the authoritative human-readable account of the SKOS Core Vocabulary at the time of publication. It also describes the policies for ownership, naming, persistence and change by which the SKOS Core Vocabulary is managed.
Multi-lingual labels for concepts

ex:shrubs

prefix ex: <http://www.example.com/concepts#>
prefix skos: <http://www.w3.org/2004/02/skos/core#>
Important Update Regarding the XML format of the NASA Taxonomy - Jan 9, 2007

The next version of the NASA taxonomy will be in the SKOS format.

The SKOS Core is a model and an RDF vocabulary proposed by the W3C for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, other types of controlled vocabulary.

The SKOS Core Vocabulary is an application of the Resource Description Framework (RDF), that can be used to express a
Principle 1: semantic annotation

- Description of web objects with "concepts" from a shared vocabulary

<table>
<thead>
<tr>
<th>Description:</th>
</tr>
</thead>
<tbody>
<tr>
<td>recordnumber: 23727;</td>
</tr>
<tr>
<td>timestamp: 2001-08-22;</td>
</tr>
<tr>
<td>type: cultural;original;</td>
</tr>
<tr>
<td>collector: Johannes Frederik van Oo Zeldzaamheden; Utgawa</td>
</tr>
<tr>
<td>series: 360;</td>
</tr>
<tr>
<td>Culture: Japans;</td>
</tr>
<tr>
<td>Date: 1800-1829;</td>
</tr>
<tr>
<td>Description: 1893 JAPAN aankoop;</td>
</tr>
<tr>
<td>Identifier in Current Repository: 360-4664;</td>
</tr>
<tr>
<td>Creation Site: Japan;</td>
</tr>
<tr>
<td>Current Repository: RMV;</td>
</tr>
<tr>
<td>Material: papier;</td>
</tr>
<tr>
<td>Measurements: format: oban, 25.5 cm x 37.5 cm; Edo;</td>
</tr>
<tr>
<td>Style/Period/Period: Edo junisho;</td>
</tr>
<tr>
<td>Title: prenten;</td>
</tr>
<tr>
<td>Type: Work;</td>
</tr>
</tbody>
</table>
Principle 2: semantic search

- Search for objects which are linked via concepts (semantic link)
- Use the type of semantic link to provide meaningful presentation of the search results

Query “Paris”

Paris

PartOf

Montmartre
Term disambiguation is key issue in semantic search

• Post-query
  – Sort search results based on different meanings of the search term
  – Mimics Google-type search

• Pre-query
  – Ask user to disambiguate by displaying list of possible meanings
  – Interface is more complex, but more search functionality can be offered
Principle 3: vocabulary alignment

“Tokugawa”

AAT style/period
Edo (Japanese period)
Tokugawa

AAT is Getty’s
Art & Architecture Thesaurus

SVCN period
Edo

SVCN is local in-house ethnology thesaurus
The myth of a unified vocabulary

- In large virtual collections there are always multiple vocabularies
  - In multiple languages
- Every vocabulary has its own perspective
  - You can’t just merge them
- But you can use vocabularies jointly by defining a limited set of links
  - “Vocabulary alignment”
- It is surprising what you can do with just a few links
Learning alignments

• Learning relations between art styles in AAT and artists in ULAN through NLP of art historic texts – “Who are Impressionist painters?”

<table>
<thead>
<tr>
<th>Artist Name</th>
<th>IS</th>
<th>In GS</th>
</tr>
</thead>
<tbody>
<tr>
<td>edgar degas</td>
<td>0.0699</td>
<td>1</td>
</tr>
<tr>
<td>edouard manet</td>
<td>0.0548</td>
<td>1</td>
</tr>
<tr>
<td>pierre-auguste renoir</td>
<td>0.0539</td>
<td>1</td>
</tr>
<tr>
<td>morisot, berthe</td>
<td>0.0393</td>
<td>1</td>
</tr>
<tr>
<td>gogh, vincent van</td>
<td>0.0337</td>
<td>0</td>
</tr>
<tr>
<td>cassatt, mary</td>
<td>0.0318</td>
<td>1</td>
</tr>
<tr>
<td>cezanne, paul</td>
<td>0.0302</td>
<td>1</td>
</tr>
</tbody>
</table>
From metadata to semantic metadata

- Thesaurus schema mapping
- Meta-Data schema mapping
- Meta-Data mapping
- Thesaurus alignment
Example textual annotation

<inm:Record>
  <inm:NUMMER>6</inm:NUMMER>
  <inm:TITEL>Delftse Bijbel...</inm:TITEL>
  <inm:TITEL_EN>Delft Bible...</inm:TITEL_EN>
  <inm:MAKER>Yemantszoon, Mauricius : d</inm:MAKER>
  <inm:OBJECT>tekstbladzijde</inm:OBJECT>
  <inm:TECHNIEK>boekdruk</inm:TECHNIEK>
  <inm:DATERING>10 jan. 1477</inm:DATERING>
  <inm:CLASSIFICATIE>D</inm:CLASSIFICATIE>
  <inm:ORIGINEEL>Bijbel. Oude Testament...</inm:ORIGINEEL>
</inm:REPRODUCTIE>
  <inm:TWAAM/>
  <inm:TWOND>typografische vormgeving</inm:TWOND>
  <inm:TWOND>bijbels</inm:TWOND>
  <inm:TWGED>Delft</inm:TWGED>
  <inm:OMSCHRIJVING>Eerste bijbel die in het Nederlands verscheen...</inm:OMSCHRIJVING>
  <inm:OMSCHRIJVING_EN>The first Bible to appear in the Dutch language...</inm:OMSCHRIJVING>
  <inm:AFMETINGEN>27 x 20 cm</inm:AFMETINGEN>
  ...
</inm:Record>
Resulting semantic annotation (rendered as HTML with RDFa)

Description:

classificatie: Geschiedenis van de boekdrukkunst;
drukker: Meer, Jacob Jacobszoon van der, Yemantszoon, Mauricius;
origineel: Bijbel. Oude Testament - Delft: Jacob Jacobszoon van der Meer en Mauricius Yemantszoon, 10 jan. 1477, dl. 2, p. 1;
Date: 10 jan. 1477;
Description: The first Bible to appear in the Dutch language, known as the Delft Bible. It consists of the Old Testament only and is an anonymous adaptation of the - again anonymous - History Bible of 1380. It is an example of an incunabulum where the hand-written book still served as an example for lay-out and design. Contrary to many other incunabula, the place of origin, the names of the printers and even the day of its completion are mentioned in the colophon.;

Measurements. Dimensions: 27 x 20 cm;
rights.copyright: Den Haag Koninklijke Bibliotheek;
Source: Bibliopolis;
Subject: bibles, incunabula; initials, ornamental borders, rubrications; typographical design;
subject.geographicPlace: Delft;
Technique: letterpress printing;
Title: Delft Bible, printed in Delft by Jacob Jacobszoon van der Meer and Mauricius Yemantszoon, 1477;
Type: tekstbladzijde;
type: Work;

Used as value to describe other resources:
BBB_168E56_1477_P1.JPG; relation.depicts
Perspectives

• Basic Semantic Web technology is ready for deployment

• Web 2.0 facilities fit well:
  – Involving community experts in annotation
  – Personalization, myArt

• Social barriers have to be overcome!
  – "open door" policy
  – Involvement of general public => issues of "quality"
Caveats for museum software

• Be wary of Flash
  – Accessibility

• Make sure you can connect others and other can connect to you
  – “Don’t buy software which does not support standard open API’s”

• Export facilities to common formats (XML, ...)

http://e-culture.multimedian.nl

- Part of the Dutch knowledge-economy project MultimediaN
- Partners: VU, CWI, UvA, DEN, ICN
- People:
  Alia Amin, Lora Aroyo, Mark van Assem, Victor de Boer, Lynda Hardman, Michiel Hildebrand, Laura Hollink, Marco de Niet, Borys Omelayenko, Marie-France van Orsouw, Jacco van Ossenbruggen, Guus Schreiber Jos Taekema, Annemiek Teesing, Anna Tordai, Jan Wielemaker, Bob Wielinga
- Artchive.com, Rijksmuseum Amsterdam, Dutch ethnology musea (Amsterdam, Leiden), National Library (Bibliopolis)