Semantic Web and Ontologies

"People are starting to realize that their information outlives their software"

Tim Berners-Lee

"Information meaning is too tightly coupled to its initial use or application. Thus it is very difficult for either (a) machines to reuse information or (b) for people to query on concepts (instead of just on terms).

Jeffrey. T. Pollock

J. A. Gulla/NTNU

Jon Atle Gulla

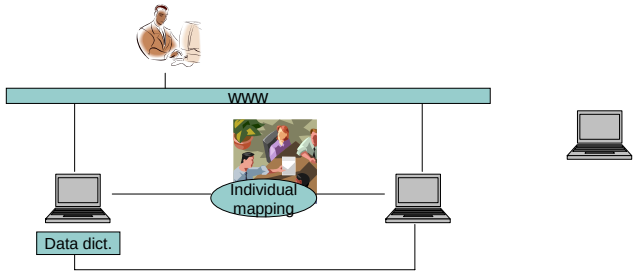
- *Professor of Information Systems, NTNU*
- *Siv.ing/Dr.ing. Cand.philol. Sloan Fellow*
- *Email: jag@idi.ntnu.no*

- Fast Search & Transfer, Munich
- Norsk Hydro, Brussels
- German National Research Center for Information Technology, Darmstadt

Areas of research

Semantic Web technologies
Ontology engineering
Process and document mining
Semantic search
Enterprise business applications
Business processes

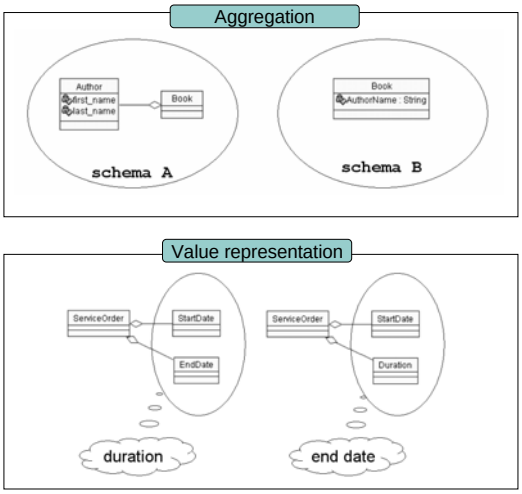
The Interoperability Problem



SEMANTIC STANDARDS

No scalability, not machine processable, limited flexibility & expressiveness

Silly Semantic Conflicts



From SICoP white paper: Introducing Semantic Technologies and the Vision of the Semantic Web

Semantic Conflicts

Semantic conflicts	Description	Example
Data type	Different primitives or abstract types for same information	SSN as a VARCHAR vs. a NUM
Labeling	Synonyms/antonyms have different text labels	When ORG_NAME and COMPNY tables have data that mean the same thing
Aggregation	Different conceptions about the relationships among concepts in similar data sets.	Does a "motorcycle" have 1, 2, 3, 4 or more wheels, how are the constraints modeled in your schema?
Generalization	Different abstractions are used to model same domain	Are "cars" and "trucks" kinds of "vehicles" or are they top-level classes themselves?
Value representation	Different choices are made about what concepts are made explicit	"StartTime" plus "Duration" equals "EndTime"
Impedance mismatch	Fundamentally different data representations are used	Relational to Object mappings (key migrations, multiplicity, etc.)
Naming	Synonyms/antonyms exist in same/similar concept instance values	"Company" table has many entries: "DaimlerBenz", "Mercedes", etc. but they refer to the same thing
Scaling and unit	Different units of measures with incompatible scales	km vs. English mile
Confounding	Similar concepts with different definitions	"EarningsPerShare" object for a NASD application vs. a NYSE system
Domain	Fundamental incompatibilities in underlying domain	"MainAssembly" object in a Ford product system vs. a brake supplier system
Integrity	Disparity among the integrity constraints	Does an airline ticket have a primary key that uniquely IDs a passenger?

From Jeff Pollock, December 2004

Outline

- What is the Semantic Web?
- Semantic Web Principles
- Semantic Markup Languages
- Ontologies
- Key Capabilities
- Challenges and Conclusions

- (OWL ontology examples)

What is the Semantic Web?

What is 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, James Hendler and Ora Lassila, The Semantic Web, Scientific American, May 2001

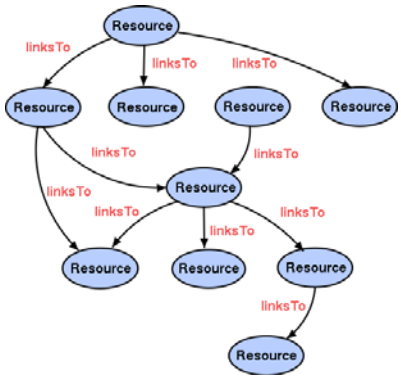
- *Definition*: The **Semantic Web** is the representation of **data** on the World Wide Web. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming, and Web Ontology Language (OWL).

Semantic + Interoperability = Intelligence

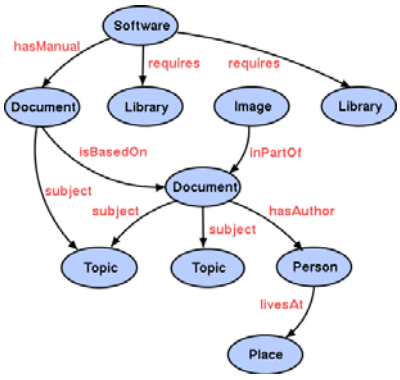
What is the Semantic Web not?

- Not a new and distinct set of web sites
- Not being constructed with just human accessibility in mind
- Not built upon radical untested theories
- Not a drastic departure from current data modeling concepts
- Not a magical piece of artificial intelligence
- Not an existing entity, ready for users to make use of it

The Current Web is not Enough



- Homogeneous resources
- Semantically empty
- Needs human interpretation



- Identified resources
- Meaningful links
- Machine processable

HTML for Interoperability

<H1>
<The Rhyme of the Ancient Mariner>
</H1>
The Rhyme of the Ancient Mariner, by Samuel Coleridge, is available for the low price of $9.99. This Dover reprint is beautifully illustrated by Gustave Dore.
<p>
Julian Schnabel recently directed a movie, Pandemonium, about the relationship between Coleridge and Wordsworth.

Can you devise an algorithm that will retrieve the price and author of the book?
AND that's likely to work correctly for **ALL** book descriptions?

HTML for Search

Doc #15432

The*5
Of*3
Ancient*2
By*2
Coleridge*2
Mariner*2
Rhyme*2
A
available
Between
Directed
Dover
….

When to retrieve this document?
What is the content?

XML for Interoperability

<book>
<title> The Rhyme of the Ancient Mariner </title>
<author> Coleridge </author>
<price> $9.99 </price>
</ book>

- Need to know the price? Just look inside the price tag.
- How do I know that you mean the same thing by <price> that I do?
  - *Does that include tax? shipping? surcharges?*
- This is critical in B2B e-commerce.

Limits of XML

- If the computers of two companies are negotiating, they need to know that they truly understand each other.
  - Computer 1: Do you sell heavy duty crowbars?
    [thinks: *I need crowbars that can withstand 10,000 lbs. Pressure*]
  - Computer 2: Yes.
    [thinks: *Our crowbars are good to 5,000 lbs.*]
- XML provides *syntactic* interoperability. There is a need for *semantic* interoperability.
- The semantic web provides this added layer of interoperability through the use of *shared ontologies.*

Semantic Web Principles

Semantic Web Principles

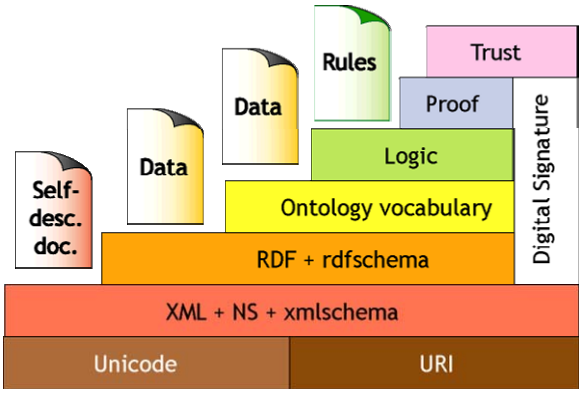
- For a computer program to reason, it must have a conceptual understanding of the world. This understanding is provided by us. That is, we must provide the computer with an *ontology.*
- An ontology is specified using a semantic markup language

- Ontology is the branch of philosophy that answers the question "what is there?"
- An ontology is an *explicit specification of a conceptualization of a domain*
- An ontology is typically a hierarchical collection of classes, permissible relationships amongst those classes, and inference rules/contraints that is shared by a community

- *An ontology reflects a community's common understanding of a domain*

Semantic Markup Languages

Semantic Markup Languages

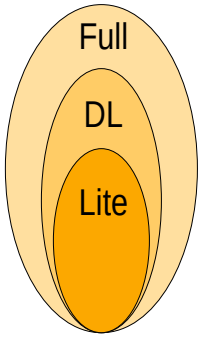


Semantic Web LayerCake (Berners-Lee, 99;Swartz-Hendler, 2001)

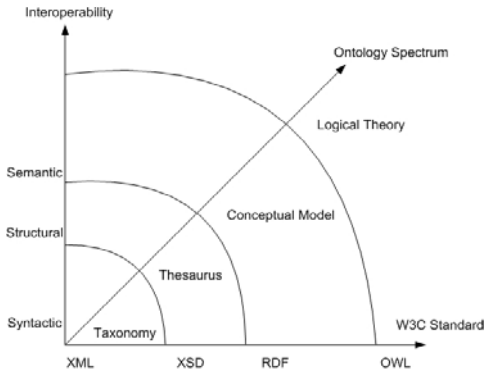
Web Ontology Language (OWL)

- OWL: Ontology language standard for Semantic Web

- **OWL Full:**
  - Extension of RDF
  - Expressive!
  - Theoretical properties not well understood
- **OWL DL:**
  - The part of OWL Full that fits in the Description Logic framework
  - Computational completeness and decidability
- **OWL Lite:**
  - Subset of OWL DL
  - Easier for frame-based tools to deal with
  - easier reasoning
  - Classification hiearachies and simple constraint features



The Ontology Spectrum



- Web Ontology Language (OWL)
  - Recommended by W3C
  - Logic language (based on Description Logic) with XML syntax
  - Supported by numerous tools (e.g. Protege)

Ontologies

Content of Ontologies

Ontologies typically have two distinct components:

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

Ontologies Come in Many Flavors

Informal definition of concepts...

• PLANNING: an ACTIVITY whose major EFFECT is to produce a PLAN
• STRA
• STRA

Part

Pizza OWL ontology from Protege

<?xml version="1.0" ?>
<rdf:RDF xmlns:xsp="**http://www.owl-ontologies.com/2005/08/07/xsp.owl#**"
xmlns="**http://www.co-ode.org/ontologies/pizza/2005/10/18/classified/pizza.owl#**"
xmlns:protege="**http://protege.stanford.edu/plugins/owl/protege#**" xmlns:rdf="**http://www.w3.org/1999/02/22-rdf-syntax-ns#**"
xmlns:xsd="**http://www.w3.org/2001/XMLSchema#**" xmlns:rdfs="**http://www.w3.org/2000/01/rdf-schema#**"
xmlns:owl="**http://www.w3.org/2002/07/owl#**" xmlns:daml="**http://www.daml.org/2001/03/daml+oil#**"
xmlns:dc="**http://purl.org/dc/elements/1.1/**" xml:base="**http://www.co-ode.org/ontologies/pizza/2005/10/18/classified/pizza.owl**">
<owl:Ontology rdf:about="">
<owl:versionInfo rdf:datatype="**http://www.w3.org/2001/XMLSchema#string**">**classified**</owl:versionInfo>
<protege:defaultLanguage rdf:datatype="**http://www.w3.org/2001/XMLSchema#string**">**en**</protege:defaultLanguage>
<rdfs:comment xml:lang="**en**">**An example ontology that contains all constructs required for the various versions of the Pizza Tutorial run by Manchester University (see http://www.co-ode.org/resources/tutorials/)**</rdfs:comment>
<owl:versionInfo rdf:datatype="**http://www.w3.org/2001/XMLSchema#string**">**version 1.3**</owl:versionInfo>
<owl:imports rdf:resource="**http://protege.stanford.edu/plugins/owl/protege**" />
</owl:Ontology>
<owl:Class rdf:ID="**DomainConcept**">
<owl:disjointWith>
<owl:Class rdf:ID="**ValuePartition**" />
</owl:disjointWith>
</owl:Class>
<owl:Class rdf:ID="**FishTopping**">
<owl:disjointWith>
<owl:Class rdf:ID="**MeatTopping**" />
</owl:disjointWith>

How to Use Ontologies?

- Develop domain ontology in some well-defined language
- Semantically annotate the objects in the system

A web page describes objects:
Datasets, human beings, services, items for sale, etc.

The semantics of an object are defined by the place it occupies in some domain ontology.
Objects are self-describing

- Set up structure/procedure for ontology engineering and maintenance
- Provide ontologically driven services
  - Intelligent search, Reasoning and composition
  - Mining
  - Agents
  - Interoperable applications
  - …

Semantic Annotation Example

- You ask the system
  "*Show me all universities near the beach*."
- The NTNU page doesn't say anything about the beach, but it does say (through semantic markup) that it's near the Trondheimsfjord.
- NTNU makes use of a geography ontology which includes the rule "Fjord(x) →hasBeaches(x)".
- When your search agent parses the NTNU page, it loads in the relevant ontologies, deduces that NTNU is near the beach, and returns the page.

Capabilities and Challenges

Key Capabilities

- Near term:
  - Semantic web services
    Automatically compose services according to user needs
  - Information integration and/or interoperability
    Application understands other applications' data
  - Intelligent search
    Search on concepts, not term frequencies
- Long-term:
  - Model-driven applications
  - Adaptive and autonomic computing
  - Intelligent reasoning

Semantic Web Challenges (1)

- Political/human side
  - Do we accept standards?
  - Early wins?
  - Can we agree on the exact content on the ontology?

  Hierarchies of ontologies, ontology alignment, ontology enrichment
- Ontology Engineering Process
  - Need experts in both domain and ontology engineering
  - Stakeholder involvement
  - How to deal with complexity?
- Quality
  - One ontology to rule them all?
  - Level of granularity/precision
  - How to assess the correctness/usability/value of ontologies?
  - Tailoring to information retrieval/mining
  - Personalization

Semantic Web Challenges (2)

- Technology maturity
  - OWL Full reasoning not Turing complete, intractable
  - Performance issues
  - Limited industrial experience
  - Lack of Semantic Web applications
- Practical application of Semantic Web
  - Who & how to maintain ontologies and make them accessible?
  - Who & how to annotate information objects with semantic descriptions?

Conclusions

- Semantic Web technology for interoperability and semantics
  - Information sharing
  - Application intelligence
- Semantic markup languages
- Ontologies
- Capabilities and challenges

OWL Ontology Example

OWL Header

```
<rdf:RDF
    xmlns:owl = "http://www.w3.org/2002/07/owl#"
    xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs = http://www.w3.org/2000/01/rdf-schema#>

<owl:Ontology rdf:about" ">
    <rdfs:comment>An example ontology</rdfs:comment>
    <owl:imports rdf:resource="http://www.mydomain.org/persons"/>
    <rdfs:label>University ontology</rdfs:label>
```

Namespace declarations provide a mechanism for abbreviating URI's
Import statement includes definitions from referenced ontology

OWL Classes

```
<owl:Class rdf:ID="associateProfessor">
    <rdfs:subClassOf rdf:resource="academicStaffMember"/>
</owl:Class>

<owl:Class rdf:about="associateProfessor">
    <owl:disjointWith rdf:resource="professor"/>
</owl:Class>
```

An associate professor is a member of the academic staff.
The class of associate professors is disjoint from the class of professors

OWL Properties

```
<owl:ObjectProperty rdf:ID="isTaughtBy">
    <owl:domain rdf:resource="#course"/>
    <owl:range rdf:resource="academicStaffMember"/>
    <rdfs:subPropertyOf rdf:resource="#involves"/>
</owl:ObjectProperty>
```

A course is taught by a member of the academic staff
The property "isTaughtBy" is a subproperty of "involves"

OWL Property Restrictions

```
<owl:Class rdf:about="#firstYearCourse">
    <rdf:subClassOf>
        <owl:Restriction>
            <owl:onProperty rdf:resource="#isTaughtBy"/>
            <owl:allValuesFrom rdf:resource=#Professor"/>
        </owl:Restriction>
    </rdf:subClassOf>
</owl:Class>
```

First year courses are taught by professors only

OWL Instances

```
<academicStaffMember rdf:ID="949352"/>

<course rdf:about="tdt4215">
    <isTaughtBy rdf:resource="949352">
</course>
```

Employee 949352 is a member of the academic staff and teaches course tdt4215