# The Vision of the Semantic Web

Michael Sintek
sintek@dfki.de

*DFKI GmbH Kaiserslautern &*
*Stanford University*

MedCIRCLE Workshop
Brussels, September 12, 2002



## The Semantic Web

- coined by Tim Berners-Lee (1997)
- "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."
  - *T. Berners-Lee, J. Hendler, O. Lassila, "The Semantic Web",* Scientific American, May 2001

## The Conventional Web

- Invented by Tim Berners-Lee in 1989
- Central idea: combination of
  - Hypertext
  - Internet
- URI: global identifiers
- HTTP / HTML: protocol for / representation of hypertext documents
- Allows browsing of (mainly static) information with help of search engines (like Google) and portals/directories

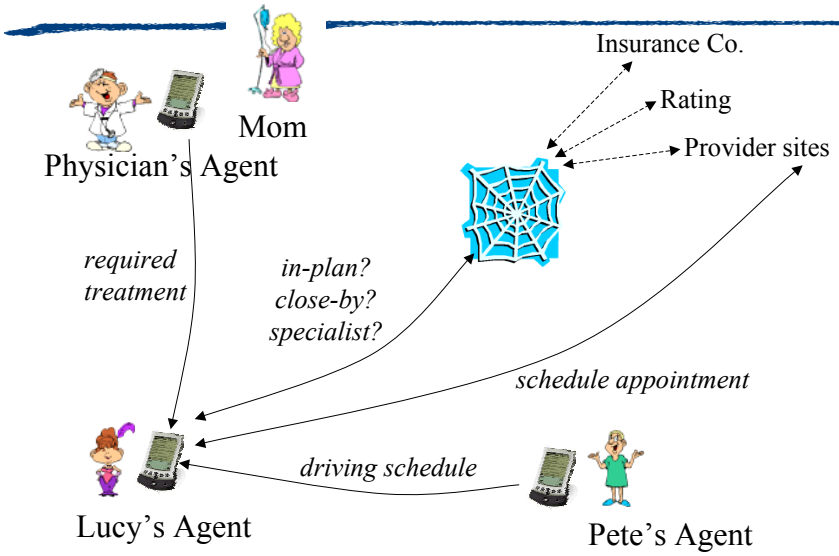


## Goals of the Semantic Web

- Machines “understand” and process information
- “The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where *software agents* roaming from page to page can readily *carry out sophisticated tasks* for users.”
  - *T. Berners-Lee, J. Hendler, O. Lassila, “The Semantic Web”,* Scientific American, May 2001

# Example: Doctor's Appointment

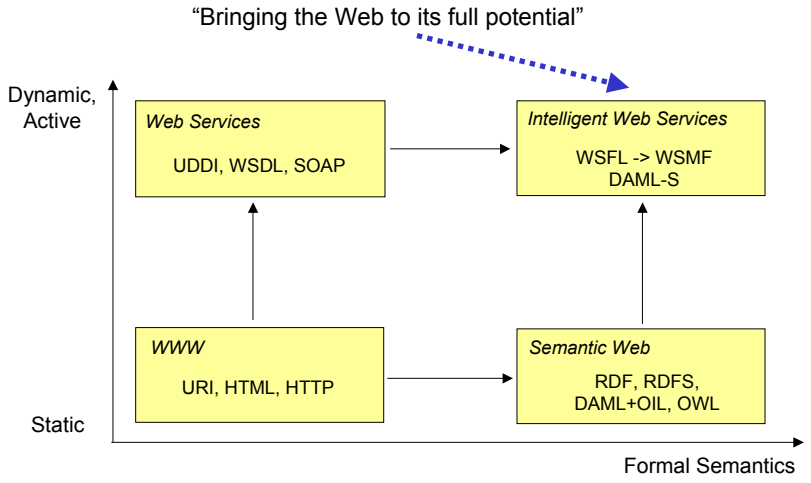
***"The Semantic Web"*, Scientific American, May 2001**



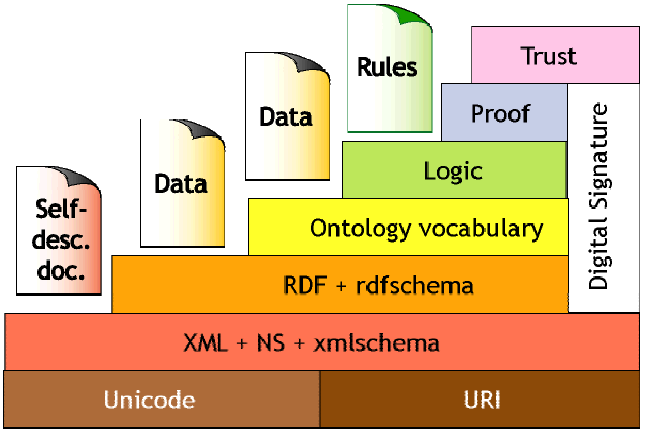



# Semantic Web Enabled Web Services (SWWS)

**Fensel, Bussler: IST Project, Start August 2002**



## The Semantic Web "Layer Cake"






## Web Languages for Knowledge Capturing

- Human knowledge is (partially) captured on the Web as *informal texts*, *semiformal documents*, and *structured metadata*
- Each kind of *knowledge* has its (preferred) markup *language*

| | | | |
|---|---|---|---|
| **Knowledge:** | Informal | Semiformal | Metadata |
| **Language:** | HTML | XML | RDF |

## Web Languages for Machine Interpretation

- **XML (Extensible Markup Language):** Semiformal documents range between *non-formatted texts* and *fully formatted databases*
- **RDF (Resource Description Framework):** Structured metadata describe arbitrary *heterogeneous* Web pages/objects in a *homogeneous* manner

*Machines (e.g., search engines) can analyze XML or RDF markups better than full HTML*



## Address Example: External to HTML to XML

*external*

*Xaver M. Linde*
Wikingerufer 7
**10555 Berlin**

*HTML*

```
<em>Xaver M. Linde</em><br>
Wikingerufer 7<br>
<strong>10555 Berlin</strong>
```

*XML*

```
<address>
  <name>Xaver M. Linde</name>
  <street>Wikingerufer 7</street>
  <town>10555 Berlin</town>
</address>
```

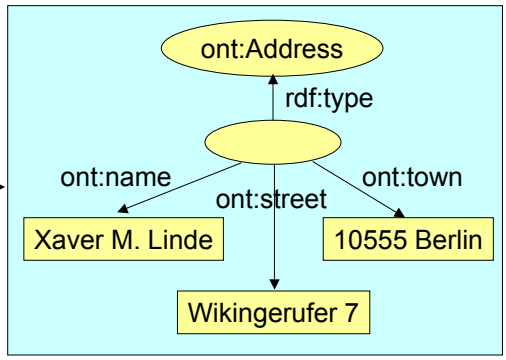
*Presentation vs. Structure*

## Address Example: XML to RDF

*XML*

```
<address>
  <name>Xaver M. Linde</name>
  <street>Wikingerufer 7</street>
  <town>10555 Berlin</town>
</address>
```

*RDF*

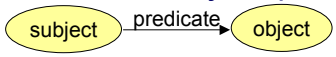



*Structure vs. Semantics*



## RDF—Resource Description Framework

- Statements of the form <subject, predicate, object>




- subject, predicate, object are URIs, objects can also be literals (strings)
- Has graph representation and several XML representations/serializations, e.g.:

```
<ont:Address>
   <ont:name>Xaver M. Linde</ont:name>
   <ont:street>Wikingerufer 7</ont:street>
   <ont:town>10555 Berlin</ont:town>
</ont:Address>
```

## Ontologies, RDFS

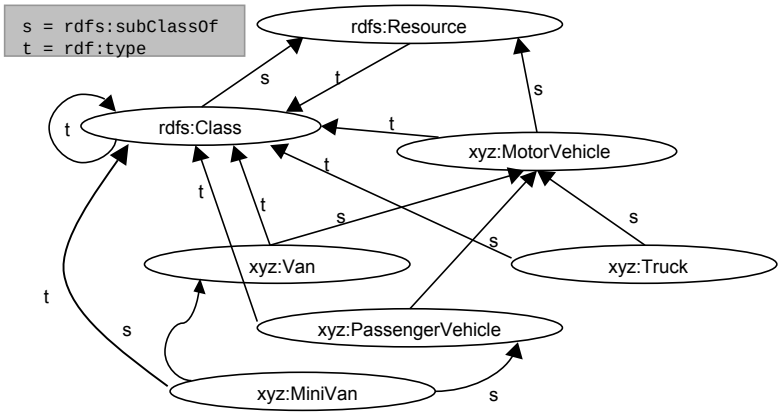
- RDF just defines the data model
- Need for definition of vocabularies for the data model—an *ontology* language!
- What is an ontology?
  „An ontology is a specification of a conceptualization.“
  - Tom Gruber, 1993
- Ontologies are social contracts
  - Agreed, explicit semantics
  - Understandable to outsiders
  - (Often) derived and maintained in a community process
- Ontologies require knowledge representation
  - Is-a hierarchy, part-of, attributes, axioms, …
- RDFS: RDF Schema as simple ontology language



## RDFS: Simple OO/Frame Language

- Similar to *class system* in OO programming languages (like Java or C++) or UML
- Classes = sets of objects (“instances”):
  - `rdfs:Class`
- Instances = members of classes:
  - `rdf:type` (already in RDF)
- Specialization: a class is a subclass of another:
  - `rdfs:subclassOf`
- Properties (attributes, slots) attached to classes:
  - `rdfs:Property`
  - `rdfs:domain` & `rdfs:range`
- `rdfs:label`, `rdfs:comment`, etc.

# RDFS: Example






# RDF/S: Jump Starters I

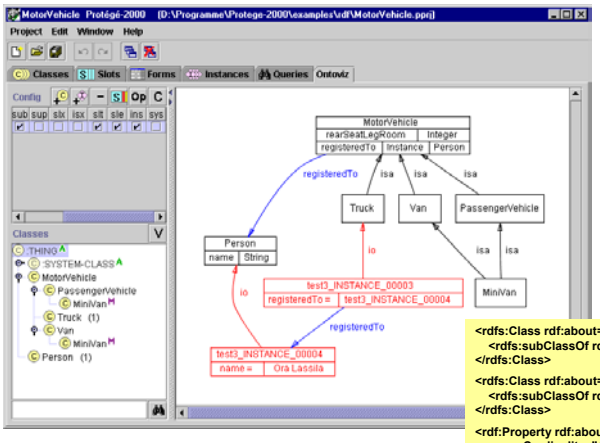
- dmoz open directory project.org (human-edited directory)
- MusicBrainz.org (music encyclopedia)
- RSS (RDF Site Summary)
- Adobe (embedded metadata)
- CC/PP (Composite Capability/Preference Profiles)
- P3P (Platform for Privacy Preferences)

# RDF/S: Jump Starters II

- B2B Vocabulary Projects
  - PapiNet.org: Vocabulary for paper industry
  - BPMI.org: Vocabulary for exchanging Business Process Models
  - XML-HR: Vocabularies for human resources (HR)
  - DMTF (Distributed Management Task Force): vocabularies for managing enterprises
  - …
- Research Vocabulary Projects
  - GeneOntology Working Group (http://www.geneontology.org/)
  - HIDDEL
  - MathNet
  - …



# Protégé-2000 as RDF/S-Editor




```
<rdfs:Class rdf:about="&mv;MotorVehicle">
   <rdfs:subClassOf rdf:resource="&rdfs;Resource"/>
</rdfs:Class>

<rdfs:Class rdf:about="&mv;PassengerVehicle">
   <rdfs:subClassOf rdf:resource="&mv;MotorVehicle"/>
</rdfs:Class>

<rdf:Property rdf:about="&mv;rearSeatLegRoom"
   a:maxCardinality="1"  a:range="integer">
   <rdfs:domain rdf:resource="&mv;MotorVehicle"/>
   <rdfs:range rdf:resource="&rdfs;Literal"/>
</rdf:Property>
```

## DAML+OIL and OWL

- DARPA DAML project: *DAML+OIL*
- Web Ontology Working Group: *OWL*
- Higher expressiveness than RDF Schema:
  - Class Expressions (Intersection, Union, Complement)
  - XML Schema Datatypes
  - Property restrictions
    - Cardinality constraints
    - Value restrictions
  - Axioms: equality, transitivity, …

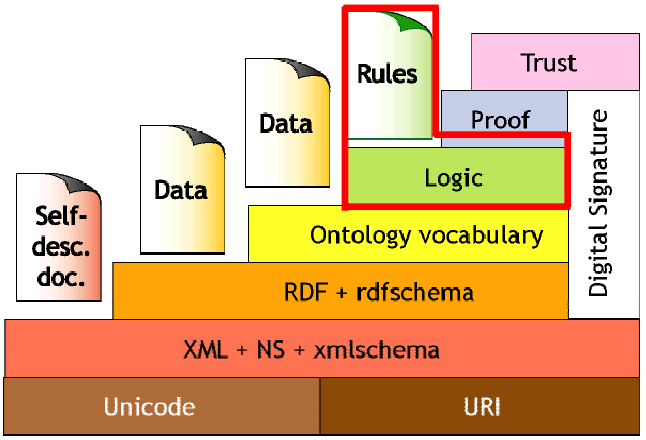


## DAML+OIL/OWL: Informal Examples

- mother = woman with child
- herbivore = animal
  which is not a carnivore
  and which eats things
  which are plants or parts of plants

$$\text{herbivore} \doteq_c \text{animal} \sqcap \neg\text{carnivore} \sqcap \forall\text{eats}.(\exists\text{is-part-of}.\text{plant} \sqcup \text{plant})$$

## Queries, Rules, Inferences






## RDF Query and Rule Languages

- Motivation: "The Semantic Web is an extension of the current web …, better enabling computers … to work in cooperation."
- → Agents exchange RDF data/schemas and DAML+OIL/OWL ontologies etc. and have to *query/process/map between* them
- Various query/rule languages on top of RDF exist:
  - SiLRI, RDQL, RQL, DQL, N3/cwm, Squish, TRIPLE

# Need for a *Combined* RDF Query, Inference, and Transformation Language: TRIPLE

| *RDF characteristics* | *Need for* |
|---|---|
| – meta-data for data/documents on the web (e.g., Dublin Core) | – search, intelligent information retrieval, navigation |
| – ontologies (vocabularies for meta-data) | – checking (inst. – ontology), ontology mapping |
| – exchanged between "agents" (e.g., e-commerce, Web Services) | – data transformation and integration |

*query* *inference* *transformation*

- Needed: *query, inference, and transformation* language for RDF
- TRIPLE: joint work with Stefan Decker (ISI)



# TRIPLE Example: Dublin Core

```
dc := "http://purl.org/dc/elements/1.0/".
dfki := "http://www.dfki.de/".
....
@dfki:documents {
   dfki:d_01_01 [
      dc:title → TRIPLE;
      dc:creator → "Michael Sintek";
      dc:subject → RDF;
      dc:subject → triples; ... ].
   ∀N  p(N)[ rdf:type →  xyz:Person;
             xyz:name → N ] ←
          ∃D D[dc:creator → N].
}
∀N ← ∃P P[rdf:type → xyz:Person;
     xyz:name → N]@dfki:documents.
```

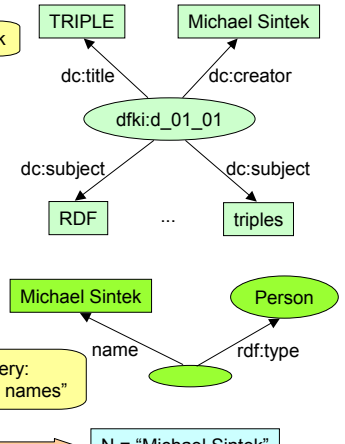
namespace abbreviations

model block

fact

rule

query: "find all names"




N = "Michael Sintek"

## (Semantic) Web Services

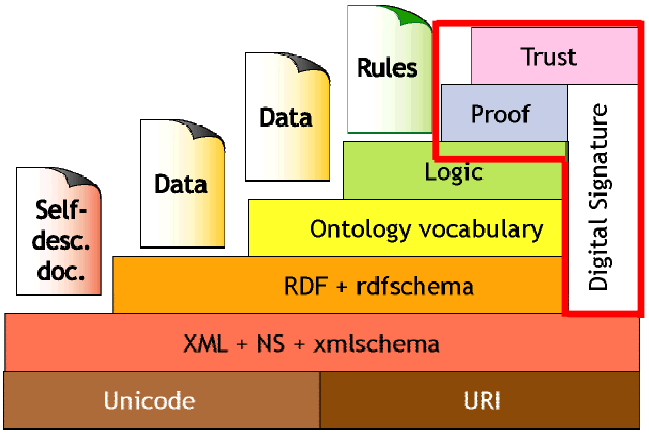
- Public process description and advertisement
  - WSDL: Web Services Description Language
  - UDDI: Universal Description, Discovery, and Integration of Business for the Web
  - (SOAP: Simple Object Access Protocol)
- Discovery and Composition of Services
  - DAML-S
  - WSFL: Web Services Flow Language



## Web Services: Future Work

- Ontologies & Inferences
  - State of the art: person browses web service descriptions and composes manually
  - Goal: automatic discovery and composition
  - Solution: semantically rich descriptions (via ontologies) and inferences
- Decentralization, P2P
  - State of the art: UDDI = central registry
  - Decentralization needed to avoid
    - Single point of failure
    - High costs
    - Slow reaction (info out of date)
  - Solution: P2P (Peer-to-peer) network of Web Services

## Web of Trust






## Web of Trust: Technologies

- P3P: Platform for Privacy Preferences Project
  - Web sites describe their policy for handling private information
  - P3P-enabled browsers compare this to the user's preferences
- PGP: Pretty Good Privacy (private/public key)
- Future:
  - assign degree of trust to relations + (partial) transitivity
  - rules/inferences/proofs

## Evolution of Knowledge

- "If properly designed, the Semantic Web can assist the evolution of human knowledge as a whole."
  - *T. Berners-Lee, J. Hendler, O. Lassila, "The Semantic Web",* Scientific American, May 2001
- Small groups innovate rapidly, but produce subcultures whose concepts are not understood by others
- Semantic Web will help in joining together these subcultures