



Kundenprojekt Web Technologien

Semantic Web

Malgorzata Mochol
Freie Universität Berlin
Institut für Informatik
AG Netzbasierte Informationssysteme

mochol@inf.fu-berlin.de

<http://page.mi.fu-berlin.de/mochol/>

- Semantic Web (SW) Einführung
 - Erfinder
 - SW Stack
 - SW Technologien
- SW Anwendungsbeispiel



Semantic Web – Einführung

“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.”

Berners-Lee, Hendler, und Lassila, 2001.



Foto: W3C



Foto: Homepage



Foto: Homepage

Das Ziel des Semantic Web ist es
WWW-übertragende Daten
durch Menschen mit
Bedeutungsinformationen (Semantik)
anzureichen für
die **Verarbeitung durch Maschinen**
und **Nutzung durch Menschen**.

Bildersuche: „Apache“



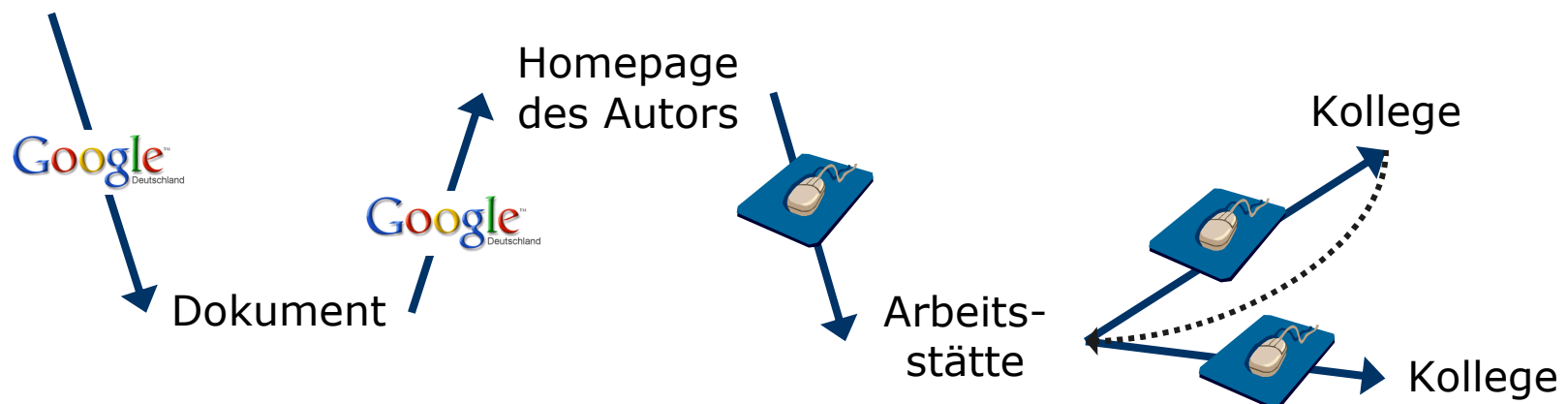
- Maschinen fehlt dieser Kontext aus Begriffen und Zusammenhängen
- Kontext muss Maschinen zusätzlich bereitgestellt werden

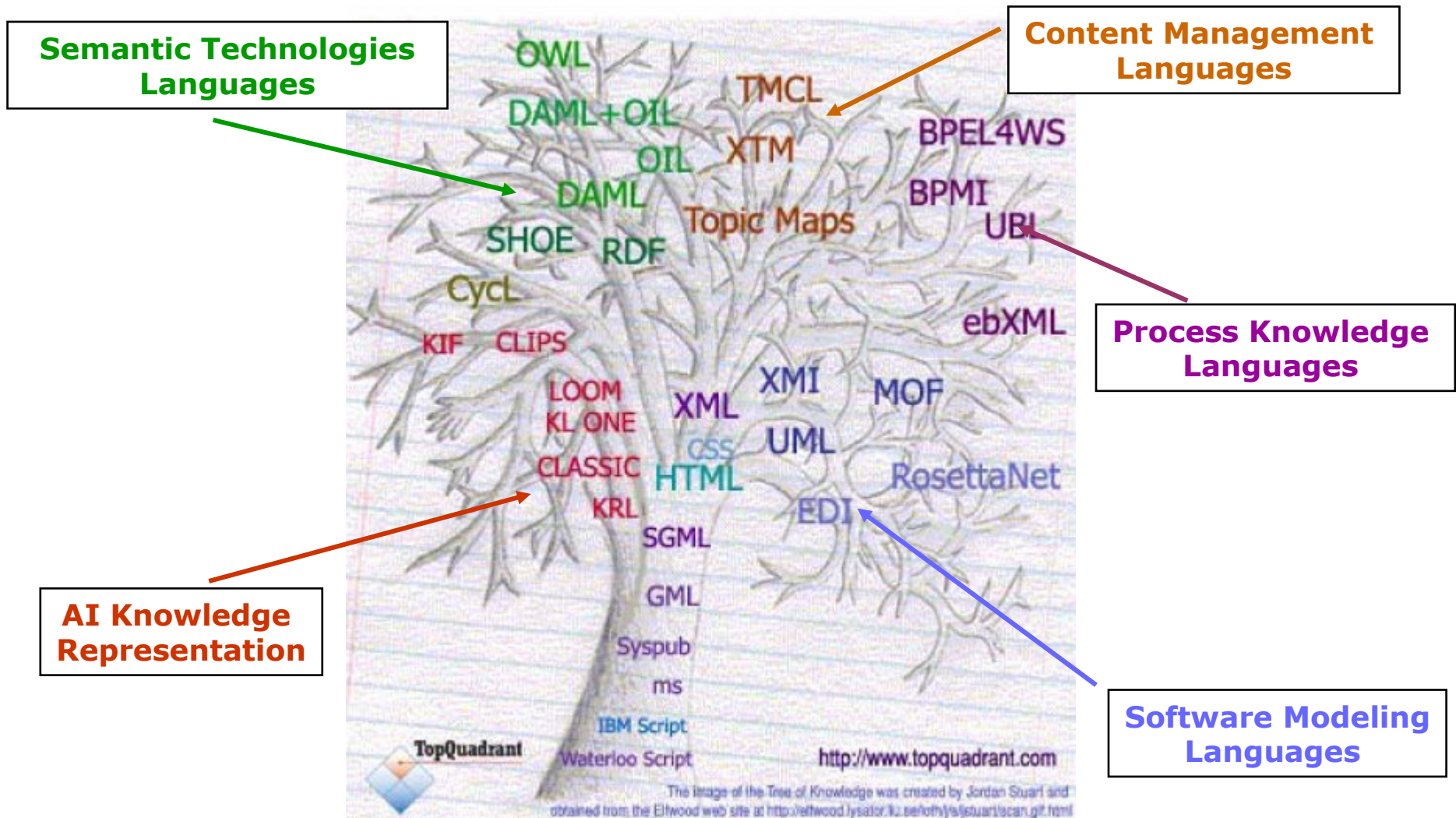
- Damit Metadaten nutzbar sind
 - muss der Informationsanbieter sich so ausdrücken, dass Informationsnutzer ihn verstehen
 - muss der Informationsnachfrager so fragen, dass er etwas finden kann
- Gemeinsame Benutzung von Konzepten
- Gemeinsame Sprache
- **Ontologie** zur Definition einer gemeinsamen Sprache
 - Es gibt Konzepte, die wir mit „Bank“ und „Sparkasse“ benennen
 - Es gibt ein Konzept, das wir „Geldinstitut“ nennen und das die Konzepte „Bank“ und „Sparkasse“ umfasst

- Webinhalte und ihre Vernetzung werden für Maschinen verständlich.
- Auch komplexe Anfragen können ans Web gestellt werden.

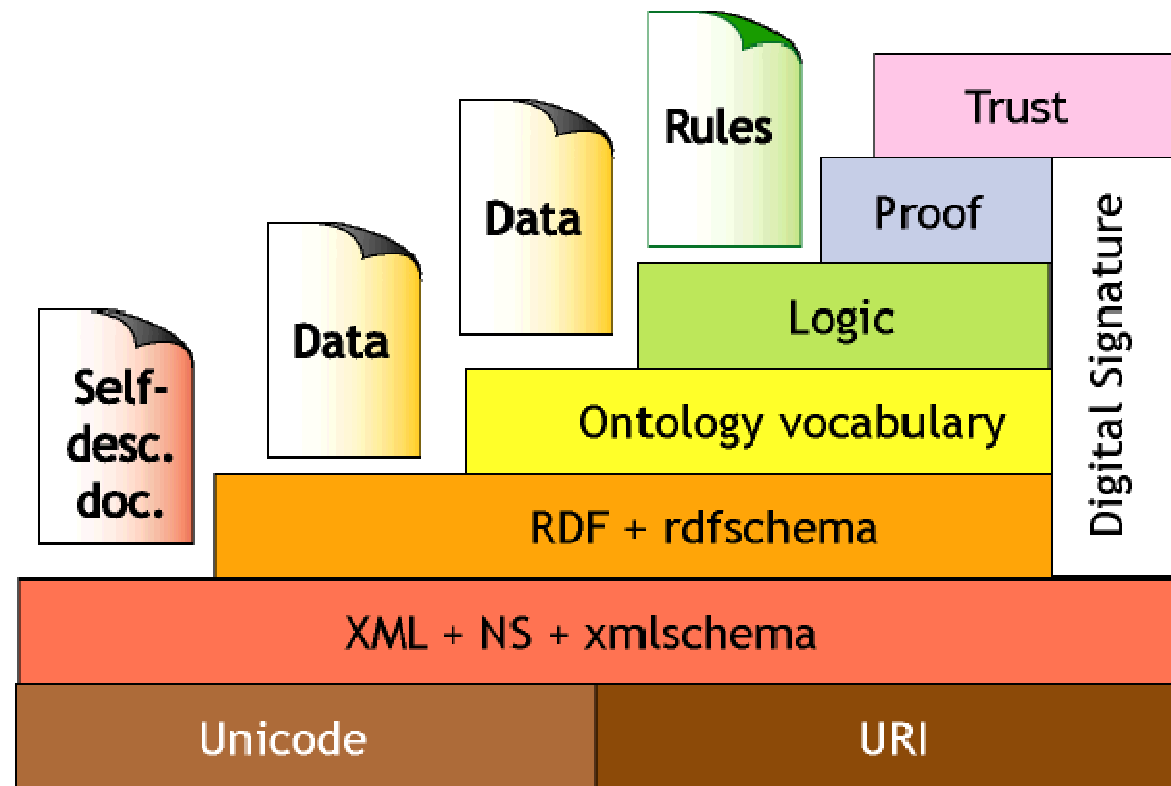


- Beispiel: Mit welchen Kollegen arbeitet der Autor eines bestimmten Dokumentes zur Zeit zusammen?





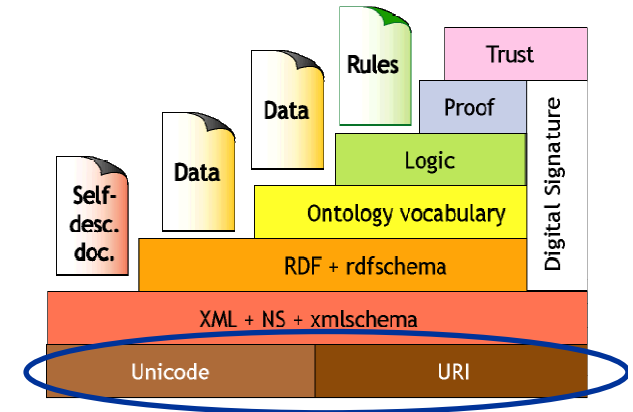
Quelle: TopQuadrant Technology Briefing: Semantic Technology, Version 1.2, March 2004



auch „Semantic Web Layer Cake“ oder „Semantic Web Tower“ genannt

Unicode

- jedes Zeichen eigene Nummer (system-, programm- und spracheunabhängig)
- Unicode-Codierung – Zeichensätze für fast jede natürliche Sprache

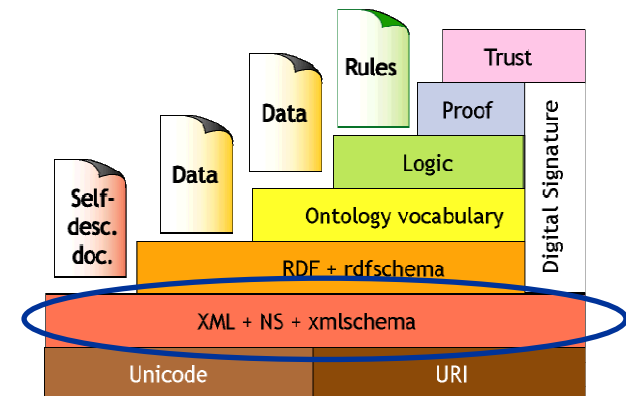


URI – Uniform Resource Identifier

- eindeutige Identifikation einer Quelle/Ressource → jedes beliebige Objekt verfügt über einen URI
- Mechanismus um Daten verteilt repräsentieren zu können
- URLs – Untergruppe von URIs
- Syntax vom W3C standardisiert

XML + Namensräume + XML-Schema

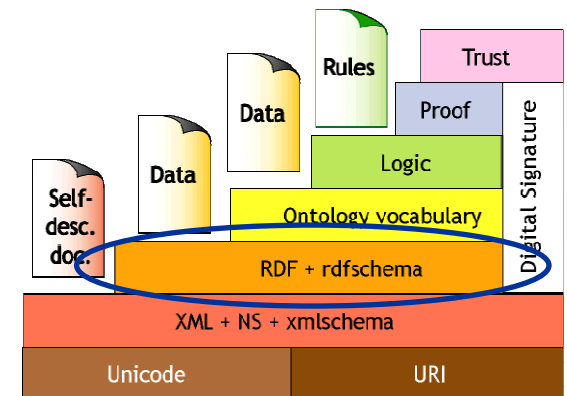
- hierarchisch strukturierte, medienneutrale **Daten**
- Vokabular kann mit **XML-Schema** definiert werden
- Bedeutung des Vokabulars kann mit **Namensräumen** festgelegt werden
- XML-Daten können mit **XLink** verlinkt werden: Links können Namen, aber keinen Namensraum haben



⇒ maschinenverarbeitbare verlinkte Daten, Links jedoch nicht maschinenverarbeitbar

RDF + Namensräume + RDF-Schema

- Web als Menge vernetzter Ressourcen
- Vokabular für Beziehungen kann mit **RDF-Schema** definiert werden
- Bedeutung des Vokabulars wird mit **Namensräumen** festgelegt
- **RDF Modell** bietet eine syntaxunabhängige Darstellung



⇒ maschinenverarbeitbares
Netzwerk von Beziehungen

- RDF/XML Syntax Specification – W3C Recommendation seit Feb. 2004
- verschiedene Versionen:
 - Tripel: kompakt, lesbar
 - RDF/XML: für maschinelle Verarbeitung
- Tripel setzen bel. Web-Ressourcen URI-s und URI-o miteinander in Beziehung:
$$\langle \text{URI-s}, \text{URI-p}, \text{URI-o} \rangle$$

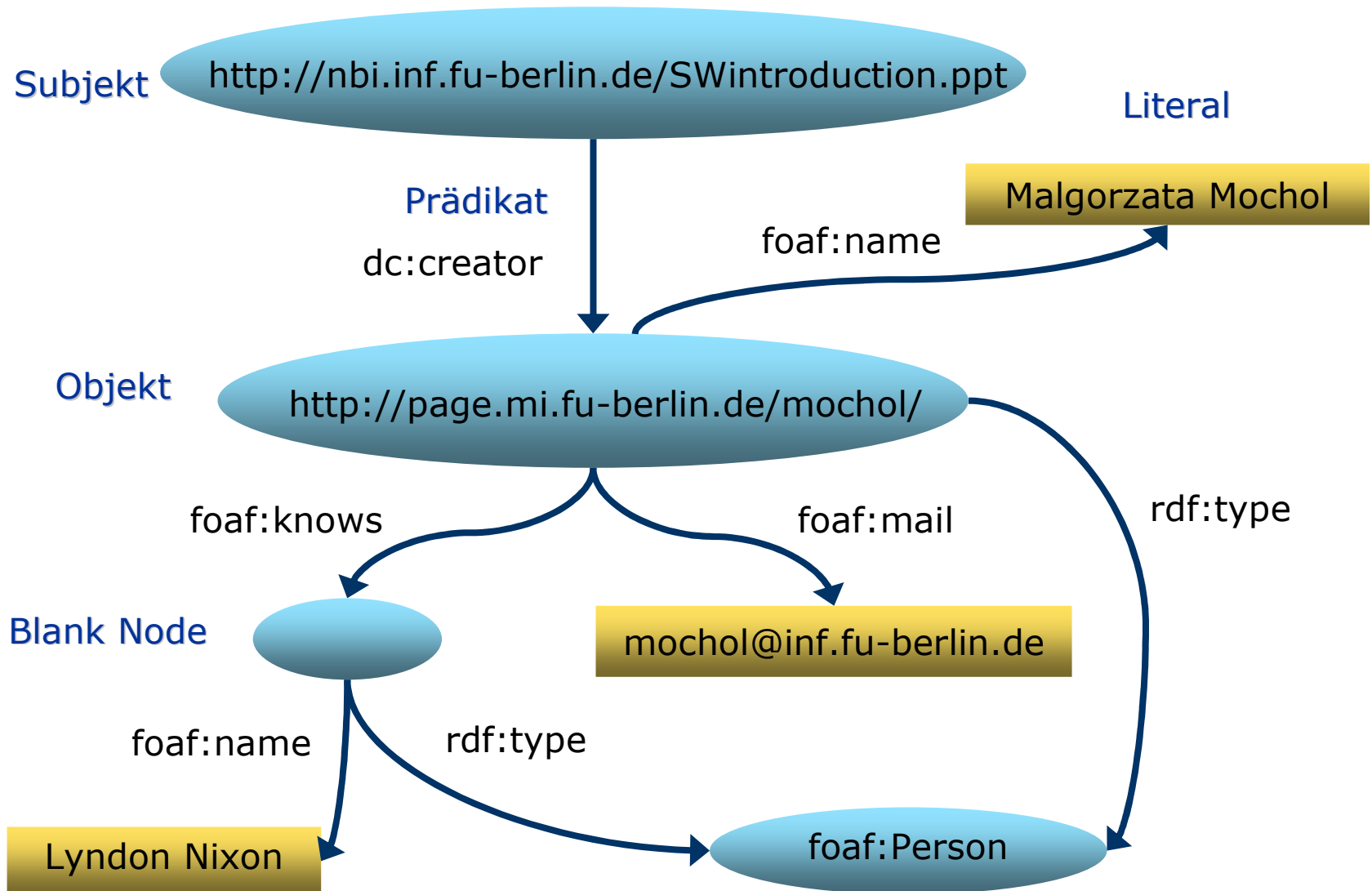
URI-s steht zu URI-o in der Beziehung URI-p

- RDF Statement – die kleinste Informationseinheit, die ein Fakt darstellt

- Beispiel:

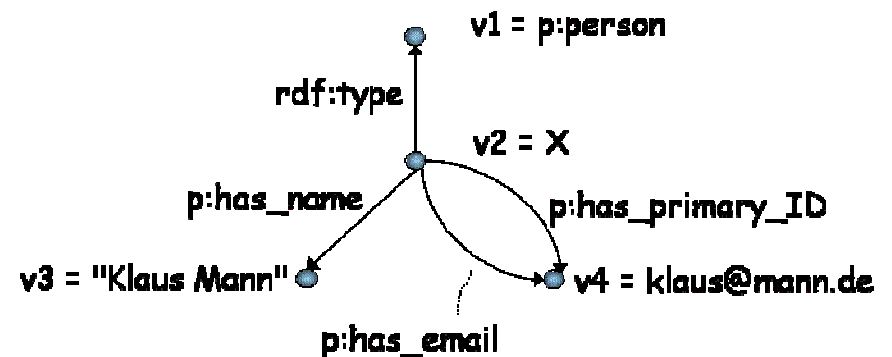
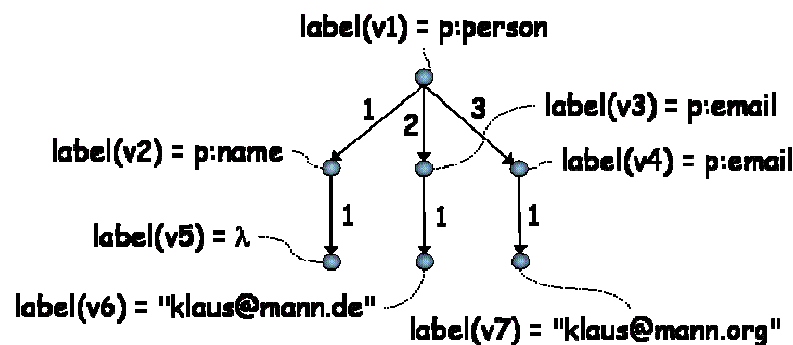
This presentation was ***created*** by ***Malgorzata Mochol***

- Subject (Ressource): ***This presentation***
 - Predicate (Property): ***creator***
 - Object (Wert): ***Malgorzata Mochol***
-
- RDF benutzt URIs :
 - Subject: ***<http://nbi.inf.fu-berlin.de/SemWeb.ppt>***
 - Predicate: ***<http://purl.org/dc/elements/1.1/creator>***
 - Object: ***<http://page.mi.fu-berlin.de/mochol/>***



Vergleich der Datenmodelle

	XML	RDF
Datenmodell	hierarchisches Modell	Netzwerkmodell
Was wird dargestellt?	einzelner Baum: <ul style="list-style-type: none">■ benannte Knoten■ unbeschriftete, aber geordnete Kanten	möglicherweise unendlich viele gerichtete Multi-Graphen: <ul style="list-style-type: none">■ benannte Knoten■ benannte Kanten■ Knoten = Name



Vergleich der Schema-Sprachen

	XML Schema	RDF Schema / OWL
Abstraktions- ebene	~ Datenbankschema	~ ER-Diagramm
Prinzip	nur zulässig, was explizit erlaubt: Closed World Assumption (CWA)	alles zulässig, was Randbedingungen erfüllen kann: Open World Assumption (OWA)
Validierung bzgl. Schema	möglich	nicht möglich
Berechnungs- komplexität	polynomial	RDF Schema: NP-vollständig

CWA:

nicht zulässig

explizit
erlaubt

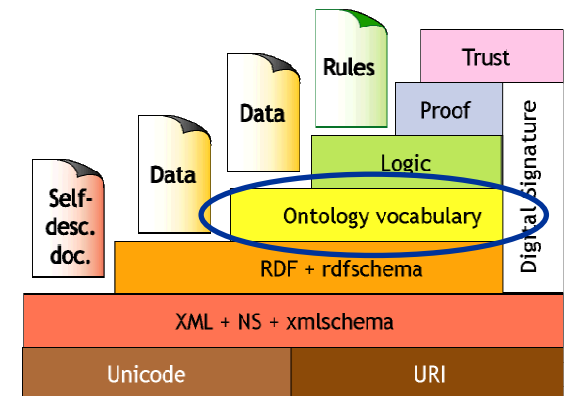
OWA:

zulässig

Randbedingungen
nicht erfüllbar

Ontologien

- Vokabulare
- Begriffsbeziehungen (Unterklasse, Untereigenschaft, Wertebereiche, ..., selbstdefinierte)
- Sprachen für Web-Ontologien:
 - DAML+OIL
 - OWL – Web Ontology Language
 - Erweiterte Beschreibungsmöglichkeiten
 - In unterschiedlichen Mächtigkeiten/Komplexitäten (OWL-Lite, OWL-DL, OWL-Full)



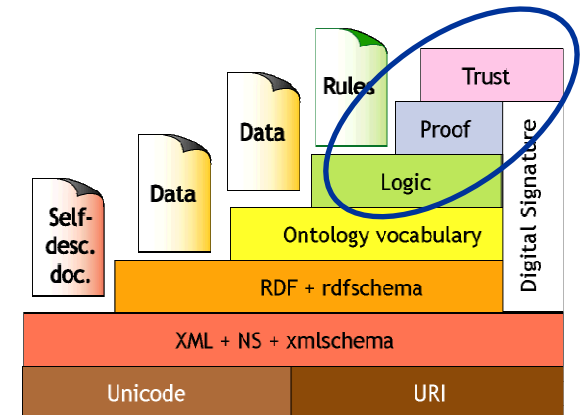
Logik

- Semantik auf logischer Basis
- Ableitungsregeln

Proof

- Konsistenz
- Ableitung (Inferenz)

Trust



→ Beginn der Entwicklung

XML

- XML heute omnipräsent, wenn auch nicht immer sichtbar

RDF

- HTML-Seiten und XML-Dokumente werden erstellt, aber kaum RDF.
- neue XHTML-Version wird RDF integrieren: Jedes XHTML-Element kann dann RDF-Meta-Informationen haben.
- Woher kommen diese Meta-Informationen?



Anfragesprachen

- XML-basierte Anfragesprachen (XPath, XQuery) nicht ausreichend
 - Anweisungen orientieren sich an der Syntax der jeweiligen XML-Dokumente
- In RDF
 - kein Wurzelement → Pfade können an jedem Knoten starten
 - Knoten und Kanten haben Namen
 - Unterscheidung zwischen Arten von Relationen
- Eine Pfad-basierte Anfragesprache muß diese Aspekte unterstützen

- **Unterstützung und Verständnis für das Datenmodell** von RDF/OWL und die semantischen Eigenschaften des Vokabulars
Beispiel:
 - Anfrage: „alle Instanzen von **printer**“
 - Ausgabe: nicht nur direkte Instanzen von **printer** sondern auch alle Unterklassen des Konzeptes **printer**, wie **laserJetPrinter** oder **deskJetPrinter**.
- Verständnis für **Pfadausdrücke**
 - Pfade im RDF Graphen (nicht im XML Modell)
 - Spezifikation von Knoten und Kanten
- Fähigkeit der **semantischen Integration** von Dokumenten bzw. von den zugrunde liegenden Ontologien
- Fähigkeit der **Kompositionalität**
 - Komplexe Anfrage können auf einfacheren aufbauen

- SQL-basierte
 - RQL – RDF Query Language (<http://139.91.183.30:9090/RDF/RQL/>)
 - RDQL - A Query Language for RDF (<http://www.w3.org/Submission/2004/SUBM-RDQL-20040109/>)
 - RdfDB (<http://www.guha.com/rdfdb/query.html>)
 - SeRQL ("circle") – Sesame RDF Query Language (<http://www.openrdf.org/doc/sesame/users/ch06.html>)
- XPath-ähnliche (<http://esw.w3.org/topic/RdfPath>)
 - Versa (<http://www.xml.com/pub/a/2005/07/20/versa.html>)
 - RDFPath (<http://infomesh.net/2003/rdfpath>)
- Regel-basierte
 - Triple (<http://www.dfki.uni-kl.de/frodo/triple/iswc2002/TripleReport.pdf>)
 - Versa (<http://www.xml.com/pub/a/2005/07/20/versa.html>)
 - N3QL N3-basierte Anfragesprache für RDF (<http://www.w3.org/DesignIssues/N3QL.html>)

→ SQL-basierte Sprachen am meisten verbreitet

→ Zur Zeit Entwicklung eines Sprachstandards **SPARQL**

- SPARQL = **S**imple **P**rotocol **A**nd **R**DF **Q**uery **L**anguage
- Anfragesprache und Daten-Zugriffs-Protokoll für das Semantic Web
- W3C Candidate Recommendation – 14.Juni 2007
<http://www.w3.org/TR/rdf-sparql-query/>
- Tutorial:
 - <http://www.xml.com/pub/a/2005/11/16/introducing-sparql-querying-semantic-web-tutorial.html>

- SQL-basierte Sprache → SQL-ähnliche Syntax
- Reine Anfragesprache → SELECT
- Keine Datenmanipulation → UPDATE, DELETE
- Viele unabhängige Implementierungen / Anwendungen

SPARQL – einfache Anfrage

```
PREFIX person: <http://example/person/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT ?name
WHERE {
    ?x foaf:name ?name
}
```

Beispielerggebnis

name
"Lyndon"
"Malgorzata"

- Variablen mit **?** Gekennzeichnet
- Triple Muster stellen Anfrage-Bedingungen dar
- XML-Namespace ähnliches PREFIX
- Lösungen der Anfrage in Tabellenform
- Eine Zeile entspricht einer Lösung

SPARQL – Anfrage mit FILTER

Anfrage: Alle Bücher mit Titel die mind. 1 mal vorhanden sind und weniger als 25 kosten

```
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX stock: <http://example.org/stock#>
PREFIX inv: <http://example.org/inventory#>
SELECT ?book ?title
WHERE {
    ?book dc:title ?title .
    ?book inv:price ?price . FILTER ( ?price < 25 )
    ?book inv:quantity ?num . FILTER ( ?num > 0 )
}
```

Beispielergebnis

book	title
book1	„Harry Potter“

- WHERE ist optional
- * gibt alle Variablen aus Graphen Muster aus

Beispiel basiert auf: http://axel.deri.ie/%7Eaxepol/publications/ESWC2007_SPARQL_Tutorial.pdf

- ARQ - SPARQL Prozeßor für Jena (<http://jena.sourceforge.net/ARQ/>)
- Rasqal (<http://librdf.org/rasqal/>)
 - Open Source C library für RDF
 - Unterstützt SPARQL, RDQL
- RDF::Query (<http://kasei.us/code/rdf-query/>)
 - Erlaubt SPARQL and RDQL Anfragen



Semantic Web – Anwendungsbeispiel

Organisatorisch:

- Stellenanbieter nutzen gemeinsames kontrolliertes Vokabular für die Annotierung von Stellenangeboten
- Stellensuchende nutzen gleiches Vokabular für Stellengesuche/Bewerberprofile

Technisch:

- Einfache Annotation → Reichere Annotation → Ersatz von Freitext durch RDF
- Stelleangebote direkt auf der Web-Seite des Unternehmens
- Semantische Suchmaschinen :
 - sammeln Informationen
 - Vergleich auf Basis von semantischen Informationen (Semantic Matching)

- Mit RDF und Bezug auf gemeinsames Vokabular (z.B. abgeleitet von HR-XML)

```
<html>
```

```
<head>
```

```
<rdf:RDF xmlns:rdf="...#" xmlns:jpp="...#">  
  <jpp:JobPositionPosting  
    rdf:about="http://www.example.org/jp1.html"/>  
</rdf:RDF>
```

```
</head>
```

```
<body>
```

```
  ...Job posting in free text...
```

```
</body>
```

```
</html>
```

- Suchmaschinen können so Stellenangebote identifizieren

```
<html>
  <head>
    <rdf:RDF xmlns:rdf="...#" xmlns:jpp="...#" xmlns:skills="...#">
      <jpp:JobPositionPosting
        rdf:about="http://www.example.org/jp1.html"/>
      <jpp:requiredCompetence>
        <skills:Java>
          <skills:hasCompetenceLevel rdf:resource="...#expert"/>
        </skills:Java>
      </jpp:requiredCompetence>
    </rdf:RDF>
  </head>
  <body>
    ... Job posting in free text ...
  </body>
</html>
```

```
<?xml version="1.0" encoding="UTF-8" ?>
<rdf:RDF xmlns:rdf="...#" xmlns:jpp="...#" xmlns:skills="...#">
  <jpp:JobPositionPosting rdf:about="#JobPositionPostingId-inf-44">
    <jpp:hasHiringOrganisation>
      <org:Organisation>
        <org:name>Freie Universität Berlin</org:name>
      </org:Organisation>
    </jpp:hasHiringOrganisation>...
    <jpp:requiredCompetence>
      <skills:Java>
        <skills:hasCompetenceLevel rdf:resource="...#expert"/>
      </skills:Java>
    </jpp:requiredCompetence>...
  </jpp:JobPositionPosting>...
</rdf:RDF>
```



Literaturhinweise

- W3C Semantic Web Activity <http://www.w3.org/2001/sw/>
- Resource Description Framework (RDF), <http://www.w3.org/RDF/>
- T. Berners-Lee, J. Hendler, O. Lassila: *The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*, Scientific American, 2001, <http://www.sciam.com/article.cfm?chanID=sa006&colID=1&articleID=00048144-10D2-1C70-84A9809EC588EF21>
- T. B. Passin: *Explorer's Guide to the Semantic Web*, ISBN1932394206, Manning Publications, 2004
- P. Hitzler, M. Krötzsch, S. Rudolph, Y. Sure: *Semantic Web.Grundlagen*, ISBN3540339930, Springer, 2007
- A. Gomez-Perez, O. Corcho, M. Fernandez-Lopez: *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. ISBN 1852335513, Springer, 2004

- J. Baumeister, D. Seipel: *Anfragesprachen für das Semantic Web*; Informatik-Spektrum, Springer Verlag; 28 (1), 40-44, Springer, 2005
<http://www.springerlink.com/content/yu3cra6lt8g65u0p/fulltext.pdf>
- *A Comparison of RDF Query Languages*
<http://www.aifb.uni-karlsruhe.de/WBS/pha/rdf-query/>
- SPARQL Tutorial:
http://axel.deri.ie/%7Eaxepol/publications/ESWC2007/SPARQL_Tutorial.pdf