Introduction to Semantic Web

Thomas Baker, Fraunhofer-Gesellschaft
CASHMERE-int Workshop
Göttingen, Germany
2005-02-28
This Talk

- The Semantic Web Vision
- Languages of the Semantic Web
  - Dublin Core to RDF and Ontologies
- Semantic Web Applications

Sources for this talk:
- http://www.w3.org/2004/Talks/1013-semweb-em/talk
- http://www.w3.org/2004/Talks/0120-semweb-umich/
The Semantic Web Vision
World Wide Web Consortium

- Directed by Tim Berners-Lee, “Inventor” of Web
- Mission: “Lead the Web to its full potential”
- Hosts: MIT, ERCIM, Keio University
- Develops Web Standards
  - HTML, CSS, XML, Security
  - Web Accessibility Initiative (WAI)
  - Web Services (SOAP, WSDL, etc.)
  - Semantic Web languages (RDF, OWL…)
World Wide Web Consortium

Method

- Technical specifications developed with Working Groups and extensive public review
- Advanced Development to chart long term architectural directions
- Building infrastructure to address technical and social needs of the Web
Semantic Web Vision
Web of Data – the file level

- Circa 1993
- FTP, Gopher, Archie: sharing resources on the Internet
- Stopped at file level
Web of Data – the text level

β Circa 1994
β HTML and URLs
β Below file level
β But stopped at text level
Semantic Web of Data

- And now
- XML, RDF, OWL, URI
- Below file level
- Below text level
- At data level
Semantic Web Vision

 Scenario: Someone needs to see a doctor.

 Tim Berners-Lee, Scientific American 2001

 Handheld devices with “agents” negotiate:

 Doctor’s agent knows prescribed treatment

 List of doctors, rated at least “very good”, within 20 miles, accepted by insurance plan.

 Devices match calendars to fix appointment.

 Semantic Web challenge: provide language that expresses both data and rules for reasoning about that data
Semantic Web Vision

- Web of data
  - Integrating multiple sources to draw new conclusions
  - Architecture for describing all kinds of things (items, collections, services, processes, etc.)
  - Effective management and reuse of data at various scales (personal, group, enterprise, community, web)
Semantic Web Vision

§ URI – Foundation of Semantic Web

§ Identify **anything** distinctly and uniquely by an opaque string of characters in global URI space

§ Physical things – books, cars, people…

§ Digital objects – Web pages, digital images…

§ Conceptual things – colors, subjects, metadata terms!
URIs as anchors for merging data
URIs as anchors for merging data
Languages of the Semantic Web
Simple Resource Description

- Mid-1990s: Web takes off – flood of information
  - “How can we describe this simply?”
- “Dublin Core”: Title, Subject, Date…
  - Pidgin-like “metadata” for describing Web pages
- Library Science meets Computer Science
  - Not enough to agree on words – also need grammar!
  - Machine processing requires agreement on a simple data model
Simple Linked Data Model

Machine-processable statements about **things** (Web pages, organisations, people, concepts, products, etc) and the **links** between them
“Resource Description Framework”

β 1997: Define metadata vocabularies (like Dublin Core) and use them to make statements

β Resource: anything identifiable with a URI

β Description: statements about properties of resources

β Framework: a common model for statements using diverse vocabularies
RDF “triples”

A simple model for “statements”
- Subject: what the statement is about
- Predicate: a property of the subject
- Object: the value of the property

“A natural way to describe the vast majority of the data processed by machines”
Statement about a document using Dublin Core in RDF/XML

```xml
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:dc="http://purl.org/dc/elements/1.1/" >
    <rdf:Description about="http://docs"
                     dc:creator="Joe Smith"/>
</rdf:RDF>
```
URIs to identify properties

- Connecting information to its definitions and its context
  - Dublin Core element Creator is denoted by a URI
    - http://purl.org/dc/elements/1.1/creator
  - Not just any concept of “color”, but one defined in an identifiable context:
    - http://pantone.example.com/2004/std#color
Statements defining the Dublin Core property itself on the Web

<rdf:Property rdf:about="http://purl.org/dc/elements/1.1/creator">
  <rdfs:label xml:lang="en-US">Creator</rdfs:label>
  <rdfs:comment xml:lang="en-US">An entity primarily responsible for making the content of the resource.</rdfs:comment>
  <dc:description xml:lang="en-US">Examples of a Creator include a person, an organisation, or a service. Typically, the name of a Creator should be used to indicate the entity.</dc:description>
  <rdfs:isDefinedBy rdf:resource="http://purl.org/dc/elements/1.1/"/>
  <dcterms:issued>1999-07-02</dcterms:issued>
  <dcterms:modified>2002-10-04</dcterms:modified>
</rdf:Property>
Semantic Web languages today

- RDF – Resource Description Framework
  - See [http://www.w3.org/TR/rdf-primer/](http://www.w3.org/TR/rdf-primer/)
- OWL Web Ontology Language
  - See [http://www.w3.org/TR/owl-features/](http://www.w3.org/TR/owl-features/)
- Core specifications are W3C
  Recommendations as of February 2004:
  - See [http://www.w3.org/2001/sw/](http://www.w3.org/2001/sw/)
Challenge: Independent vocabularies working together

- Need to formally declare how vocabularies relate to each other
  - “MARC:illustrator refines DC:contributor”
  - “RRS:title refines DC:title”
- Relationships can be declared in RDF on Web
- Will require well-understood conventions for declaring and managing vocabularies
Defining Best Practice for Vocabulary Management

- Identify terms with URI references
- Articulate maintenance policies for your vocabulary (semantic stability, persistence)
- Identify historical versions
- Declare your terms using a formal, machine-processable schema language
Semantic Web Applications
Evolution towards a Semantic Web

- Not revolution
- Emphasis on methods for extracting existing data from documents, servers and databases
- Translate into common form using Semantic Web languages
Personal Information Environments

- **Haystack** - User configurable universal information client
  - Benefits from RDF’s universal information model
- Uses RDF for personalization, data, layout, preferences, etc.

http://haystack.lcs.mit.edu/
Integrating Enterprise Data

- Tucana - Enterprise Information Integration
- Expose diverse data sources as RDF
- Scalable back-end storage

http://www.tucanatech.com/
Aggregating Knowledge Bases

- **TAP**
- Simple tools that treat the Web a giant distributed database.
- Local, independently managed knowledge bases can be aggregated

http://tap.stanford.edu/
Visualizing Social Networks

FoafCorp: visualizing corporate boards of directors

http://rdfweb.org/foafcorp/intro.html
Web’ifying Thesauri

- SKOS – an RDF vocabulary for thesauri
  - Encoding and mapping of thesauri, controlled vocabularies
  - Bridges library classification and Web technologies
  - [http://www.w3.org/2004/02/skos/core/guide/](http://www.w3.org/2004/02/skos/core/guide/)

- A product of SWAD-Europe - Supporting Semantic Web standards in Europe
  - Targeted research, software applications, outreach
  - Topics: calendaring, social networks, images, geospatial, internationalization, RDF storage, etc.
Bridging Institutional Repositories

- Project Simile - Semantic Web meets Digital libraries and personal information management
  - Partners: W3C, HP, MIT Libraries, MIT CSAIL
- Implement a digital asset management architecture based on Web standards
  - Add useful “views” to a digital artifact and bind those views to consuming services
- Leverage and extend DSPACE, enhancing its support for managing heterogeneous data
“RDF Browsers”

- Common interface, framework for navigation
- Architecture supports integration of heterogeneous data sources
- Tools for exposing content collections in RDF

http://simile.mit.edu/
Integrating Life Science Data

- Connecting information: gene, diseases, cures
- Scientists in different locations, working on different problems, integrating results into coherent whole
- Recognized need for effective data integration from heterogeneous collections
- Increasingly available datasets in RDF
- Increasing scientific / vendor interest
W3C Semantic Web Activity

Phase 2

Feb 2004: Core specs are W3C Recommendations
- Open Standards and Open Source tools, technologies for modeling real world resources; sharing these models across the Web.

Mar 2004: Phase 2 launched
- RDF Data Access - “Joining the Web”
- Best Practices and Deployment
- Advanced Development

Deployment / Facilitating 'Network Effect'
W3C Semantic Web Activity
RDF Data Access (DAWG)

To define an HTTP and/or SOAP protocol for selecting instances of RDF

‘Join’ Web data as easily as merging tables in a local relational database.

Use Cases

Personal Information Management, transportation, tourism, product life-cycle data management, publishing…

Outputs

Use Cases and Requirements - Aug 2004
SPARQL - Query Language Specification - Oct 2004
To provide guidance for developers of Semantic Web applications.

Best Practice notes for ontology engineering, vocabulary development...

Educational material and demo applications.

Support transformation of ontologies and thesauri to RDF/OWL

http://www.w3.org/2001/sw/BestPractices/
W3C Semantic Web Activity
Advanced Development

- Collaborative development
  - Creation of core components (e.g. libwww) that will form the basis for the Semantic Web.

- Facilitate Semantic Web deployment and identify futures areas of standardization

http://www.w3.org/2000/01/sw/
A Common Framework

- Many organizations are starting to realize they need ‘digital libraries’
- Even if they don’t call it that
- Common data representation
  - Common description framework and architecture reduces (technical / social) costs and is more efficient
- Everyone benefits
Towards the technology of trust
Conclusions

- Core Specifications in place
- More applications / toolkits / software every day
- A new class of Semantic Web applications at individual, enterprise, and Web scale
- Semantic Web based on a new type of human and machine language
Additional Information

- W3C World Wide Web Consortium
  - http://www.w3.org
- Semantic Web Initiative Home Page
  - http://www.w3.org/2001/sw/
- Dublin Core Metadata Initiative
  - http://dublincore.org/
- Eric Miller, W3C Semantic Web Activity Lead
  - http://www.w3.org/People/EM/
- W3C Semantic Web Best Practices and Deployment Working Group
  - http://www.w3.org/2001/sw/BestPractices/
thomas.baker@izb.fraunhofer.de