ulm university universität uulm

Semantic Web Grundlagen

Einführung und XML

Birte Glimm
Institut für Künstliche Intelligenz | 17. Okt 2011

Agenda

- Vorstellung der Dozentin
- Organisatorisches zur Vorlesung
- Was ist das “Semantic Web”?

Vorstellung der Dozentin

Juniorprof. Dr. Birte Glimm

Seit Jul 11 Juniorprofessorin am Institut für Künstliche Intelligenz, Universität Ulm

Aug 08 – Jun 11 Wissenschaftliche Mitarbeiterin Information Systems Group, Department of Computer Science, University of Oxford

Apr 08 – Jul 08 Senior Software Engineer, sd&m AG, Hamburg

Okt 07 – Mär 08 Wissenschaftliche Mitarbeiterin Information Systems Group, Department of Computer Science, University of Oxford

Sep 04 – Sep 07 PhD Studentin, Information Management Group, School of Computer Science, University of Manchester

Vorstellung der Dozentin

Feb 2004 – Aug 2004 Austauschstudentin, Information Management Group, School of Computer Science, University of Manchester

Sep 2001 – Sep 2004 BSc. Informatik, Hochschule für angewandte Wissenschaften, Hamburg

Jan 2000 – Aug 2001 Softwareentwicklung, Elephant Seven GmbH, Hamburg

Apr 1998 – Dez 1999 Geschäftsführende Gesellschafterin der JEN.net GmbH

Apr 1995 – Mär 1998 Berufsfachschule Hamburger Technische Kunstschule, Kommunikationsdesign

Stipendiatin der Stiftung der Deutschen Wirtschaft

Vorstellung der Dozentin

Wissenschaftliche Interessen

- Beschreibungslogiken
- Abfragesprachen (z.B. SPARQL)
- Komplexitätstheorie
- Logik
- Semantic Web

Agenda

- Vorstellung der Dozentin
- Organisatorisches zur Vorlesung
- Was ist das “Semantic Web”?

Organisatorisches: Zeit und Ort

Vorlesung: Montag, 14:00 – 16:00

Übung: Donnerstag, 14:00 – 16:00

Ort: Raum H21, Gebäude O28

Webseite: `http://www.uni-ulm.de/in/ki/lehre/winter-2011-2012/semweb11.html`

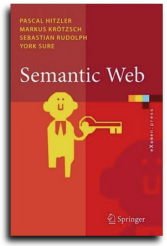
ILIAS: `https://elearning.uni-ulm.de/goto.php?target=crs_28608&client_id=Uni_Ulm`

Organisatorisches: Inhalt

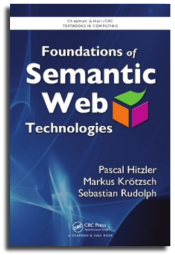
Einleitung und XML	17. Okt	SPARQL Syntax	12. Dez
Einführung in RDF	20. Okt	Übung 4	15. Dez
RDF Schema	24. Okt	SPARQL Semantik	19. Dez
fällt aus	27. Okt	SPARQL 1.1	22. Dez
Logik – Grundlagen	31. Okt	Übung 5	9. Jan
Übung 1	3. Nov	SPARQL Entailment	12. Jan
Semantik von RDF(S)	7. Nov	SPARQL Implemetierung	16. Jan
RDF(S) & Datalog Regeln	10. Nov	Abfragen & RIF	19. Jan
OWL Syntax & Intuition	14. Nov	Übung 6	23. Jan
Übung 2	17. Nov	Ontology Editing	26. Jan
OWL & BLs	21. Nov	Ontology Engineering	30. Jan
OWL 2	24. Nov	Linked Data	2. Feb
Tableau	28. Nov	Übung 7	6. Feb
Übung 3	1. Dez	SemWeb Anwendungen	9. Feb
Blocking & Unravelling	5. Dez	Wiederholung	13. Feb
Hypertableau	8. Dez	Übung 8	16. Feb

Literatur

Hitzler, Krötzsch, Rudolph, Sure
“Sematic Web Grundlagen”
Springer-Verlag



Hitzler, Krötzsch, Rudolph, Sure
“Foundations of Semantic Web Technologies”
CRC Press



Agenda

- Vorstellung der Dozentin
- Organisatorisches zur Vorlesung
- Was ist das “Semantic Web”?

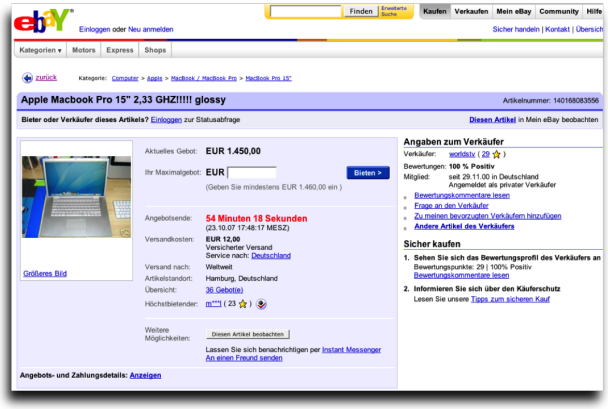
Das Web

Das Web flankiert den Übergang von der Industrie- zur Informationsgesellschaft und bietet die Infrastruktur für eine neue Qualität des Umgangs mit Information hinsichtlich Beschaffung wie auch Bereitstellung.

- hohe Verfügbarkeit
- hohe Aktualität
- geringe Kosten

Das Web

Kommerzialisierung in allen Größenordnungen



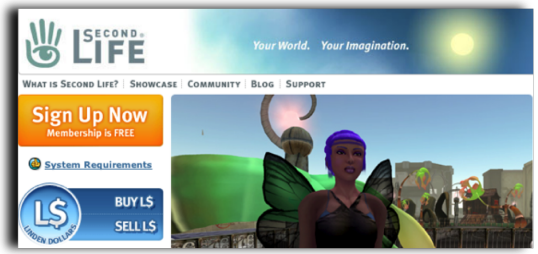
Das Web

Kommerzialisierung in allen Größenordnungen

Das Web

Weitere Lebensbereiche werden “webisiert”:

- Behörden, Verwaltung (eGovernment)
- Ausbildung (eLearning, eEducation)
- Sozialkontakte (Social-Networking-Plattformen, Partnerbörsen)
- Alltag?



Warum Semantic Web?

Exkurs: Syntax vs. Semantik

Syntax (von grch. $\sigma\upsilon\nu\tau\alpha\xi\iota\varsigma$ – Zusammenstellung, Satzbau) steht für die (normative) Struktur von Daten, d.h. sie charakterisiert, was “wohlgeformte” Daten sind.

Semantik (grch. $\sigma\eta\mu\alpha\upsilon\tau\iota\kappa o\varsigma$ – zum Zeichen gehörend) steht für die Bedeutung von Daten, d.h. sie charakterisiert beispielsweise, welche inhaltliche Schlussfolgerungen sich ziehen lassen.

$4+) = ($	$3 + 4 = 12$	$3 + 4 = 7$
syntaktisch falsch	syntaktisch richtig	syntaktisch richtig
–	semantisch falsch	semantisch richtig

Probleme des Web

- Fülle an Informationen
- Ausgerichtet auf Menschen als Endnutzer

Guilin Qi

Professor
School of Computer Science and Engineering
Southeast University
China

Tel : +86 (0) 25 52090910
Fax : +86 (0) 25 52090880
E-mail : gqi@seu.edu.cn

Curriculum Vitae (PDF)

Brief Introduction : Dr. Guilin Qi is a professor working at Southeast University in China. His research topics include knowledge representation and reasoning, semantic Web, uncertainty reasoning. His current research interests include the areas of

- **Knowledge representation**: belief merging, belief revision, inconsistency handling, nonmonotonic reasoning, information fusion, argumentation, paraconsistent logic

Nur Computer-lesbare Layout Informationen

```
<h1>Ian Horrocks</h1>
<table><tr>
  <td class="personImg">
    <img src="IH005-1.jpg" alt="Photo Ian Horrocks"/>
  </td>
  <td>
    <div class="personinfo">
      <div>Professor Ian Horrocks FRS</div>
      <div>Professor of Computer Science</div>
      <div>Fellow, <a href="http://www.oriel.ox.ac.uk">Oriel Col
      <div>ian.horrocks@cs.ox.ac.uk</div>
      <div>+44 1865 273939</div>
      <div>+44 1865 273839 (fax)</div>
    </div>
    <p>Wolfson Building, Parks Road, Oxford OX1 3QD</p>
  </td>
</tr></table>
```


Probleme des Web

- Lokalisierung von Information problematisch heutige
- Suchmaschinen gut, aber stichwortbasiert
- wünschenswert: inhaltliche, semantische Suche

Probleme des Web

- Heterogenität der vorhandenen Information auf verschiedensten Ebenen:
  - Zeichenkodierung (z.B. ASCII vs. Unicode)
  - verwendete natürliche Sprachen
  - Anordnung von Information auf Webseiten
- Informationsintegration aber stichwortbasiert
- wünschenswert: inhaltliche, semantische Suc

Probleme des Web

- **implizites Wissen**, d.h. Informationen, sind nicht explizit spezifiziert, folgen aber aus der Kombination gegebener Daten
- formallogische Methoden erforderlich
- *automatisches Schlussfolgern*

Probleme des Web

Lösungsansätze:

1. Ad hoc: Verwendung von KI-Methoden zur Auswertung bestehender unstrukturierter Informationen im Web
2. A priori: Strukturierung der Web-Informationen zur Erleichterung der automatisierten Auswertung

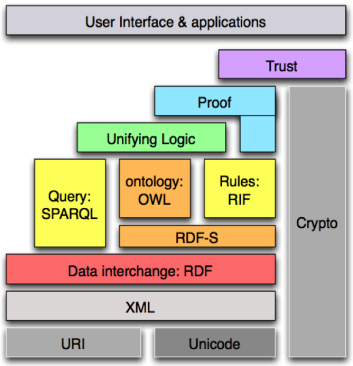
$\Rightarrow$ Semantic Web

Probleme des Web

Zwei essentielle Voraussetzungen zur Realisierung:

1. offene Standards zur Beschreibung von Informationen
   - klar definiert
   - flexibel
   - erweiterbar
2. Methoden zur Gewinnung von Informationen aus derlei Beschreibungen

Semantic Web – Standards



1994	First public presentation of the Semantic Web idea
1998	Start of standardization of data model (RDF) and a first ontology languages (RDFS) at W3C
2000	Start of large research projects about ontologies in the US and Europe (DAML & Ontoknowledge)
2002	Start of standardization of a new ontology language (OWL) based on research results
2004	Finalization of the standard for data (RDF) and ontology (OWL)
2008	Standardization of a query language (SPARQL)
2009	Extension of OWL to OWL 2.0
2010	Standard Rule Interchange Format (RIF)

Einführung in XML

- Einleitung und XML
- Einführung in RDF
- RDF Schema
- Logik – Grundlagen
- Semantik von RDF(S)
- Regeln
- OWL – Syntax und Intuition 1
- OWL – Syntax und Intuition 2
- OWL Semantik
- Abfragesprache SPARQL
- SPARQL Semantik
- SPARQL Entailment Regimes
- Reasoning Verfahren 1
- Reasoning Verfahren 2
- Reasoning Verfahren 3
- Konj. Abfragen & Regeln
- Linked Data
- SemWeb Anwendungen
- Wiederholung

Agenda

- XML – Motivation/Idee
- XML – Syntax
- URIs
- Namensräume

Annotation mit Markup-Sprachen

- Grundidee des Markup: versehen von (unstrukturiertem) Text mit zusätzlicher Information (bzw. Struktur)
- synonym: auszeichnen, auch: annotieren von Text

Text	= Daten
Zusatzinformation	= Metadaten

Annotation mit Markup-Sprachen

- häufige Markup-Strategie: Einschließen des zu annotierenden Textes in sogenannte tags (engl.: Etikett, Schild):

```
<Tag-Bezeichner> ...Text... </Tag-Bezeichner>
```

↑ öffnendes Tag ↑ schließendes Tag

- Zusatzinformation wird von verarbeitenden Programmen gelesen und interpretiert

Annotation mit Markup-Sprachen

- prominentestes Beispiel: HTML Tags kodieren Darstellungsinformationen:
  ```
  <i>Dieses Buch</i> hat den Titel
  <b>Semantic Web Grundlagen</b>.
  ```
- Darstellung im Browser:
  Dieses Buch hat den Titel **Semantic Web Grundlagen**.
- Strategie auch geeignet zur inhaltlichen Annotation, z.B.:
  ```
  <Vorname>Birte</Vorname>
  <Nachname>Glimm</Nachname> arbeitet in
  <Stadt>Ulm</Stadt>.
  ```

Annotation mit Markup-Sprachen

```
<Vorlesung>
  <Titel>SemWeb</Titel>
  <Dozent>
    <Titel>Dr.</Titel>
    <Vorname>Birte</Vorname>
    <Nachname>Glimm</Nachname>
  </Dozent>
</Vorlesung>
```

- Verschachtelung von Tags erlaubt
- Doppelte Nutzung von Tags erlaubt
- XML Tags bilden eine Baumstruktur

Agenda

- XML – Motivation/Idee
- XML – Syntax
- URIs
- Namensräume

XML

- eXtensible Markup Language
- Ursprung: strukturierter Text (HTML4.0 $\in$ XML $\subset$ SGML)
- Web-Standard (W3C) zum Datenaustausch:
  - Ein- und Ausgabedaten von Anwendungen können mittels XML beschrieben werden
  - Industrie muss sich nur noch auf standardisierte Beschreibung (= Vokabular) einigen
- Komplementärsprache zu HTML:
  - HTML beschreibt die Präsentation
  - XML beschreibt den Inhalt
- Datenbank-Sichtweise: XML als Datenmodell für semistrukturierte Daten

XML-Syntax – Präambel

- XML-Dokument ist Textdokument
- Beginnt mit Deklaration
  - Enthält Versionsnummer des verwendeten Standards
  - Optional: die Zeichenkodierung

```
<?xml version="1.0" encoding="utf-8"?>
```

XML-Syntax – XML-Element

XML-Element (engl. element):

- Beschreibung eines Objekts, die durch passende Markierungen (tags) geklammert ist
- Inhalt eines Elements: Text und/oder weitere Elemente (beliebige Schachtelung möglich)
- Leere Elemente: `<year></year>` kurz: `<year/>`
- "äußerstes" Element heißt Wurzelelement (nur eines pro XML-Dokument)

Startmarkierung
Unterelemente
Freitext
Endmarkierung

```
<author>
   <firstname> Serge </firstname>
   <lastname> Abiteboul </lastname>
   <email> sab@abc.com </email>
   email address may be wrong!
</author>
```

Element author

XML-Syntax – XML-Attribute

XML-Attribut (engl. attribute):

- Name-Zeichenkettenwert-Paar in Start- oder selbstschließendem Tag
- Assoziiert mit einem Element
- Alternative Möglichkeit, Daten zu beschreiben

Attribut email

```
<author email="sab@abc.com">
   <firstname> Serge </firstname>
   <lastname> Abiteboul </lastname>
</author>
```

Weitere denkbare Beschreibung derselben Daten:

```
<author firstname="Serge" lastname="Abiteboul" email="sab@abc.com"/>
```


HTML vs. XML

- HTML: festes Vokabular (Menge von tags) und Semantik (die Darstellung von Text)
- XML: freie Bezeichner zur Beschreibung von anwendungsspezifischer Syntax und Semantik
- XML $\subset$ SGML

```
<h1> Bib </h1>
<p>
   <i> Foundations of Databases </i>
   Serge Abiteboul
   </br> Addison Wesley, 1997
<p>
...
```

HTML

```
<Bib id="o1">
   <paper id="o12">
      <title> Foundations of Databases </title>
      <author>
         <firstname> Serge </firstname>
         <lastname> Abiteboul </lastname>
      </author>
      <year> 1997 </year>
      <publisher> Addison Wesley </publisher>
   </paper>
   ...
</Bib>
```

XML

Agenda

- XML – Motivation/Idee
- XML – Syntax
- URIs
- Namensräume

URIs – Idee

- URI = Uniform Resource Identifier
- Dienen zur weltweit eindeutigen Bezeichnung von Ressourcen
- Ressource kann jedes Objekt sein, was (im Kontext der gegebenen Anwendung) eine klare Identität besitzt (z.B. Bücher, Orte, Menschen, Verlage, Beziehungen zwischen diesen Dingen, abstrakte Konzepte usw.)
- In bestimmten Domänen ähnliches bereits realisiert: ISBN für Bücher

URIs – Syntax

- Erweiterung des URL-Konzeptes; nicht jede URI bezeichnet aber ein Webdokument (umgekehrt wird als URI für Webdokumente häufig deren URL verwendet)
- Beginnt mit dem sogenannten URI-Schema, das durch einen Doppelpunkt (:) vom nachfolgenden Teil getrennt ist (z.B.: http, ftp, mailto)
- Häufig hierarchisch aufgebaut

URIs – Selbstdefinierte URIs

- Nötig, wenn für eine Ressource (noch) keine URI existiert (bzw. bekannt ist)
- Strategie zur Vermeidung von (ungewollten) Überschneidungen: Nutzung von http-URIs einer eigenen Webseite
- Ermöglicht auch Ablegen einer Dokumentation zur URI an dieser Stelle

Beschreibendes vs. Beschriebenes

- Trennung von URI für Ressource und deren Dokumentation durch URI-Referenzen (durch “#” angehängte Fragmente) oder content negotiation
- Z.B.: als URI für Shakespeares “Othello” http://de.wikipedia.org/wiki/Othello nicht geeignet, besser http://de.wikipedia.org/wiki/Othello#URI

Agenda

- XML – Motivation/Idee
- XML – Syntax
- URIs
- Namensräume

XML-Namensräume: Motivation

- XML-Dokumente besitzen Element- und Attributnamen (“Markup Vocabulary”) mit allgemeiner Gültigkeit
- Eine XML-Anwendung basiert auf allgemeiner Interpretation dieser Namen
- Ein XML-Dokument soll Markup-Vokabular aus mehreren “Dictionaries” enthalten können. (Erinnerung: XML-Dokument muss keine DTD haben.)
- Namespaces zur Vermeidung von Namenskonflikten

XML-Namensräume

- XML Namespaces sind ähnlich zu Modul-Konzepten in Programmiersprachen
- Disambiguierung von Tag-Namen durch Verwendung unterschiedlicher “Prefixe”
- Ein Prefix wird vom lokalen Namen separiert durch einen Doppelpunkt (:), so entstehen `prefix:name` Tags
- Namespace-Bindungen werden von manchen Werkzeugen ignoriert, sog. “flache Namespaces”

Namensraum-Bindungen

- Prefixe werden belegt mit Namespace URIs, indem ein Attribut `xmlns:prefix` bei dem relevanten Element oder einem seiner Vorgängerelemente eingefügt wird: $\texttt{prefix:name}_1, \ldots, \texttt{prefix:name}_n$
- Der Wert des `xmlns:prefix`-Attributes ist eine URI, welche (für XML Schemata) auf eine Beschreibung der Namespace Syntax verweisen kann aber nicht muss
- Ein Element kann Bindings nutzen für mehrere (unterschiedliche) Namespaces durch Verwendung separater Attribute $\texttt{xmlns:prefix}_1, \ldots, \texttt{xmlns:prefix}_m$

Beispiel: Ohne Namensräume

```
<Vorlesung>
  <Titel>SemWeb</Titel>
  <Dozent>
    <Titel>Dr.</Titel>
    <Vorname>Birte</Vorname>
    <Nachname>Glimm</Nachname>
  </Dozent>
</Vorlesung>
```

Titel ist mehrdeutig verwendeter Tagname

Zwei verschiedene Namensräume

```
<Vorlesung  xmlns:lv="http://www.semantic-web-Grundlagen/Lehrveranstaltungen"
            xmlns:person="http://www.semantic-web-Grundlagen/Person">
  <lv:Titel>SemWeb</lv:Titel>
  <lv:Dozent>
    <person:Titel>Dr.</person:Titel>
    <person:Vorname>Birte</person:Vorname>
    <person:Nachname>Glimm</person:Nachname>
  </lv:Dozent>
</lv:Vorlesung>
```

`Titel` wurde disambiguiert durch Verwendung der Prefixe `lv` und `person`

Agenda

- XML – Motivation/Idee
- XML – Syntax
- URIs
- Namensräume