Revision

- Last lecture looked at Metadata, in particular raised some issues about various schema.
- This lecture considers some fundamentals behind schema.

Outline

- Thinking about schemas
- Some contexts…
- Semiotics
- Ontology
- OWL and other Web stuff
- What these mean for classification and metadata

Reading

- Eco, Umberto [1997]. Kant and the platypus; essays on language and cognition. Harcourt Brace, N.Y. Chapter 2
- OWL Web Ontology Language Reference http://www.w3.org/TR/owl-ref/

Bitzer’s definition of a Horse

‘Bitzer,’ said Thomas Gradgrind. ‘Your definition of a horse.’
‘Quadruped. Graminivorous. Forty teeth, namely twenty-four grinders, four eye-teeth, and twelve incisive. Sheds coat in the spring; in marshy countries, sheds hoofs, too. Hoofs hard, but requiring to be shod with iron. Age known by marks in mouth.’ Thus (and much more) Bitzer.  
Dickens Hard Times

Thinking about the world, then…

- The 18th and 19th centuries were times of re-thinking the order of things.
- Scientific revolution on a grand scale leading to/as part of an industrial revolution and a [subsequent] social revolution.
- [see Jenny Uglow’s The Lunar Men]  
- Have you seen Master & Commander?
Utilitarianism

- John Stuart Mill and the listing of usefulness [remember Bitzer’s definition of a horse]
- A world that can be organised [well, perceived] as facts.
- Extended to economic analysis [e.g. Marx] and production [Ford, mass production techniques, etc]

And now

- Information revolution
- Needing to organise information
- Need to understand how the world can be explained to machines if AI can come to fruition
- So some philosophical ideas are being revisited

Carl Linneus (1707-1778)

- His Systema Naturae [1735] grew from a slim pamphlet to a multivolume work, as his concepts were modified and as more and more plant and animal specimens were sent to him from every corner of the globe.
- For Linnaeus, species of organisms were real entities, which could be grouped into higher categories called genera.
- Eg. the
  - kingdom Animalia contained the
  - class Vertebrata, which contained the
  - order Primates, which contained the
  - genus Homo with the
  - species sapiens

Carl Linneus (1707-1778)

- His most famous student, Daniel Solander, was the naturalist on James Cook’s first round-the-world voyage.
- Anders Sparman, another of Linnaeus’s students, was a botanist on Cook’s second voyage.

Aristotle

Substance, in the truest and primary and most definite sense of the word, is that which is neither predicable of a subject nor present in a subject; for instance, the individual man or horse. But in a secondary sense those things are called substances within which, as species, the primary substances are included; also those which, as genera, include the species.

For instance, the individual man is included in the species ‘man’, and the genus to which the species belongs is ‘animal’; these, therefore – that is to say, the species ‘man’ and the genus ‘animal’, - are termed secondary substances.

Aristotle

It is plain from what has been said that both the name and the definition of the predicate must be predicable of the subject. For instance, ‘man’ is predicted of the individual man. Now in this case the name of the species ‘man’ is applied to the individual, for we use the term ‘man’ in describing the individual; and the definition of ‘man’ will also be predicated of the individual man, for the individual man is both man and animal. Thus, both the name and the definition of the species are predicable of the individual.

Aristotle The Categories Trans. E. M. Edghill
Immanuel Kant

German philosopher – 1724-1804
Writings are essentially against metaphysics, and relate to humanity in a real world – Critique of Pure Reason [1781], Critique of Practical Reason [1788], and works dealing with pure science, as well as ethics and aesthetics.

Emmanuel Kant - categories

Kant intends to provide a principle to identify the most fundamental concepts of thought, the categories of the understanding, and then to show that our knowledge of any object always involves these categories.

The key to his argument is the claim that knowledge is always expressed in a judgment; he then argues that there are certain characteristic forms or ‘logical functions’ of judgment, and that in order for our judgments to be about objects, these logical functions of judgments must also provide the basic concepts for conceiving of objects.

The logical functions of judgment are based on the premise that every judgment has a quantity, quality, relation and modality.

Kant – induction and deduction

- A priori knowledge is knowledge that is absolutely independent of all experience (E.g. the sum 2 + 2 = 4). Even though we may do the sum on our fingers in another language on the moon this makes no difference to the outcome (2 + 2 always equals 4).
- A priori means knowledge that is ‘prior to’ (comes before) experience. It does not depend upon the evidence of experience for its authority.
- A posteriori knowledge (people, places and things) depends on the evidence of experience for verification. This knowledge is only possible through experience (e.g. assessment/testing).

Kant

The point of all this is to acknowledge that Kant is trying to establish structures through which we apprehend the world, its objects, ourselves, and from which we derive abstract constructs.

This implies a logical structure of everything, into which everything can be fitted.

Charles Pierce

- In Pierce’s semiotic theory, our knowledge of the external world, in fact all thinking of any kind, is composed linked representations or signifiers ('signs' for short).
- According to Peirce, semiotics is
  - the science that studies the use of signs by "any scientific intelligence," By that term, he meant "any intelligence capable of learning by experience," including animal intelligence and even mindlike processes in inanimate matter.
  - Computer techniques for processing knowledge bases and databases could be called computational semiotics.

Where this is going

- It is possible to use philosophy to devise structures to think about the world.
- Such structures are either
  - Built from experience, from information we receive or
  - Devised in the abstract, and objects are able to be fitted to such a structure

PDF created with FinePrint pdfFactory trial version http://www.pdffactory.com
Schemas built from experience

- Reflect a version of the “real world”
- Are solutions to the puzzle presented by nature as the creation of a higher being [a higher intelligence]. i.e. they are premised on rationality, rather than an organic process, such as evolution.
- But this is also where Pierce’s semiotics fits in.

Ontology

- This is the philosophy of “the study of what there is”
- It is typically both
  - Abstract and
  - General

What is Ontology in computing?

- Ontology is about the exact description of things and their relationships.
- For the web, ontology is about the exact description of web information and relationships between web information.
- In general computing, ontologies have not been designed for ease of interconnection

Ontology and Computing

- Two major areas
  - Artificial intelligence, where it can be used to generate general structures so machines can model the world
  - Schemas for information management, that are so general that materials can be transferred between systems and can be managed by automated systems [knowbots]

AI requirement

- Among many things, for AI to work there is a need for
  - a framework into which facts can be inserted
  - But such a framework may well need to be dynamic in its structure
  - So it can incorporate new concepts/facts

The Semantic Web

- The Semantic Web is a mesh of information linked up in such a way as to be easily processable by machines, on a global scale.
- Think of it as being an efficient way of representing data on the World Wide Web, or as a globally linked database.
- RDF [Resource Description Framework] Schema is very simple, and yet allows a system to build up knowledge bases of data
**OWL - Web Ontology Language**

- OWL is built on top of RDF
- OWL is for processing information on the web
- OWL was designed to be interpreted by computers
- OWL was not designed for being read by people
- OWL is written in XML
- OWL is a web standard [W3C]

**OWL as part of Semantic Web**

- OWL is a part of the "Semantic Web Vision" - a future where:
  - Web information has exact meaning
  - Web information can be processed by computers
  - Computers can integrate information from the web

**So what's it about**

- **Combining**
  - Philosophy
  - Artificial Intelligence approaches
  - The Web

To produce a smart web that can automatically link materials that are related.