

# Semantic Web and Vocabulary Control

## An OntoLex-oriented Introduction

謝舒凱 Shu-Kai Hsieh

Lab of Linguistic Ontology, Language Processing and e-Humanities  
Department of English, NTNU

Semantic Web workshop, Academia Sinica, 2008

- 1 Semantic Web and Controlled Vocabulary
- 2 Controlled Vocabularies (Basic Lexicon) and Ontologies
- 3 Representing Controlled Vocabularies
  - LMF: Lexical Markup Framework
  - Bootstrapping a controlled vocabulary into an ontology
- 4 Applications
  - NEDO: An Infrastructure to enhance Language Technologies for Asian language
  - KYOTO: Knowledge Yielding Ontologies for Transition-based Organization
- 5 Conclusion

## Background: the call for the Semantic Web

- WWW has changed the way we communicate with each other, how information is stored, disseminated and retrieved, how business is conducted (e.g., Google, P2P).
- Cooperative Information Systems: next generation of Information systems ► the call for the Semantic Web.
- Web semantics: to present Web content in a form that is more easily *machine-processable* and to use AI/NLP technologies to organize and take advantage of these representations.
- Personal agents: a future scenario.

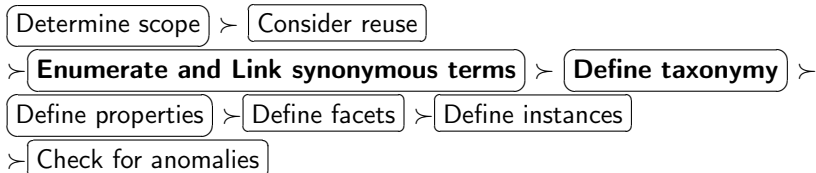
## Technologies involved in the Semantic Web agents

([Antoniou and van Harmelen, 2004])

- **Metadata** (後設資料) will be used to identify and extract information from Web sources
- **Ontologies** (知識本體) will be used to assist in web searches, to interpret retrieved information, and to communicate with other agents.
- **Logic** (推論邏輯) will be used for processing retrieved information and for drawing conclusions.

# Ontology construction

Following [Noy and McGuinness, ] the main stages in the ontology development process can be distinguished as follows:



# Controlled Vocabulary for the Semantic Web

so, in the context of Semantic Web,

- an important step toward ontology construction is to write down all the relevant **terms** that are expected to appear in the ontology.
- This task is similar to the construction of **controlled vocabulary** for indexing and searching document collections that has been long used in library and information science.

# Controlled Vocabulary Impact on IR and Semantic Web

- Information Retrieval = location finder?
- CV can help on both recall and precision rate. Example.
- Representing the Web content in a *machine-processable/understandable* form.

## Controlled Vocabularies (Basic Lexicon) and Ontologies: From the OntoLex perspective

A **lexicon** is a list of words in a language - a **vocabulary** - along with some knowledge of how each word is used. [Hirst, 2004]

- From ontologies to the lexicon and back
- Ontologizing the lexicon vs. Lexicalizing the Ontology
- A **formal ontology** is thus defined as a controlled vocabulary expressed in an ontological representation language.

## Principles of Controlled Vocabularies

(based on Z39.19-2005): Four important design principles of Controlled Vocabularies

- eliminating **ambiguity**. Example.
- controlling (**near**)-**synonyms**. Example.
- establishing **relationships** among terms where appropriate. Example.
- **testing and validation** of terms. Example.

A major goal of vocabulary control is to ensure that each distinct concept refers to a unique linguistic form. (like design of WordNet synset).

## Organizational Structures of Controlled Vocabularies

(based on Z39.19-2005)

- List: a limited set of terms arranged as a simple alphabetical list or in some other logically evident way. Example.
- Synonym ring: a set of terms that are considered equivalent for the purposes of retrieval. Example.
- Taxonomy: a collection of controlled vocabulary consisting of preferred terms, all of which are connected in a hierarchy or polyhierarchy. Example.
- Thesaurus: a collection of controlled vocabulary arranged in a network-like structure. Example.
- Others

## Relationships used in Controlled Vocabularies

- Lexical Semantic Relationships are also kind of Lexical Knowledge.
- Classification of Semantic Relations in different ways
  - structural view: *paradigmatic* (concerning substitution) and *syntagmatic* (concerning positioning). Example.
  - logical view: Identity:  $LU_1 = LU_2$ , Inclusion:  $LU_2$  is included into  $LU_1$ , Overlap:  $LU_1$  and  $LU_2$  have a non-empty intersection, but either one is not included in the other, and Disjunction:  $LU_1$  and  $LU_2$  have no common element. Example.
  - lexical configurational view: *hierarchical* and *non-hierarchical* relations. Example.
- Lexical Relations and Ontological Relations

## LMF: Lexical Markup Framework

- LMF is a work in progress (currently is in FDIS status) within ISO to define a standard for Natural Language Processing (NLP) lexicons.
- LMF is a big step in the standardization of representation of multilingual lexicons.

## LMF: Lexical Markup Framework: Extracts of LMF Chinese lexicons

```
***** <?xml version="1.0" encoding="UTF-8"?> <!DOCTYPE LexicalResource SYSTEM "kyoto_w.n.dtd" [
  <LexicalResource >< GlobalInformationlabel = "exampleencodingofChineseWords" LMFbyAS" / >< LexiconlanguageCoding = "ISO639-3" label = "ChineseWords"
  "zho" owner = "AcademiaSinica" version = "1.4" >
  <LexicalEntry id="自然"> <Lemma writtenForm="自然" partOfSpeech="Noun" >
  <Sense id="自然_1" synset = "zho-14-03059301-Na" >< /Sense ><
  Senseid = "6_2" synset = "zho-14-03059302-VH" >< /Sense ><
  Senseid = "6_3" synset = "zho-14-03059303-VH" >< /Sense ><
  Senseid = "6_4" synset = "zho-14-03059304-D" >< /Sense ><
  Senseid = "6_5" synset = "zho-14-03059305-D" >< /Sense ><
  Senseid = "6_6" synset = "zho-14-03059306-D" >< /Sense ><
  /LexicalEntry >
```

## Bootstrapping methodology in converting a controlled vocabulary into an upper level ontology

- Multilingual Swadesh List ▷ SUMO/MILO
- 廣群芳譜 ▷ SUMO/MILO

## NEDO Grant Project : Query Expansion

[Tokunaga et al., 2008]

We propose to:

- lexical specification and data categories relevant for building multilingual lexical resources for Asian languages
- a core upper-layer ontology needed for ensuring multilingual interoperability, and
- the evaluation platform used to test the entire architectural framework.

System demo

<http://www.cl.cs.titech.ac.jp/NEDO/search/index.jsp>

## NEDO Grant Project : Query Expansion demo

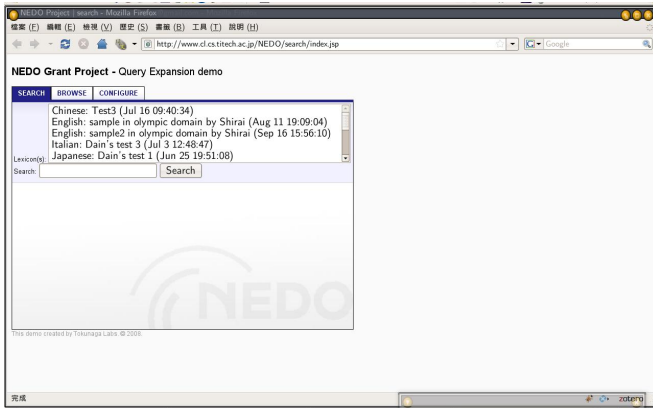


Figure: Project website

# KYOTO

[Vossen et al., 2008]





- Semantic Web is already among the key action lines of the European Union's Sixth Framework Programme.
- In the 7th Framework Programme we propose *KYOTO* as a system that allows people in communities to define the meaning of their words and terms in a shared Wiki platform so that it becomes anchored across languages and cultures but also so that a computer can use this knowledge to detect knowledge and facts in text. Whereas the current Wikipedia uses free text to share knowledge. *KYOTO* will represent this knowledge so that a computer can understand it

# KYOTO



Figure: System demo

Thank you

-  Antoniou, G. and van Harmelen, F. (2004).  
*A Semantic Web Primer*.  
The MIT Press.
-  Hirst, G. (2004).  
Ontology and the lexicon.  
*Handbook on Ontologies*.
-  Noy, N. F. and McGuinness, D. L.  
Ontology development 101: A guide to creating your first ontology.  
[semanticweb.org/wiki/Ontology\\_Development\\_101](http://semanticweb.org/wiki/Ontology_Development_101).
-  Tokunaga, T., Sornlertlamvanich, V., Charoenporn, T., Calzolari, N., Monachini, M., Soria, C., Huang, C.-R., Hsieh, S.-K., and Kiyooki Shirai, a. Y. X. (2008).

An infrastructure to enhance language technologies for asian language. poster paper.

*In The 6th International Conference on Language Resources and Evaluation, Marrakech: Morocco.*



Vossen, P., Agirre, E., Calzolari, N., Fellbaum, C., Hsieh, S.-K., Huang, C.-R., Isahara, H., Kanzaki, K., Monachini, A., Neri, F., Raffaelli, R., Rigau, G., Tescon, M., VanGent, J., and Marchetti, A. (2008).

Kyoto: A system for mining, structuring, and distributing knowledge across languages and cultures. poster paper.

*In The 6th International Conference on Language Resources and Evaluation. Marrakech: Morocco.*