Towards the Semantic Web

Ora Lassila

Research Fellow, Nokia Research Center (Boston)
Chief Scientist, Nokia Venture Partners LLP
Advisory Board Member, W3C

W3C Semantic Tour, London, June 2003

Towards the Semantic Web

• Motivation
• Technologies
  • representation formats
  • ontologies
  • Web Services
• Future
• Q & A
Motivation for the Semantic Web

• **Original driver: Automation**
  - it would be nice if computers could do more (on the Web)
  - solution: make information on the Web more “machine-friendly”
  - origins of the Semantic Web are in web metadata

• **Short term goal: Interoperability**
  - combining information from multiple sources
  - Web Services: discovery, composition
  - “serendipitous” interoperability

• **Long term goal: “Departure from the Tool Paradigm”**
  - instead of using computers like tools, make them work *on our behalf*
  - removing humans from the loop to the extent possible

More About the Motivation

• **Problem: Web was built for humans**
  - human interpretation needed to “understand” content (this does not scale)
  - consequently, automation is difficult
  - it is particularly difficult to automate “unforeseen” situations

• **Rough solution: make the Web friendlier for machines**
  - we need “machine-understandable” content (not “machine-readable”, we already have that)
  - “machine-understandable” means content with accessible formal semantics

• **The Web is more than just a “library”**
  - think of it as infrastructure for services & functionality

• **Drivers**
  - automation (e.g., in search), interoperability (e.g., in e-commerce)
  - but: compelling business models may still be missing
WWW: an Architecture for Linkages

- Current Web architecture essentially gives us a framework for “pointing”
- Problem is that this pointing has no meaning
  - (except sometimes through human interpretation)

Can we improve on this?

Note: for us (humans), separating our own interpretation from (largely syntactic) representation is hard

Linkages on the “Old Web”
Linkages on the “Semantic Web”

Alice

trusts

meaning of “trusts”

works-with

meaning of “works-with”

Bob

Ora

Linkages on the “Semantic Web” (2)

• **Semantic Web resources (the “nodes”) can**
  • stand alone, or
  • denote other things (e.g., physical entities)

• **Hypertexts become “semantic” networks**
  • this is good for agents and automation
  • e.g., semantic navigation of hypertexts
  • how does one “name” the semantic links and nodes?
Semantics via Sharing

• **Controlled vocabularies**
  • interoperability improves if the same term is always used to denote the same thing (e.g., instead of arbitrary keywords, choose from a list)

• **What is an “ontology”**
  1. a controlled vocabulary
  2. a concept taxonomy
  3. other relations between concepts
  • definition: “A specification of conceptualization” (Gruber)

• **Library scientists are good with this stuff**
  • e.g., Dewey Decimal System is an ontology

More about Ontologies

• **How to build ontologies?**
  • we could form committees…
    • the Dublin Core initiative took years to decide on 15 core metadata elements
    • (this doesn’t mean that DC wouldn’t be a good standard)
  • my preference is the “Darwinian” approach
    • good and/or popular ontologies will prevail
    • we must have a framework which allows ontology extension (RDF does)
  • probably some combination of official standards and de-facto standards is the way to go

• **Some “upper ontology” projects underway**

• **Ontologies enable reasoning**
  • this allows the move from “syntactic” to “semantic” processing
  • but: where does “semantic data” come from?
Resource Description Framework

- Originally conceived as W3C’s metadata model
  - document metadata for digital libraries, content rating, site maps, etc.

- RDF has
  - a data model of directed labeled graphs (DLGs)
  - an XML-based syntax for serializing DLGs

- Nodes & arcs in an RDF DLG are named by URIs
  - important for robust vocabulary creation

“It’s a Model, Stupid!”

- Simple data model
  - think of it either as directed labeled graphs or in object-oriented terms
  - more powerful than the trees XML gives you

- Graphs decompose into object/attribute/value -triples
  - “subject/predicate/object” = a statement
  - (in RDF parlance, nodes are called “resources” and arcs “properties”)

- Everything in an RDF graph is named by URIs
  - when naming is not based on mere words, name conflicts can be avoided
  - graphs can span multiple hosts (servers, etc.)

- RDF is followed by more powerful languages
  - DAML+OIL (from the DARPA Agent Markup Language program)
  - OWL (from W3C’s WebOnt working group)
W3C Web Ontology Language (OWL)

- Layered on RDF, offers more expressive power
- Comes in three “flavors” (layered on one another)
  - OWL Full
  - OWL DL
  - OWL Lite
- OWL DL is particularly attractive for “real world” applications
  - complete & decidable
  - good reasoning engines exist (e.g., from Network Inference)
- OWL Lite is what RDF Schema should have been
  - simple yet expressive language for concept taxonomies
- Standardization work almost done

Stepping Towards the Semantic Web

- Semantic Web is built in a layered manner
- Not everybody needs all the layers
Isn’t It Enough to Just Use XML?

• Short answer: no
  • the typical answer – albeit incorrect – is “yes”

• Long answer: XML offers a way to introduce new syntax (new names, tags, …), but no way of introducing or coordinating semantics

• XML has a tree-like data model
  • if your (representational) problem does not lend itself to be a tree, you lose (sorry – this is even before we get to the “semantics” part)

• Hype (from a major IT company white paper): “The industry is clearly focusing in on [XML] as the lingua franca to enable Web services…”
  • not only is XML not a lingua franca, it is not even a lingua

XML ≠ Machine Accessible Meaning (1)

• (thanks to Frank van Harmelen, VUA)
XML ≠ Machine Accessible Meaning (2)

XML ≠ Machine Accessible Meaning (3)
Reasoning and Inference

- Reasoning allows one to draw inferences based on generalized "rules"
  - generation of "more" semantic information
  - simplest practical form: polymorphism in OO systems

- Enabled by ontologies

- Reasoning eases interoperability
  - relationships between different but compatible ontologies & data could be inferred

Reasoning example:
1. X is a Cat
2. a Cat is a Mammal
3. a Mammal gives birth to live young

- X gives birth to live young

Note: This is AI
Semantic Web: Characterizations

Ontological approaches (RDF, OWL, etc.)

“RDF chauvinism”

“Weak” Semantic Web (uniform data models, useful manipulation)

“Strong” Semantic Web (logic & reasoning)

Syntactic approaches (“plain” XML)

Interoperability of Services

• Semantic Web, via ontologies and reasoning, will improve interoperability of information systems

• This can be applied to “services”
  • semantic description of service interfaces enables automatic discovery, composition, etc.
    • DARPA’s DAML-S activity (Stanford, CMU, Yale, SRI, BBN, Nokia)
    • analog to “Tower of Babble” (from Genesis 11:1-9)
    • will Web Services succeed without the Semantic Web? (I think not)

• Substitution of “equivalent” services

• Web Services are a good abstraction of all kinds of functionality
Tools & Beyond (examples)

Tools
- hammer & nails
- calendaring software
- e.g., Google
- almost any software today

Beyond tools
- building contractor
- automated “secretary”
- answers from a “search agent”
- various personal assistants…

Fulfillment of the Vision (Using AI)

- Knowledge representation
  - (obvious: the Semantic Web is all about KR)
  - formal semantics as “the Manifest Destiny of AI”

- Automated planning
  - enables autonomous operation
  - useful in many tasks (e.g., service composition)

- Machine learning
  - enables adaptivity
  - could be used in bootstrapping semantic annotations for existing content

- The “AI Paradox”
  - well-understood things stop being AI (e.g., OOP, rules, logic)
  - parallels between AI and the Semantic Web: the latter also has aspects which, once adopted, will stop being “Semantic Web”
Summary

• Use of human interpretation does not scale

• We need to
  • move from tools to autonomous systems that work on our behalf
  • introduce formal semantics (machine-understandable content)

• Ontologies • Reasoning • Agents
  • we have only done the first step and started on the second…
  • (business models for all this are needed)

• We may need artificial intelligence to ultimately fulfill the Semantic Web vision
  • (some of you may have been misinformed about this earlier)

Questions?

• http://www.nokia.com/research/semanticweb
• mailto:ora.lassila@nokia.com