Current Status and Future Promise of the Semantic Web

James Hendler   and   Ora Lassila

University of Maryland
College Park, MD

Nokia Research Center
Cambridge, MA

NOKIA
Connecting People

Before We Get Started…

- Two views of the Semantic Web:
  - implementing <u>SEMANTIC</u> applications using <u>web</u> technologies
  - using <u>semantic</u> technologies to support new <u>WEB</u> applications
- Exploring the relationship (and *tension*) between the two over time
- This talk is
  - a retrospective, a status report
  - an "interpretation"
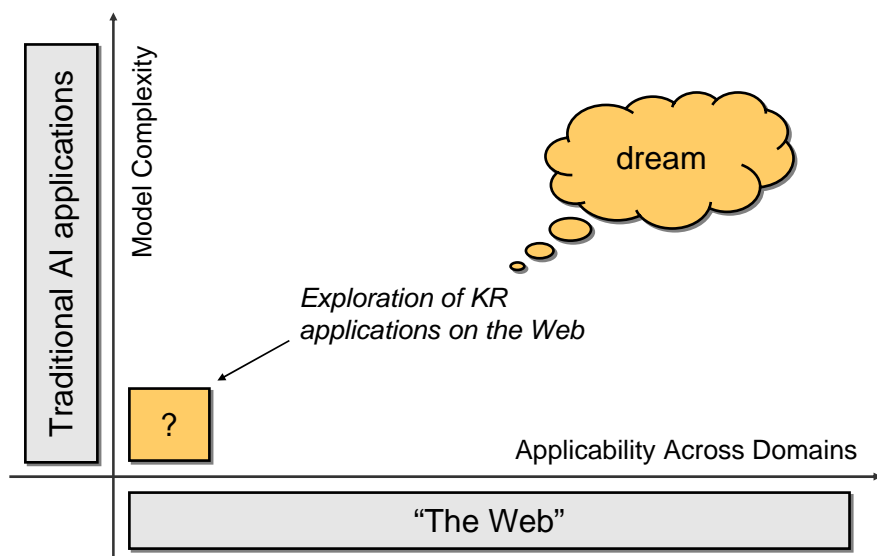 - some thoughts on the future…

NOKIA
Connecting People

1990's: "Pre-history"

- Rebirth of Artificial Intelligence (end of "AI Winter")
  - "big" AI applications
    - Deep Blue, Mars Rover, Deep Space 1, …
    - embedded vs. stand-alone
  - Web AI
    - IR, statistical NLP, machine learning
    - lots of data!
- Emergence of the Web
  - new ways of doing things
    - new business models (even new *social* models)
    - new technology
  - "dot-com" boom
- Early forays into "meta-content"

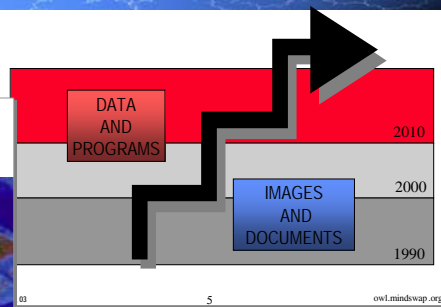**NOKIA**
Connecting People

1990's: "Pre-history"

Traditional AI applications

Model Complexity

dream

Exploration of KR applications on the Web

?

Applicability Across Domains

"The Web"

2000-2001: What Did We Believe?

- Jim: Semantic Web and the advent of pervasive computing (March 2000)

What about the rest of the Web?

TetherlessDAML
The Mobile Semantic Web

Get Info
Find Restaurant
Buy Car
Buy TV
Find Job
Buy Airline Ticket

http://www.daml.org

DATA AND PROGRAMS

IMAGES AND DOCUMENTS

2010
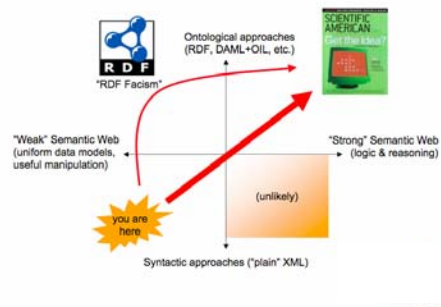2000
1990

owl.mindswap.org

- Jim: Roadmap from the "old" Web to the Semantic Web (October 2001)

2000-2001: What Did We Believe?

- Ora: Semantic Web and the advent of pervasive computing (June 1999)

Semantic Web: Characterizations

Ontological approaches (RDF, DAML+OIL, etc.)

RDF
"RDF Facism"

SCIENTIFIC AMERICAN

"Weak" Semantic Web (uniform data models, useful manipulation)

"Strong" Semantic Web (logic & reasoning)

(unlikely)

you are here

Syntactic approaches ("plain" XML)

Why Your Microwave Oven Needs (to Surf) the WWW

Ora Lassila

Research Manager
Agent Technology Group, Nokia Research Center / Boston
&
Member of the Advisory Board
World Wide Web Consortium (MIT/LCS)

June 1999

- Ora: Roadmap from the "old" Web to the Semantic Web (October 2001)
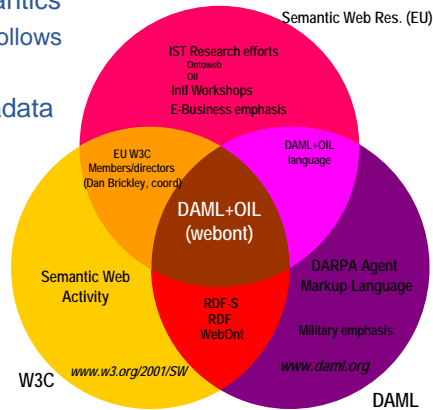
NOKIA
Connecting People

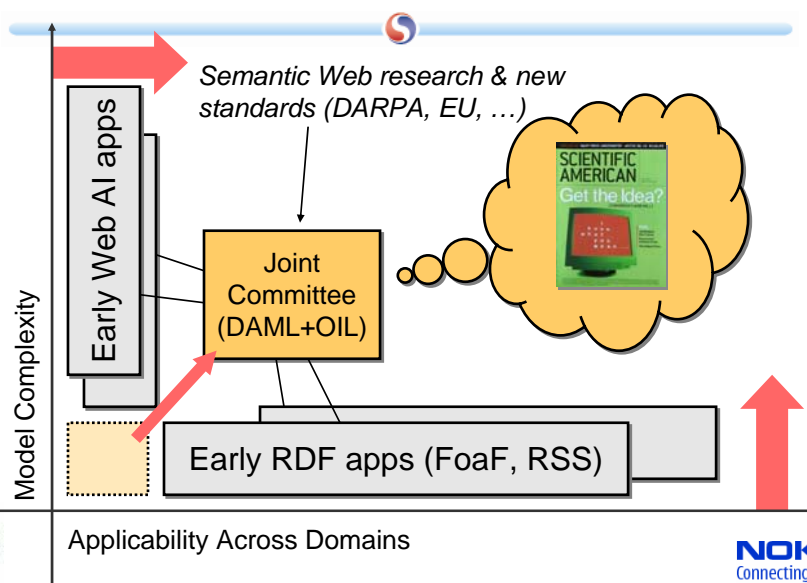# 2000-2001: "Early Years"

- "Dot-Com" optimism still prevails:
  easy to explore new directions
  - Government meddles with semantics
    - DARPA's DAML program; EU follows
    - DAML+OIL
  - Web community discovers metadata
    - W3C Metadata Activity
    - RDF

"3-pronged" attack:
- DARPA
- EU IST
- W3C

Semantic Web Res. (EU)

IST Research efforts
Ontoweb
Oil
Intl Workshops
E-Business emphasis

EU W3C
Members/directors
(Dan Brickley, coord)

DAML+OIL
language

DAML+OIL
(webont)

DARPA Agent
Markup Language

Semantic Web
Activity

RDF-S
RDF
WebOnt

Military emphasis

W3C

www.w3.org/2001/SW

www.daml.org

DAML

2000-2001: "Early Years"

Semantic Web research & new standards (DARPA, EU, …)

Early Web AI apps

Joint
Committee
(DAML+OIL)

SCIENTIFIC
AMERICAN
Get the Idea?

Model Complexity

Early RDF apps (FoaF, RSS)

Applicability Across Domains

NOKIA
Connecting People

4

Original Outline (July 2000)

Scientific American Article notes

[Joint starting place:]

I. Semantic Web Vision (TBL)

II. What are the enablers? (in sequence)
Screen Scraping (Ora and TBL)
Data on Web (Ora and TBL)
Zip code link between Data Bases (TBL)
Ontology Independence (JAH)

Effect of Scale (TBL)

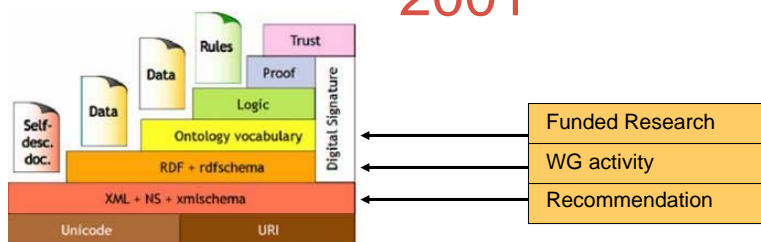**"Then, a miracle occurs"**

III. What can you do with it? (not necessarily in sequence)
Self-describing documents (JAH)
Logic to encode... (TBL)
Services and Advertising (Ora)
Devices (Ora)
Digital Signatures, Authentication, and Trust (TBL)

SCIENTIFIC AMERICAN
Get the Idea?
[TOMORROW'S WEB WILL]
i know what you mean...

NOKIA
Connecting People

2001

Trust
Rules
Proof
Data
Logic
Self-desc. doc.
Data
Ontology vocabulary
Digital Signature
RDF + rdfschema
XML + NS + xmlschema
Unicode | URI

Funded Research
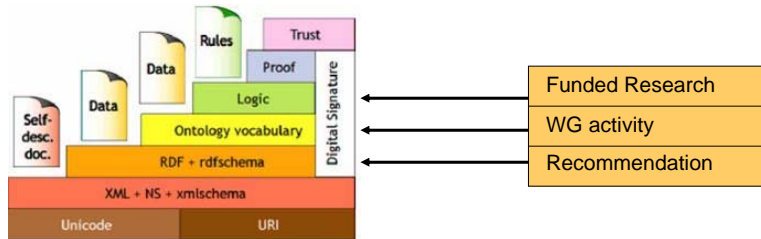
WG activity

Recommendation

Semantic Web Today

• The Semantic Web of 2002 resembles the early days of the World Wide Web
  ◆ Development funded primarily by Govt, but emerging corporate interest
  ◆ A lot of excitement, but confusion as to business case
  ◆ Open source tools and "geeks in control"
  ◆ Standards starting to stabilize to point where they permit deployment
  ◆ Developer tools, libraries, languages

10

• Research, experimentation, early demonstrations
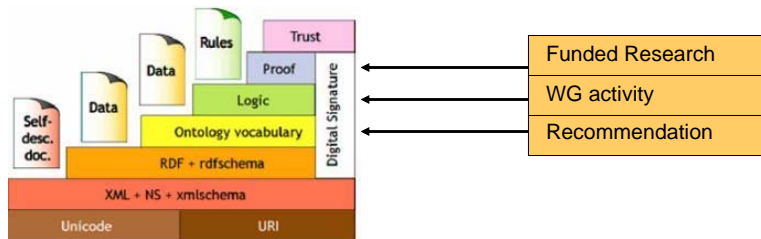• Reminiscent of the early days of the Web

2003

Trust
Rules
Data
Proof
Data
Logic
Self-desc. doc.
Digital Signature
Ontology vocabulary
RDF + rdfschema
XML + NS + xmlschema
Unicode
URI

Funded Research
WG activity
Recommendation

- Early government adoption
- Emerging corporate interest

Semantic Web Today

The Semantic Web of 2002 resembles the World Wide Web

- Development funded primarily by Government
- A lot of excitement, but confusion as well
- Open source tools and "geeks in control
- Standards starting to stabilize to point
- Developer tools, libraries, languages
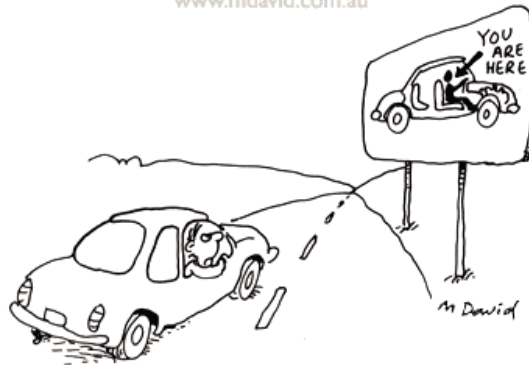
10

"Our" Semantic Web

- Jan 1, 03: Crawler finds 5.8M+ DAML statements on 20,000+ web pages
  - Doesn't include many instance KBs tied to ontologies
  - Doesn't include many very large RDFS-based KBs that include some OWL
- Ontology library at http://www.daml.org has 195 ontologies (March 2003)
  - Open for anyone to create
  - Open for anyone to use
- OWL is being supported by large corporation labs
  - Web tool developers: IBM, HP, Sun, Intel, Fujitsu
  - Content providers: Daimler-Chrysler, Nokia, Motorola, EDS, Agfa
- OWL is starting to be used by thesaurus developers
  - C.f. National Cancer Institute metathesaurus released in OWL Lite
  - United Nations Standard Product Codes available in DAML
  - NASA thesaurus available in DAML
- Use of semantic markup for Web Services beginning to move beyond basic research
  - DAML-S cited as required reading for Web Services Composition WG

Sanken, 03   23   www.mindswap.org

2005

Trust
Rules
Data
Proof
Data
Logic
Self-desc. doc.
Digital Signature
Ontology vocabulary
RDF + rdfschema
XML + NS + xmlschema
Unicode
URI

Funded Research
WG activity
Recommendation

- Commercial tools
- Lots of open source software
- Scalability

2006: You Are Here!

Then a Miracle Occurs…
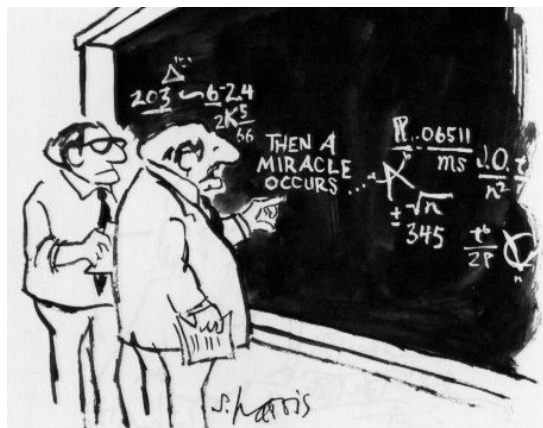
"I think you should be more explicit here in step two."

from *What's so Funny about Science?* by Sidney Harris (1977)

Significant Corporate Activity

- Semantic (Web) technology companies starting & growing
  - Joost, Radar Networks, MetaWeb, Siderean, SandPiper, SiberLogic, Ontology Works, Intellidimension, Intellisophic, TopQuadrant, Data Grid, …
- Bigger players buying in
  - Adobe, Cisco, HP, IBM, Microsoft, Nokia, Oracle, Sun, Vodaphone… announcements/use in 2006-2007
  - integrator and contractor uptake: Northrop Grumman buys Tucana, Lockheed-Martin uses SiberLogic in FCS, WebMethods buys Cerebra, …
  - tools being announced: AllegroGraph, TopBraid, …
- Government projects in and across agencies
  - US, EU, Japan, Korea, China, …
- Life sciences/pharma an increasingly important market
  - Health Care and Life Sciences Interest Group at W3C
- Many open source tools available
  - Kowari, RDFLib, Jena, Sesame, Protégé, SWOOP, Onto(xxx), Wilbur, …

NOKIA
Connecting People

Significant Corporate Activity

50+ Semantic Web press releases within one month!

# Growing Government Activity (US&EU)

- Agencies moving beyond the "talk" phase
  - primarily prototyping, but first acquisitions starting
- Example:
  - NASA is developing an enterprise data strategy around using existing data via Semantic Web integration

Lots of activities across NASA

- Science, Engineering, and Mission all have SWT production or development efforts in place
- Now focus in on re-using the data systems we already have in place
- Agency wide integration planning is underway for building a federation of models into an integrated information service across all disciplines
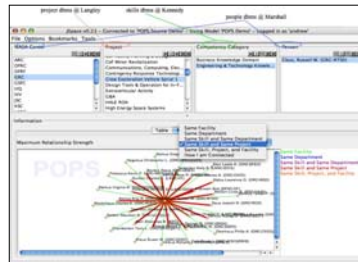
(A. Schain, 3/06)

There's a Lot Out There!

2,120,000 hits on "RDF filetype:rdf"

Paid ads

13,600 hits on "ontology filetype:owl"

(March, 2006)

9

More OWL Use

http://swoogle.umbc.edu
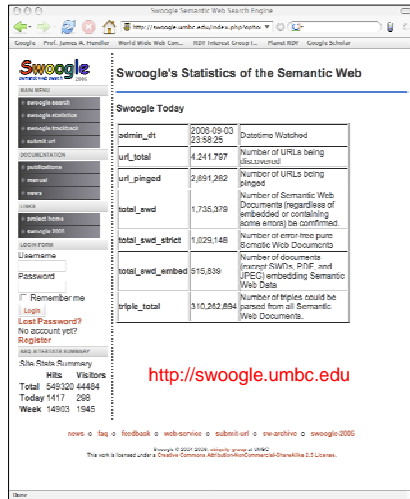
- **The OWL namespace has been declared by 113,000 SWDs (8%) and actually used by 108,000 (7%)**
- **The RDFS namespace enjoys more use, being declared by 677,000 (47%) and used by 538,000 (37%) SWDs**
- **owl:Class is the most used term from the OWL namespace with about 1,800,000 instantiations in 68,000 SWDs**
- **significant use of two OWL equality assertions: owl:sameAs (280,000 assertions in 17,000 SWDs) and owl:equivalentClass (70,000 assertions in 4,300 SWDs) – their common use may be an indication of increased ontology alignment**

(from Ebiquity blog, Sept 1, 2006)

Semantic WEB
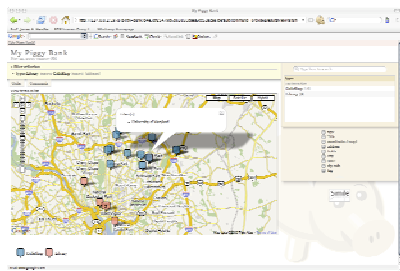
Rich metadata

Data harvesting & visualization

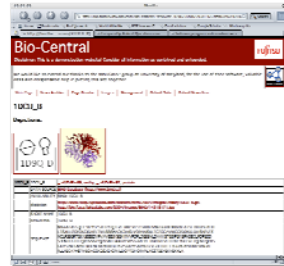A little **Semantics** goes a long way

Web-based social networks

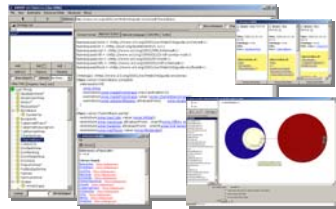# SEMANTIC Web
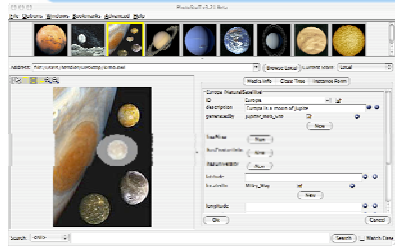
Digital asset management

Scientific portals

Tools for developers

A little **Web** goes a long way

NOKIA Connecting People

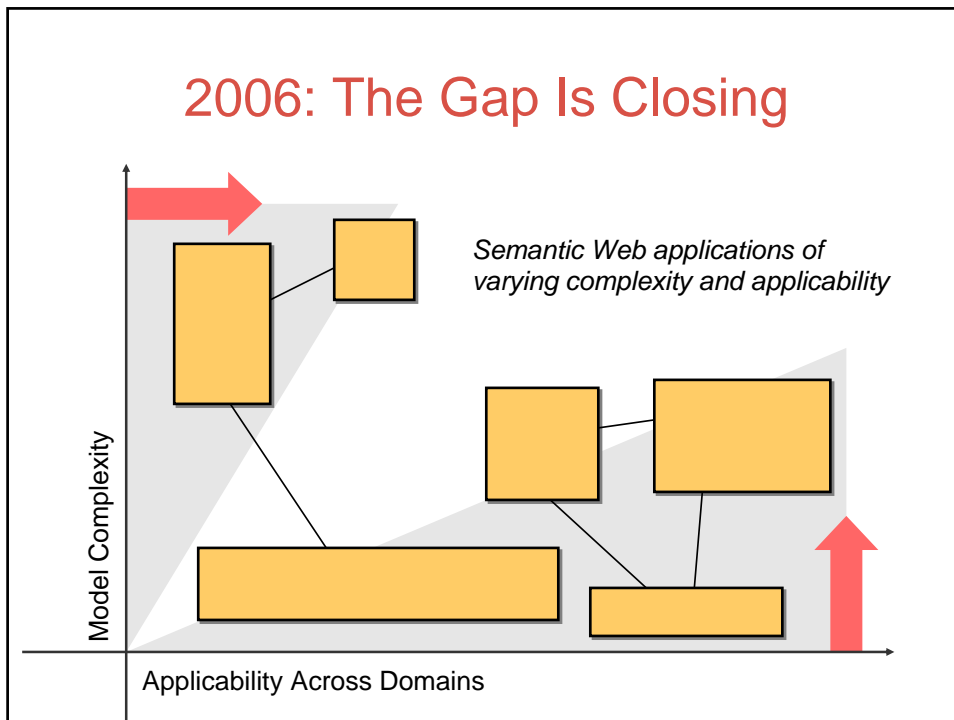# Enterprise Information Integration

- Deployment of semantic technologies is easier in a "controlled" environment
  - such as a corporate intranet
- Key benefits from Semantic Web Technology:
  - reuse of installed clients and servers
    - careful design of SW languages for Web compatibility
  - leave data in place, integrate through an RDF store
    - analogous to 3-tiered Web application
  - heterogeneity supported by ontologies

"Corporate Semantic Web", Gartner "hot pick" for 2006

NOKIA Connecting People

2006: The Gap Is Closing

Semantic Web applications of varying complexity and applicability

Model Complexity

Applicability Across Domains

SEMANTIC Web Lessons

- What we learned from AI…
  - embedded AI succeeded, stand-alone did not
  - tools are hard to sell
  - reasoners are a *means*, not an end
  - knowledge engineering bottleneck
- …applied in the Web context
  - futureproofing
    - URIs are important
  - good standards evolve
    - languages (RDFS, OWL, RIF, …)
    - content!

NOKIA
Connecting People

Semantic <u>WEB</u> Lessons

- Web needed high value sites
  - personal (homepages, pets)
  - public (hobbyists, govt)
- As these linked up, new functionality emerged
  - Yahoo, Alta Vista, …
- New business models followed…
  - "give it away" (Netscape)
  - marketplace (Amazon)
  - advertising (Yahoo, Google)
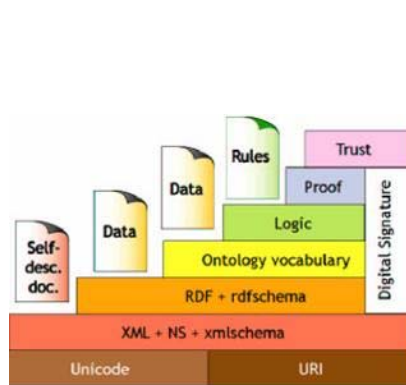- Semantic Web?
  - **SHARE; GIVE IT AWAY!**
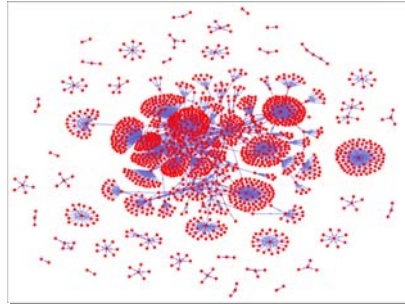
- What do we need?
  - Open Source Tools
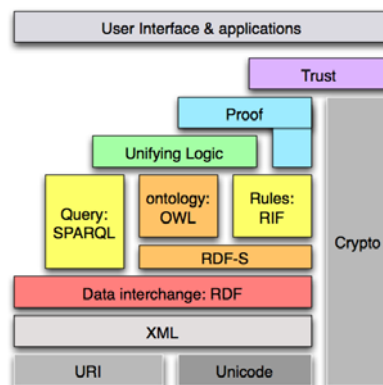 - Open Source Datasets
  - Open Source Harvesters

NOKIA Connecting People

The "Layer Cake" is Evolving…


(Tim Berners-Lee)
2001

(Tim Berners-Lee)
2006

NOKIA Connecting People

New Languages Underway

- RIF: Rules Interchange Format
  - representing rules on the Web
  - linking rule-based systems together
- SPARQL: Query language for (distributed) triple stores
  - the "SQL of the Semantic Web"
- GRDDL/RDFa: Integration of HTML and Semantic Web
  - "embedding" RDF-based annotation on traditional Web pages
- OWL: New features, specialized subsets
  - RDF++/OWL Mini – simplification, identity, scaling to large datasets
  - OWL 1.1 – additional expressivity for OWL constructs
- And more…
  - multimedia annotation, Web-page metadata annotation, Health Care and Life Sciences (LSID), privacy, etc.
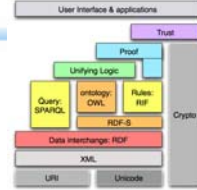
NOKIA
Connecting People

Linking Is Power!

Model Complexity

Applicability Across Domains
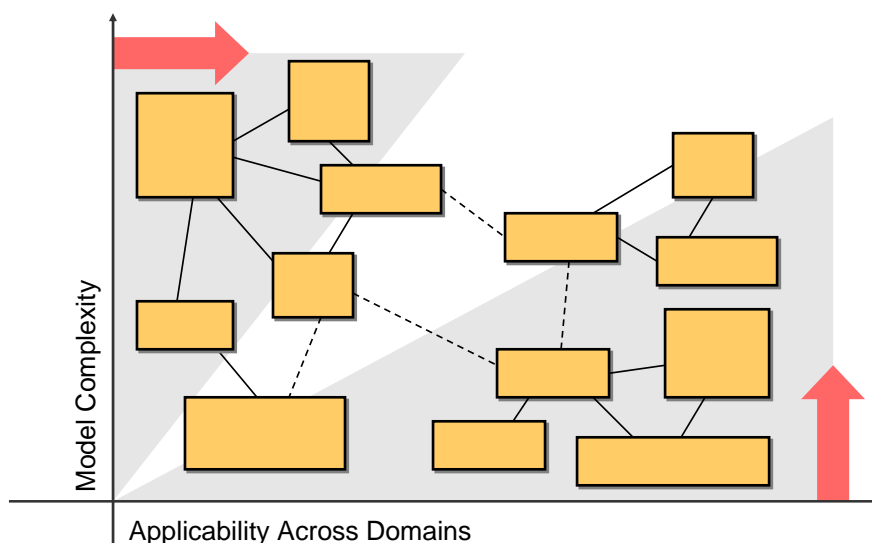
Semantic Web vs. "Web 2.0"

- Data with formal semantics
  - RDF, OWL
  - SPARQL, RIF
- Spontaneous information integration (finally!)
- Semantic Web services, agents
- Strong emphasis on open standards

- New social phenomena: blogs, wikis, tagging, folksonomies
- New user interfaces
  - AJAX (or: "Rich User Experience")
- "New" kinds of data
  - microformats, RSS
  - "mash-ups"
  - Web services
- Plays "fast & loose" with standards

NOKIA
Connecting People

Semantic Web & "Web 2.0"

- What is their relationship?
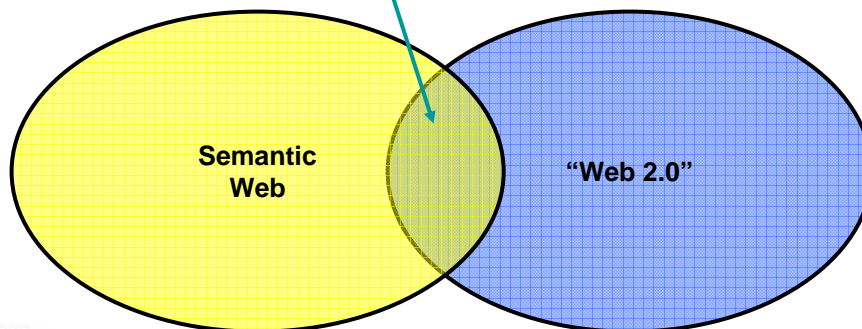- Will they stay separate? Does that even make sense?

Semantic Web   vs.   "Web 2.0"
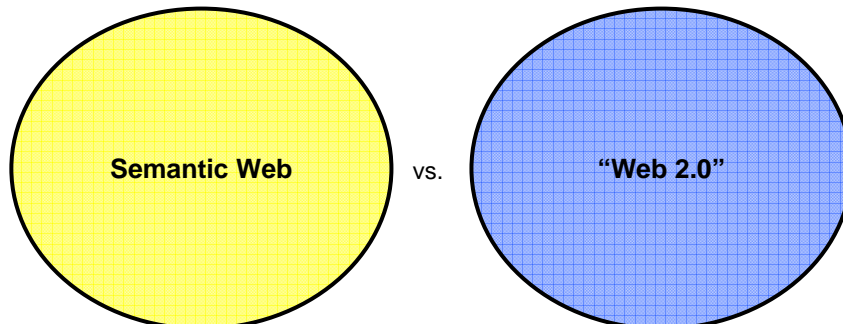
NOKIA
Connecting People

Semantic Web & "Web 2.0"

- NO! Considerable synergies exist

Semantic Web | "Web 2.0"

Exploiting "Web 2.0"

- Vast amounts of "semi-engineered" knowledge
  - Flickr: tens of millions of keyword-tagged photos
  - microformatted Web documents
  - Wikipedia: thousands of carefully documented subjects (in a hierarchy, with disambiguation, …)
- Generate "persistent" URIs
  - "Tank" http://en.wikipedia.org/wiki/Tank (armament)
  - "Tank" http://en.wikipedia.org/wiki/Tank%2C_Pakistan (small town in Pakistan)
- Remember: Anything with a URI can be linked to the Semantic Web!

Linking of "Web 2.0" & Semantic Web

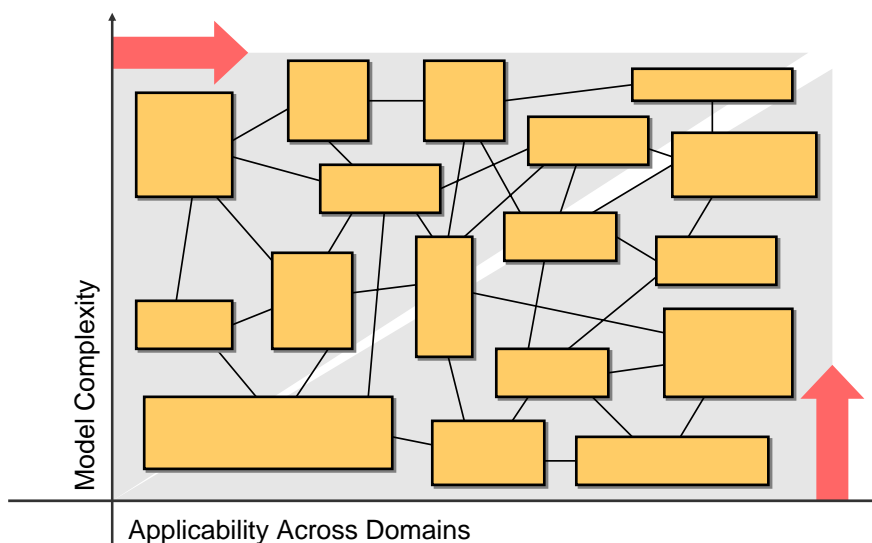- Using informal Knowledge Engineering (KE) to bootstrap "formal" KE
- Extending formal KE from tags/wiki

Looking Further Out

Model Complexity

Applicability Across Domains

Where Are the Agents?

- "Brave New Applications"
  - operate autonomously in "unanticipated" situations
  - exhibit robustness in the face of
    - changing, inconsistent and unexpected data
    - variations in reliability, trust
  - capable of serendipitous behavior, opportunism
- Move from the "tool use" of personal computing to systems that *work on our behalf*
- (Semantic) Web services as "plumbing" for agents
  - emerging as we speak…

Pervasive Computing & Semantic Web

- Pervasive Computing is an interoperability nightmare!
  - instead of sometimes connecting a handful of devices, dynamically connect/disconnect/reconnect possibly hundreds of devices
- Today, high cost of ensuring interoperation
  - any interaction has to be specifically designed/engineered
  - heavy emphasis on application-specific standardization
  - spontaneous interoperability is next to impossible
- The vision is largely contingent on getting unanticipated "encounters" of devices to work
  - how do you behave in a situation not covered by a standard?
  - not "future-proof"
- Semantic Web is a good match
  - (it is an "interoperability technology")

NOKIA
Connecting People

Other Emerging Trends

- Semantic Web Services
  - crucial for linking "programs" into the mix
  - "plumbing" for agents…
- Scaling Semantic Web stores to database sizes
- Information extraction and semantics ("Web 3.0")
  - can we "retrofit" semantics on the existing Web?
- Semantic Web information creation
  - can we make it so we don't have to retrofit in the future?
    - tools that help embed the semantics as a document is created
    - better dynamic integration of structured data into the Semantic Web
  - "Semantic Desktop"

NOKIA
Connecting People

Summary

- Most things we predicted have happened
  - (or are happening at the moment…)
- Some things happened faster than we anticipated
  - triple store scaling
  - reasoner performance *actually matters*
  - ontologies are there (but very little linking)
- Some things are yet to materialize (but we are hopeful)
  - public information sources (as RDF, OWL, …)
  - digital convergence, pervasive computing just emerging
  - little progress on agents

Now go out there and make some money off this…!

Any Questions?

The Semantic Web

"A Little **Web** Goes A Long Way"

Model Complexity
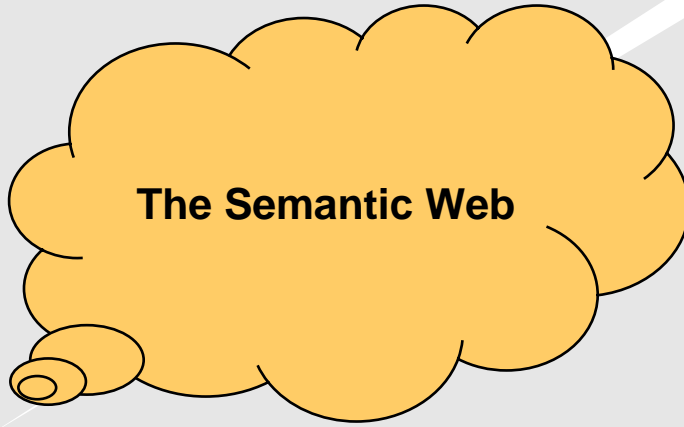
Applicability Across Domains

"A Little **Semantics** Goes A Long Way"

NOKIA
Connecting People