The Dublin Core and The Semantic Web

October 29, 2001

Stuart Weibel
OCLC Office of Research
Director,
Dublin Core Metadata Initiative

What is the Semantic Web?

- An Idea
- A Vision of Possibilities
- Enabling Infrastructure
- A Community of common standards
- Diverse motivations and objectives

The Idea

- Improve the ability of people to communicate using machines
  - Have machines do more of our work in an automated fashion
  - Conventions about infrastructure
  - Conventions about semantics
  - Use the web to enable it - connectivity
- The W3C activity is an important (but not the only) expression of this idea.

The Possibilities

- “The Semantic Web is a 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.”

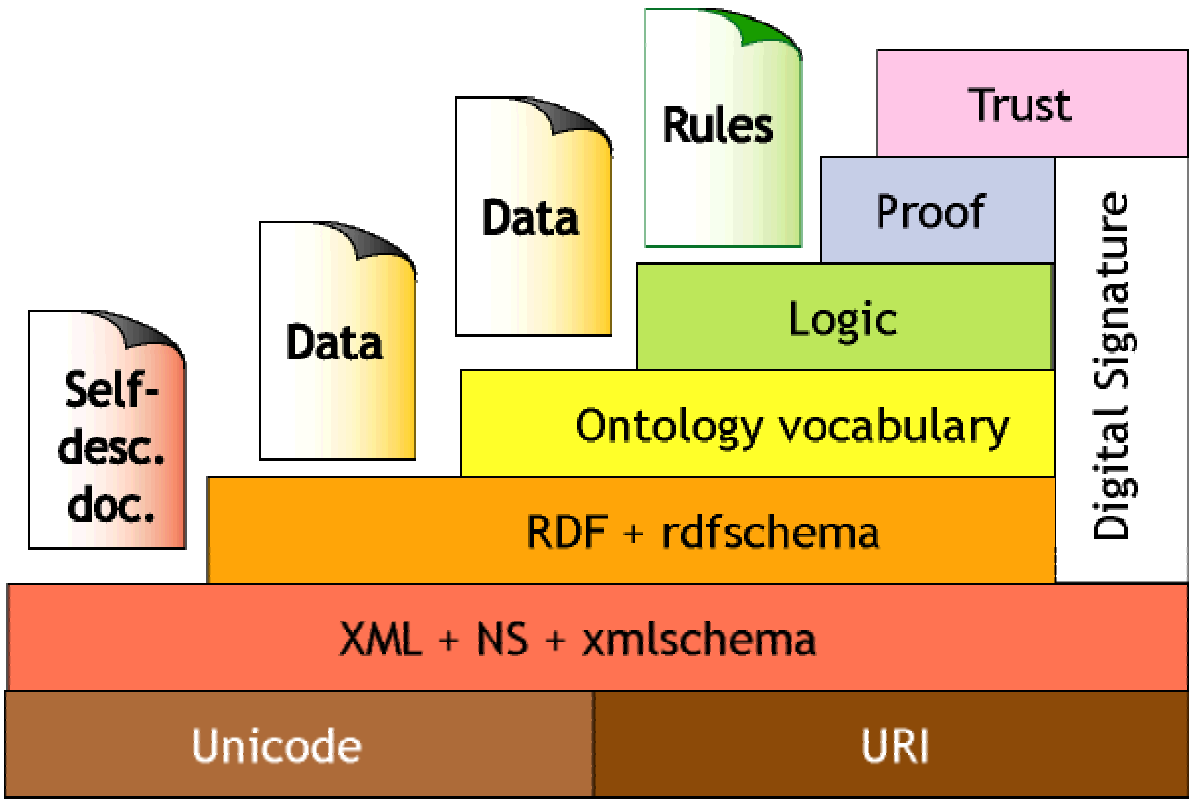
W3C Semantic Web Activity Statement

http://www.w3.org/2001/sw/Activity

W3C Semantic Web Activity Objectives

1. Continue the work of the RDF Interest Group
2. Undertake revisions to the RDF Model and Syntax Recommendation
3. Complete work on the RDF Schema specification including reconciliation of RDFS and XML Schemas.
4. Coordinate with W3C Web technology initiatives (P3P, CC/PP, XML Protocol, WAI, and others)
5. Coordinate with selected non-W3C initiatives DCMI, DAML, OIL, and SHOE….
6. Advanced development and design of supporting XML and RDF technologies

Semantic Web Architecture (by Tim Berners-Lee)



http://www.w3.org/2000/Talks/1206-xml2k-tbl/Overview.html

Semantic Web Challenges

- Machines talking to machines
  - Universal syntactic conventions necessary
  - Semantics need to be defined and declared
  - Agreements about what we are saying and how to structure it are important

- Communication is hard: is there anything about the Web and computers that makes it easier for us to communicate?

Semantic Web Challenges (continued)

- People talking to machines
  - How will people effectively interact with knowledge representation of this type?
- Federating domains - enabling complex tasks based on information from various sources
  - How will a person or machine know where to find appropriate information?
  - How will machines rationalize complicated data that is previously unknown to them?

Semantic Web Challenges (continued)

- Wide scope - from home to global
  - The Semantic Web must be useful and effective over a broad range of devices and networking environments
- Robust - imperfect understanding
  - How will machines do reliable inferencing across data of uncertain quality, created with diverse objectives and motivation?

Parts of the Solution

- Metadata languages
- Controlled vocabularies
- Encoding conventions
- Metadata registries (online metadata dictionaries)
- Schema languages
- Knowledge representation languages
- Ontologies

Dublin Core Metadata Initiative

- The mission of DCMI is to make it easier to find resources using the Internet through the following activities:
  - Developing metadata standards for discovery across domains (example: the Dublin Core)
  - Defining frameworks for the interoperation of metadata sets
  - Facilitating the development of community or disciplinary specific metadata sets

A Grammar of Dublin Core

- http://www.dlib.org/dlib/october00/baker/10baker.html
- Simpler than natural languages, but easy to learn and useful in practice
- Pidgins: small vocabularies (Dublin Core: fifteen special nouns and some optional adjectives)
- Simple grammars: sentences (statements) follow a simple fixed pattern...

The fifteen special nouns (properties)

Creator	Title	Subject
Contributor	Date	Description
Publisher	Type	Format
Coverage	Rights	Relation
Source	Language	Identifier

Varieties of qualifiers: Element Refinements

- Make the meaning of an element narrower or more specific.
  - a *Date Created* versus a *Date Modified*
  - an *IsReplacedBy Relation* versus a *Replaces Relation*
- If your software does not understand the qualifier, you can safely ignore it.

Varieties of Qualifiers: Value Encoding Schemes

- Says that the value is
  - a term from a controlled vocabulary (e.g., Library of Congress Subject Headings)
  - a string formatted in a standard way (e.g., "2001-05-02" means May 3, not February 5)
- Even if a scheme is not known by software, the value should be "appropriate" and usable for resource discovery.

Project SCHEMAS

http://www.schemas-forum.org/

- Bring together metadata designers
- Investigate solutions across domains
- Develop understanding how to mix and match elements sets
- Pioneer machine-readable expression of Application Profiles in RDF/XML
- Develop guidelines

Interoperability

- Standards need to work together to realize the vision of a Semantic Web
- Human interoperability - understanding that “None of us is as smart as all of us”
- Machine interoperability - technical infrastructure - RDF, XML - Registries

Dublin Core & SCHEMAS

Helping to build the Semantic Web by:

- Opening channels to other communities and standardization activities
- Building RDF registries with schemas and application profiles
- Supporting development and validation of tools

DAML Project Page at Darpa

http://dtsn.darpa.mil/iso/programtemp.asp?mode=347

The goal of the DAML program is to create technologies that will enable software agents to dynamically identify and understand information sources, and to provide interoperability between agents in a semantic manner. This goal will be pursued by a research plan that includes … six tasks:

DAML
Project Objectives

1. Create an **Agent Mark-Up Language** (DAML) built upon XML … machine-readable semantic annotations….
2. **Create tools** that embed DAML markup on to web pages and other information sources
3. … build up, instantiate, operate, and **test sets** of agent-based programs that markup and use DAML.
4. Measure, via empirical experimentation, the **productivity improvements** provided by these tools.
5. Apply these tools to … military-specific problems, and support for the intelligence community…
6. **Transition DAML** to the commercial and military markets

OIL
Ontology Inference Layer

- European ontologies research activity
- Frame-based knowledge representation and inferencing technology

- http://www.ontoknowledge.org/oil/

Web Ontology Working Group Objectives

- A Web ontology language, that builds on current Web languages (such as RDFS)… but which extends RDFS to allow more complex relationships between entities…
  - Scoping, typing, rigorous inference
- Support the development and linking of ontologies… in a web-like manner
- Definition of formal semantics allowing language designers, tool builders, and other "experts" to be able to precisely understand the meaning and "legal" inferences for expressions in the language.
- Use XML syntax and datatypes wherever possible, and designed for maximum compatibility with XML and RDF

But… have we been here before?

- Remember the Fifth Generation Project?
- How do we distinguish among realistic objectives and intractable problems?
- Is there anything about the Web that brings universal knowledge representation suddenly within reach?
- Can reliable inferencing be done over data of diverse origins and intent?
- Will standards converge or proliferate?

Strategies for Semantic Web Development

- Be mindful of distinctions between speculative research and practical deployment
- Stay tightly coupled to realistic goals and objectives
- Semantic Web infrastructure may not do everything we hope, but it probably can help us solve practical problems for organizing, discovering, and re-using data and information

Links

- Semantic Web
  http://www.w3.org/2001/sw/
- Dublin Core Metadata Initiative
  http://dublincore.org/
- Project SCHEMAS
  http://www.schemas-forum.org/
- DAML
  http://www.daml.org/
- OIL
  http://www.ontoknowledge.org/oil/

Questions or Comments?

weibel@oclc.org

http://DublinCore.org

Arigatou goziemasu!