Web Science – Investigating the Future of Information and Communication

Helmi Ben Hmida
Hannover University, Germany
Summarizing the Problem: Computers don’t understand Meaning

“My mouse is broken. I need a new one...”
The Semantic Web Vision

“... the idea of having data on the Web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications”

http://www.w3.org/sw/
The Semantic Web

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

-- Tim Berners-Lee

“The wedding cake”
Distributed hypertext/hypermedia
Information accessed via (keyword based) search and browse
Browser tools render information for **human consumption**

**Where we are Today:**
**The Syntactic Web**
Where we are Today: The Syntactic Web
Hard Work using “Syntactic Web”

Find images of Peter Patel-Schneider, Frank van Harmelen and Alan Rector...

What is the Problem?

Consider a typical web page:

- Markup consists of:
  - rendering information (e.g., font size and colour)
  - Hyper-links to related content
- Semantic content is accessible to humans, but not (easily) to computers...
What is the (Proposed) Solution?

Add semantic annotations to web resources
What is the (Proposed) Solution?

Now... *that* should clear up a few things around here
Solution: XML markup with “meaningful” tags?

```xml
<name>...</name>
<location>...</location>
<date>...</date>
<slogan>...</slogan>
<participants>...</participants>
<introduction>...</introduction>
<speaker>...</speaker>
<bio>...</bio>
```
What is the Semantic Web?

Web was “invented” by Tim Berners-Lee (amongst others), a physicist working at CERN.

His vision of the Web was much more ambitious than the reality of the existing (syntactic) Web:

“… a set of connected applications … forming a consistent logical web of data …”

“… an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation …”

This vision of the Web has become known as the Semantic Web.
Having a web that knows “what you want” or “what you mean” is accomplished by semantics…. specifically using semantic annotation on web resources
Ontologies
In Philosophy, fundamental branch of metaphysics

- Studies “being” or “existence” and their **basic categories**
- Aims to find out what **entities** and **types of entities** exist

---

**Ontology: Origins and History**

**Supreme genus:**
- Substance
  - Material
  - Immaterial
  - Spirit

**Subordinate genera:**
- Body
  - Animate
  - Inanimate
  - Mineral

- Living
  - Sensitive
  - Insensitive
  - Plant

- Animal
  - Rational
  - Irrational
  - Human

- Species:
  - Socrates
  - Plato
  - Aristotle
  - Etc.

**Differentiae:**
- Supreme genus:
- Subordinate genera:
- Proximate genera:
- Species:
- Quali:
- Individuals:
Ontology

“Specification of concepts and their meanings”

Shared and common understanding of knowledge concerning domain of interests
Ontology Definition

Formal, explicit specification of a shared conceptualization

- **unambiguous terminology definitions**
- **conceptual model of a domain (ontological theory)**
- **machine-readability with computational semantics**
- **commonly accepted understanding**

[Gruber93]
Example Ontology (Protégé)
RDF and RDFS

**RDF** stands for **Resource Description Framework**

is a W3C standard, which provides tools to describe Web resources provides interoperability between applications that exchange machine-understandable information
The RDF Data Model

Statements are
\(<\text{subject}, \text{predicate}, \text{object}>\) triples:
\(<\text{Ian, hasColleague, Uli}>\)

Can be represented as a graph:

- Statements describe properties of resources
- A resource is any object that can be pointed to by a URI:
  - a document, a picture, a paragraph on the Web;
  - isbn://5031-4444-3333
  - ...
- Properties themselves are also resources (URIs)
Linking Statements

The subject of one statement can be the object of another. Such collections of statements form a directed, labelled graph.

Note that the object of a triple can also be a “literal” (a string)
RDF Syntax

**Subject** of an RDF statement is a resource

**Predicate** of an RDF statement is a property of a resource

**Object** of an RDF statement is the value of a property of a resource
RDF Schema (RDFS)

RDF Schema allows you to define vocabulary terms and the relations between those terms

- it gives “extra meaning” to particular RDF predicates and resources
- this “extra meaning”, or semantics, specifies how a term should be interpreted
RDFS Examples

RDF Schema terms (just a few examples):
- Class
- Property
- type
- subClassOf
- range
- domain

These terms are the RDF Schema building blocks (constructors) used to create vocabularies:

- `<Person, type, Class>`
- `<hasColleague, type, Property>`
- `<Professor, subClassOf, Person>`
- `<Carole, type, Professor>`
- `<hasColleague, range, Person>`
- `<hasColleague, domain, Person>`
Ontology Languages for the Web

Semantic Web effort led to development of “resource description” language(s)
  - E.g., RDF, and later RDF Schema (RDFS)

RDFS is recognisable as an ontology language
  - Classes and properties
  - Sub/super-classes (and properties)
  - Range and domain (of properties)

But RDFS too weak to describe resources in sufficient detail, e.g.:
  - No existence/cardinality constraints
  - No transitive, inverse or symmetrical properties
  - No localised range and domain constraints
  - ...

And RDF(S) has “higher order flavour” with non-standard semantics
  - Difficult to provide reasoning support
From RDF to OWL

- OWL is a language for defining Web Ontologies and their associated Knowledge Bases

- The OWL language is a revision of the DAML+OIL web ontology language incorporating learning from the design and application use of DAML+OIL.
OWL became standard

10 February 2004 the World Wide Web Consortium announced final approval of two key Semantic Web technologies, the revised Resource Description Framework (RDF) and the Web Ontology Language (OWL).
Ontology Elements

- Concepts (classes) + their hierarchy
- Concept properties (slots/attributes)
- Property restrictions (type, cardinality, domain)
- Relations between concepts (disjoint, equality)
- Instances
There are two types of animals, Male and Female.

```
<rdfs:Class rdf:ID="Male">
  <rdfs:subClassOf rdf:resource="#Animal"/>
</rdfs:Class>
```

The `subClassOf` element asserts that its subject - Male - is a subclass of its object -- the resource identified by #Animal.

```
<rdfs:Class rdf:ID="Female">
  <rdfs:subClassOf rdf:resource="#Animal"/>
  <owl:disjointWith rdf:resource="#Male"/>
</rdfs:Class>
```

Some animals are Female, too, but nothing can be both Male and Female (in this ontology) because these two classes are disjoint (using the `disjointWith` tag).
How to build an ontology?

Steps:
- determine domain and scope
- enumerate important terms
- define classes and class hierarchies
- define slots
- define slot restrictions (cardinality, value-type
Ontology Example

**Concept**
conceptual entity of the domain

**Attribute**
property of a concept

**Relation**
relationship between concepts or properties

**Axiom**
coherent description between Concepts / Properties / Relations via logical expressions

\[ \text{holds(Professor, Lecture)} \Rightarrow \text{Lecture.topic} \in \text{Professor.researchField} \]
Description Logics
What Are Description Logics?

A family of logic based Knowledge Representation formalisms

- Descendants of **semantic networks** and **KL-ONE**
- Describe domain in terms of **concepts** (classes), **roles** (properties, relationships) and **individuals**
- **Operators** allow for composition of complex concepts
- **Names** can be given to complex concepts, e.g.:

\[ \{\text{HappyParent} \equiv \text{Parent} \sqcap \forall \text{hasChild} (\text{Intelligent} \sqcup \text{Athletic})\} \]
Semantics and Reasoning

- Provision of **reasoning services**
  - Decision procedures for key problems (satisfiability, subsumption, etc)
  - Implemented systems (highly optimised)

\[
\{ \text{HappyParent } \equiv \text{Parent} \land \\
\forall \text{hasChild( Intelligent } \lor \text{ Athletic)} \}
\]

John : HappyParent
John hasChild Mary
Mary : ¬Athletic
→ Mary : Intelligent
Why Description Logic?

OWL exploits results of 15+ years of DL research

- Well defined (model theoretic) **semantics**
- **Formal properties** well understood (complexity, decidability)
- Known **reasoning algorithms**
- **Implemented systems** (highly optimised)
Summary

1 know what you mean...

A new form of Web content that is meaningful to computers will unleash a revolution of new abilities
Summary

**Semantic Web** aims to make web content more accessible to automated processes
- Adds semantic annotations to web resources

**Ontologies** provide vocabulary for annotations
- Terms have well defined meaning

**OWL** ontology language based on (description) logic
- Exploits results of basic research on complexity, reasoning, etc.

Many **research challenges** remain
- Including expressive power, scalability and tools

*"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."*  
-- Tim Berners-Lee
A new form of Web content that is meaningful to computers will unleash a revolution of new abilities.