



# Information und Struktur

## Einführung in XML

Fachhochschule Darmstadt

2.12.1999

Dr. Burkhardt Renz



# Gliederung

## – was sind unsere Ziele?

- ▶ Was ist die Idee von XML?  
Ein Beispiel, die Idee, Vergleiche
- ▶ Wie ist ein XML-Dokument aufgebaut?  
Elemente, Attribute, Entities  
"Ordentliche" und "gültige" Dokumente
- ▶ Anwendungen von XML  
Medienneutrale Dokumentendrehscheibe  
Dokumentenaustausch  
Web-Suche  
Komponententechnologie
- ▶ Ausblick  
Stichworte mit X, Literatur, Links



# Information – Form I

## – Unformatierter Text: ASCII

- ▶ Hennig Behme Stefan Mintert XML in der Praxis  
Addison-Wesley München 1998 Frank Boumphrey Jon  
Duckett Joe Graf Paul Houle Trevor Jenkins  
Pete  
r Jones Adrian Kingsley-Hughes Kathie  
Kingsley-Hughes Craig McQueen Stephen Mohr XML  
Applicati  
ons Wrox Press 1998 Charles F. Goldfarb Paul  
Pr  
escod XML Handbuch Prentice-Hall New York 1999



## Information – Form II

### – Formatierter Text: HTML

- ▶ `<p><b>Hennig Behme, Stefan Mintert</b> XML in der Praxis <i>Addison-Wesley München 1998</i></p>`  
`<p><b>Frank Boumphrey, Jon Duckett, Joe Graf, Paul Houle, Trevor Jenkins, Peter Jones, Adrian Kingsley-Hughes, Kathie Kingsley-Hughes, Craig McQueen, Stephen Mohr</b> XML Applications <i>Wrox Press o.O. 1998</i></p>`  
`<p><b>Charles F. Goldfarb, Paul Prescod</b> XML Handbuch <i>Prentice-Hall New York 1999</i></p>`



# Information – Form III

## – Information mit Semantik: XML

```
▶ <?xml version="1.0" encoding="ISO-8859-1"?>
  <Biblio>
    <Buch xml:lang="de">
      <Autor><Name>Behme</Name>
        <Vornamen>Henning</Vornamen>
      </Autor>
      <Autor><Name>Mintert</Name>
        <Vornamen>Stefan</Vornamen>
      </Autor>
      <Titel>XML in der Praxis</Titel>
      <Jahr>1998</Jahr>
      <Ort>München</Ort>
      <Verlag>Addison-Wesley</Verlag>
    </Buch>
    <Buch xml:lang="en">
      <Autor>
        ...
    </Buch>
  </Biblio>
```

1 Idee

2 XML-Dokument

3 Anwendungen



## Fazit des Beispiels

### – Informationsgehalt für Verarbeitung

- ▶ **Unformatierter Text: ASCII**
  - strukturlose Information
  - Bedeutung extern
  - spezielle Software notwendig
  
- ▶ **Formatierter Text: HTML**
  - Struktur für Präsentation
  - Bedeutung extern
  - Browser können HTML anzeigen
  - für Verarbeitung spezielle Software notwendig
  
- ▶ **Information mit Semantik: XML**
  - Struktur für Bedeutung, nicht Präsentation
  - Parser können XML verarbeiten und
  - Information standardisiert für spezielle Programme bereitstellen



# Was ist XML?

## – Idee und Prinzipien

- ▶ **Strukturbeschreibungssprache**
  - logische, semantische Auszeichnung
  - hierarchisch strukturierte Daten
  - beliebige Komplexität
  
- ▶ **Erweiterbarkeit**
  - anwendungsspezifische Sprachen
  
- ▶ **Verständlichkeit**
  - leicht verständlich für Menschen
  - und für verarbeitende Programme (Parser)
  
- ▶ **Austauschbarkeit**
  - plattformunabhängig
  - frei zugängliches Format
  - Standard



## Was XML nicht ist – Vergleiche

- ▶ XML ist keine Programmiersprache  
wie etwa C++ oder Java
- ▶ XML ist keine Skriptsprache  
wie etwa Perl oder JavaScript
- ▶ XML ist kein Satzsystem  
wie etwa T<sub>E</sub>X oder 3B2
- ▶ XML ist keine Seitenbeschreibungssprache  
wie etwa Postscript oder HTML





# Die Geschichte von XML

## – SGML, HTML, XML

### ▶ SGML

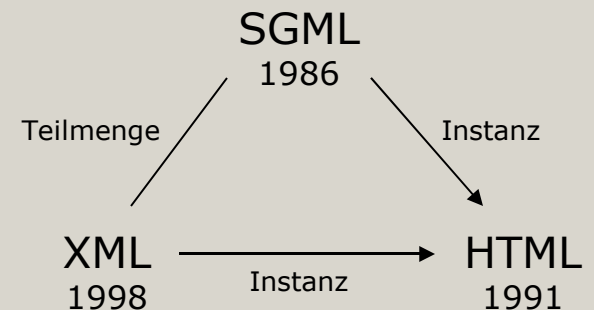
Standard Generalized Markup Language  
Goldfarb, Mosher, Lorie  
1986 ISO-Standard

### ▶ HTML

Instanz von SGML  
Berners-Lee  
1991 W3C Standard

### ▶ XML

Teilmenge von SGML  
Bosak, Bray, Paoli, Clark et al.  
1996 Design Prinzipien für XML  
1998 XML Version 1.0 W3C Standard





# XML und HTML

## – Probleme mit HTML und ihre Lösung

### ▶ HTML

- fixe Menge von Tags
- Präsentationstechnologie
- flache Struktur
- große Übertragungsmenge im Internet

### ▶ XML

- frei definierbare Tags
- Präsentation getrennt: XML + Stylesheet = HTML++
- Baumstruktur
- Information kann gefiltert werden



# XML-Dokument

## – der Aufbau im Überblick

▶ Prolog

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

Document Type Definition (optional)

```
<!DOCTYPE Biblio SYSTEM "biblio.dtd">
```

▶ Element

```
<Biblio>  
  <Buch xml:lang="de">  
    <Autor><Name>Behme</Name>  
      <Vornamen>Henning</Vornamen>  
    </Autor>  
    ...  
  </Buch>  
</Biblio>
```



## XML-Bausteine – Elemente, Attribute, Entities

### ▶ Elemente

```
<Titel>XML Applications</Titel>
```

### ▶ Attribute

```
<Buch xml:lang = "en">
```

### ▶ Entities

```
&lt; entspricht <  
vordefinierte Character Entity
```

```
&Abb1; entspricht "abb1.jpg"  
<!ENTITY Abb1 SYSTEM "abb1.jpg" NDATA JPEG>
```

### ▶ Kommentare

```
<!-- hier steht ein Kommentar -->
```

1 Idee

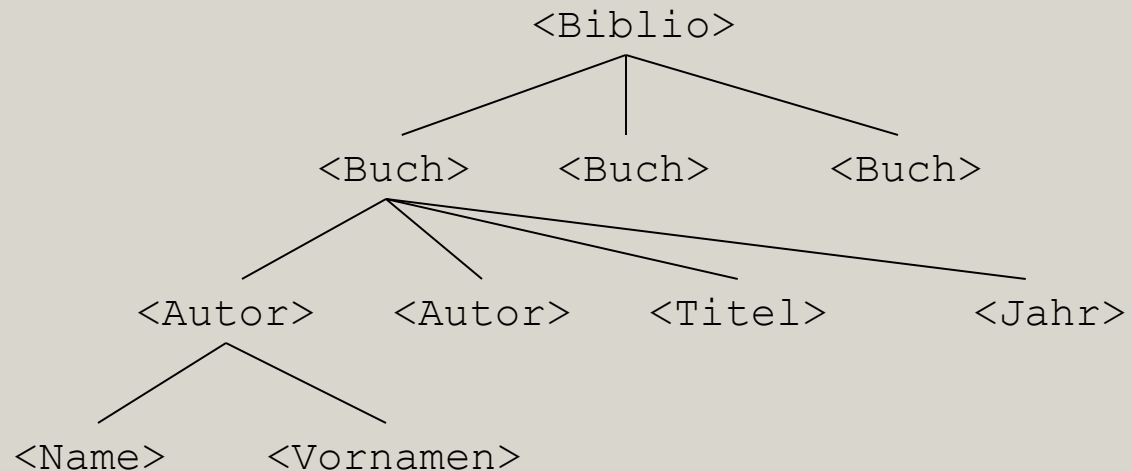
2 XML-Dokument

3 Anwendungen



# XML – Aufbau des Dokuments – Baumstruktur

## ▶ Struktur eines XML-Dokuments



## ▶ Darstellung dieser Struktur

Teil der Document Type Definition DTD  
Weiterentwicklung zu XML Schema

- 1 Idee
- 2 XML-Dokument
- 3 Anwendungen



## XML-Dokumentarten – well-formed und valid

- ▶ "Ordentliche" (well-formed) XML-Dokumente
  1. Mindestens ein Element
  2. Ein Element für das gesamte Dokument, das Wurzelement
  3. Korrekte Verschachtelung der Elemente  
also eine Baumstruktur  
keine DTD nötig
  
- ▶ "Geprüfte" (valid) XML-Dokumente
  - DTD
  - Inhalt und Struktur des XML-Dokuments entspricht der DTD



# XML – Dokument Typ Definition

## – das Wichtigste zu DTDs

### ▶ Element-Definition

```
<!ELEMENT Buch( Autor+, Titel, Jahr, Ort? )>  
<!ELEMENT Autor( #PCDATA )>
```

### ▶ Operatoren

- , Reihenfolge
- | Auswahl
- + Wiederholung, mindestens 1
- \* Wiederholung, 0 möglich
- ? optional
- () Gruppierung

### ▶ Attribut-Liste

```
<!ATTLIST Buch xml:lang CDATA "en">
```

### ▶ Entity-Deklaration

```
<!ENTITY fh-da "Fachhochschule Darmstadt">
```



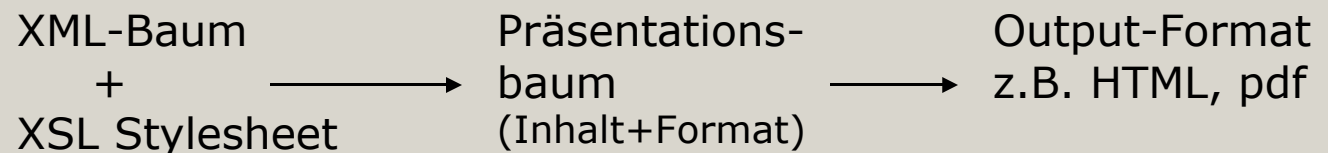
# XML – Links und Stylesheets

- ▶ XLink  
mehrere Ziele, externe Links, Verhalten des Links
- ▶ XPointer  
absolute Position im XML-Baum  
relative Position

- ▶ XSL extensible Stylesheets Beispiel

```
<xsl:template match = "Autor">  
  <fo:block font-size = "12pt">  
    <xsl:apply-templates/>  
  </fo:block>  
</xsl:template>
```

- ▶ XSL Das Prinzip







## **XML-Anwendungen** **– einige wenige Beispiele**

- ▶ Medienneutrale Dokumenten-Drehscheibe  
XML als Basisformat des Elektronischen Publizierens
- ▶ Dokumentenaustausch  
XML als Standardformat für den Datenaustausch
- ▶ CML, MathML, etc.  
XML als Sprache der chemischen Dokumentation, für mathematische Formeln etc.
- ▶ Suche in XML-Dokument  
XQL XML Query Language  
RDF Resource Description Framework
- ▶ Komponententechnologie  
Software zusammengestellt aus Komponenten  
Technologien CORBA, COM, JavaBeans  
XML als Protokoll-Sprache



## Fazit

### – XML und Informationstechnologie

- ▶ XML ist genial einfach  
Für viele Probleme des Informationsmanagements ist XML eine offensichtliche Lösung.
- ▶ XML-Werkzeuge sind erhältlich  
Rasante Entwicklung im Internet, als ob viele nur auf die Idee gewartet hätten.  
Beispiel: [www.apache.org](http://www.apache.org)
- ▶ Aber: XML ist nur eine Sprache  
Die Kunst, die Information zu strukturieren, wird durch XML nicht ersetzt — im Gegenteil.



## Literatur

Hennig Behme, Stefan Mintert: XML in der Praxis, Addison Wesley München 1998

Frank Boumphrey et al.: XML Applications, Wrox Press 1998

Charles F. Goldfarb, Paul Prescod: XML Handbuch, Prentice-Hall New York 1999

## Links

[www.w3c.org/XML/](http://www.w3c.org/XML/) ⇒ W3C

[www.oasis-open.org](http://www.oasis-open.org) ⇒ Anwendungen von XML für Dokumentation

[www.megginson.com](http://www.megginson.com) ⇒ The Simple API to XML

[www.apache.org](http://www.apache.org) ⇒ OpenSource-Initiative für XML/XSL-Prozessoren

[www.jclark.com/xml/](http://www.jclark.com/xml/) ⇒ XML-Parser in Java