Wissenschafterinnenkolleg Internettechnologien

Technische Universität Wien
Vienna University of Technology

Semantic Web: Introduction
1.4.2004, 15:00 – 16:30

AK aus Knowledge Engineering 1

Elke Michlmayr

Übersicht

- Organisatorisches
- „Geschichtliches"
  - Entstehung des Forschungsgebiets Semantic Web
- Die Begriffe Metadaten, Semantik, Ontologie
- Semantic Web Schichtenmodell (Cake)
- Zusammenfassung
- Diskussion/Gruppenarbeit

Semantic Web

"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."

Tim Berners-Lee, James Hendler, Ora Lassila,
The Semantic Web, Scientific American, May 2001

"The Semantic Web is a major research initiative of the World Wide Web Consortium (W3C) to create a metadata-rich Web of resources that can describe themselves not only by how they should be displayed (HTML) or syntactically (XML), but also by the meaning of the metadata."

W3C Semantic Web Activity Page

Vision (aus Scientific American Artikel)

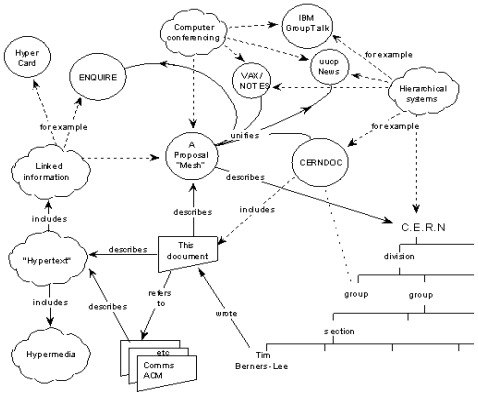
- Softwareagenten (=Maschinen)
  - Handeln autonom nach Auftrag eineR UserIn und durchsuchen das Web nach Informationen
  - Fungieren als persönliche Suchmaschine
  - Kommunizieren miteinander
  - Übernehmen komplexe Aufgaben, wie z.B. Buchen einer Urlaubsreise (Hotel, Flug)

Schnittmenge zweier Forschungsgebiete

- WWW
  - Stellt die Basis zur Verfügung (Netzwerk, Protokolle)
  - Hypertext: Sehr viele im WWW nicht verwirklichte Forschungsideen und –ansätze zur Verknüpfung von Information, wie z.B. bidirektionale Links
  - Markup Languages
- Künstliche Intelligenz
  - Wissensbasierte Systeme (Fakten + Regeln)

Projektvorschlag WWW (1989)

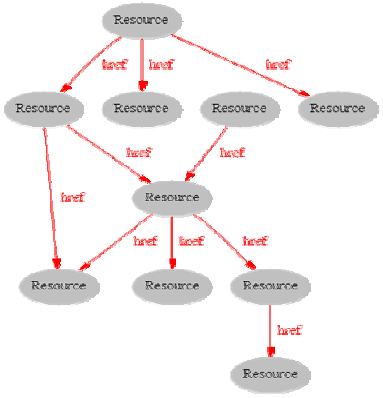
" ... how information about a Web of relationships amongst named objects could unify a number of information management tasks."



http://www.w3.org/History/1989/proposal.html

WWW: Probleme

- Maschinen können die Inhalte von Webseiten nicht verstehen
  - Informationssuche
    - Basiert auf Keywords
    - Nur Syntax kann berücksichtigt werden
    - Nicht ausreichend (Beispiel: Bildersuche)
- Maschinen können die Bedeutung von Links nicht verstehen



http://www.w3.org/Talks/2001/1102-semweb-fin/web.gif

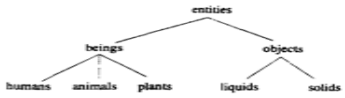
Notwendig: Metadaten und Semantik

- Daten mit semantischen Metadaten auszeichnen
  - „Metadaten sind maschinenlesbare Informationen über elektronische Ressourcen und andere Dinge." (Tim Berners-Lee)
  - Semantik ist die „Lehre von der Bedeutung der Wörter" (Griechisch: "semantikos")
- Problem: Wie wird die Semantik der Metadaten spezifiziert?
  - Beispiel: "Y ist AutorIn dieser Webseite."
    Wie wird die Bedeutung von "AutorIn" maschinen-verständlich definiert?

-> Verknüpfung der Metadaten mit Ontologien

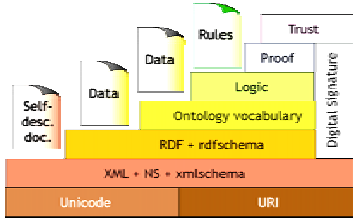
Ontologien: Formale semantische Modelle

- Aus der KI (bzw. Philosophie)
- Dienen zur Standardisierung von Bedeutungen
- Modellierung von abgegrenzten Wissensgebieten (=Domänen)
- Begriffe werden in hierarchische Strukturen geordnet und miteinander über definierte Assoziationen in Verbindung gebracht
- Ein Begriff wird durch seine Relation zu anderen Begriffen definiert



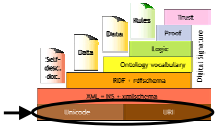
Schichtenmodell: "Semantic Web Cake"

- Besteht aus aufeinander aufbauenden Schichten
- Für jede Schicht wurde/wird vom W3C ein Standard definiert
- Jede Schicht verwendet die Funktionalitäten der darunterliegenden Schichten
- Jede Schicht bietet erweiterte Funktionen

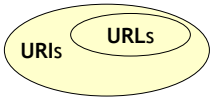


http://www.w3.org/2001/09/06-ecdl/slide17-0.html

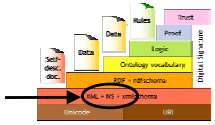
Unicode + URIs



- Unicode
  - Stellt Zeichensätze für (fast alle) natürlichen Sprachen zur Verfügung
- URIs (Unique Resource Identifiers)
  - Prinzip: "Everything has a URI"
  - Zur Identifikation von Ressourcen ("Ein Name für ein Ding")
  - Keine Adressierung!
    - Ressource *kann*, muss aber nicht über die URI erreichbar sein
  - URLs sind eine Untergruppe von URIs
    - Identifikation + Adressierung



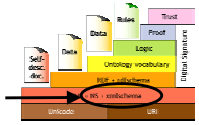
XML Namespaces



- Vermeidung von „name clashes"
  - Eindeutige Identifizierung von Elementen und Attributen einer XML Anwendung
- Erlauben gleichzeitige Verwendung mehrerer XML-Anwendungen in einem XML Dokument
- Beispiel
  - In welchem Namespace liegt `<sentence>`?

```
<sentence xmlns="http://example.org/xml/documents/"
          xmlns:c="http://animals.example.net/xmlns/">
  <c:person>I</c:person> just got a new pet
  <c:animal>dog</c:animal>.
</sentence>
```

XML Schema (1/2)



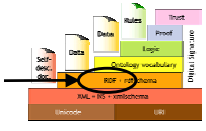
- Nachfolger für DTDs
- Erlaubt die Definition von syntaktischen Regeln für XML-Dokumente
  - Namen für Elemente und Attribute
  - Reihenfolge und Anzahl von Elementen
  - Erlaubte Attribute für Elemente
  - Erlaubte Datentypen für Werte von Elementen und Attributen
- Ein XML-Dokument ist <u>gültig</u>, wenn es auf ein Schema verweist und die in diesem Schema definierten syntaktischen Regeln einhält

XML Schema (2/2)

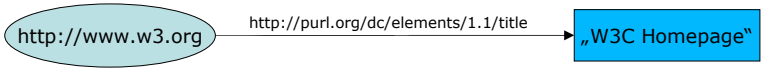
```
<?xml version="1.0" encoding="UTF-8"?>
<course xmlns:xsi="http://www.w3c.org/2001/XMLSchema-Instance"
        xsi:noNameSpaceSchemaLocation="http://www.abc.org/my.xsd">
  <student>Gustav Sielmann</student>
  <student>Arnold Rummer</student>
</course>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema xmlns:xsd = "http://www.w3c.org/2001/XMLSchema">
<xsd:element name="course">
  <xsd:complexType>
    <xsd:sequence>
      <xsd:element name="student" type="xsd:string"
                                   maxOccurs = "unbounded"/>
    </xsd:sequence>
  </xsd:complexType>
</xsd:element>
</xsd:schema>
```


RDF (1/3)



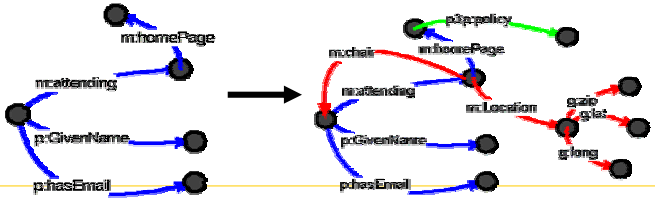
- Resource Description Framework
  - Seit 1999 entwickelt (PICS)
  - Seit Februar 2004 W3C Recommendation
- Modell und Syntax zur Beschreibung von (Meta-)Daten
- zur Beschreibung beliebiger "Ressourcen" (Webseiten, Dinge, Personen, ...)
- RDF Statements: Subjekt, Prädikat, Objekt



Subjekt: Ressource (Resource)	Prädikat (Eigenschaft): (Property)	Objekt: Literal oder Ressource (Property Value)

RDF (2/3)

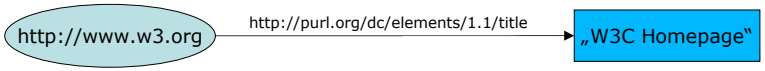
- Graph als Basismodell
  - Einfach in logische Aussagen überführbar (wichtig für obere Schichten des Schichtenmodells)
- Flexibilität und Erweiterbarkeit
  - Statements können miteinander verknüpft werden
  - Namespaces!



RDF (3/3)

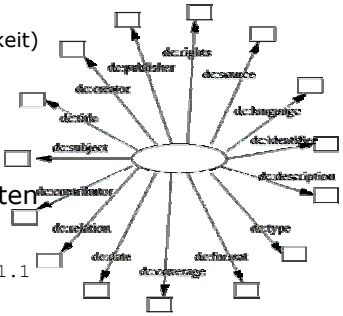
- Serialisierungsformat: RDF/XML

```
<?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 rdf:about="http://www.w3.org">
    <dc:title>W3C Homepage</dc:title>
  </rdf:Description>
</rdf:RDF>
```



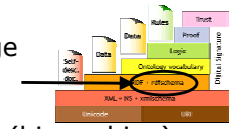
Exkurs: Dublin Core Metadata Initiative

- Ziele
  - Interoperabilität (Austauschbarkeit) von Metadaten
  - Entwicklung maschinell verarbeitbarer Semantiken
- Standardset von Dublin Core Elementen zur Beschreibung von Dokumenten
  - in RDF Schema definiert
  - Namespace: `http://purl.org/dc/elements/1.1`



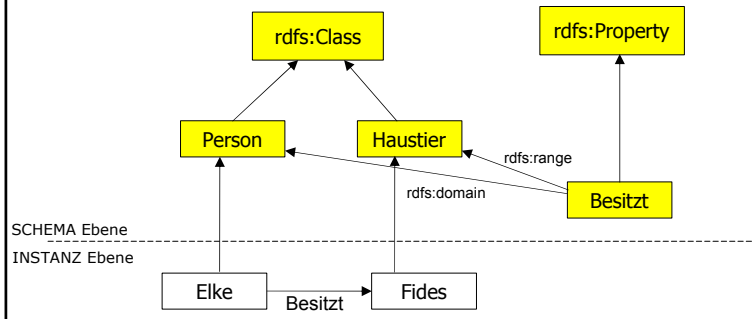
RDF Schema

- RDF Vocabulary Description Language
  - Ermöglicht die Definition von RDF Vokabularen

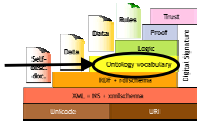


- Mit RDF Vokabularen werden Klassen(hierarchien), Properties und ihre Beziehungen zueinander definiert
  - rdfs:Class, rdfs:subClassOf
  - rdfs:Property, rdfs:SubPropertyOf
  - rdfs:domain (Gültige Klassen für Properties)
  - rdfs:range (Gültige Werte für Properties)
- Beispiel für ein RDF Vokabular:
  - CC/PP für Mobiltelefone (Composite Capability/Preference Profiles)

RDF Schema



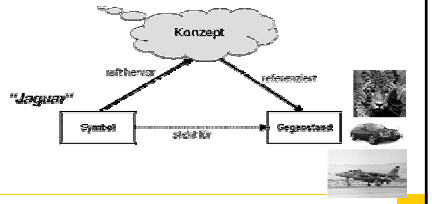
Ontologien (1/4)



- Davenport, 1998: "People can't share knowledge if they do not speak a common language"
- Definitionen des Begriffs Ontologie
  - Gruber, 1993: An ontology is an explicit specification of a conceptualization.
  - Uschold, 1996: "An ontology is a shared understanding of some domain of interest."
  - Eine Ontologie ist ein Vokabular, das die Entitäten einer Domäne und die Zusammenhänge unter den Entitäten in systematischer und formaler Weise beschreibt.
- Wissensbasis = Ontologie + Instanzen
- Mit RDF Schema können nur einfache Ontologien erstellt werden, „richtige" Ontologiesprachen bieten mehr
  - z.B.: Ausdrücken, dass zwei Properties zueinander gegenteilig (invers) sind

Ontologien (2/4)

- Zu jedem Thema besteht unterschiedliches Hintergrundwissen, Sprachvokabular und Sichtweise
- Ontologien stellen Wissensmodelle bereit, die konzeptuelle und terminologische Klarheit schaffen
  - Ist das immer möglich?
  - Abgegrenzte Domänen
- Semiotisches Dreieck
  - Symbol
  - Konzept
  - Gegenstand



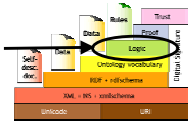
Ontologien (3/4)

- kontrolliertes Vokabular:
  - Sammlung von Symbolen (Wortschatz)
  - die eindeutigen Bedeutungen zugeordnet sind,
  - so dass keine Homonyme auftreten.
- Taxonomie:
  - kontrolliertes Vokabular,
  - dessen Symbole in hierarchischer Beziehung zueinander stehen
- Thesaurus:
  - Wie Taxonomie, aber:
  - Symbole sind durch einfache Relationen miteinander verbunden.
    - Ist-ähnlich, Ist-Synonym
- Ontologie:
  - Klassen und Eigenschaften
  - Klassen stehen in hierarchischer Beziehung zueinander
  - (komplexe) Relationen zwischen Klassen
  - Instanzen!

Ontologien (4/4)

- Entstehungsgeschichte der Ontologiesprachen für das Semantic Web: DAML, OIL, DAML+OIL, OWL
  - DAML und OIL 1999 parallel gestartet
    - USA: DARPA Agent Markup Language (DAML) Projekt
    - EU: Ontology Inference Layer (OIL) Projekt
  - Dezember 2000: DAML+OIL
    - DAML+OIL baut auf RDF Schema auf (wurde als RDF Vokabular spezifiziert)
  - Dezember 2001: DAML+OIL beim W3C eingereicht
  - DAML+OIL Ausgangsbasis für die Web Ontology Language (OWL)
  - OWL: W3C Recommendation seit Februar 2004

Logic (1/2)



- Reasoning (logische Schlussfolgerungen):
  - Aus bestehendem Wissen wird neues (nicht explizit definiertes) Wissen abgeleitet
  - OWL basiert auf descriptions logics
- Logic-Layer (noch nicht realisiert)
  - Soll XML-basierte Rule Markup Languages zur Definition von Regeln bieten
  - Prädikatenlogik kann mehr als description logics, z.B.:
    `sisterOf(X,Y) :- childOf(X,Z), childOf(Y,Z), Woman(X), X \== Y.`
  - Ansätze: SWRL, RuleML

Logic (2/2)

- Beispiele
  - 1)

```
(motherOf subProperty parentOf)
(Mary motherOf Bill)
```

Reasoner schliesst:

```
(Mary parentOf Bill)
```

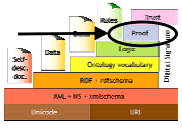
  - 2)

```
(Mary motherOf Bill)
(parentOf inverse childOf)
(Bill childOf ?X)
```

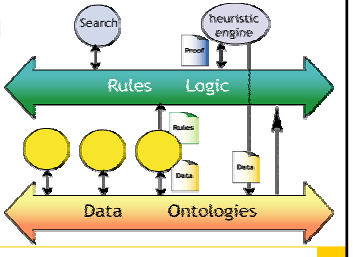
Reasoner schliesst:

```
?X = Mary
```

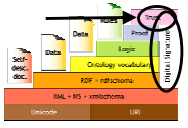

Proof



- Logic Layer ermöglicht allen, Regeln zu veröffentlichen
- Proof Layer (noch nicht realisiert):
  - „Auswerten der Regeln"
  - Deduktion von *Proofs*
    - eine bestimmte Aussage wird aus einer oder mehreren anderen Aussagen abgeleitet
    - Proof: „Das ist so, weil …"
  - Repräsentation der Proofs (Beweise) in XML-basierter Sprache
  - Proof validation



Trust + Digital Signatures



- „Alle können alles aussagen."
  - Welchen Aussagen ist zu trauen?
- Wem glauben?
  - Public Key Infrastructure notwendig
  - Digitale Signaturen werden zur Signierung von Aussagen verwendet
  - BenutzerInnen geben ihren Agents ihre vertrauenswürdigen Quellen bekannt
  - Web of Trust
    - „Trust the friends of your friends."
  - Auch Distrust (Misstrauen) ist sinnvoll!
- Proof + Trust Layer gemeinsam ermöglichen den Austausch von Proofs zwischen Agenten

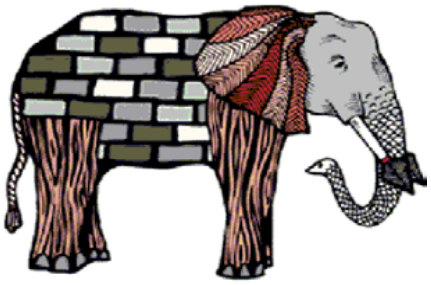
Semantic Web Forschungsbereiche

- Ontologies
  - creation, merging, linking, …
- Knowledge Portals
- Visualization
- User Interfaces
- Data Semantics
- Searching and Querying the Semantic Web
- Semantic Web Services
- Semantics in P2P networks
- Semantic Grid

Zum Abschluss: SW Prinzipien

- Dezentrale Wissensrepräsentation in einem maschinenverständlichen Informationsraum
- The Semantic Web must allow the independent work of diverse communities to be combined effectively.
  - Provide communities the ability to resolve ambiguities and clarify inconsistencies.
- Support the ability to add new information without insisting that the old be modified.
- Minimalist Design
  - Make the simple things simple, and the complex things possible.
  - Standardize not more than is necessary.
- The Semantic Web is unbounded.
  - Anyone can say anything about anything.
  - There will always be more to discover.

Danke fürs Zuhören!



http://www.w3.org/Talks/2001/1102-semweb-fin/all.htm