# Building a More Meaningful Web: From Traditional Knowledge Organization Systems to New Semantic Tools

Dagobert Soergel
College of Information Studies
University of Maryland

The **6th Networked Knowledge Organization Systems/Sources (NKOS) Workshop**, organized by Gail Hodge (gailhodge@aol.com), Dagobert Soergel (dsoergel@umd.edu), and Marcia Zeng (mzeng@kent.edu) was held on May 31, 2003 in Houston, TX as part of JCDL 2003. It focused on

- transforming traditional knowledge organization systems (KOSs), such as classification schemes and thesauri, to new forms of knowledge representation, such as ontologies and topic maps, that can be used for AI and semantic Web applications and for sophisticated retrieval and learning, thus
- Leveraging the considerable intellectual capital available in existing KOS for lowering the cost of developing knowledge-intensive applications

The following papers were presented. (http://nkos.slis.kent.edu)

**From legacy knowledge organization systems to full-fledged ontologies**
Dagobert Soergel, U. of MD
**Reengineering AGROVOC to Ontologies. Towards better semantic structure**
F. Fisseha, A. Liang, J. Keizer, FAO
**From semantic networks, to ontologies, and concept maps: knowledge tools in digital libraries**.
M. A. Gonçalves, Digital Library Res. Lab., VATech
**Using the NASA Thesaurus to Support the Indexing of Streaming Media**
Gail Hodge, Janet Ormes, Patrick Healey, NASA Goddard
**Concept-based Learning Spaces. Apply domain-specific KOS principles for organizing collections/services for given applications**
Terence Smith, UC Santa Barbara, Marcia Lei Zeng, Kent State Univ.; Alexandria Digital Library Project
**Web Services and Terminology**.
Adam Farquhar, SchlumbergerSema
**Update on Revision to the NISO Z39.19 Thesaurus Standard and Other Terminology Standards**
Amy Warner, Lexonomy, Inc./consultant to NISO

This report gives highlights from four papers, two on reengineering legacy KOS and two on applications of the rich structure of ontologies.

**Reengineering thesauri to ontologies**

Reengineering thesauri, classification schemes etc. into ontologies means to build on the information contained in them and refining it as needed. Consider the following statements in a hypothetical ontology (Soergel):

| | | |
|---|---|---|
| Reading instruction | *isa* | Instruction |
| Reading instruction | *has domain* | Reading |
| Reading instruction | *governed by* | Learning standards |

| | | |
|---|---|---|
| Reading ability | *isa* | Ability |
| Reading ability | *has domain* | Reading |
| Reading ability | *supported by* | Perception |

and the following rules

**Rule 1**

**If** X isa (type of) instruction and X *has domain* Z
and Y *isa* ability and Y *has domain* Z
**Then** X *should consider* Y

**Rule 2**

**If** X *should consider* **Y**
**and Y *is supported by*** W
**Then** X *should consider* W

A system could infer that the designer of reading instruction should consider reading ability and perception. This inference relies on the detailed semantic relationships given in the ontology. But the ERIC Thesaurus (ERIC = Educational Resources Information Center) gives us only the following:

**Reading instruction**
BT Instruction
RT Reading
RT Learning standards

**Reading ability**
BT Ability
RT Reading
RT Perception

The relationships are not differentiated enough to support inference. A few examples from FAO's project to convert the AgroVoc Thesaurus into a richer ontology illustrate the point further (Fisseha):

## Broader Term (BT) and Narrower Term (NT) relations in AGROVOC

**BT** and **NT** are typical ***hierarchical*** relations in a thesaurus. However, their semantics is not explicitly defined. It is common for BT/NT relations within thesauri to include at least the following:

**Is-A** (e.g. Milk/ Cow's Milk;
Development Agency/IDRC)

**Ingredient of** (e.g. Milk/ Milk Fat)
Milk fat is an ingredient of milk

**Property of** (e.g. Maize/Sweet corn)
Sweetness is a property of corn

**Some examples from AGROVOC**

**MILK**
- NT Milk Fat
- NT Colostrum
- NT Cow's Milk

**Development Agencies**
- NT development banks
- NT voluntary agencies
- NT IDRC

**MAIZE**
- NT dent maize
- NT flint maize
- NT popcorn
- NT soft maize
- NT sweet corn
- NT waxy maize

## Related Term (RT) in AGROVOC

**RT** represents the ***associative*** relation. The RT usually involves the most ambiguous semantics. RT can include the following.

- causality
- agency or instrument
- hierarchy - where polyhierarchy has not been allowed the missing hierarchical relationships are replaced by associative relationships
- sequence in time or space
- constituency
- characteristic feature
- object of an action, process or discipline
- location
- similarity (in cases where two near-synonyms have been included as descriptors)
- antonym

**Some examples from AGROVOC**

**DEGRADATION**

- RT chemical reactions
  RT discoloration
  RT hydrolysis
- RT shrinkage

causality

**IDRC**
- RT Canada
-

location

## Some ideas for reengineering AGROVOC

**Most of the problems could be solved by:**

1. **Re-analyzing the existing relations to introduce explicit semantics. For instance,**
   **BT/NT** relationship could be resolved to **'Is-A'** relation as the default and refined to more specific relationships as needed
   **RT** relationship could be refined to more specific relationships (such as *produces*, *used by*, *made for*).
2. **Specifying composite concepts in terms of basic concepts that can be represented unambiguously. For instance,**
   **perishable product** could be represented as **product** *with attribute* **perishable**
   **fencing sword** could be represented as **sword** ***used for*** **fencing**
   **mother** could be represented as **parent** *with attribute* **female**

From these ideas on reengineering a thesaurus, a three-step general procedure for converting legacy Knowledge Organization Systems (KOS) into ontologies can be abstracted (Soergel):

1. **Define the ontology structure**
2. **Fill in values from** one or more **legacy KOS** to the extent possible
3. **Edit manually** using an ontology editor:
   - make existing information more precise
   - add new information

It was suggested that he laborious process of refining relationship could be streamlined by intelligent conversion using "**rules as you go**" (Soergel). The idea is to have the editor watch out for patterns and based on these patterns formulate rules that can be immediately applied to all similar cases as illustrated in the following example:

If an editor has determined (or it is known from another source that there is a relationship

**animal** ***has-part*** **milk**

it can be concluded that

**cow** ***NT*** **cow's milk** should become **cow** ***has-part*** **cow's milk**

since **cow** is an animal and **cow's milk** contains the word **milk**.

Similarly: **rice** *RT* **moist soil** becomes **rice** *grows in* **moist soil** (**plant** *grows in* **soil type**)

This indicates that the reengineering effort should start with the topmost concepts.

**Ontology applications**

Zeng presents an educational environment built around a rich ontology>

## Science learning spaces: Concept KOS

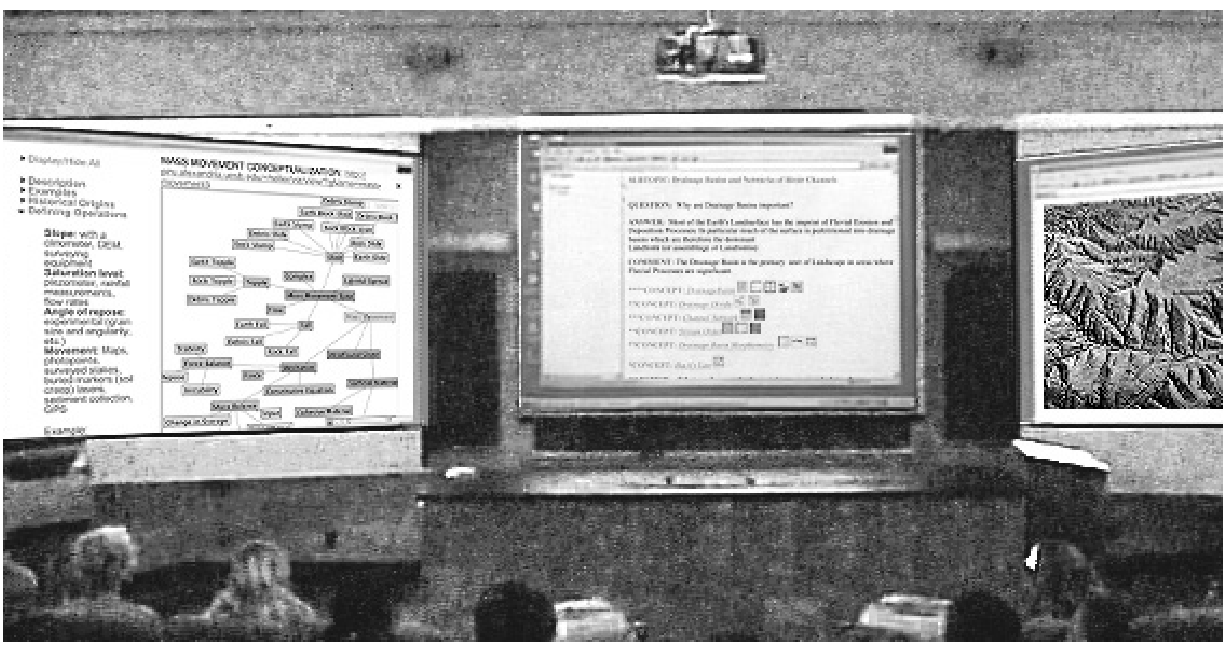
**Concepts of science as basic knowledge granules**

- Sets of concepts form bases for scientific representation
- DL and KOS technology can support organization of science learning materials in terms of concepts
  - Collections of models of science concepts (knowledge base)
  - Collections of learning objects (LO) cataloged with concepts
  - Collections of instructional materials organized by concepts

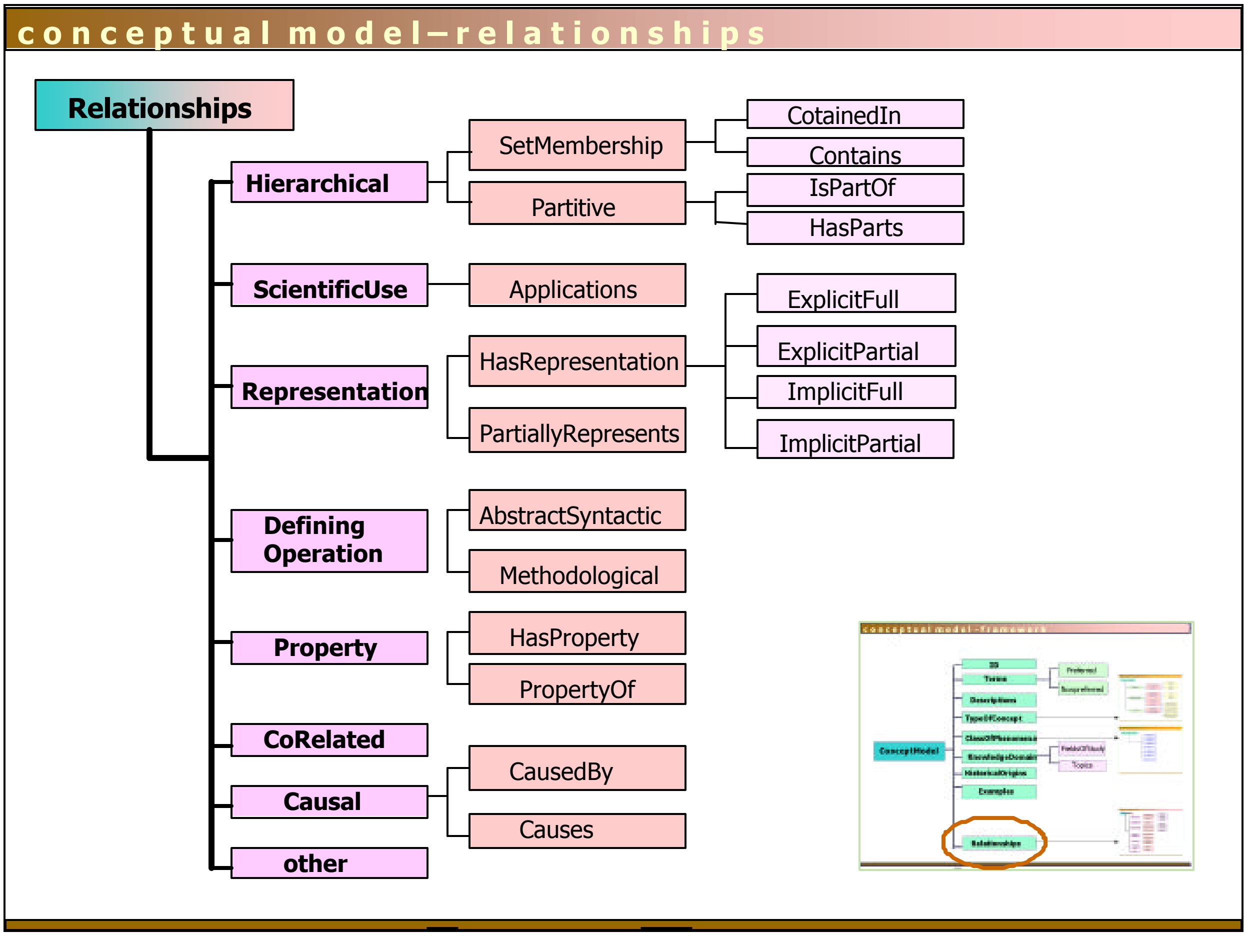
**Organize learning materials as "trajectory through concept space"**

- Lecture, lab, self-paced materials
- Services for creating/editing/displaying such materials

## Learning environment display (lecture mode)



Concept window Lecture window Object window




Farquhar describes the use of an ontology in an organization.

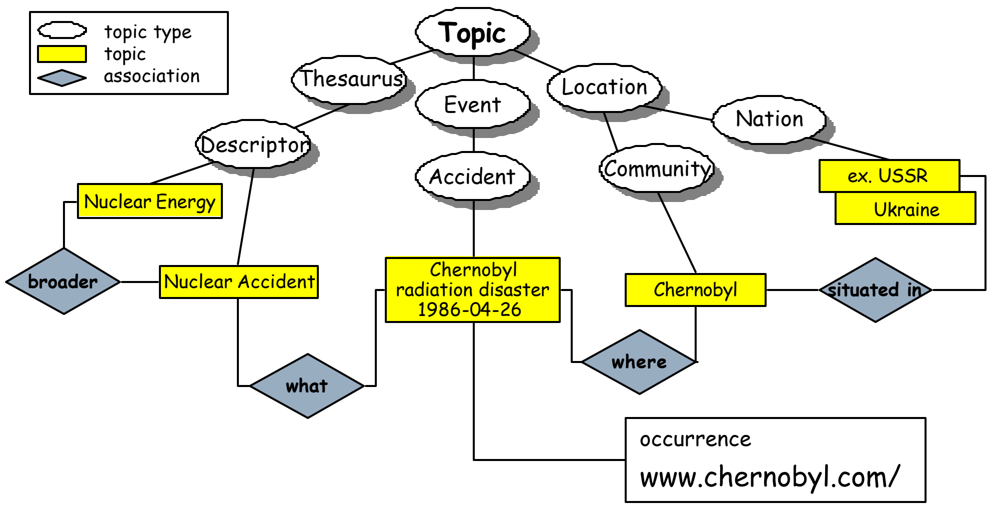
**Semantic Network Services**

**Sharing an integrated Ontology using Topic Maps and Web Services**

**Adam Farquhar (presenter)**
KM Architect, Schlumberger, Austin, TX
**Thomas Bandholtz**
KM Solution Manager, SchlumbergerSema, Cologne (DE)
Member, OASIS TC Published Subjects & GeoLang (Topic Maps)

Research project UFOPLAN-Ref. No. 20111612, promoted by BMU/Federal Environmental Agency, Germany

# Integration in a Topic Map




# sns: findTopics

search term

```
<findTopics>
        <queryTerm>Mauersegler</queryTerm>
        <searchType>contains</searchType>
        <lang>de</lang>
        <path>/event</path>
        <fields>names</fields>
</findTopics>
```

search method

topic type path

fields to search

**results in a list of matching topics**

**Conclusions**

The workshop participants identified several follow-on activities:

- Identify a core set of relationships that would further define the traditional RT with extensibility for specific domain relationships.
- Develop data exchange/interchange formats in XML/RDF.
- Develop use cases to better define terminology Web services.
- Define a tool suite for converting traditional tools into intermediate formats that can support the development of new semantic tools.
- Evaluate the cost and benefits of reusing traditional tools versus building new tools "from scratch", particularly with regard to machine-to-machine in the KM environment.
- Investigate licensing versus public domain business models for terminology.

In summary, the workshop presented papers on how to convert legacy KOS to systems with richer, precisely defined semantics (ontologies) and papers on applications of such rich ontologies and shows a direction the field should move in.

About NKOS

NKOS is a community of researchers, developers and practitioners seeking to enable knowledge organization systems, such as classification systems, thesauri, gazetteers, and ontologies, as networked interactive information services to support the description, retrieval, presentation, and use of diverse information resources through the Internet. Presentations are available from the NKOS Web site (http://nkos.slis.kent.edu).