



XML Technologien – SS 2011

Prof. Dr.-Ing. Robert Tolksdorf
& Markus Luczak-Rösch
Freie Universität Berlin
Institut für Informatik
Netzbasierte Informationssysteme

tolk@ag-nbi.de
markus.luczak-roesch@fu-berlin.de

- Organisatorisches
- Was ist XML?
- Anwendungen von XML
- Überblick über die Vorlesung
- Literatur



Organisatorisches

- Vorlesung:
Prof. Dr.-Ing. Robert Tolksdorf, tolk@ag-nbi.de
Markus Luczak-Rösch, markus.luczak-roesch@fu-berlin.de
- AG Netzbasierte Informationssysteme
- Büro: Königin-Luise-Str. 24-26, 1.OG, Raum 118
(NICHT Takustr. 9)
- Sprechstunde: Di. 15-17 Uhr
 - Termine per Mail abstimmen
 - oder bei Herrn Tolksdorf via Form anmelden
<http://flp.cs.tu-berlin.de/%7Etolk/sprechstunde>

- <http://blog.ag-nbi.de/2011/04/04/vorlesung-xml-technologien/>
- hier finden sich
 - Folien der Vorlesungen
 - Termine der Vorlesungen und Projektarbeit
 - Hinweise/Links auf Literatur

- >60 Teilnehmer haben sich im Online-KVV angemeldet (Stand: 06.04.2011)
- **Master- und Bachelor-Studierende:**
 - zusätzlich verbindliche Anmeldung mit Unterschrift notwendig
 - Ohne diese Anmeldung dürfen keine Leistungen erbracht werden.
 - verbindliche Anmeldung für Msc-Studierende in der nächsten Woche

- Es gibt keine verpflichtenden Übungsaufgaben
- Stattdessen: **Projektarbeit** in Gruppen zu ≤ 10 Personen
- Was erwartet Sie?
 - Wir präsentieren zu Beginn der Praxisphase zwei Projektaufgaben und bilden Projektgruppen, die diese bearbeiten
 - Präsenztermine zur Vorlesungszeit sind Betreuungstermine
 - 10 Minuten Zeitfenster je Gruppe
 - Sie fragen, ich antworte
 - Meilensteinpräsentation zum Fortschritt Mitte der Projektarbeitsphase (alle Gruppen)
 - Abschlusspräsentation am Ende des Semesters (alle Gruppen)

- Ab 24.05. bieten wir zwei Tutorien an (Di. und Mi. zur Vorlesungszeit)
 - Behandlung von Vorlesungsstoff in Übungsaufgaben
 - Präsentation einer Musterlösung

- Werden erst zum Auftakt der Projektphase bekanntgegeben
- Simulieren ein kleines Kundenprojekt
- Werden inhaltlich alle Bereiche des Vorlesungsstoffs abdecken

Scheinkriterien

- aktive Teilnahme an Projektarbeit
- Klausur bzw. Nachklausur erfolgreich bestanden
- Schein wird grundsätzlich benotet

Klausur

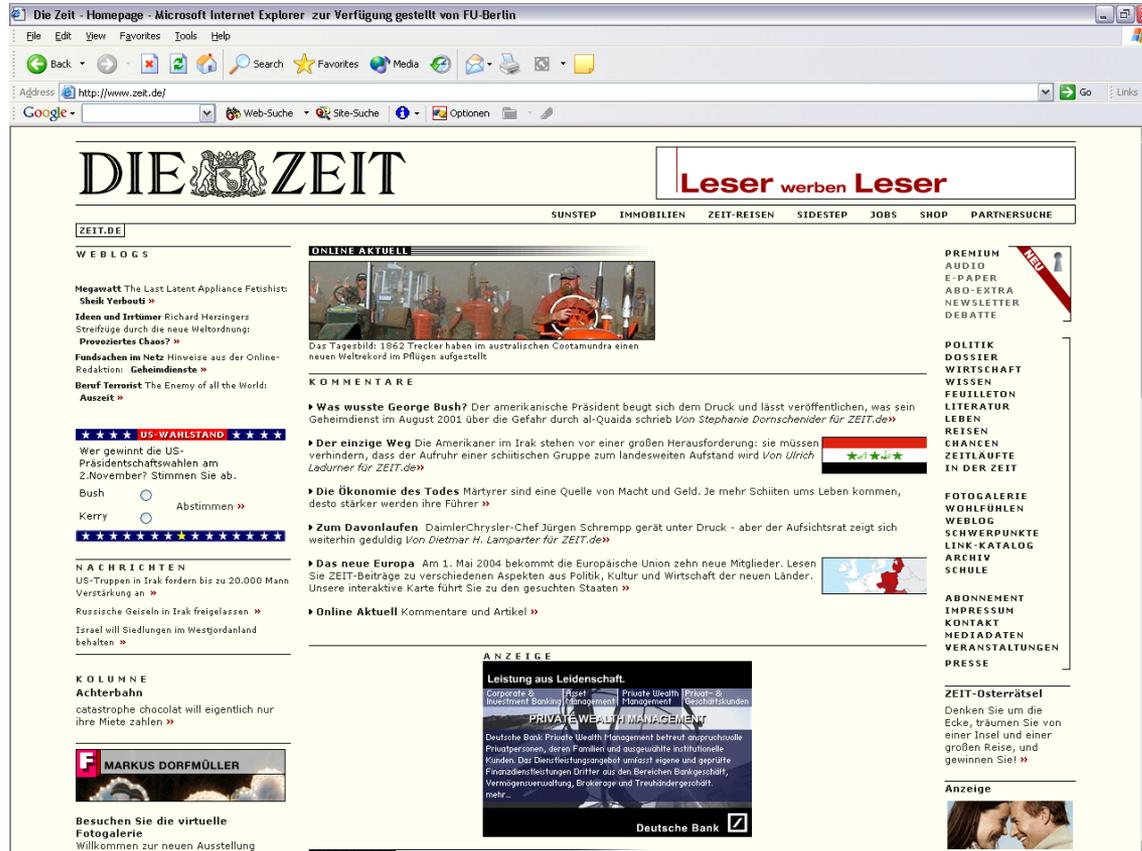
- Teilnahmevoraussetzung: Anmeldung im Online-KVV
- Klausurtermin: 13.07.2011 (letzter Vorlesungstermin)

Nachklausur

- Teilnahmevoraussetzung: Klausur nicht bestanden
- einzige Ausnahme: ärztliches Attest für Klausurtermin
- Termin: wird noch bekannt gegeben



Was ist XML?

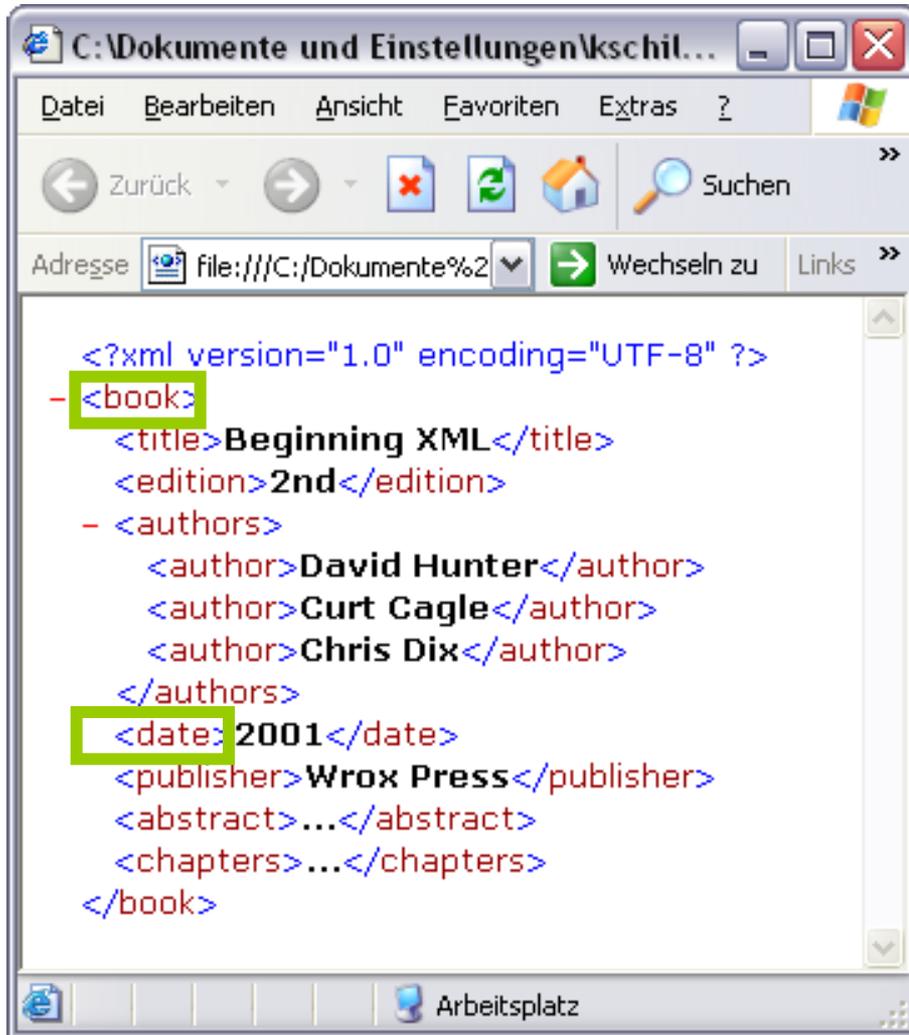


HTML hat sich für die Präsentation von Inhalten bewährt.

immer häufiger medienneutrale Darstellung nötig:

- Vielfalt von Endgeräten und Bandbreiten macht Trennung Inhalt von Präsentation nötig
- Austausch von Daten und Dokumenten zwischen Computern
 - ⇒ z.B. Übermittlung eines Bestellformulars
 - ⇒ z.B. Web Services

HTML: *keine* layoutunabhängige Darstellung von Inhalten



```
<?xml version="1.0" encoding="UTF-8" ?>
- <book>
  <title>Beginning XML</title>
  <edition>2nd</edition>
  - <authors>
    <author>David Hunter</author>
    <author>Curt Cagle</author>
    <author>Chris Dix</author>
  </authors>
  <date>2001</date>
  <publisher>Wrox Press</publisher>
  <abstract>...</abstract>
  <chapters>...</chapters>
</book>
```

- Extensible Markup Language
- erlaubt Strukturieren von Inhalten
- Unterschiede zu HTML:
 - Medienneutral
- Tag-Namen
<name>...</name> beliebig
- generische
Auszeichnungssprache

- textbasierte Sprachen, die Dokumente mit zusätzlichen Tags („Markierungen“) versehen:

`<tag-name>ausgezeichneter Text</tag-name>`



- dadurch zusätzliche Information (Metainformationen)
- Beispiel: Hypertext Markup Language (HTML)
- kombinieren Vorteile von Binärdateien mit denjenigen von Textdateien:
 - anwendungsunabhängige Dateiformate, die reichhaltige Metadaten enthalten können

HTML

- vorgegebene Auswahl von Tags, keine anderen dürfen verwendet werden

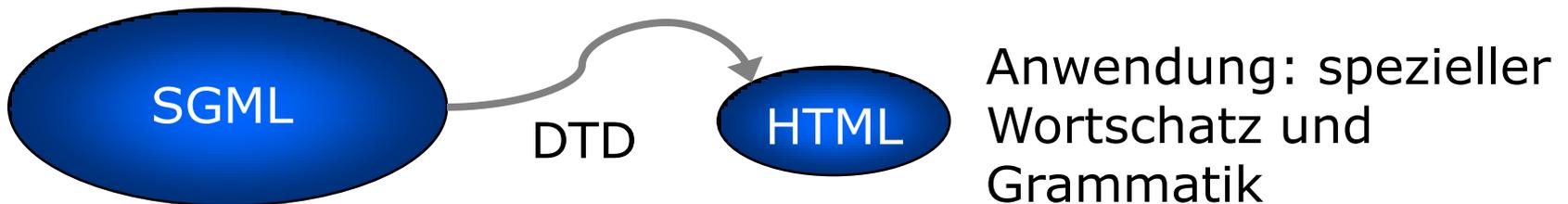
generische Auszeichnungssprache (generalized markup language)

- keine Tags vorgegeben, beliebige Tags erlaubt
- Vorteil: beliebige Metainformationen darstellbar
- Nachteil: Bedeutung der Metainformationen (Tags) offen
- Beispiele: SGML und XML

- 60'er** Generische Kodierung – Graphic Communication Association (GCA)
- 1969** Charles Goldfarb entwickelt bei IBM die Generalized Markup Language (GML).
- 1980** ANSI veröffentlicht ersten Entwurf von SGML.
- 1986** ISO verabschiedet SGML.
- 1989** Bernes-Lee schlägt SGML-basiertes Hypertext-System vor.
- 1990** Beners-Lee entwickelt HTML, HTTP und URL. World Wide Web nimmt Betrieb mit zwei Maschinen am CERN auf.
- 1995** HTML 2.0
- 1998** XML 1.0 (einschl. DTDs)
- 2000** XML 1.0, 2nd Edition
XHTML 1.0 (Reformulierung von HTML in XML)
- 2004** XML 1.1
XML 1.0, 3rd Edition
- 2006** XML 1.1, 2nd Edition
XML 1.0, 4th Edition

- Standard Generalized Markup Language
- 1969 von Charles Goldfarb und zwei seiner Kollegen bei IBM für das Dokumentenmanagement entwickelt.
- seit 1986 ein internationaler Standard
- keine vorgegebenen Tags, auch keine für das Layout von Dokumenten
- Vorgänger von XML

- gibt zwar keine konkreten Tags vor
- Mit Document Type Definitions (DTDs) können aber spezielle Auszeichnungssprachen mit konkreten Tags definiert werden:
 - werden Anwendungen von SGML genannt
 - bekannteste Anwendung von SGML: HTML



- Anwendung selbst kann keine Anwendung definieren

- HTML
 - erst XHTML (XML) → heute HTML 5
- DocBook – Dokumentformat (vornehmlich für Bücher, Artikel und Dokumentationen)
 - heute XML
- TEI – Text Encoding Initiative
 - heute XML
- EAD – Encoded Archival Description
 - heute XML

Vor- und Nachteile von SGML

- + kombiniert Vorteile von Binärdateien mit denjenigen von Textdateien
- + beliebig erweiterbar
- + erlaubt die Definition von konkreten Auszeichnungssprachen wie HTML
- sehr komplex: Spezifikation über 600 Seiten lang
- SGML-Parser schwierig zu implementieren

- **HTML**

- für Präsentation von Web-Inhalten bewährt
- keine medienneutrale Darstellung von Inhalten

- **medienneutrale Darstellung**

- generische Auszeichnungssprachen (wie SGML) geeignet

- **SGML**

- für das Web SGML viel zu komplex

XML: konsequente Vereinfachung von SGML, die für Web-Anwendungen hinreichend **allgemein** ist.

Was bedeutet Erweiterbarkeit?

- X in XML steht für erweiterbar (extensible).
- Was bedeutet Erweiterbarkeit? → Vergleich HTML vs. XML hilfreich:
- **HTML**
 - vorgegebene Auswahl an Tags
 - Neues Tag kann nur eingeführt werden, wenn sich das W3C auf eine neue HTML-Version einigt!
- **XML**
 - beliebige Tags können benutzt werden
 - Anwender des entsprechenden Tags müssen sich auf eine gemeinsame Interpretation des Tags einigen

Die XML-Familie: Der Kern

- **XML 1.0 / 1.1**
 - Syntax wohlgeformter XML-Dokumente
 - Definition von Anwendungen (Untermengen) mit DTDs
- **Namensräume**
 - gleichzeitige Verwendung unterschiedlicher Vokabularien
 - z.B. Unterscheidung Titel einer Person vom Titel eines Buches
 - Festlegung der Bedeutung von Tags
- **XML-Schema**
 - gleiche Aufgabe wie DTDs
 - jedoch wesentlich mächtiger

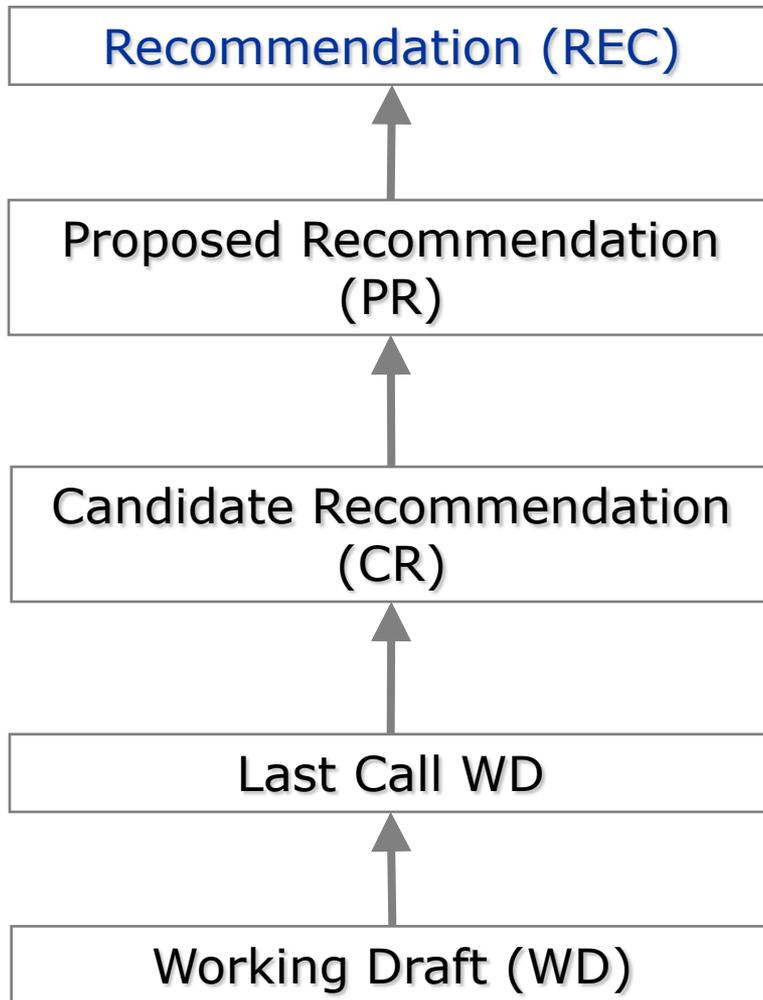
- **Extensible Stylesheet Language (XSLT)**
 - Transformation von XML-Dokumenten in beliebige Text-Formate:
XML → HTML / WML / XML / ASCII / ...
- **XPath**
 - Zugriff auf beliebige Teile eines XML-Dokuments
 - z.B. Zugriff auf alle Buchtitel
- **XPointer**
 - Erweiterung der X-Path-Spezifikation
- **XQuery**
 - Abfragesprache

- **XLink**
 - Syntax zur Definition von Links in XML-Dokumenten
- **XInclude**
 - Verweis auf Teile von XML- oder Textdateien
 - wird beim Auflösungsprozess durch den Inhalt der Datei ersetzt
- **Document Object Model (DOM)**
 - Parsen, Modifizieren und Erstellen von XML-Dokumenten

gesamte XML-Familie besteht aus
lizenzfreien W3C-Standards



- 1994 als Projekt am MIT gegründet
- keine Normierungsorganisation im klassischen Sinn
- kann Einhaltung von Normen nicht auf rechtlichem Wege einklagen
- definiert deshalb lediglich Empfehlungen (recommendations)
- W3C-Recommendations lizenzfrei



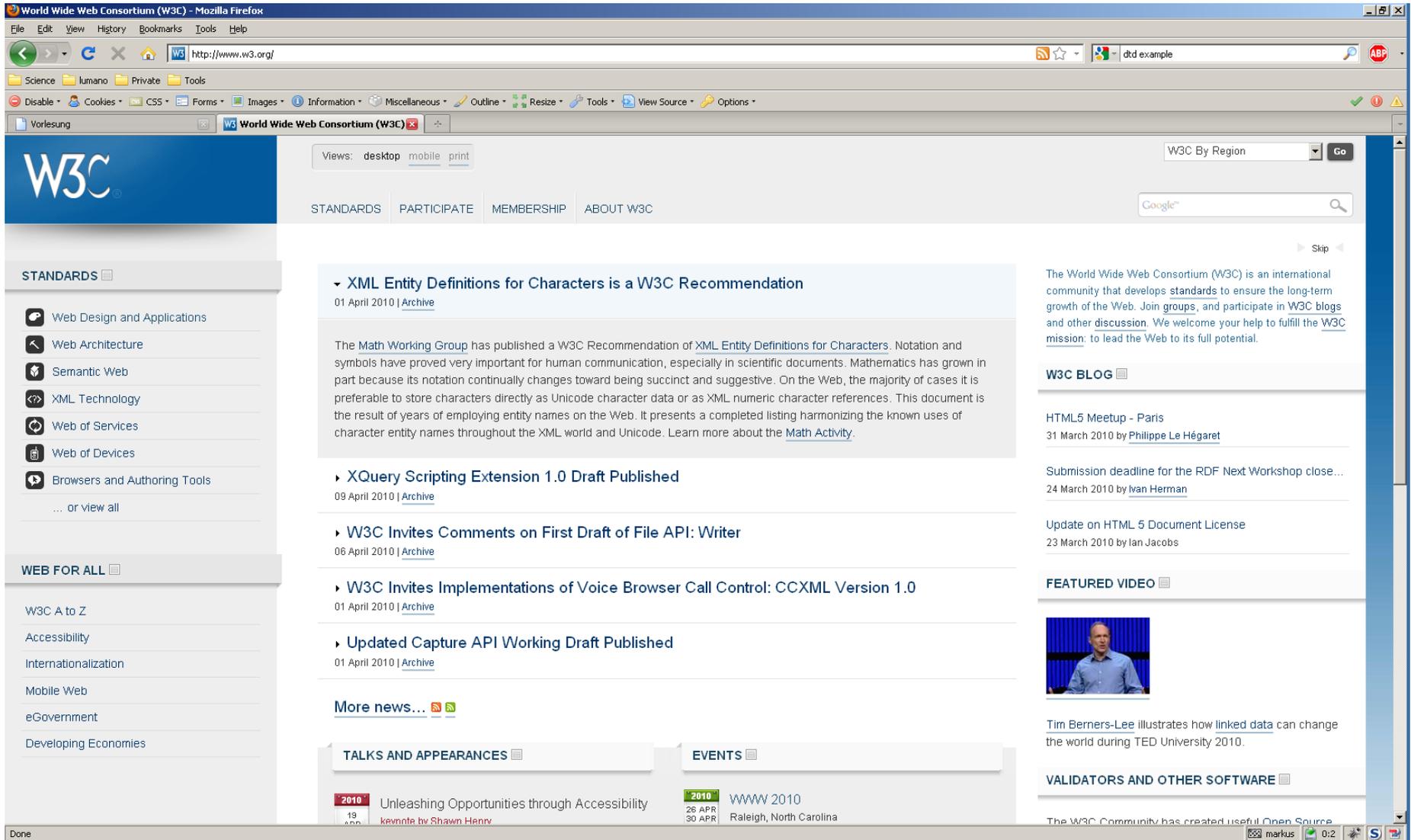
offizieller W3C-Standard

offizieller Konsens der betreffenden AG dar, wird dem Advisory Committee übergeben

Direktor: definierte Ziele erreicht, von entsprechender Community begutachtet

letztes WD, definierte Anforderungen erreicht

aktueller Diskussionsstand einer AG

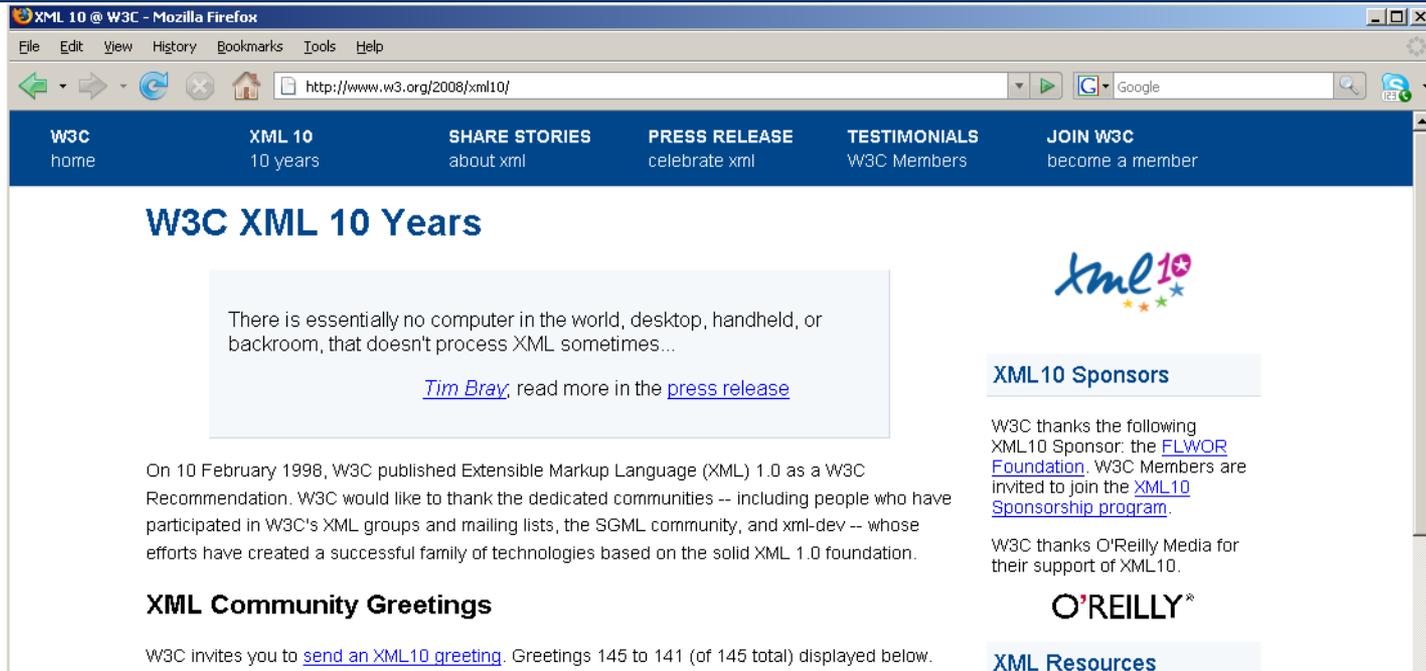


The screenshot shows the W3C website in a Mozilla Firefox browser window. The browser's address bar shows the URL <http://www.w3.org/>. The website features a blue header with the W3C logo and navigation links for STANDARDS, PARTICIPATE, MEMBERSHIP, and ABOUT W3C. A search bar is located in the top right corner. The main content area is titled "XML Entity Definitions for Characters is a W3C Recommendation" and includes a sub-header "01 April 2010 | Archive". The text describes the Math Working Group's publication of a W3C Recommendation of XML Entity Definitions for Characters. Below this, there are several other news items, including "XQuery Scripting Extension 1.0 Draft Published", "W3C Invites Comments on First Draft of File API: Writer", "W3C Invites Implementations of Voice Browser Call Control: CCXML Version 1.0", and "Updated Capture API Working Draft Published". A "More news..." link is also present. On the right side, there is a "W3C BLOG" section with entries for "HTML5 Meetup - Paris" and "Submission deadline for the RDF Next Workshop close...". Below the blog is a "FEATURED VIDEO" section featuring a video of Tim Berners-Lee. At the bottom, there are sections for "TALKS AND APPEARANCES" and "EVENTS", with specific events listed for 2010, such as "Unleashing Opportunities through Accessibility" and "WWW 2010".



Anwendungen von XML

W3C XML wurde 10!



**„There is essentially no computer in the world,
desktop, handheld, or backroom, that
doesn't process XML sometimes... „**

Tim Bray

[Reply to this greeting](#)

- Michael Kay
- Eve Meler

XML ist fast überall (versteckt)!

- **HTML**
 - XHTML basiert auf XML
- **Internet Explorer/Firefox**
 - können XML verarbeiten
- **Microsofts .net-Architektur**
 - basiert auf XML-Standards: SOAP und WSDL
- **Microsoft Office (ab MSOffice2003)**
 - stellt XML-Schnittstelle bereit: WordML und SpreadsheetML

1. anwendungsspezifische Standards
2. Trennung von Inhalt und Präsentation
3. Web-Dienste (Web Services)

1. Anwendungsspezifische Standards

- XML hat uneingeschränkten Wortschatz:
<xyz>David</xyz>, <aβγ>Hunter</aβγ>
- ⇒ XML **kein** Standard für Publishing oder E-Business
- für spezielle Anwendungen kann jedoch spezifischer Wortschatz und Grammatik festgelegt werden

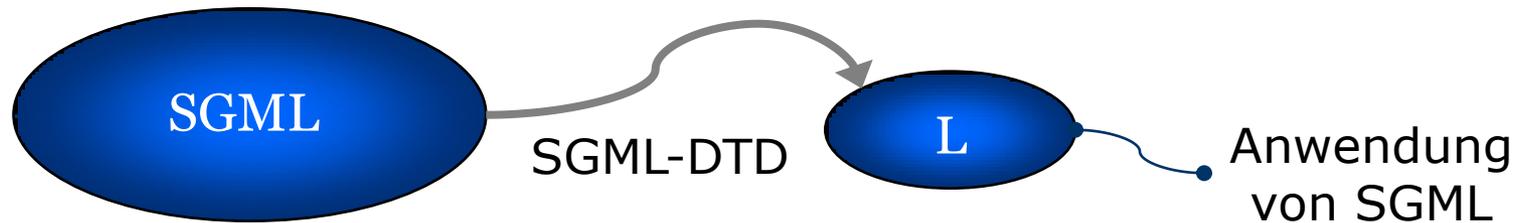
```
<book>
  <title> STRING </title>
  <authors>
    <author> STRING </author>+
  </authors>
  <date> DATE </date>
  <ISBN> STRING </ISBN> ?
  <publisher> STRING </publisher>
</book>
```

- sog. XML-Sprachen (oder Anwendungen von XML)
- mit DTDs und XML-Schemata

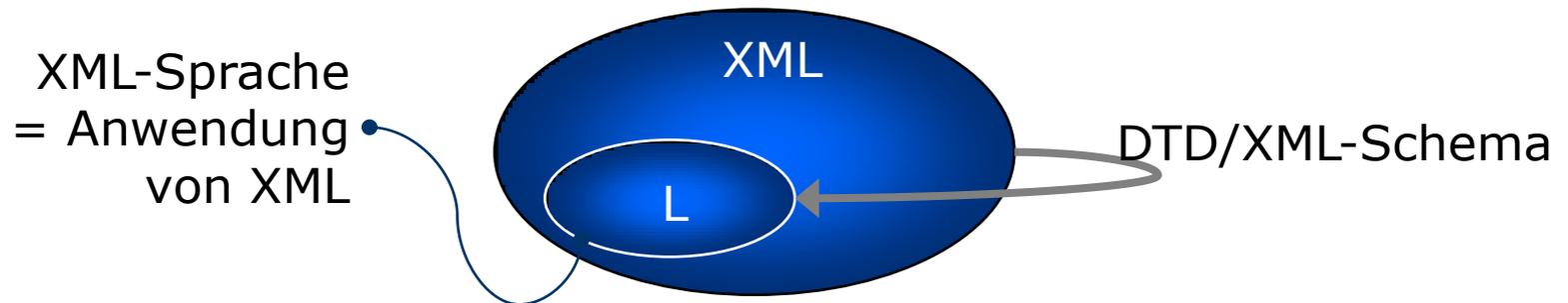
Beispiel SVG

```
<?xml version="1.0"?>  
<svg xmlns="http://www.w3.org/2000/svg"  
  version="1.1"  
  xml:lang="en">  
  <title>Hello World</title>  
  <text x="50.5" y="80.25" font-size="20">Hello World</text>  
</svg>
```





- L muss *nicht* Teilsprache von SGML sein.
- L kann *keine* neue Sprache definieren.
- Beispiel: HTML

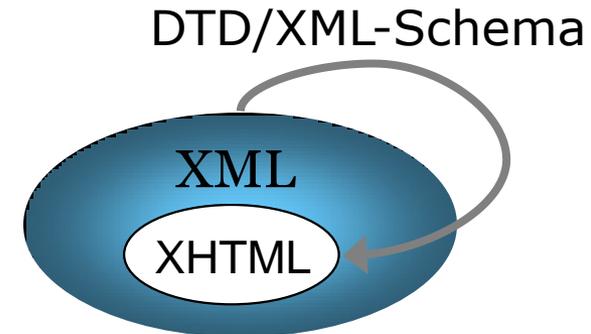
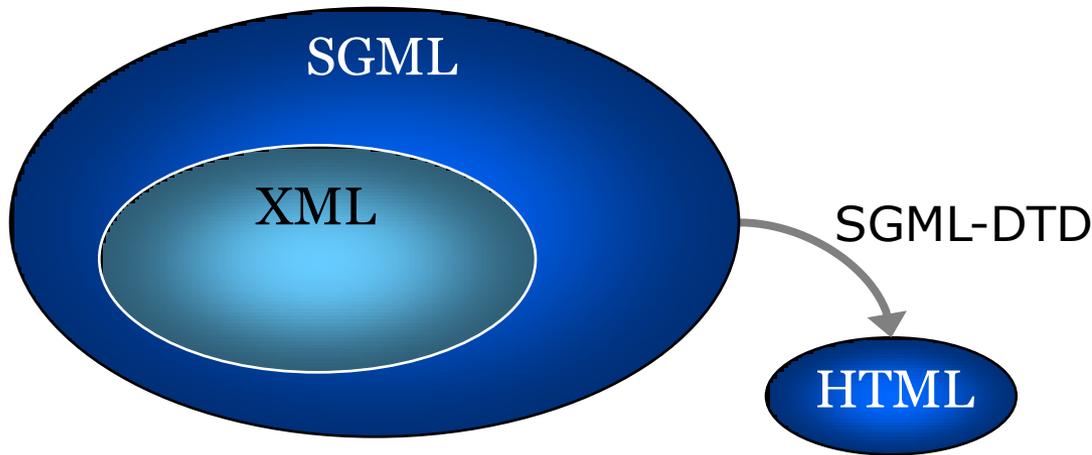


- L immer Teilsprache von XML
- L kann *keine* neue Sprache definieren.
- Beispiel: XHTML

- XHTML: Reformulierung von HTML 4 in XML
- WML: Präsentation von Inhalten für mobile Endgeräte
- SVG: Vektorgraphiken
- DocBook: strukturierte Darstellung von Bücher/Artikel
- XML/EDI: E-Business-Standard EDI als XML
- XSLT: Transformation & Layouten von XML-Dokumenten
- XMLSchema: Definition von XML-Standards

- unterschiedliche Anwendungen, einheitliche Syntax
- unglaubliche Erfolgsgeschichte!

SGML, HTML, XML, XHTML?!



HTML

- Anwendung von SGML

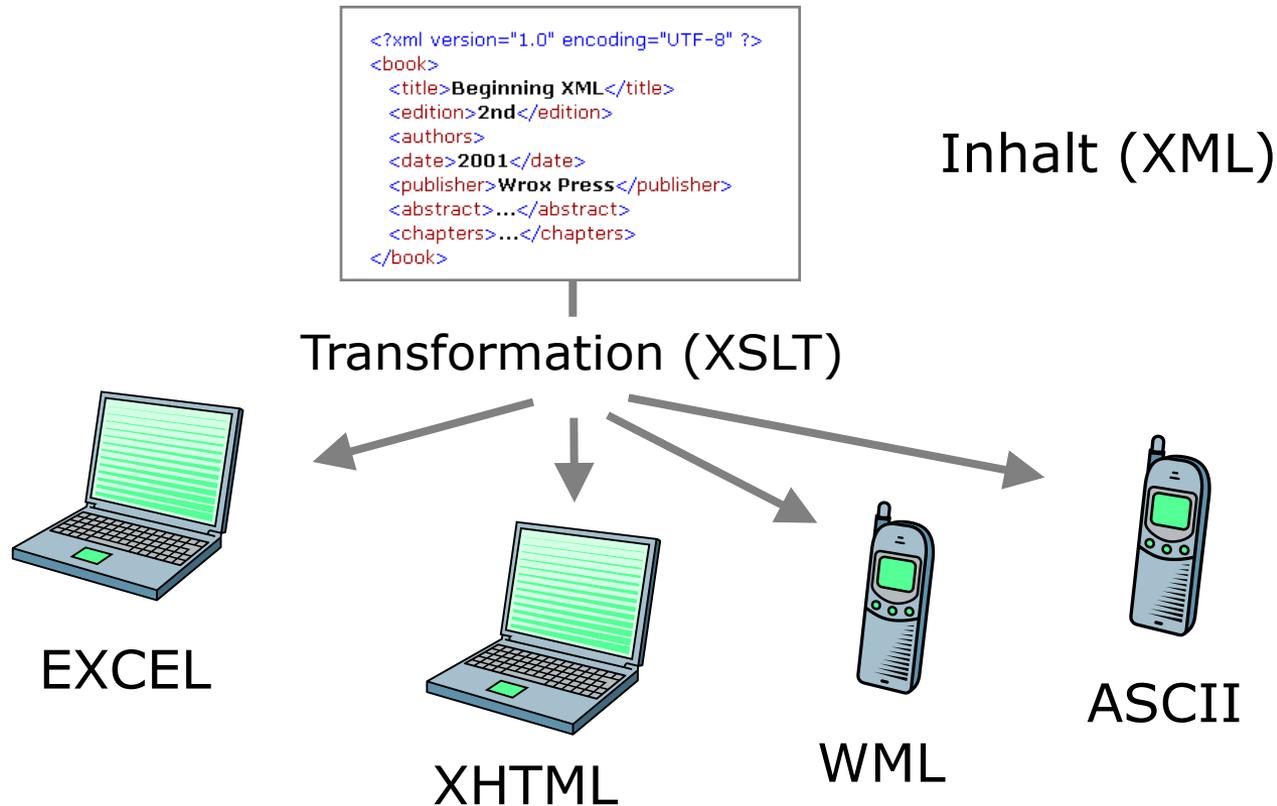
XML

- Teilsprache von SGML

XHTML

- XML-Sprache = Anwendung von XML
- alle XHTML-Dokumente immer wohlgeformte XML-Dokumente

2. Trennung Inhalt von Präsentation



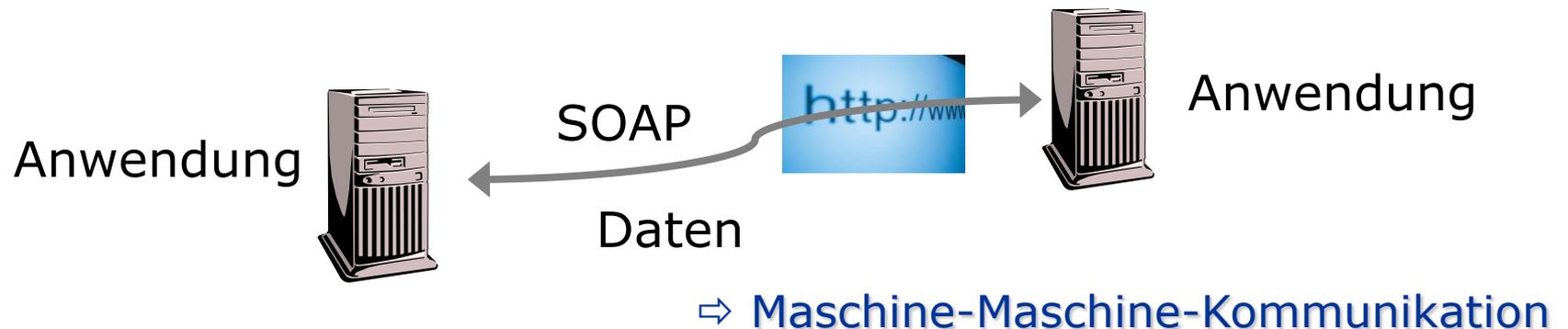
- **Multi-Delivery:** Trennung Inhalt von Präsentation
- weit verbreitet, aber nicht sichtbar!

3. Web-Dienste (Web Services)

traditionelle Web-Anwendung



Web Service



- implementieren keine neuen Systeme
- Fassade für bestehende Systeme, um diese einfach zuzugreifen
- unabhängig von Programmiersprachen und Betriebssystemen
- keine revolutionär neue Technologie
- nutzen gängige Internet-Protokolle wie HTTP(S) und SMTP
- erstmals jedoch alle bedeutenden IT-Unternehmen auf gemeinsame Standards geeinigt: SOAP/WSDL

Beispiel:



ohne Browser

Freie Universität



Berlin

Google Code

SOAP Search API
Getting Help

Google SOAP Search API (No Longer Available)

Google Code Home > Google SOAP Search API > Google SOAP Search API Reference

Google SOAP

API Reference

your computer can do the searching for you.

Google code blog

A well earned retirement for the SOAP Search API

Monday, August 31, 2009

There's a time for everything in life: a time for playing, learning & growing up; a time for maturing, working & performing, and a time for retiring, relaxing & handing the reins over to the next generation. This is true for products too, and this is why, six months ago, we [announced](#) our Labs program for Google Code. This program provides clear distinction between graduate developer products where you'll find mature products with transparent deprecation policies which you can count on for the long run, and labs developer products where you can explore our newest products and get started with them early.

- Suchanfrage
- Suchergebnis
- Anwendung: in periodischer Weise für ein bestimmtes Thema nach neuen Veröffentlichungen suchen lassen

Beispiel:



ohne Browser

Freie Universität



Berlin



Google AJAX APIs

<http://code.google.com/intl/de-DE/apis/ajax/>

- Javascript-Snippets und API
- Asynchrone Javascript-Aufrufe
- XML oder JSON Rückgabe
- ...



- REST API
- Aufruf REST URI
- XML als Antwort
 - oder JSON

Chapter 2. Web Search

The BOSS examples in this section contain reserved characters that need to be escaped. However, for legibility the examples do not display the escape values. Therefore, if you use these examples verbatim to test queries, they will not work. For a list of all reserved characters and their escape values, see [Reserved Characters and Escape Values](#).

Service URL Syntax

```
view plain print ?
01. http://boss.yahooapis.com/ysearch/web/v1/{query}?appid={yourBOSSappid}
    {&param1=val1&param2=val2&etc}
```

SIMPLE QUERY EXAMPLE

```
view plain print ?
01. http://boss.yahooapis.com/ysearch/web/v1/foo?appid={yourBOSSappid}&
    format=xml
```

SIMPLE XML RESPONSE

```
view plain print ?
01. <ysearchresponse responsecode="200">
02.   <nextpage>/ysearch/web/v1/foo?appid={yourBOSSappid}&amp;
03.   format=xml&amp;start=10]]&gt;&lt;/nextpage&gt;
04.   &lt;resultset_web count="10" start="0" totalhits="29440998" deephits="8810"&gt;
05.     &lt;result&gt;
06.       &lt;abstract&gt;<![CDATA[World &lt;b&gt;soccer&lt;/b&gt; coverage
07.       from ESPN, including Premiership, Serie A, La Liga, and Major
08.       &lt;b&gt;Soccer&lt;/b&gt;. Get news headlines, live scores, stats, and
09.       tournament information.]]&gt;&lt;/abstract&gt;
10.       &lt;date&gt;2008/06/08&lt;/date&gt;
11.       &lt;dispurl&gt;<![CDATA[www.&lt;b&gt;soccernet.com&lt;/b&gt;]]&gt;&lt;/dispurl&gt;
12.       &lt;clickurl&gt;http://us.lrd.yahoo.com
13.       /_ylc=X3odMTFkNXVldGJyBGFwcGlka2Jvc3NkZW1vBHBvcwMwBHN1cnZpY2UDWVNIYXJjaARzc
14.       /SIG=10u3e8260/**http%3A//www.soccernet.com/&lt;/clickurl&gt;
15.       &lt;size&gt;94650&lt;/size&gt;
16.       &lt;title&gt;ESPN Soccernet&lt;/title&gt;
17.       &lt;url&gt;http://www.soccernet.com/&lt;/url&gt;
18.     &lt;/result&gt;
19.   &lt;/resultset_web&gt;
20. &lt;/ysearchresponse&gt;</pre></div><div data-bbox="788 243 965 518" data-label="Table-Of-Contents"><ul><li>BOSS API Guide<ul><li>Getting Started</li><li>Web Search<ul><li>Service URL Syntax</li><li>Simple Query Example</li><li>Simple XML Response</li><li>Optional Arguments, in addition to Universal BOSS API Arguments for Web, Images and News</li><li>Response Fields</li><li>Recommended Region and Language Usage</li></ul></li><li>BOSS Site Explorer</li><li>Image Search</li><li>News Search</li><li>Spelling Suggestions</li><li>Appendix and FAQs</li></ul></li></ul></div><div data-bbox="171 850 818 893" data-label="Text"><p><a href="http://developer.yahoo.com/search/boss/">http://developer.yahoo.com/search/boss/</a></p></div><div data-bbox="29 938 628 970" data-label="Page-Footer"><p>AG Netzbasierete Informationssysteme <a href="http://www.ag-nbi.de">http://www.ag-nbi.de</a></p></div><div data-bbox="915 938 957 970" data-label="Page-Footer"><p>44</p></div>
```

- **Syntax** – die Art und Weise, wie Worte in einem Satz zusammengesetzt wurden.
- **Semantik** – Informationen, die in diesem Sinne kodiert wurden.
- **Pragmatik** – Implikationen aus den Informationen in einem Kontext.

“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, and Lassila, 2001.



Foto: W3C



Foto: Homepage



Foto: Homepage

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

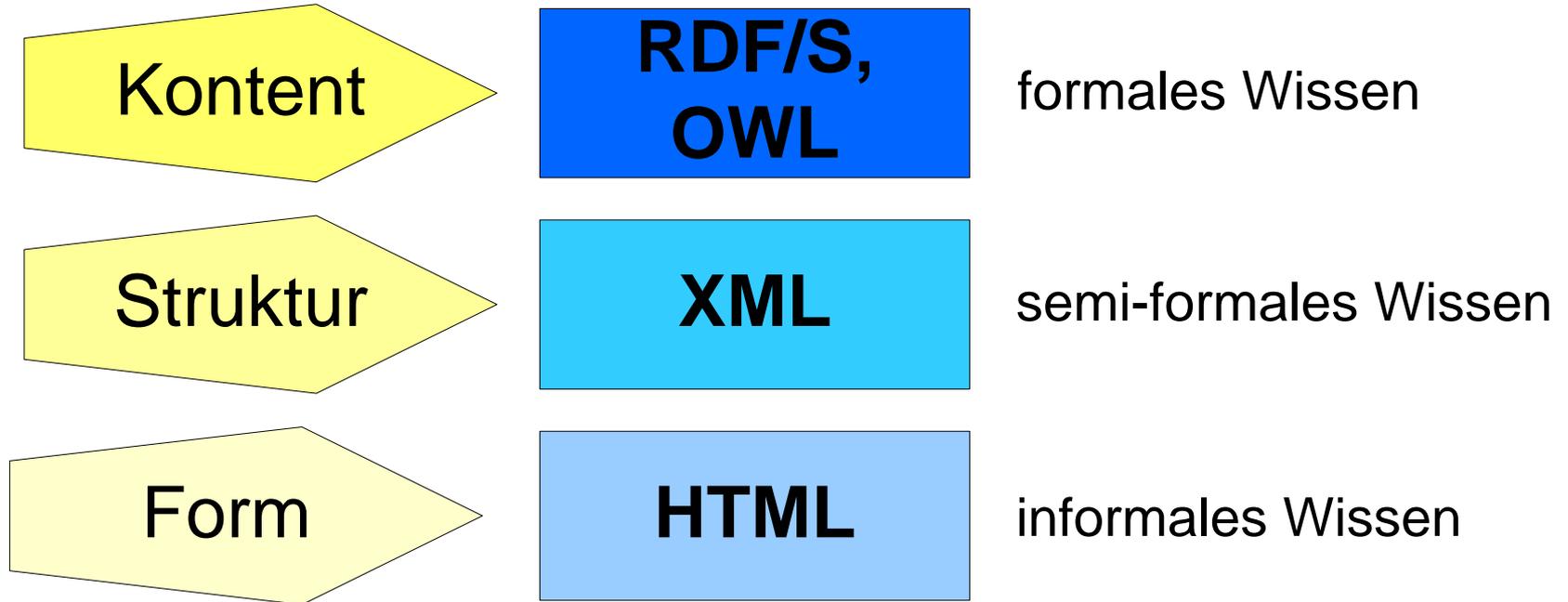
Webinhalte und ihre Vernetzung werden für Maschinen verständlich.

Auch komplexe Anfragen können ans Web gestellt werden.

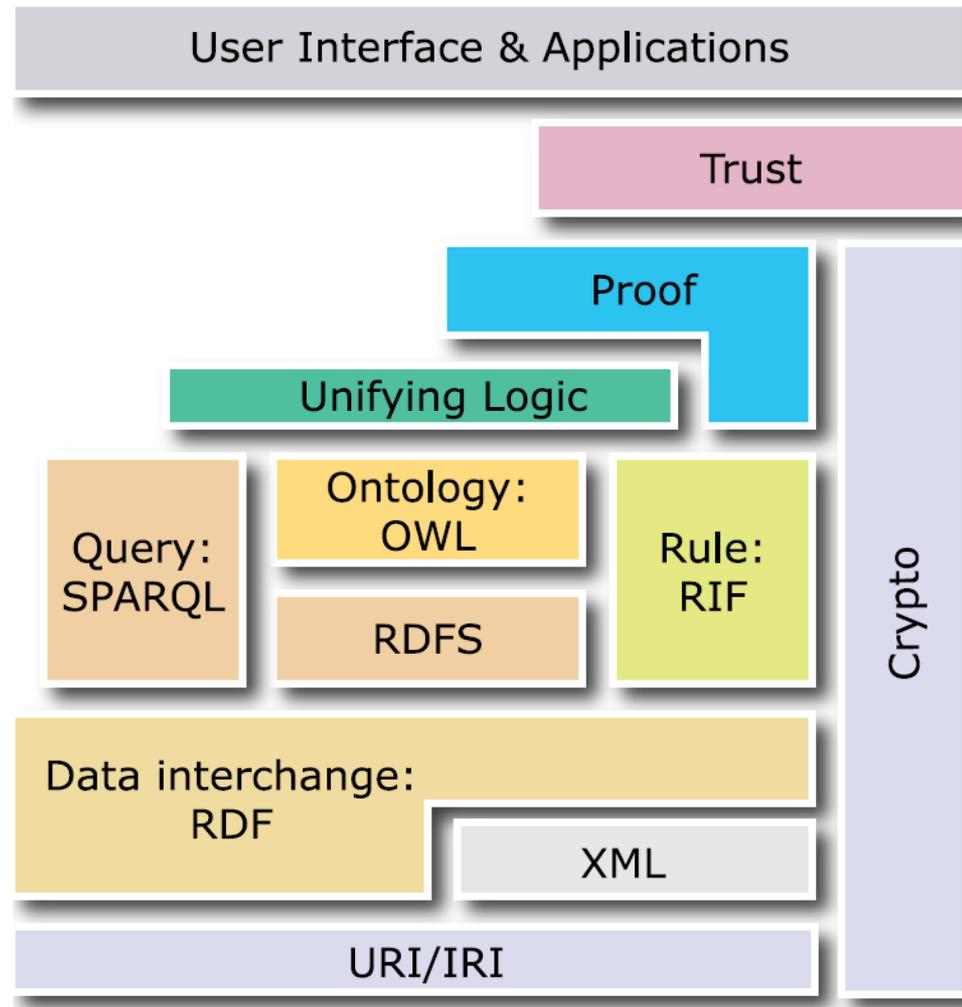


Beispiel: Mit welchen Kollegen arbeitet der Autor eines bestimmten Dokumentes zur Zeit zusammen?

3 Levels von Markup im Web



Semantic Web Stack (W3C, 2000)



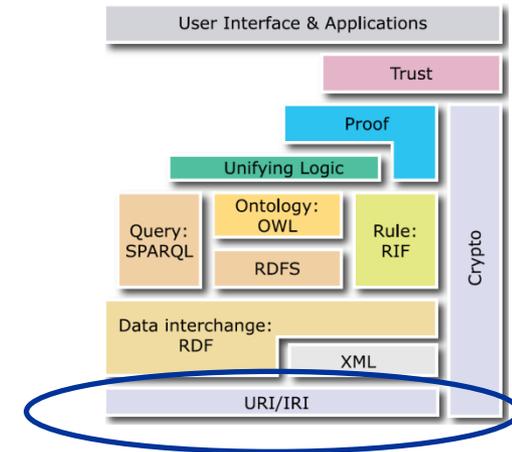
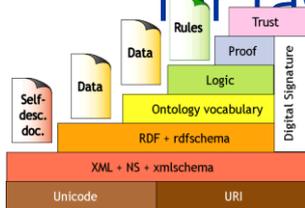
auch „Semantic Web Layer Cake“ genannt

Quelle: <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>

Unicode

jedes Zeichen eigene Nummer (system-, programm- und sprachunabhängig)

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

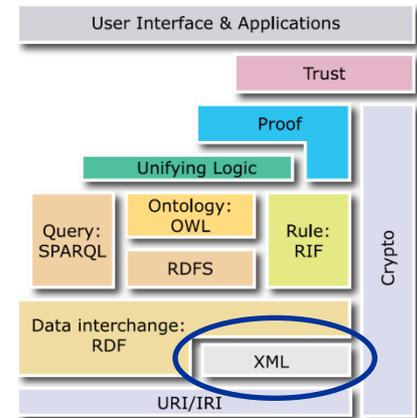
XML + Namensräume + XML-Schema

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

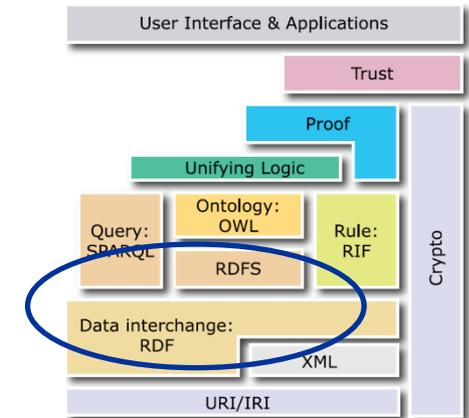
RDF + Namensräume + RDF-Schema

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:

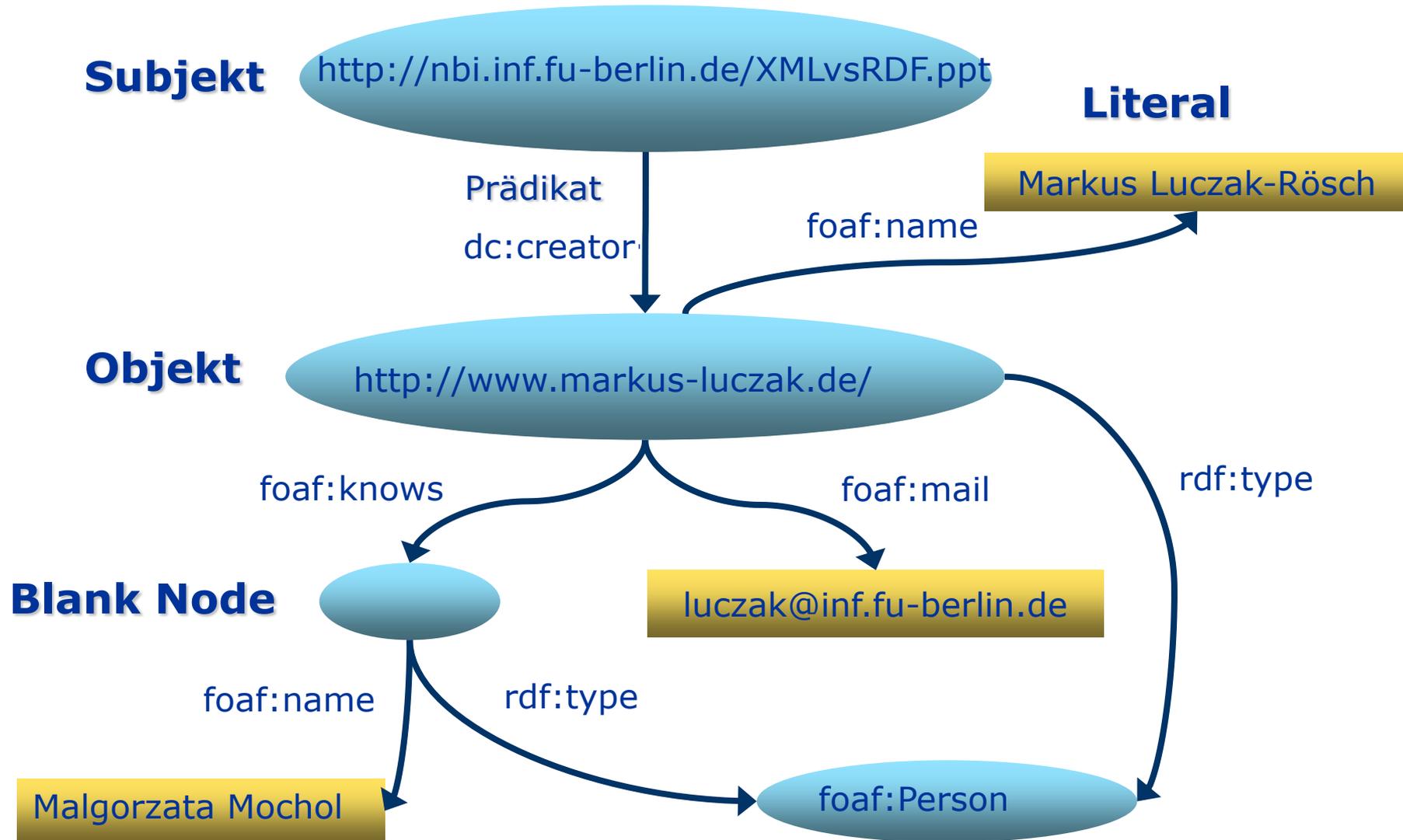
- Tripel: kompakt, lesbar
- RDF/XML: für maschinelle Verarbeitung

Tripel setzen bel. Web-Ressourcen URI-s und URI-o miteinander in Beziehung:

<URI-s, URI-p, URI-o>

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 **Markus Luczak-Rösch
 - Subject (Ressource): ***This presentation***
 - Predicate (Property): ***creator***
 - Object (Wert): ***Markus Luczak-Rösch***
- RDF benutzt URIs :
 - Subject: ***<http://nbi.inf.fu-berlin.de/SemWeb.ppt>***
 - Predicate: ***<http://purl.org/dc/elements/1.1/creator>***
 - Object: ***<http://www.markus-luczak.de/>***



Ontologien

Vokabulare

Begriffsbeziehungen (Unterklasse, Untereigenschaft, Wertebereiche, ..., selbstdefinierte)

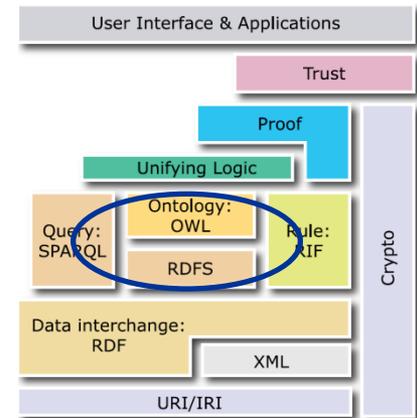
Sprache für Web-Ontologien:

OWL – Web Ontology Language

Erweiterte Beschreibungsmöglichkeiten

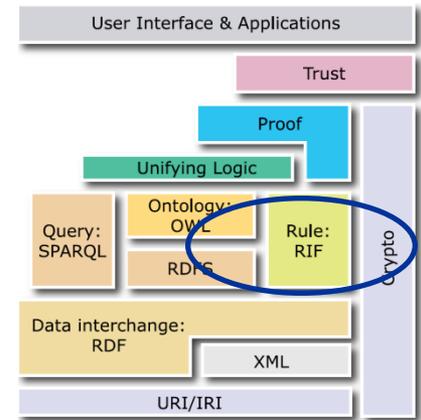
In unterschiedlichen Komplexitäten
(OWL-Lite, OWL-DL, OWL-Full)

mittlerweile OWL 2 mit feinerer Unterscheidung der Komplexität



Regelsprachen

bilden die Grundlage für das logische
schließen auf Basis semantischer Daten
früher SWRL (echte Regelsprache für OWL)
als Teil des Layer Cakes
heute RIF als ein Austauschformat
zwischen unterschiedlichen Regelsystemen

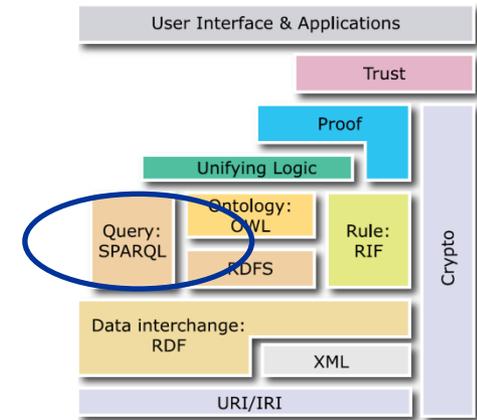


Anfragesprache SPARQL

Dient zur Abfrage von Instanzdaten in einer RDF-Datenbank

„Gib mir alle Menschen, die vor 1900 in Berlin geboren wurden“

```
SELECT ?name ?birth ?death ?person
WHERE {
  ?person dbpedia2:birthPlace <http://dbpedia.org/resource/Berlin> .
  ?person dbo:birthDate ?birth .
  ?person foaf:name ?name .
  ?person dbo:deathDate ?death
  FILTER (?birth < "1900-01-01"^^xsd:date) .
}
ORDER BY ?name
```





Überblick über die Vorlesung

- XML-Basistechnologien
 - 6 Termine
- XML und Markupssprachen in der Anwendung
 - 3 Termine
- Einführung Projektmanagement
 - 1 Termin
- Rückblick & Ausblick
 - 1 Termin
- Klausur
 - 1 Termin

- XML-Basistechnologien - 6 Termine

- XML-Syntax, einschl. Namensräume
- DTDs und XML-Schemata
- XML-Parser
- XSLT, XPath, etc.



Fundamentals

- nicht behandelt werden:

- XML-Technologien zur Präsentation von Dokumenten wie XHTML oder WML
- anwendungsspezifische XML-Standards wie SVG oder VoiceXML

Image: <http://www.morguefile.com/archive/display/211651>

- XML und Markupssprachen in der Anwendung

- Web Services - 1 Termin
 - dienstorientierte Architekturen
 - SOAP & WSDL
 - REST



Applications

- Semantic Web Grundlagen und RDF - 1 Termin
- Moderne Markuptechnologien und Microformate – 1 Termin
 - HTML 5
 - Microformats
 - RDFa

Image: <http://www.morguefile.com/archive/display/24026>

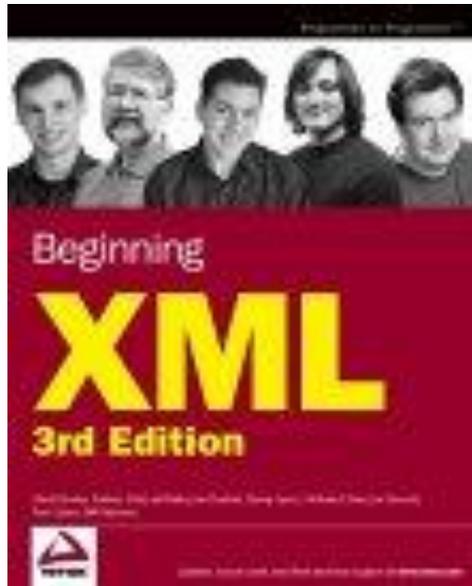
- **Projektmanagement: 1 Termin**
 - Einführung in die Praxis des Projektmanagements
- **Rückblick: 1 Termin**
 - kleine Wiederholung
 - Schwerpunkt → Klausurfragen



Image: <http://www.morguefile.com/archive/display/564796>



Literatur



Hunter et al., Beginning XML
(3rd Edition), Wrox Press, 2004.

ca. 41 €

Im Semesterapparat!

- XML

- 1.0, W3C Recommendation, Sept. 2006, <http://www.w3.org/TR/xml/>
- 1.1, W3C Recommendation, Sept. 2006, <http://www.w3.org/TR/2006/REC-xml11-20060816/>

- XML-Schema

- XML Schema Part 0: Primer Second Edition, W3C, 2004

- XSLT

- XSL Transformations (XSLT) Version 1.0, W3C, Nov. 1999 XSL Transformations (XSLT) Version 2.0, W3C, Jan. 2007

- **Web Services**

- SOAP Version 1.2 Part 0: Primer (2nd Edition), W3C, April 2007 (<http://www.w3.org/TR/2007/REC-soap12-part0-20070427/>)
- Web Services Description Language (WSDL)
 - WSDL Version 1.1, W3C, 2001
 - WSDL Version 2.0, W3C Recommendation, Juni 2007

- **Semantic Web Grundlagen und RDF**

- W3C RDF Primer: <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>
- W3C Semantic Web Standards: <http://www.w3.org/RDF/>
- Linked Data Design Issues: <http://www.w3.org/DesignIssues/LinkedData.html>

- **Moderne Markuptechnologien und Microformate**

- HTML 5 W3C Working Draft: <http://dev.w3.org/html5/spec/Overview.html>
- W3C RDFa Primer: <http://www.w3.org/TR/xhtml-rdfa-primer/>

Wie geht es weiter?

- ☑ Organisatorisches
 - ☑ Was ist XML?
 - ☑ Anwendungen von XML
 - ☑ Überblick über die Vorlesung
 - ☑ Literatur
-
- XML-Syntax
 - Namensräume
 - Semantik von XML-Tags



Exkurs - Abschlussarbeiten

- Fortlaufend unter <http://blog.ag-nbi.de/>
- Themenschwerpunkte:
 - Netzbasierte Informationssysteme & Web-Technologien
 - Netzwerke
- Außerdem Abschlussarbeiten bei der AG CSW:
<http://www.inf.fu-berlin.de/groups/ag-csw/Teaching/Theses/index.html>