Vocabulary management, SKOS, and the semantic web

Bob DuCharme

2011 Semantic Web Technology Conference
June 7, 2011
Introductions

- Presentation and all its URLs: http://www.snee.com/semtech/2011
- Me: Solutions Architect at TopQuadrant; formerly XML, SGML guy at RIA, Moody’s, LexisNexis, Innodata Isogen
- Weblog: http://www.snee.com/bobdc.blog
SKOS: the W3C’s OWL ontology for creating thesauruses, taxonomies, and controlled vocabularies.
Terminology

- Controlled vocabulary
- Taxonomy
- Thesaurus
- Ontology
### Controlled vocabularies

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mr.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ms.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mrs.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miss</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>State</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AL</td>
<td></td>
</tr>
<tr>
<td>AK</td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td></td>
</tr>
<tr>
<td>CA</td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td></td>
</tr>
<tr>
<td>CT</td>
<td></td>
</tr>
<tr>
<td>WI</td>
<td></td>
</tr>
<tr>
<td>WY</td>
<td></td>
</tr>
</tbody>
</table>
Terminology

- Controlled vocabulary
- Taxonomy
- Thesaurus
- Ontology
Taxonomies

Taxonomies: business value
Taxonomies: business value
Terminology

- Controlled vocabulary
- Taxonomy
- Thesaurus
- Ontology
Thesaurus

Mammal

- Horse
- Cat
- Dog
  - Bulldog
  - Collie

Related term

(use for: mutt, cur)

Building

- Residential
  - Doghouse
- Commercial
  - House
Terminology

- Controlled vocabulary
- Taxonomy
- Thesaurus
- Ontology
Ontology

- Define terms and structures
- And relationships: Requires, contraindicates...
- Classes of things to keep track of, potential properties they might have
- OWL: W3C RDF-based standard for specifying ontologies
Simple Knowledge Organization System

- Controlled vocabulary
- Taxonomy
- Thesaurus
- Ontology

SKOS: the W3C’s OWL ontology for creating thesauri, taxonomies, and controlled vocabularies.
Managing terms?

- mammal
- dog
- perro
- chien
- mutt
- chucho
- gato
- cat
- bulldog
- dachshund
No, managing concepts

preferred label (English): “mammal”

preferred label (English): “cat”
Preferred label (Spanish): “gato”

preferred label (English): "dog"
preferred label (Spanish): "perro"
preferred label (French): "chien"
alternative label (English): "mutt"
alternative label (Spanish): "chucho"

preferred label (English): “bulldog”

preferred label (English): “dachshund”
More metadata for concepts

- Standardized
  - Definition
  - Related terms
  - History note
  - More…

- Unstandardized
  - Whatever you want!
http://myCompany.com/animals/c43209101

preferred label (English): "dog"
preferred label (Spanish): "perro"
preferred label (French): "chien"

alternative label (English): "mutt"
alternative label (Spanish): "chucho"

history note: "Edited by Jack on 5/4/11"

related term: http://myCompany.com/shelters/c3048293

product: http://myCompany.com/vaccines/c2197503
Who is using SKOS?

- AGROVOC
- New York Times: People, Organizations, Locations, Subject Descriptors
- Library of Congress subject headers
- AGFA drug admin. forms
- NASA: many categories
Other taxonomy standards


- **ISO 2788:1986** Guidelines for the Establishment and Development of Monolingual Thesauri

- **ISO 5964:1985** Guidelines for the Establishment and Development of Multilingual Thesauri

- **ISO 25964** (combines 2788 and 5964)

- **Zthes**

- **MADS/RDF** (“Metadata Authority Description Schema in RDF”) Library of Congress
http://myCompany.com/animals/c43209101

preferred label (English): "dog"
preferred label (Spanish): "perro"
preferred label (French): "chien"
alternative label (English): "mutt"
alternative label (Spanish): "chucho"
history note: "Edited by Jack on 5/4/11"
related term: http://myCompany.com/shelters/c3048293
product: http://myCompany.com/vaccines/c2197503
SKOS-XL: SKOS Extension for Labels

http://myCompany.com/animals/c43209101
preferred label: http://myCompany.com/animals/c4329201
preferred label: http://myCompany.com/animals/c4329202
alternative label: http://myCompany.com/animals/c4329205
history note: "Edited by Jack on 5/4/11"
related term: http://myCompany.com/shelters/c3048293
product: http://myCompany.com/vaccinations/c2197503

http://myCompany.com/animals/c4329201
literal form: "dog" (English)

http://myCompany.com/animals/c4329205
literal form: "chucho" (Spanish)
last edited: "2011-05-13"
editor: "Mary Jones"
<table>
<thead>
<tr>
<th>Label</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preferred label (English): &quot;dog&quot;</td>
<td></td>
</tr>
<tr>
<td>Preferred label (Spanish): &quot;perro&quot;</td>
<td></td>
</tr>
<tr>
<td>Preferred label (French): &quot;chien&quot;</td>
<td></td>
</tr>
<tr>
<td>Alternative label (English): &quot;mutt&quot;</td>
<td></td>
</tr>
<tr>
<td>Alternative label (Spanish): &quot;chucho&quot;</td>
<td></td>
</tr>
</tbody>
</table>

History note: "Edited by Jack on 5/4/11"

Related term: [http://myCompany.com/shelters/c3048293](http://myCompany.com/shelters/c3048293)

Product: [http://myCompany.com/vaccinations/c2197503](http://myCompany.com/vaccinations/c2197503)

Foo code: “5L-MN1-003”
Networked vocabularies: why

Centralized master vocabulary system?

Ad hoc, distributed, disconnected
When I use a word, it means just what I choose it to mean — neither more nor less.

- Humpty Dumpty in Lewis Carroll’s *Through the Looking Glass*
Enterprise Vocabulary Management: dream vs. reality

“Customer”

Repair Department

“Customer”

Marketing Department
Enterprise Vocabulary Management: dream vs. reality

“Customer”  →  ?  ←  “Customer”

Repair Department  Marketing Department
Networking distributed vocabularies

http://myCompany.com/vocab/repairs/c8927432
  preferred label: "customer"
  broader term: http://myCompany.com/vocab/repairs/c93402934

http://myCompany.com/vocab/marketing/c93402934
  preferred label: “customer”
Networking distributed vocabularies

http://myCompany.com/vocab/repairs/c8927432
preferred label: "customer"
broadeter term: http://myCompany.com/vocab/repairs/c93402934

http://myCompany.com/vocab/marketing/c93402934
preferred label: “customer”

Repairs department and marketing department vocabularies can be stored in the same datastore on the same server… or different datastores on different servers stored with different products.
SKOS tools

- Top Quadrant's Enterprise Vocabulary Net (EVN)
- PoolParty
- SKOSed Protégé plugin
- iQvoc
- TemaTres
Other semweb tech to use

- Data conversion tools
- Ontology editors
- SPARQL
  - Query, update, transform
  - SPARQL Rules
- Linked Data
- OWL
SPARQL Rules and AGROVOC

- SKOS spec lists six rules not implemented by ontology
  - e.g. same term can’t be preferred and alternative term in the same language
- Expressed with SPARQL Rules
- Violated by AGROVOC over 1600 times
- e.g. Slovak word for “Buds” (“púèiky”)
Getting started

- Look at existing SKOS thesauri
- Small pilot project…
- …or projects.
- Existing thesaurus editor vendor SKOS support?
- Free tools, commercial tools, trial software
Thank you

- Questions?
- bducharme@topquadrant.com