

---

# XML-Technologien

## Heutige Vorlesung

---

1. Organisatorisches
2. Was ist XML?
3. Anwendungen von XML
4. Überblick über die Vorlesung
5. Literatur

---

# Organisatorisches

## Webseite der Lehrveranstaltung

- 
- [http://www.inf.fu-berlin.de/inst/ag-nbi/lehre/05/V\\_XML/](http://www.inf.fu-berlin.de/inst/ag-nbi/lehre/05/V_XML/)
  - **Google**: Schild XML-Technologien SoSe 05
  - hier finden sich
    - Termine
    - Folien der Vorlesungen
    - Folien der Übungen
    - Hinweise/Links auf Literatur

# Anmeldung



- 140 haben sich im Online-KVV bereits angemeldet
- *keine* weiteren Anmeldungen mehr möglich!
- 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

# Übungen



- wegen zu hoher Teilnehmerzahl *keine* Übungsaufgaben
  - ⇒ bedeutet für Sie nicht weniger, sondern *mehr* Arbeit!
  - ⇒ Sie müssen den Vorlesungsstoff eigenständig vertiefen und einüben.
- Übungsaufgaben und Musterlösungen:  
[http://www.inf.fu-berlin.de/inst/ag-nbi/lehre/04/V\\_XML/](http://www.inf.fu-berlin.de/inst/ag-nbi/lehre/04/V_XML/)

- alle *zwei* Wochen eine Übung
- zwei Gruppen:
  1. Mi 16:15-17:45 im SR 005
  2. wird noch bekannt gegeben
- Beginn: nächste Woche
- Sie können frei wählen, an welchem der beiden Termine Sie teilnehmen.

## Was erwartet Sie?

- Fragen an mich zum Vorlesungsstoff
- vertiefende Themen und Beispiele
- Fragen an Sie zum Vorlesungsstoff

# Scheinkriterien

## Scheinkriterium

- Klausur bzw. mündliche Nachprüfung erfolgreich bestanden
- Schein wird grundsätzlich benotet

## Klausur

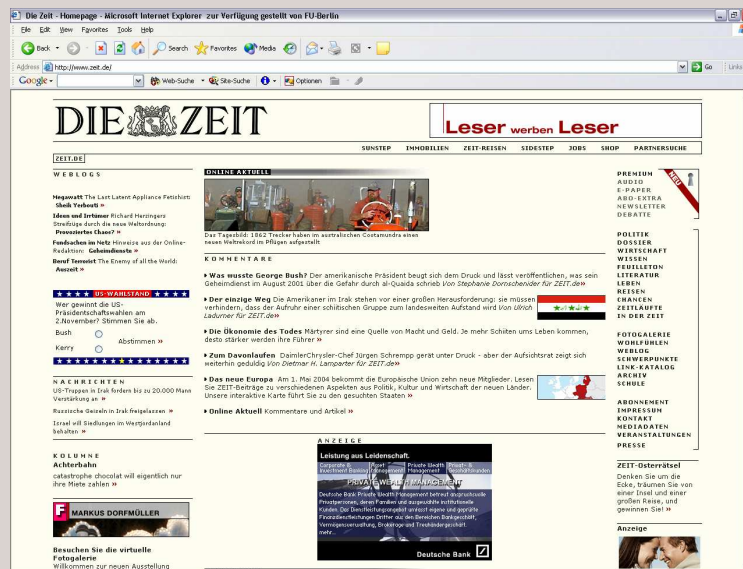
- Teilnahmevoraussetzung: Anmeldung im Online-KVV
- Klausurtermin: 13.7.2005

## mündliche Nachprüfung

- Teilnahmevoraussetzung: Klausur nicht bestanden
- einzige Ausnahme hiervon: ärztliches Attest für den Klausurtermin
- Termin: Anfang WS 2005/6

- Veranstalter: schild AT inf.fu-berlin.de
- Sprechstunde: bitte per E-Mail vereinbaren

# Was ist XML?



HTML hat sich für die Präsentation von Inhalten bewährt.

## Warum reicht HTML nicht aus?

immer häufiger **medienneutrale Darstellung** von Inhalten benötigt:

- 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**

## Auszeichnungssprachen

- **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

## SGML

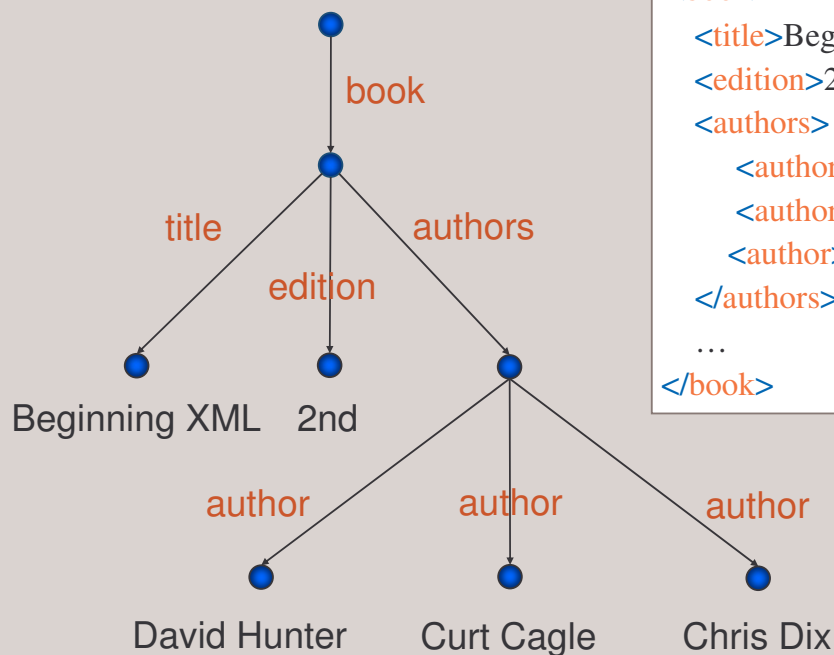
- Standard Generalized Markup Language
- 1969 von Charles Goldfarb und zwei seiner Kollegen bei IBM für das Dokumentenmanagement entwickelt.
- seit 1986 ein internationaler Standard
- *keine* vorgegebenen Tags, auch keine für das Layout von Dokumenten
- Vorgänger von XML



```
<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>
```

- sieht ganz ähnlich aus, wie XML
- erlaubt Strukturieren von Dokumenten
- keine speziellen Tags für das layout
- medienneutral

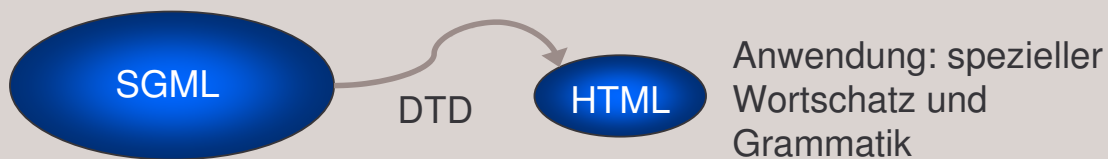
## Lauter Bäume!



```
<book>
  <title>Beginning XML</title>
  <edition>2nd</edition>
  <authors>
    <author>David Hunter</author>
    <author>Curt Cagle</author>
    <author>Chris Dix</author>
  </authors>
  ...
</book>
```

- Beachte: Reihenfolge der Kanten relevant

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



- Anwendung selbst kann keine Anwendung definieren

## 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

# Warum XML und nicht HTML oder SGML?



## 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.

# XML im Vergleich zu SGML



## Gemeinsamkeiten

- generische Auszeichnungssprache
- medienneutral
- Anwendungen können definiert werden.

## Unterschiede

- wesentlich einfacher
- wird von Web-Browsern dargestellt

```
<?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>
```

# Eine kurze Geschichte von XML



- 1969 Charles Goldfarb entwickelt bei IBM die Generalized Markup Language (**GML**).
- 1980 ANSI veröffentlicht ersten Entwurf von **SGML**.
- 1986 ISO verabschiedet **SGML**.
- 1989 Berners-Lee schlägt SGML-basiertes Hypertext-System vor.
- 1990 Berners-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, 2<sup>nd</sup> Edition** (einschl. Namensräume)  
**XHTML 1.0** (Reformulierung von HTML in XML)
- 2004 **XML 1.1**

# Was bedeutet Erweiterbarkeit?



- X in XML steht für erweiterbar (*extensible*).
- Was bedeutet Erweiterbarkeit?
- Vergleich mit HTML 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
- Nur *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

# Der Rest der XML-Familie



## Extensible Stylesheet Language (XSLT)

- Transformation von XML-Dokumenten in beliebige Text-Formate:

XML → HTML / WML / SMS / XML / ...

## Document Object Model (DOM)

- Parsen, Modifizieren und Erstellen von XML-Dokumenten

## XPath

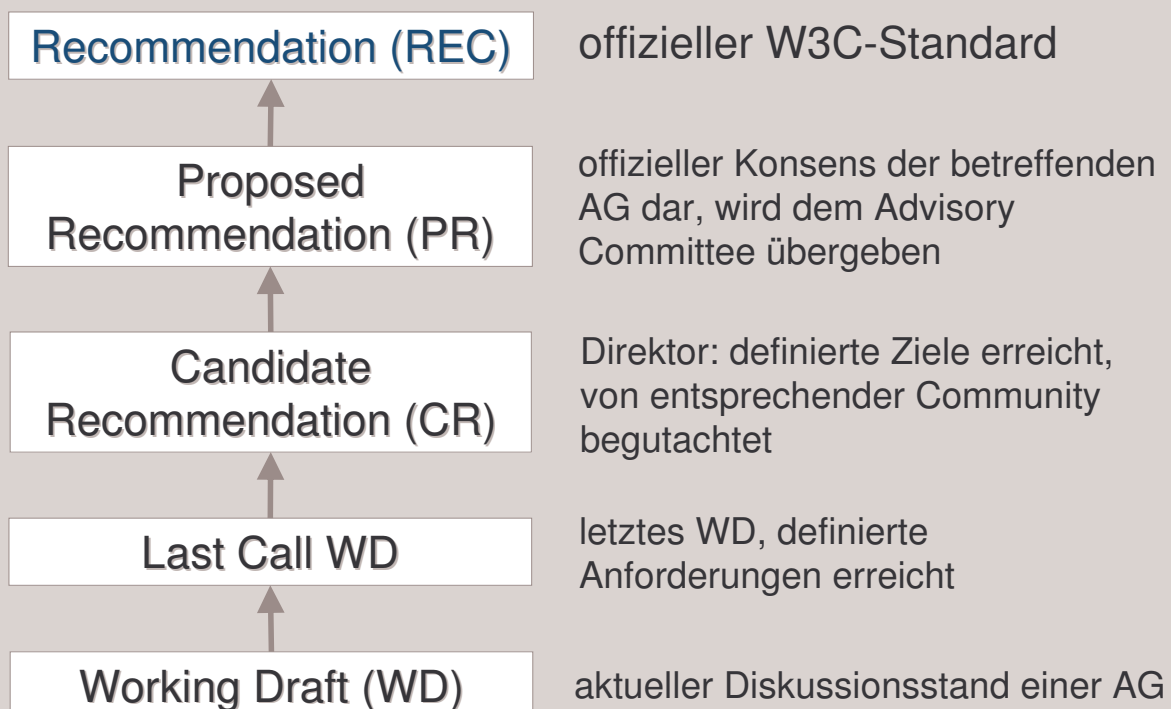
- Zugriff auf beliebige Teile eines XML-Dokumentes
- z.B. Zugriff auf alle Buchtitel

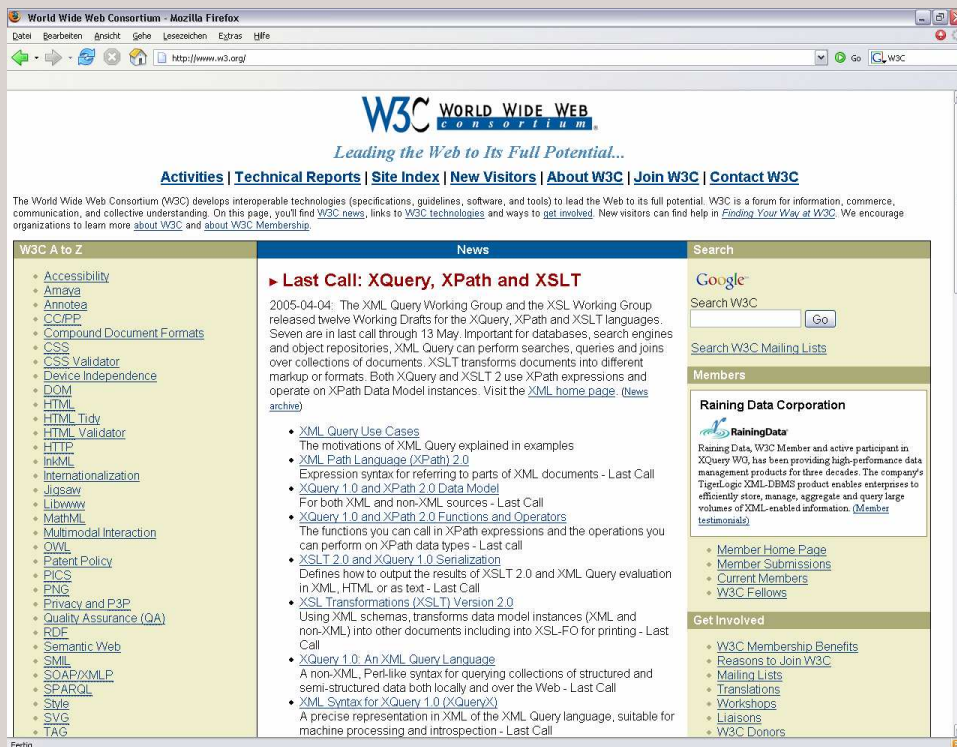
gesamte XML-Familie besteht aus  
lizenzfreien W3C-Standards



- 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

## Standardisierungsprozess des W3C





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 header features the W3C logo and the tagline "Leading the Web to Its Full Potential...". Below the header, there are navigation links: [Activities](#), [Technical Reports](#), [Site Index](#), [New Visitors](#), [About W3C](#), [Join W3C](#), and [Contact W3C](#). A paragraph of text describes the W3C's mission: "The World Wide Web Consortium (W3C) develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential. W3C is a forum for information, commerce, communication, and collective understanding. On this page, you'll find [W3C news](#), links to [W3C technologies](#) and ways to [get involved](#). New visitors can find help in [Finding Your Way at W3C](#). We encourage organizations to learn more about [W3C](#) and [about W3C Membership](#)."

The main content area is divided into three columns:

- W3C A to Z:** A list of links to various W3C technologies and standards, including Accessibility, Amaya, Annotate, CC/PP, Compound Document Formats, CSS, CSS Validator, Device Independence, DOM, HTML, HTML Tidy, HTML Validator, HTTP, InkML, Internationalization, Jigsaw, Libwww, MathML, Multimodal Interaction, OWL, Patent Policy, PICS, PNG, Privacy and P3P, Quality Assurance (QA), RDF, Semantic Web, SMIL, SOAP/XMLP, SPARQL, Style, SVG, and TAG.
- News:** A section titled "Last Call: XQuery, XPath and XSLT" dated 2005-04-04. It reports that the XML Query Working Group and the XSL Working Group released twelve Working Drafts for XQuery, XPath, and XSLT. It lists several key documents and their last call dates:
  - [XML Query Use Cases](#): The motivations of XML Query explained in examples.
  - [XML Path Language \(XPath\) 2.0](#): Expression syntax for referring to parts of XML documents - Last Call.
  - [XQuery 1.0 and XPath 2.0 Data Model](#): For both XML and non-XML sources - Last Call.
  - [XQuery 1.0 and XPath 2.0 Functions and Operators](#): The functions you can call in XPath expressions and the operations you can perform on XPath data types - Last call.
  - [XSLT 2.0 and XQuery 1.0 Serialization](#): Defines how to output the results of XSLT 2.0 and XML Query evaluation in XML, HTML, or as text - Last Call.
  - [XSL Transformations \(XSLT\) Version 2.0](#): Using XML schemas, transforms data model instances (XML and non-XML) into other documents including into XSL-FO for printing - Last Call.
  - [XQuery 1.0: An XML Query Language](#): A non-XML, Perl-like syntax for querying collections of structured and semi-structured data both locally and over the Web - Last Call.
  - [XML Syntax for XQuery 1.0 \(XQueryX\)](#): A precise representation in XML of the XML Query language, suitable for machine processing and introspection - Last Call.
- Search:** A search box with the Google logo and a "Go" button. Below it are links for "Search W3C Mailing Lists", "Members", and "Get Involved".

# Anwendungen von XML

# XML ist fast überall (versteckt)!

---



## HTML

- neueste Version basiert auf XML: XHTML

## Internet Explorer/Firefox

- können XML verarbeiten

## Microsofts .net-Architektur

- basiert auf XML-Standards: SOAP und WSDL

## Office 2003

- stellt XML-Schnittstelle bereit: WordML und SpreadsheetML

# Anwendungen von XML

---



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

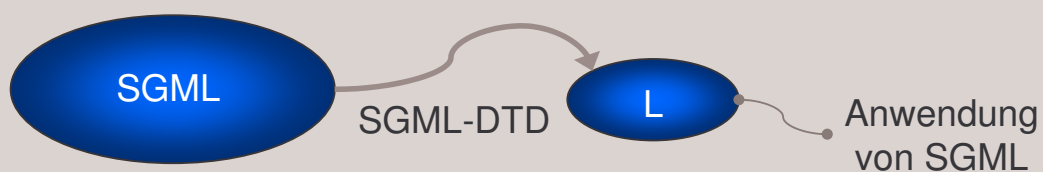


- XML hat uneingeschränkten Wortschatz:  
`<xyz>David</xyz>`, `<αβγ>Hunter</αβγ>`
- ⇒ XML *kein* Standard für Publishing oder E-Business
- für spezielle Anwendungen kann jedoch spezifischer Wortschatz und Grammatik festgelegt werden:

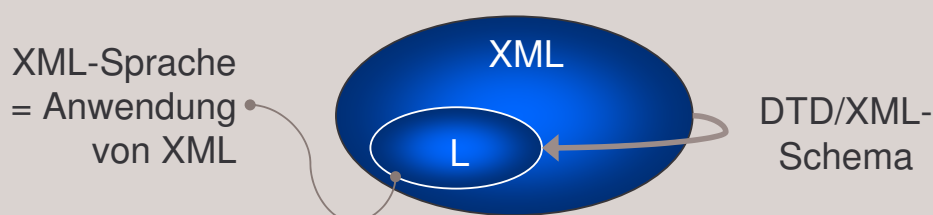
```
<book>
  <title> STRING </title>
  <authors>
    <author> STRING </author>+
  </authors>
  <date> DATE </date>
  <ISBN> STRING </ISBN>
  <publisher> STRING </publisher>
</Book>
```

- werden **XML-Sprachen** (oder **Anwendungen von XML**) genannt
- mit **DTDs** und **XML-Schemata** möglich

## SGML-Anwendungen vs. XML-Sprachen



- 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

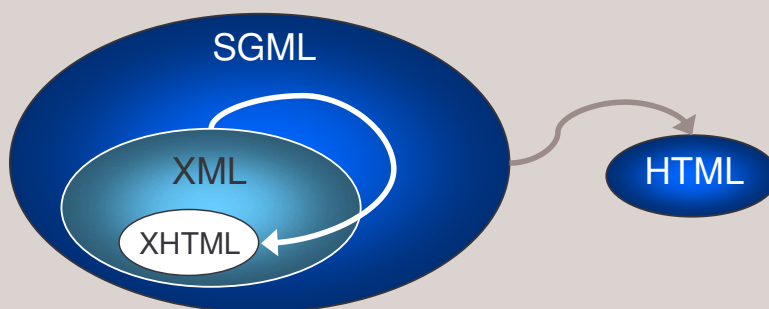
# XML-Sprachen: Beispiele



- **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

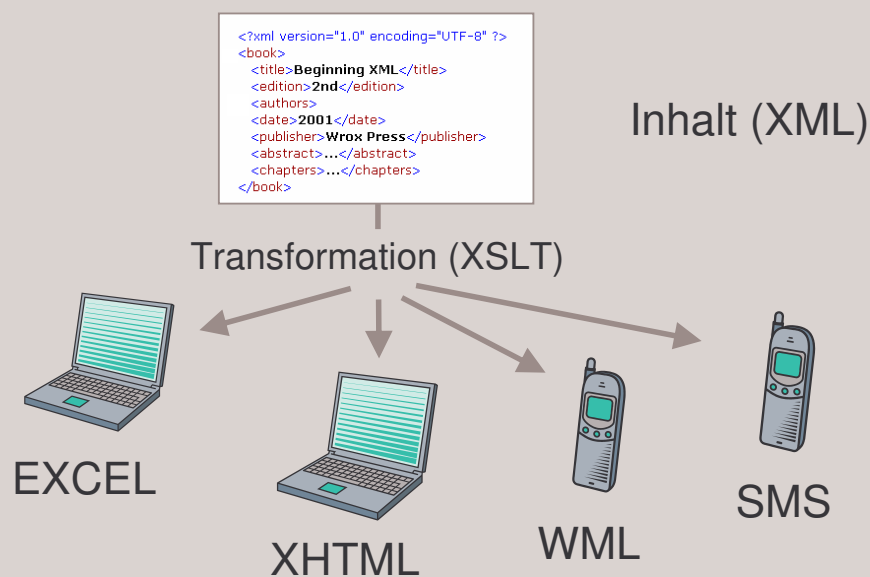
- XML-Sprache
- = Anwendung von XML
- alle XHTML-Dokumente immer wohlgeformte XML-Dokumente

# Anwendungen von XML



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

# Trennung Inhalt von Präsentation



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

# Anwendungen von XML



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

# Web-Dienste (Web Services)



## traditionelle Web-Anwendung



⇒ Mensch-Maschine-Kommunikation

## Web Service



⇒ Maschine-Maschine-Kommunikation



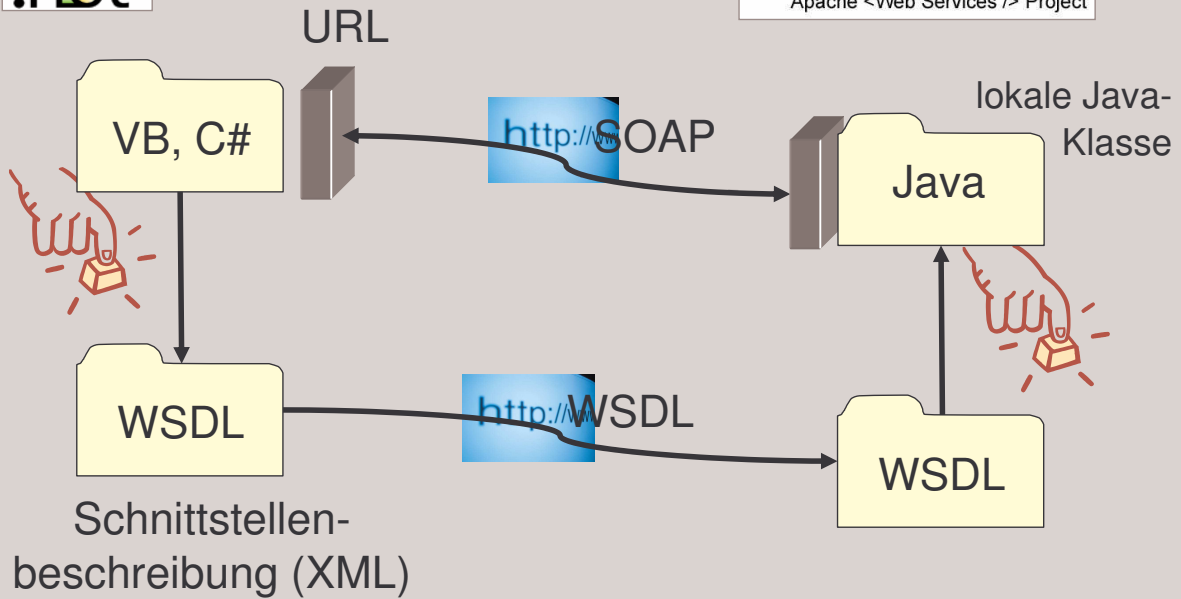
With Google Web APIs, your computer can do the searching for you.

- Suchanfrage: SOAP-Nachricht (XML)
- Suchergebnis: SOAP-Nachricht (XML)
- Anwendung: in periodischen Abständen zu bestimmten Thema nach neuen Webseiten suchen lassen

## Eigenschaften von Web Services

- 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

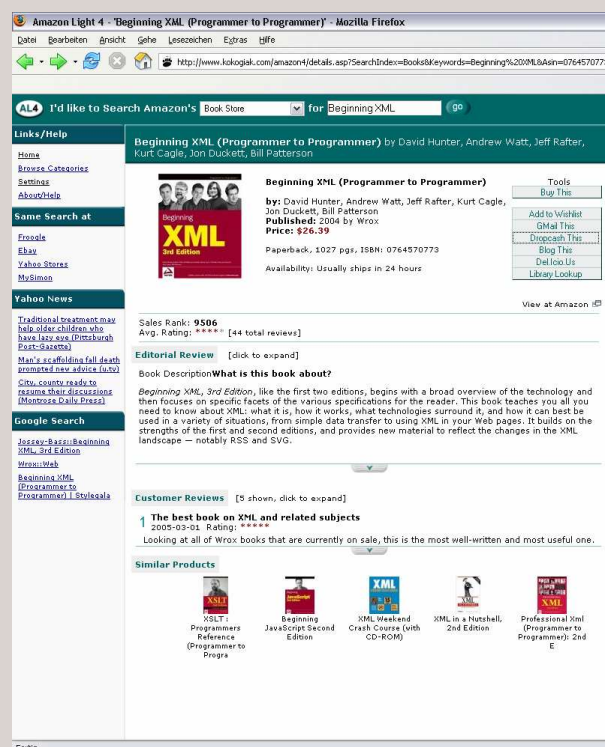
# Beispiel: Systemintegration



# Beispiel: Amazon für alle!



- Amazon bietet Web-Service-Schnittstelle an
- ⇒ einfacher Zugriff aus Anwendungen heraus
- ⇒ jeder kann neue Oberfläche für Amazon anbieten und mit anderen Diensten kombinieren: z.B. Amazon Light 4



---

# Überblick über die Vorlesung

## Vorlesungsinhalt

---

1. **XML-Basistechnologien**  
6 Termine
2. **Web-Dienste (Web Services)**  
4 Termine
3. **Ausblick**  
1 Termin

## 1. **XML-Basistechnologien:** 6 Termine

- XML-Syntax, einschl. Namensräume
- DTDs und XML-Schemata
- XML-Parser
- XSLT
- Datenmodellierung mit XML

### **nicht behandelt werden:**

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

## 2. **Web-Dienste (Web Services):** 4 Termine

- dienstorientierte Architekturen
- SOAP
- WSDL
- Web Services in der Praxis

## 3. **Ausblick:** 1 Termin

- Semantic Web



---

# Literatur

## Literaturempfehlungen



- Hunter et al., Beginning XML (3<sup>rd</sup> Edition), Wrox Press, 2004. € 34,90
- Vorlesungsskript von Mario Jeckle (160 Seiten):  
<http://www.jeckle.de/vorlesung/xml/>

## XML-Schema

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

## XSLT

- XSL Transformations (XSLT) Version 1.0, W3C, 1999

## XML-Parser

- The Java Web Services Tutorial, 2004, Kap. 1

## Web Services

- SOAP Version 1.2 Part 0: Primer, W3C, 2003
- Web Services Description Language (WSDL) 1.1, W3C, 2001
- Kaye, Loosely Coupled: The Missing Pieces of Web Services, Rds Associates Inc, 2003.



# Wie geht es nächste Woche weiter?

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