Ontologies

COMP6215 Semantic Web Technologies

Dr Nicholas Gibbins – nmg@ecs.soton.ac.uk

Knowledge Representation

Knowledge representation is central to the Semantic Web
- Data published on the Semantic Web must be structured and organised

Long-standing concern in Artificial Intelligence
- A good knowledge representation 'naturally' represents a given problem domain
- A poor knowledge representation is unintelligible

Knowledge Representation

Common KR approaches:

- Logic
- Production rules
- Semantic Networks
- Frames

The Semantic Web combines aspects of all of these schemes

Knowledge Representation

Most AI systems (and therefore SW systems) consist of:

- A knowledge base (KB)
  - Forms the system's intelligence source
  - Structured according to the knowledge representation approach taken

- An inference mechanism
  - Set of procedures that are used to examine the knowledge base to answer questions, solve problems or make decisions within the domain

Ontologies

Defining the 'O' word

Ontology, n.

1. a. Philos. The science or study of being; that branch of metaphysics concerned with the nature or essence of being or existence.

Defining the 'O' word

An ontology is a specification of a conceptualisation

- **Specification**: A formal description

- **Conceptualisation**: The objects, concepts, and other entities that are assumed to always exist in some area of interest and the relationships that hold among them

Referred to in the philosophical literature as Formal Ontology

T. R. Gruber. A translation approach to portable ontologies. Knowledge Acquisition, 5(2):199-220, 1993
N. Guarino, D. Oberle, S. Staab. What is an ontology? In: S. Staab & R. Studer (eds.). Handbook on Ontologies. 2nd revised edition. Springer, 2009.

Ontology in Computer Science

Ontologies as engineered artifacts:

- constituted by a specific vocabulary used to describe a certain reality, plus
- a set of explicit assumptions regarding the intended meaning of the vocabulary

Benefits:

- Shared understanding
- Facilitate communication
- Establish a joint terminology for a community of interest
- Normative models
- Inter-operability: sharing and reuse

Ontology Structure

Ontologies typically have two distinct components:

- Names for important concepts in the domain
  - Elephant, Adult_Elephant, African_Elephant, Herbivore, etc

- Background knowledge/constraints on the domain
  - Elephants are a kind of animal
  - Adult_Elephants are Elephants whose age is greater than 20 years
  - Adult_Elephants weigh at least 2,000 kg
  - All Elephants are either African_Elephants or Indian_Elephants
  - Herbivore are exactly those animals who eat only plants or parts of plants
  - No individual can be both a Herbivore and a Carnivore

Informal Usage

Informally, 'ontology' may also be used to describe a number of other types of conceptual specification:

- Controlled vocabulary
- Taxonomy
- Thesaurus

Study of ontology is not limited to computer scientists and philosophers

- Rich tradition of knowledge representation and ontology in library and information science…
- …but they talk about classification and metadata instead of ontologies

Controlled Vocabularies

An explicitly enumerated list of terms, each with an unambiguous, non-redundant definition

- No structure exists between terms
- A controlled vocabulary is a flat list

Examples:

- Library of Congress Subject Headings (LCSH)
- Medical Subject Headings (MeSH)

Taxonomies

A collection of controlled vocabulary terms organised into a hierarchical structure
- Each term is in one or more parent-child relationships
- May be several different types of parent-child relationship:
  - Type-instance
  - Genus-species
  - Part-whole (referred to as meronymy)

Taxonomy Examples

Library classification schemes

- Library of Congress
- Dewey Decimal
- UDC

Linnean Classification

- Kingdom, Phylum, Class, Order, Family, Genus, Species, Subspecies

MeSH Tree Structures

Taxonomy Examples

Dewey Decimal

- 5xx - Natural Sciences and Mathematics
- 53x - Physics
- 537 - Electricity and Electronics

Library of Congress

- Q - Science
- QA - Mathematics
- QA71-90 - Instruments and machines
- QA75-76.95 - Calculating machines
- QA75.5-76.95 - Electronic computers and computer science
- QA76-76.765 - Computer software

Thesauri

A thesaurus is a taxonomy with additional relations showing lateral connections

- Related Term (RT)
- See Also

Parent-child relation usually described in terms of Broader Terms (BT) and Narrower Terms (NT)

Thesauri also typically contain scope notes which define the meaning of a term

Thesaurus Example

Apples

Scope notes:	The fruit of any member of the species *Malus pumila*
Broader term:	Foodstuffs
Related terms:	Cooking Ingredients Taxable Foodstuffs Horticulture
Narrower terms:	Granny Smiths
See also:	Apple Trees
Use:	For Apple computers use Personal Computers (Apple)

Ontology

An ontology further specialises types of relationships (particularly *related term*)

A ontology typically includes:

- Class definitions and hierarchy
- Relation definitions and hierarchy

An ontology may also include the following:

- Constraints
- Axioms
- Rule-based knowledge

Summary

Controlled Vocabulary + Hierarchy = Taxonomy

Taxonomy   +   lateral relations = Thesaurus

Thesaurus   +   typed relations
                +   constraints
                +   rules
                +   axioms = Ontology

Further Reading

N. Guarino, D. Oberle, S. Staab. What is an ontology? In: S. Staab & R. Studer. *Handbook on Ontologies*. 2nd revised edition. Springer, 2009.

https://link.springer.com/chapter/10.1007%2F978-3-540-92673-3_0
https://userpages.uni-koblenz.de/~staab/Research/Publications/2009/handbookEdition2/what-is-an-ontology.pdf

Next Lecture: RDF Schema