



Birte Glimm  
Institut für Künstliche Intelligenz | 17. Okt 2011

## Semantic Web Grundlagen Einführung und XML

## 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>

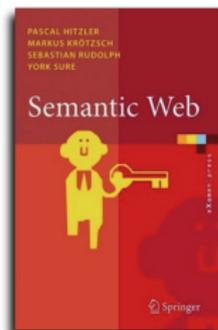
**ILIAS:** [https://elearning.uni-ulm.de/goto.php?target=crs\\_28608&client\\_id=Uni\\_Ulm](https://elearning.uni-ulm.de/goto.php?target=crs_28608&client_id=Uni_Ulm)

## Organisatorisches: Inhalt

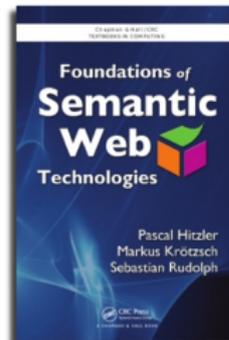
Einleitung und XML	17. Okt	SPARQL Syntax	12. Dez
Einführung in RDF	20. Okt	Übung 4	15. Dez
RDF Schema	24. Okt	SPARQL Semantik	19. Dez
fällt aus	27. Okt	SPARQL 1.1	22. Dez
Logik – Grundlagen	31. Okt	Übung 5	9. Jan
Übung 1	3. Nov	SPARQL Entailment	12. Jan
Semantik von RDF(S)	7. Nov	SPARQL Implementierung	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

**ebay** Einloggen oder Neu anmelden

Kategorien v Motors Express Shops

zurück Kategorien: Computer > Apple > MacBook / MacBook Pro > MacBook Pro 15"

**Apple Macbook Pro 15" 2,33 GHZ!!!!!! glossy** Artikelnummer: 140188083598

Bieter oder Verkäufer dieses Artikels? Einloggen zur Statusabfrage [Diesen Artikel in Mein eBay beobachten](#)



[Größeres Bild](#)

Aktuelles Gebot: **EUR 1.450,00**

Ihr Maximalgebot: EUR  [Bieten >](#)  
(Geben Sie mindestens EUR 1.460,00 ein)

Angebotsende: **54 Minuten 18 Sekunden**  
(23.10.07 17:48:17 MESZ)

Versandkosten: **EUR 12,00**  
Versicherter Versand  
Service nach: [Deutschland](#)

Versand nach: Weltweit  
Artikelstandort: Hamburg, Deutschland  
Übersicht: [36 Gebote](#)

Höchstbietender: [m...](#) ( 23 ☆ )

Weitere Möglichkeiten: [Diesen Artikel beobachten](#)

[Lassen Sie sich benachrichtigen per Instant Messenger](#)  
[An einen Freund senden](#)

Angebots- und Zahlungsdetails: [Anzeigen](#)

**Angaben zum Verkäufer**

Verkäufer: [wofidaly](#) ( 28 ☆ )

Bewertungen: **100 % Positiv**

Mitglied: seit 29.11.00 in Deutschland  
Angemeldet als privater Verkäufer

- » [Bewertungskommentare lesen](#)
- » [Frage an den Verkäufer](#)
- » [Zu meinen bevorzugten Verkäufern hinzufügen](#)
- » [Andere Artikel des Verkäufers](#)

**Sicher kaufen**

1. **Sehen Sie sich das Bewertungsprofil des Verkäufers an**  
Bewertungspunkte: 29 | 100% Positiv  
[Bewertungskommentare lesen](#)
2. **Informieren Sie sich über den Käferschutz**  
Lesen Sie unsere [Tipps zum sicheren Kauf](#)

# Das Web

## Kommerzialisierung in allen Größenordnungen

The image shows two overlapping screenshots of online marketplaces. The foreground screenshot is from Amazon.de, displaying a product page for the book "Semantic Web. Grundlagen (eXamen.press) (Taschenbuch)". The book cover features a yellow figure holding a key against a red background. The price is listed as EUR 24,95 with free shipping. The page also includes a "Verfügbarkeit" section stating the item is not yet available and a "Noch 4 Tage bis zum Erscheinungstermin von Harry Potter Band 7" notice.

The background screenshot is from eBay, showing a listing for the same book. The eBay interface includes a search bar, navigation links like "Kategorien", "Motors", "Express", and "Shops", and a sidebar with "Angaben zum Verkäufer" showing a 100% positive feedback rating and a 29/100% positive feedback score.

**Amazon.de Product Page:**

**Verfasser:** Pascal Hitzler (Autor), Markus Krötzsch (Autor), Sebastian Rudolph (Autor), York Sure (Autor)

**Preis:** EUR 24,95 **Kostenlose Lieferung.** [Siehe Details.](#)

**Verfügbarkeit:** Dieser Artikel ist noch nicht erschienen. Reservieren Sie sich Ihr Exemplar jetzt und Sie erhalten es pünktlich zum Erscheinungstermin. Verkauf und Versand durch **Amazon.de**. Geschenkverpackung verfügbar. Zustellung durch **Amazon**.

**Noch 4 Tage bis zum Erscheinungstermin von Harry Potter Band 7.** Sichern Sie sich jetzt Ihr Exemplar mit **Liefergarantie -- sonst geschenkt!**

**eBay Listing:**

**Angaben zum Verkäufer**

Verkäufer: [worldlyx](#) (28 ★)

Bewertungen: 100 % Positiv

Mitglied: seit 29.11.00 in Deutschland

Angemeldet als privater Verkäufer

- [Bewertungskommentare lesen](#)
- [Frage an den Verkäufer](#)
- [Zu meinen bevorzugten Verkäufern hinzufügen](#)
- [Andere Artikel des Verkäufers](#)

**Sicher kaufen**

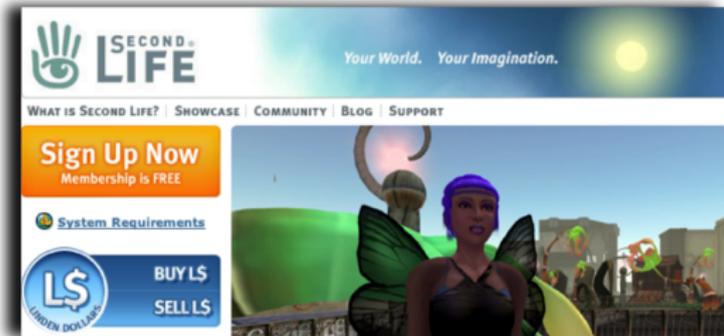
1. Sehen Sie sich das **Bewertungsprofil des Verkäufers** an
2. Informieren Sie sich über den **Käuferschutz**

Lesen Sie unsere [Tipp](#)s zum **sicheren Kauf**

## 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. *συνταξις* – Zusammenstellung, Satzbau) steht für die (normative) Struktur von Daten, d.h. sie charakterisiert, was “wohlgeformte” Daten sind.

**Semantik** (grch. *σημαυτικός* – 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](mailto: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">
    
  </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">O
      <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

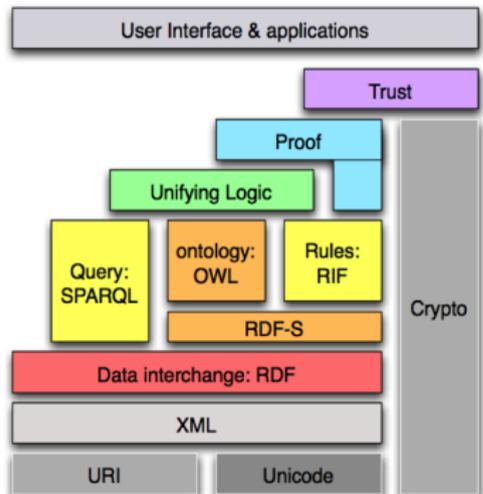
⇒ Semantic Web

## 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

- ▶ 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 ( $\text{HTML4.0} \in \text{XML} \subset \text{SGML}$ )
- ▶ 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

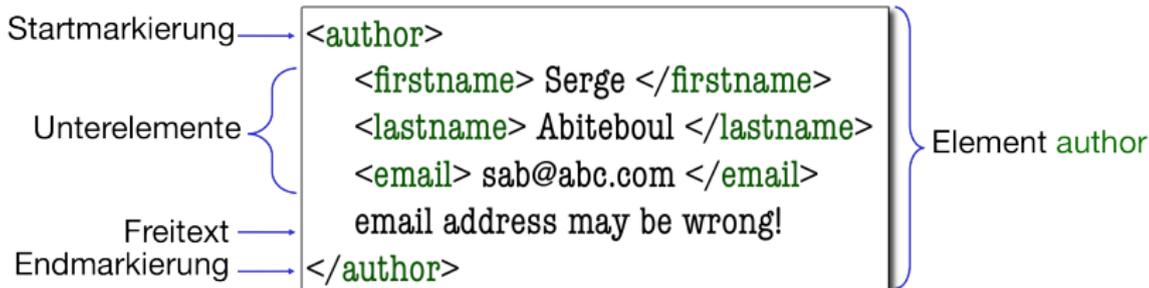
- ▶ XML-Dokument ist Textdokument
- ▶ Beginnt mit Deklaration
  - ▶ Enthält Versionsnummer des verwendeten Standards
  - ▶ Optional: die Zeichenkodierung

```
<?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 Wurzelement (nur eines pro XML-Dokument)



## 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="01">
  <paper id="012">
    <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”

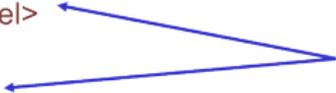
## Namensraum-Bindungen

- ▶ Prefixe werden belegt mit Namespace URIs, indem ein Attribut `xmlns:prefix` bei dem relevanten Element oder einem seiner Vorgängerelemente eingefügt wird:  
`prefix:name1, ..., prefix:namen`
- ▶ 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 `xmlns:prefix1, ..., xmlns:prefixm`

## 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