

XML-Technologien

Heutige Vorlesung

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

Organisatorisches

- Vorlesung: Malgorzata Mochol, mochol@inf.fu-berlin.de
- Übung: Lyndon Nixon, nixon@inf.fu-berlin.de
- Mitarbeiter der AG Netzbasierte Informationssysteme von Prof. Tolksdorf
- Büros: Fabeckstr. 15, 1.OG (NICHT Takustr. 9)
- Sprechstunde: bitte per Mail vereinbaren

- http://www.ag-nbi.de/lehre/07/V_XML/

- hier finden sich
 - Folien der Vorlesungen
 - Termine der Übungen
 - Übungsblätter
 - Musterlösungen
 - Hinweise/Links auf Literatur

- 90 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

Großübung

- Veranstalter: Lyndon Nixon
- Termine:
 - Di 12:15 – 13:45, Hörsaal 028, Beginn: 8.5.
 - zweiter Termin wird bekannt gegeben
- Genaue Termine & Themen auf der Webseite
- *Übung auf English*

Dienstag, 24.04. Vorlesung statt Übung

Wa

- Keine Vorlesung am Mittwochen, 25.04 !!!
- vertiefende Themen und Beispiele
- Vorstellung einer Lösung des aktuellen Übungsblattes

- 7 Übungsblätter
- wegen hoher Teilnehmerzahl werden diese *nicht* korrigiert
- Musterlösung werden in Übung vorgestellt
- in der Übung wird davon ausgegangen, dass Sie aktuelles Übungsblatt bearbeitet haben
- Lösung (Ausdruck) bitte in Übung mitbringen

Scheinkriterien

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

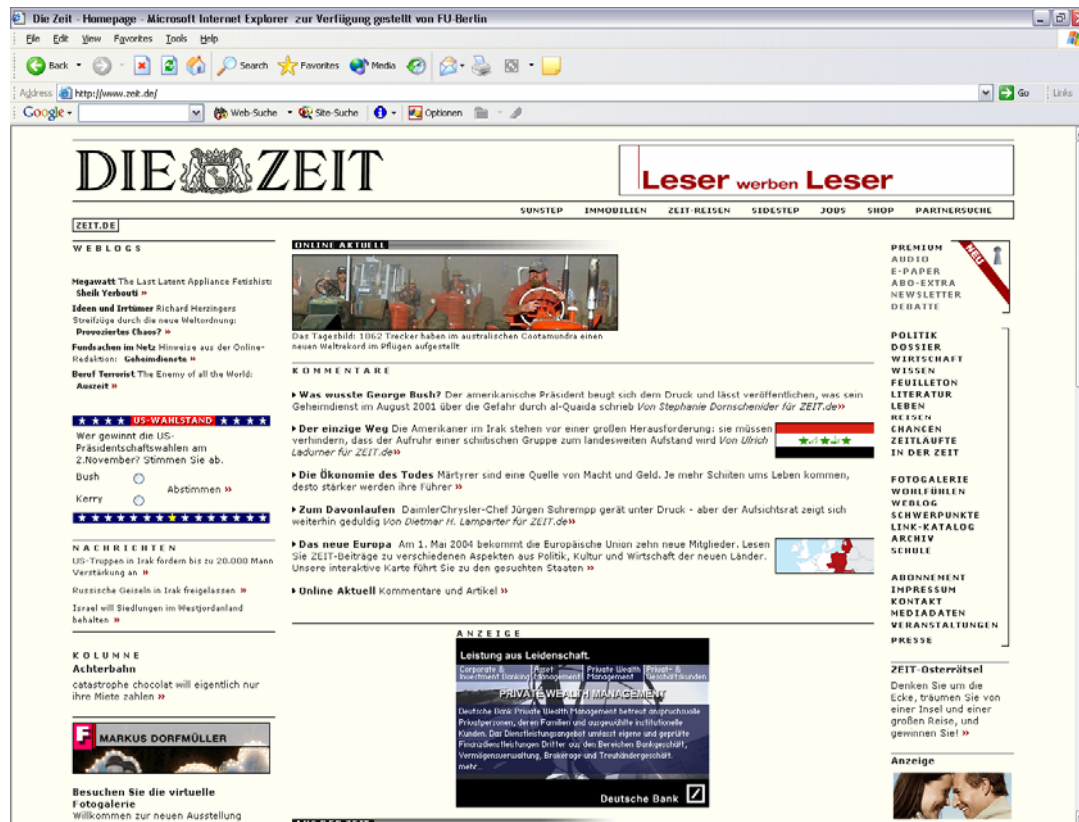
Klausur

- Teilnahmevoraussetzung: Anmeldung im Online-KVV
- Klausurtermin: 18.7.2007

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.

Warum reicht HTML nicht aus?

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

Eine kurze Geschichte von 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 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, 2nd Edition**

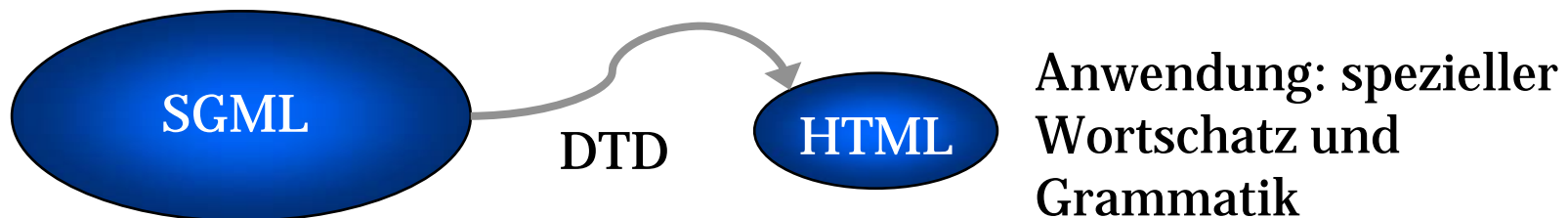
XHTML 1.0 (Reformulierung von HTML in XML)

2004 **XML 1.1**

XML 1.0, 3rd 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

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/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.

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.

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 / XHTML / ASCII / ...

XPath

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

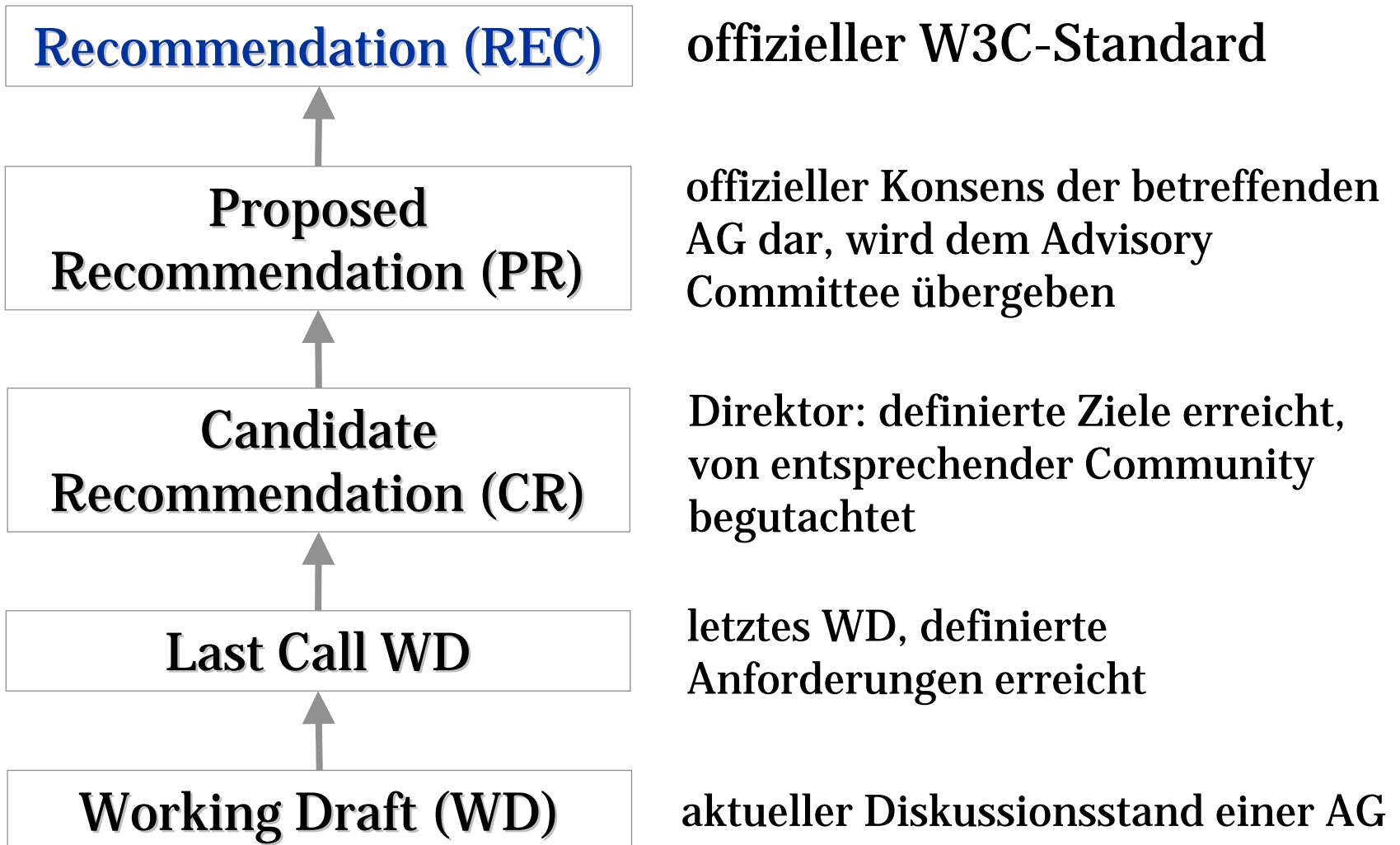
Document Object Model (DOM)

- 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






Leading the Web to Its Full Potential...

[Activities](#) | [Technical Reports](#) | [Site Index](#) | [New Visitors](#) | [About W3C](#) | [Join W3C](#) | [Contact W3C](#)

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.

W3C Supporters Help W3C by making a donation through the W3C Supporters Program .	News	Search
Mobile Web Initiative  The goal of the Mobile Web Initiative is to make Web access from a mobile device as simple, easy, and convenient as Web access from a desktop device. Become an MWI sponsor .	W3C Launches New HTML Working Group	Google™ Search W3C <input type="text"/> <input type="button" value="Go"/>
Employment Current job opportunities at W3C are HTML and WebAPI specialist , SVG and Compound Documents specialist , Voice Browser Specialist , and Web Accessibility Engineer . Current W3C Fellows Program openings are Business and Technology Communications Specialist , Web / Graphic Designer , and Software Engineer .	2007-03-07: W3C is pleased to invite participation in the new HTML Working Group , chartered to create the next HTML standard with the active participation of browser vendors, software developers, and content designers. "It's time to revisit the standard and see what we can do to meet the current community needs, and to do so effectively with commitments from browser manufacturers in a visible and open way," said Tim Berners-Lee, W3C Director. At the same time, W3C is chartering the Forms Working Group, the XHTML2 Working Group, and rechartering the Hypertext Coordination Group. Read the press release and visit the HTML Working Group home page , the Forms Working Group home page , and the XHTML2 Working Group home page . (Permalink)	Search W3C Mailing Lists
W3C A to Z <ul style="list-style-type: none">AccessibilityAmayaAnnotegCC/PPCompound Document FormatsCSSCSS ValidatorDatabindingDevice IndependenceDOMEfficient XML InterchangeGRDDLHealth Care and Life SciencesHTML	► CURIE: Working Draft for Compact URIs	Testimonials
	2007-03-07: The HTML Working Group and the Semantic Web Best Practices and Deployment Working Group jointly have published the First Public Working Draft of CURIE Syntax 1.0 . Written for markup language designers, the draft specifies the syntax and usage of CURIEs which are abbreviated or "compact URIs." Read about the HTML Activity and the Semantic Web . (Permalink)	IONA Technologies, Inc.
	► Working Draft: CSS3 Text	
	2007-03-06: The CSS Working Group has released a Working Draft of CSS Text Level 3 . Formerly titled the <i>CSS3 Text Effects Module</i> , the draft is part of the Cascading Style Sheets (CSS) language Level 3 and addresses white space, line breaks, word boundaries, text wrapping, alignment, justification and spacing. Visit the CSS home page . (Permalink)	IONA® Technologies delivers high-performance integration solutions for enterprise IT environments and joined the W3C to show its commitment to standards-based technology. (Member testimonials)
	► Incubator Group to Analyze Semantic Web	Members
		<ul style="list-style-type: none">Member Home PageMember SubmissionsCurrent MembersMeetingsFellows (New Openings)
		Get Involved
		<ul style="list-style-type: none">W3C Membership BenefitsReasons to Join W3CMailing ListsNew: Proposed Charters in ReviewTranslations

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

Microsofts Office 2003

- stellt XML-Schnittstelle bereit: WordML und SpreadsheetML

Anwendungen von XML

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

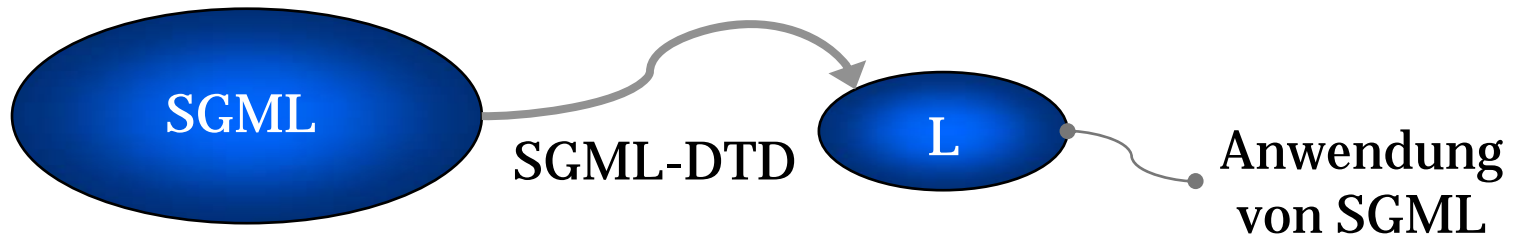
1. Anwendungsspezifische Standards

- 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:

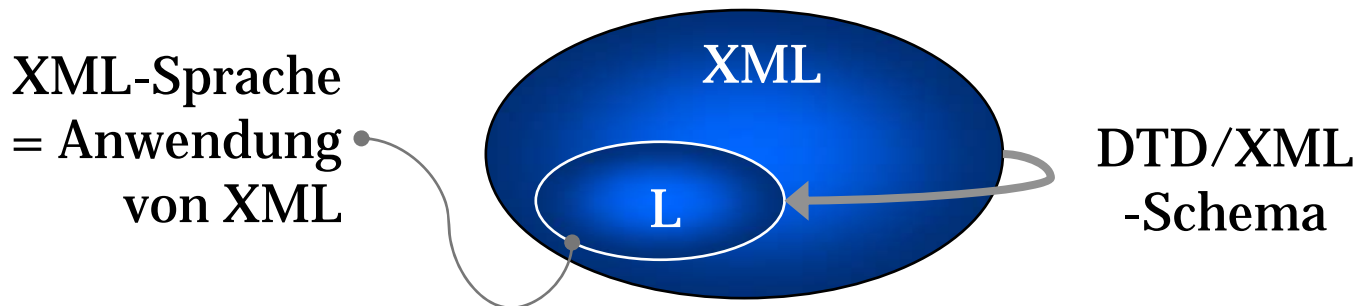
```
<book>
  <title> STRING </title>
  <authors>
    <author> STRING </author>+
  </authors>
  <date> DATE </date>
  <ISBN> STRING </ISBN> ?
  <publisher> STRING </publisher>
</book>
```

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

SGML- vs. XML-Anwendungen



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

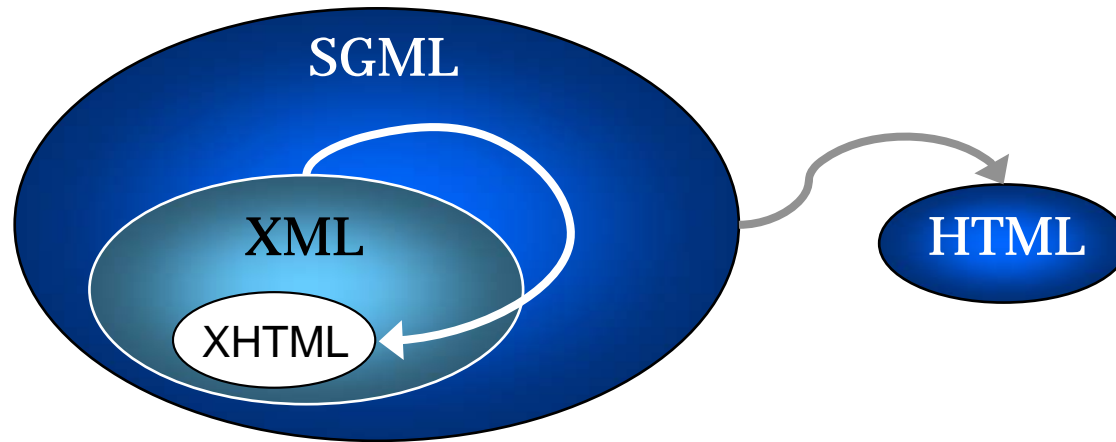


- L *immer* Teilsprache von XML
- L kann *keine* neue Sprache definieren.
- Beispiel: XHTML

- **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

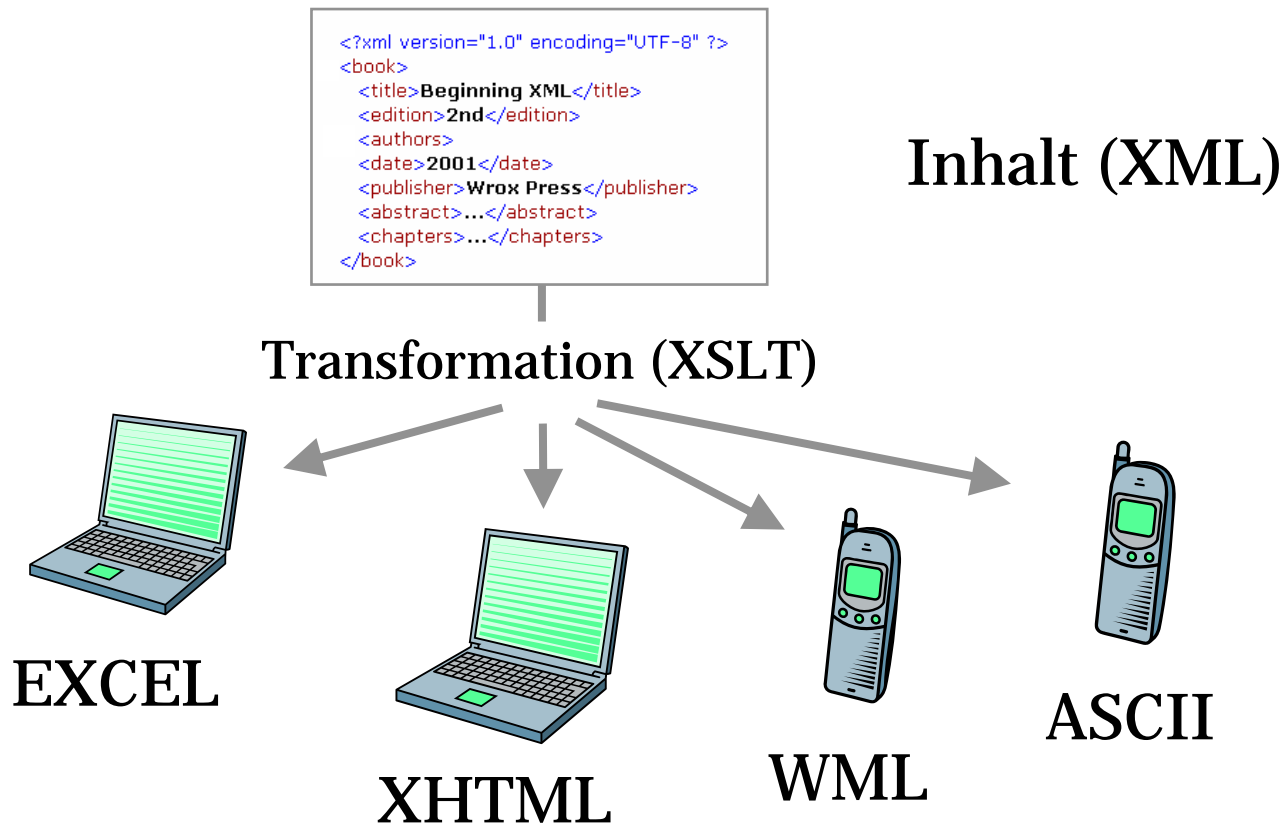
XML

- Teilsprache von SGML

XHTML

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

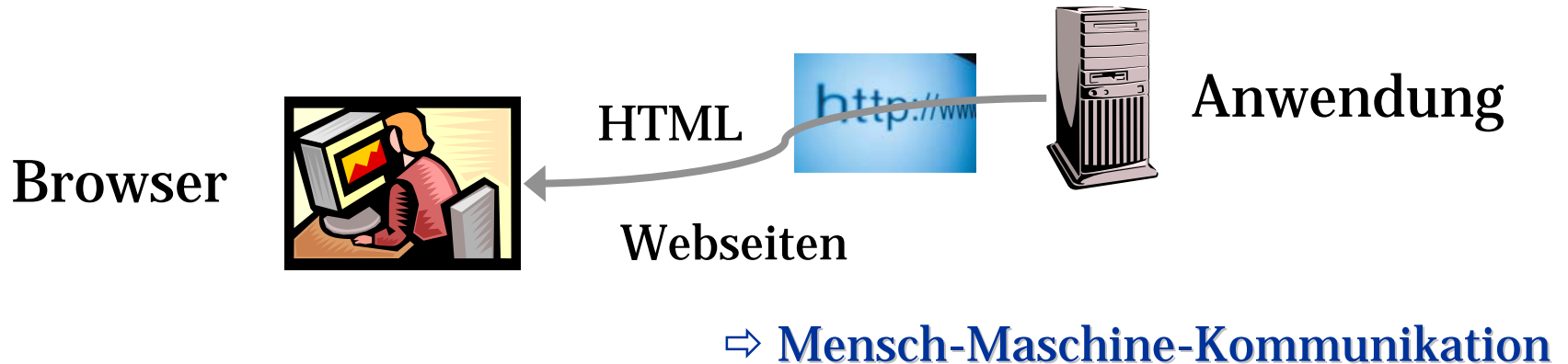
2. Trennung Inhalt von Präsentation



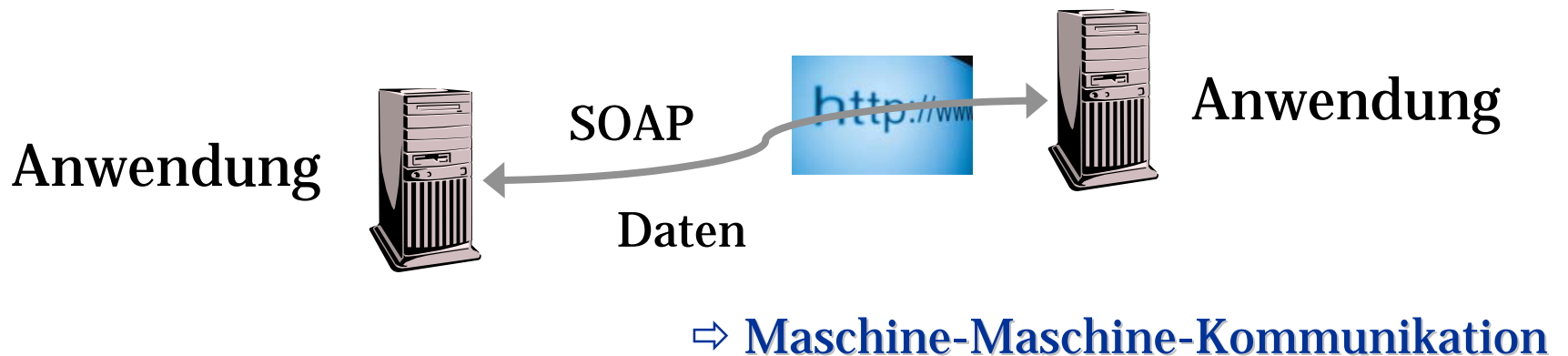
- **Multi-Delivery:** Trennung Inhalt von Präsentation
- weit verbreitet, aber nicht sichtbar!

3. Web-Dienste (Web Services)

traditionelle Web-Anwendung



Web Service





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

<http://code.google.com/apis>

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

- 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

Überblick über die Vorlesung

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

3. **Rückblick:** 1 Termin

- kleine Wiederholung

4. **Klausur:** 1 Termin

Literatur



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

ca. 41 €

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

Web Services

- SOAP Version 1.2 Part 0: Primer, W3C, 2003
- Web Services Description Language (WSDL) 1.1, W3C, 2001

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

Wann geht es nächste Woche weiter?

Nächste Vorlesung am 24.04 (Dienstag) !!!

KEINE Vorlesung am 25.04 !!!