



Where are the Semantics in the Semantic Web?

Knowledge Technologies Conference

Seattle

13 March 2002

Mike Uschold

Mathematics and Computing Technology

Boeing Phantom Works

Acknowledgements

Material from this lecture was drawn from many fruitful discussions with:

- Dieter Fensel
- Peter Clark
- Rob Jasper
- Frank van Harmlen
- Michael Gruninger
- John Thompson

Outline

- **The Evolving Semantic Web**
- **Shopping Agents**
- **What do we mean by ‘Semantics’?**
- **A Semantic Continuum**
- **Machine Processible Semantics**
- **Where are the semantics?**

The Evolving Web

- **Locating Resources**

- *free text* search → *concept-based* search

- **Web Users**

- primarily *humans* → both humans and *machines*

- **Web Tasks & Services**

- a place to *find* things → a place to *do* things

Semantics is the Core Requirement

- web content with *no semantics* → *with semantics*



Defining Feature of the Semantic Web:

Machine-Usable Web Content

What about Shopping Agents?

- Peruse Web sites for best deals
- This is **machine-usable content!**
- Is this the Semantic Web?
- Most would say **NO.**

Where are the Semantics?

What is the Problem?

- How does the machine **“know”** what to do with the Web content it encounters?
- It needs to know the **meaning** of the content.
- **How does this knowledge get into the machine?**
- Do we need a better definition of the Semantic Web, to exclude shopping agents?
- Lets take a closer look at **“semantics”**

Semantics: A Many-Splendored Thing

- “Semantics” means *meaning*.
- What has semantics? Where are they? What do they look like? How are they used?
- **Kinds of Semantics:**
 - Real-world Semantics
 - Axiomatic Semantics
 - Model-theoretic Semantics

Things that have Semantics

- **Terms, expressions, descriptions in a language**
 - Ontologies in RDF Schema, or DAML+OIL
 - RDF statements,
 - Performatives in Agent Communication Languages (ACLs).
- **Languages & Inference procedures**
 - e.g. RDF Schema & DAML+OIL

Nature and Use of Semantics

- Are the semantics **implicit**, or **explicit**?
- Are the semantics expressed **informally**, or **formally**?
- Are the semantics intended for **human** or **machine processing**?

A Semantic Continuum



Pump: “a device for moving a gas or liquid from one place or container to another”



Implicit

Informal
(explicit)

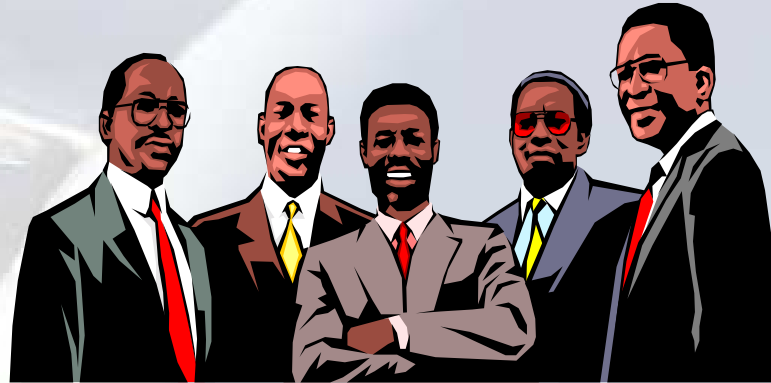
Formal
(for humans)

Formal
(for machines)

Further to the right means:

- Less ambiguity
- More likely to have correct functionality
- Better inter-operation
- Less hardwiring
- More robust to change
- **More difficult!**

1: Implicit Semantics



- E.g. XML tags *price, shipping cost*
- Exist only in people's heads
- Based on shared human consensus
- Rife with hidden ambiguities
- Inter-operation difficulties
- Mistakes and inconsistency in functionality
- Hardwiring brittle to changes

2: Explicit Informal Semantics

- **Pump:** “device for moving a gas or liquid from one place or container to another” (glossary)
- **Language specification documents**
 - HTML: `<h2>` is a second level header
 - RDFS: **Class**—generic concept of Type or Category
 - Similar for Java, UML ...
- **Written down in an informal language**
- **Only accessible to humans.**
- **Ambiguity remains, more open**
- **More inter-operation (maybe)**
- **Fewer mistakes; more consistency in functionality (maybe)**
- **Hardwiring brittle to changes**

3: Formal Semantics for Human Processing

(every pump has
(physical-parts (piston, valve, cylinder))
(device-purpose (Pumping-A-Fluid)))



- **Written down in a formal language**
- **Only used by humans.**
- **Semantics are**
 - **Hardwired into applications,**
 - **Used at runtime**
- **Reduced ambiguity**
- **Even fewer mistakes and more consistency in functionality**
- **Even more inter-operation**
- **Error-prone human still in the loop**
- **Hardwiring brittle to changes**

4: Formal Semantics for Machine Processing

(every pump has
(physical-parts (piston, valve, cylinder))
(device-purpose (Pumping-A-Fluid)))



- **Semantics are processed by machine at runtime.**
- **Least ambiguity**
- **Fewest mistakes and inconsistency in functionality**
- **Even more inter-operation**
- **Few if any good examples.**
- **Dynamic meaning discovery**
- **Reduced hardwiring of content**
- **Robust to changes in content**
- **A very challenging problem**

Machine Processible Semantics

Plethora of Difficulties;

- Language Heterogeneities;
- Incompatible Conceptualizations;
- Term Heterogeneity;
- Different Modeling Styles;
- **Cryptography, in the general case.**

Must Make Some Assumptions

- **Shared representation language**
 - e.g. RDF(S), DAML+OIL
- **Compatible Conceptualizations**
- To determine if two terms actually mean the same thing is **Computationally Intractable!**

We further assume:

- **Shared Semantic Repositories for Declaring Agreements**

Example



Simple Task:

Find documents about mechanical devices.

The purpose of this review is to remind operators of the

existence of the Operations Manual Bulletin 80-1, which provides

information regarding flight operations with low fuel quantities,

and to provide supplementary information regarding main tank

boost pump low pressure indications.

747 **FUEL PUMP** LOW PRESSURE INDICATIONS

When operating 747 airplanes with low fuel quantities for short

Example



Hey, I know this ontology, so now I know something about *Fuel Pump*.

What the heck is a *Fuel Pump*?

The purpose of this review is to remind operators of the existence of the Operations Manual Bulletin 80-1, which provides information regarding flight operations with low fuel quantities, and to provide supplementary information regarding main tank boost pump low pressure indications.747 <concept id=fuel-pump>**FUEL PUMP**</concept> LOW PRESSURE INDICATIONS

When operating 747 airplanes with low fuel quantities for short

Semantic Markup

<concept id=fuel-pump>**FUEL PUMP**</concept>

(**fuel-pump** has (superclasses **SHR**, **pump**))

Shared Hydraulics Repository (SHR)

(**pump** has

(superclasses (mechanical-device))

(text-def (“A device for ...”))

(thesaurus-term (|Pumps|)))

(every **pump** has

(physical-parts (piston, valve, cylinder))

(device-purpose (Pumping-A-Fluid)))

Dynamic Discovery of Meaning

- **Machine can learn something about meaning.**
 - No idea what kind of pump, just a kind;
 - Rules out polar bears and cars;
 - Still plenty of scope for ambiguity;
 - Definition of fuel-pump can never be complete.
- **Inference for a Practical Task?**
 - For search application looking for content about mechanical devices.
 - May be no better than conventional search...

Machine Processible Semantics

- Relies on a formal language (e.g. RDFS) semantics to infer semantics of terms and expressions in the language.
- The shared concept referred to may not be formally defined (e.g. Dublin Core terms)

Requirements for Machine-Usable Content

R1: Machine “knows” what to do with content.

R2: Human Web application developers know what to do with content the machine is expected to encounter.

R3: Humans must provide machine with ability to know the meaning of the content. Human either:

- Knows the meaning of the content and **hardwires** it into the application;
- Encodes a procedure to **discover** that meaning.

Why do Shopping Agents Work?

Issues to consider:

- What is hardwired and what is not?
Everything is hardwired.
- Are the semantics clearly specified?
Probably not specified at all, if so, informally.
- How much agreement is there on terms/concepts?
No specific agreements, strong concept overlap.
- Are the agreements publicly declared?
Probably not, although there could be industry standards.

Why do Shopping Agents Work?

Requirements for Machine Usable Content

- **Humans know the meaning of the expected content.**
Due to shared human consensus on terms like “price” and “destination”.
- **Humans know what to do with the content.**
Follows from understanding content, and knowing functional requirements of the Web application.
- **Machine knows what to do with the content.**
Humans hardwire the semantics into the application.

A Degenerate Case of the Semantic Web

The Evolving Semantic Web

Progress will occur by:

- **Moving along the semantic continuum;**
- **Reducing the amount of hardwiring by increasing machine-processing of semantics;**
- **Increasing the amount of public standards and ability to specify agreements;**
- **Developing semantic mapping and translation technologies for resolving semantic heterogeneity.**

The Evolving Semantic Web

This will result in:

- **Clearly specified semantics;**
- **Fewer misunderstandings;**
- **Consistent, correctly functioning applications;**
- **Greater inter-operation;**
- **Dynamic discovery of meaning;**
- **Robust with respect to changes in content.**

Summary: Where are the Semantics?

- Often just in the human.
- In informal specification documents.
- Hardwired in implemented code.
- Formal Documentation to help humans understand and/or write code.
- Formally encoded for machine processing.
- In the representation language specification.

Various Knowledge Technologies

Artificial Intelligence

Variety of Logics for
Knowledge Representation

W3C HTML

RDF, RDF Schema

RDF, RDF Schema
OWL, Taxonomies

Library Science

Dublin Core
Thesauri
Taxonomies

Topic Maps

Topic Maps

Document Management

SGML, XML

SGML, XML

? XML Schema ?
(datatypes)

Shopping Agents

Implicit

Informal
(explicit)

Formal
(for humans)

Formal
(for machines)

No Need for Semantics Envy

- **Machine processible semantics not necessarily better.**
- **Different communities have different needs;**
- **Will be a role for approaches on all points along the semantic continuum;**
- **Many tradeoffs.**