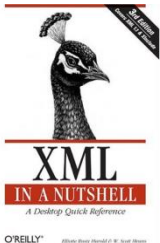


Modul 1

Semistrukturierte Datenmodellierung

Josef Altmann



Der vorliegende Foliensatz basiert vorwiegend auf:

Elliotte Rusty Harold, W. Scott Means: XML in a Nutshell: A Desktop Quick Reference, 3rd Edition, O'Reilly, 2005

W3C: XML Technology, www.w3.org/standards/xml

Inhalt

- **Strukturierte Daten**
- Semistrukturierte Daten
- Datenmodelle für semistrukturierte Daten
- Einführung in XML

Strukturierte Daten

- Daten mit gleichartigen Aufbau werden zu einer Informationseinheit zusammengefasst:

- Entität
- Objekt

[0307, Software Engineering, SE]
[0458, Medizin- und Bioinformatik, MBI]
[0456, Kommunikation Wissen Medien, KWM]
[3.009, FH3, LBS2]
[2.027, FH2, runtastic HS5]
[p20621, Josef Altmann]
[p22080, Julian Haslinger]

- Entitäten/Objekte mit gleichartigen Aufbau bzw. regelmäßiger Struktur werden gruppiert zu

- Entitätstypen
- Klassen

[0307, Software Engineering, SE]
[0458, Medizin- und Bioinformatik, MBI]
[0456, Kommunikation Wissen Medien, KWM]
[3.009, FH3, LBS2]
[2.027, FH2, runtastic HS5]
[p20621, Josef Altmann]
[p22080, Julian Haslinger]

DegreeProgramme

Room

Instructor

Strukturierte Daten

- Entitäten/Objekte des gleichen Entitätstyps bzw. der gleichen Klasse haben die gleichen Eigenschaften (typisierte Attribute)

DegreeProgramme

[0307, Software Engineering, SE]
[0458, Medizin- und Bioinformatik, MBI]
[0456, Kommunikation Wissen Medien, KWM]

{[code:integer, name:string, abbreviation:string]}

Instructor

[p20621, Josef Altmann]
[p22080, Julian Haslinger]

{[instructorNumber:integer, name:string]}

Room

[3.009, FH3, LBS2]
[2.027, FH2, runtastic HS5]

{[roomNr:integer, building:string, description:string]}

... strikte Struktur durch Schema erzwungen!

Strukturierte Daten

Datenmodell

■ Relationenmodell zur Beschreibung von strukturierten Daten

- Entität \Rightarrow Tupel
- Entitätstyp \Rightarrow Relation

DegreeProgramme

| code | name | abbreviation |
|------|-----------------------------|--------------|
| 0307 | Software Engineering | SE |
| 0458 | Medizin- u. Bioinformatik | MBI |
| 0456 | Kommunikation Wissen Medien | KWM |

} Schema

} Daten

■ Datenbankschema

- Trennung von Schema (Strukturinformation) und Daten (Instanz)
- Vollständige Strukturbeschreibung vor der Datenspeicherung

■ Daten

- sind immer Instanzen des Schemas
 - Struktur und Typisierung festgelegt, keine Abweichungen möglich
- tragen selbst keine Strukturinformationen
 - müssen mit Hilfe des Schemas interpretiert, manipuliert werden

Semistrukturierte Daten

Beispiel Instructor

```
p20621, Josef Altmann, 22610, 22699, josef.altmann@fh-hagenberg.at  
p22080, Julian, Haslinger, julian.haslinger@fh-hagenberg.at, 06641234567  
p23001, Norbert Niklas, norbert.niklas@fh-hagenberg.at  
p24001, Barbara, Traxler, barbara.traxler@fh-hagenberg.at, 22810
```

- Struktur teilweise vorhanden
 - jede Zeile repräsentiert eine Entität
 - Entitäten können gruppiert werden (Instruktoren)
- Struktur ist aber nicht strikt
 - Entitäten haben keine regelmäßige Struktur und keine strenge Typisierung
 - Struktur weiterer Entitäten nicht fixiert/vorhersagbar
 - Schema ist optional

Semistrukturierte Daten

Anforderungen

- Daten mit wechselnder und nicht streng typisierter Struktur können nicht durch ein herkömmliches (Datenbank-)Schema beschrieben werden.
- Fehlt das Schema, so muss die Bedeutung der Struktur in den Daten(sätzen) selbst wiedergegeben werden.
- Datenaustausch erfordert flexible Austauschformate, denen kein einheitliches Schema zugrunde liegt.

... WEB!

... Dokumentenorientierte Sicht (WEB) versus strikte Datenstrukturen (Datenbanken)

Semistrukturierte Daten

Datenmodell

Unstrukturiert
(Text)

Semistrukturiert
(formatierter Text)

(stark) Strukturiert
(Datenbanken)

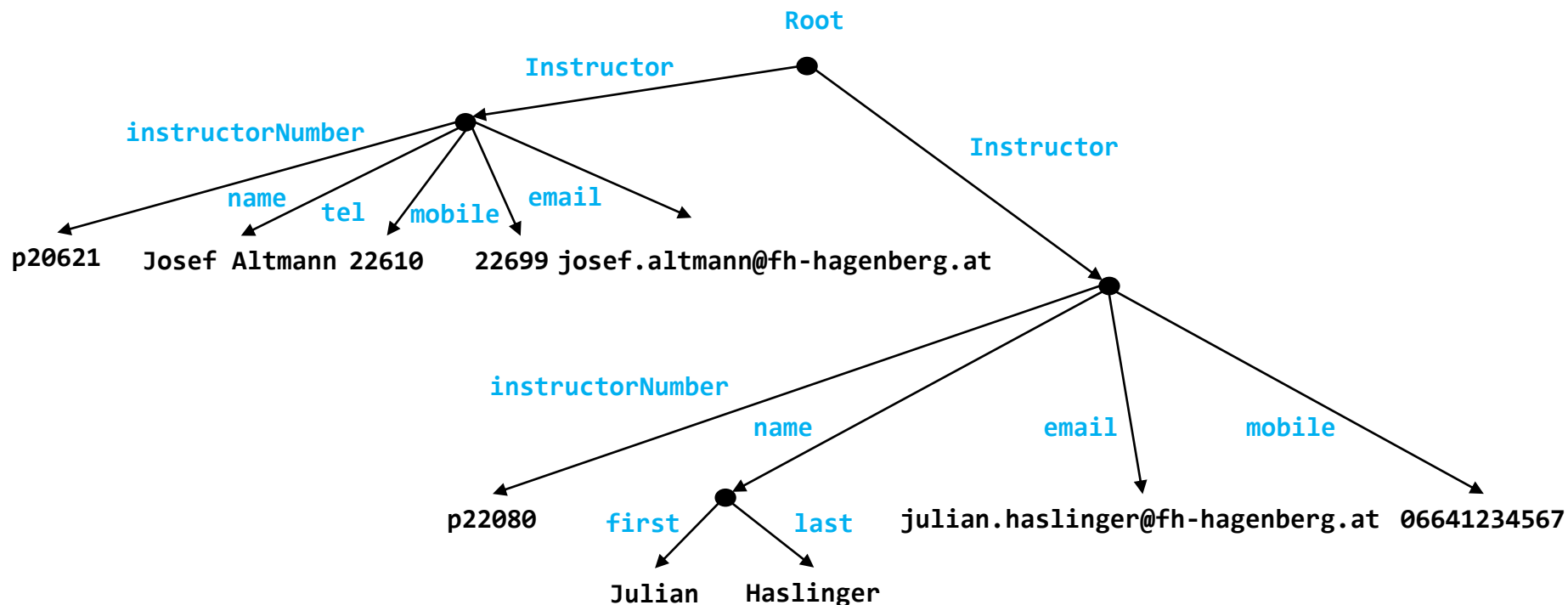
- Wie können semistrukturierte Daten beschrieben/modelliert werden?

... Ideen dazu?

Semistrukturierte Daten

Datenmodell: Gerichteter Graph

p20621, Josef Altmann, 22610, 22699, josef.altmann@fh-hagenberg.at
 p22080, Julian, Haslinger, julian.Haslinger@fh-hagenberg.at, 06641234567



- Blattknoten: Daten (Zeichenketten)
- Innere Knoten, Kanten: Struktur/Schemainformation
- Kantenbeschriftung: Attributname
- selbstbeschreibendes Datenmodell („Schemagraph“)

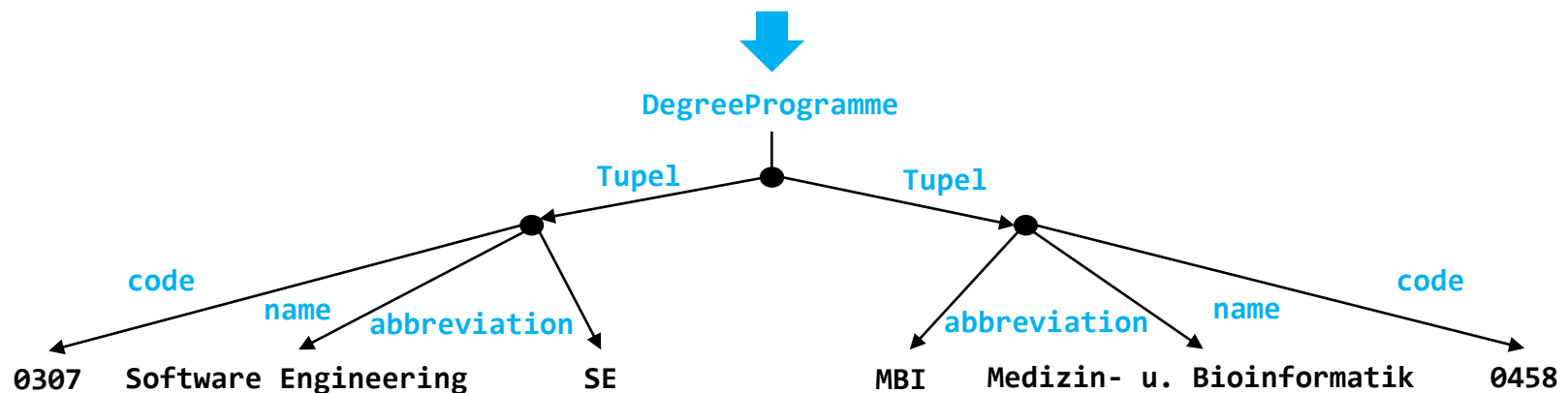
Semistrukturierte Daten

Abbildung relationaler Daten

- Strukturierte Daten sind nur ein Spezialfall von semistrukturierten Daten
- Relationale Daten können als Graph dargestellt werden

DegreeProgramme

| code | name | abbreviation |
|------|-----------------------------|--------------|
| 0307 | Software Engineering | SE |
| 0458 | Medizin- u. Bioinformatik | MBI |
| 0456 | Kommunikation Wissen Medien | KWM |



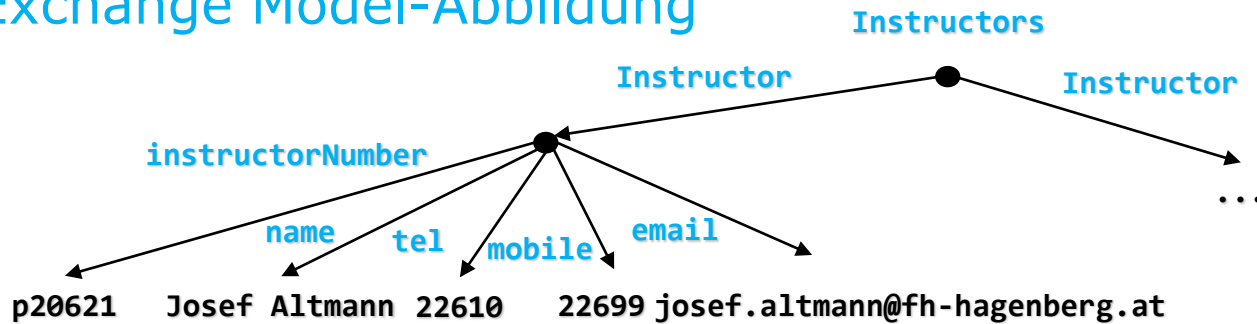
Semistrukturierte Datenmodellierung

■ Semistrukturierte Datenmodelle

- **O**bject **E**xchange **M**odel
 - OEM: Stanford University Database Group
- **eX**tensible **M**arkup **L**anguage
 - XML: w3.org/TR/xml/
- **J**ava**S**cript **O**bject **N**otation
 - JSON: json.org/
- **Y**AML Ain't Markup Language
 - YAML: yaml.org/
- ...

Semistrukturierte Datenmodellierung

Object Exchange Model-Abbildung

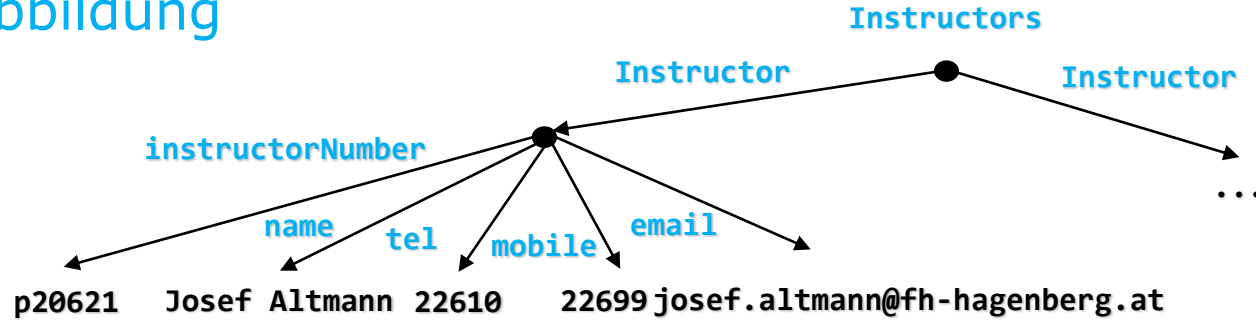


```
{
  Instructors:
    {Instructor:
      {instructorNumber: "p20621"
      name: "Josef Altmann"
      tel: 22610
      mobile: 22699
      email: "josef.altmann@fh-hagenberg.at"}}
    {Instructor:
      {instructorNumber: "p22080"
      name:
        {first: "Julian"
        last: "Haslinger"}}
      email: "julian.haslinger@fh-hagenberg.at"}
      tel: 06641234567
    }
}
```

OEM

Semistrukturierte Datenmodellierung

XML-Abbildung



```

<Instructors>
  <Instructor>
    <InstructorNumber>p20621</InstructorNumber>
    <Name>Josef Altmann</Name>
    <Tel>22610</Tel>
    <Mobile>22699</Mobile>
    <Mmail>josef.altmann@fh-hagenberg.at</Email>
  </Instructor>
  <Instructor>
    <InstructorNumber>p22080</InstructorNumber>
    <Name>
      <First>Julian</First>
      <Last>Haslinger</Last>
    </Name>
    <Email>julian.haslinger@fh-hagenberg.at</Email>
    <Tel> 0664 1234567 </Tel>
  </Instructor>
</Instructors>
  
```

XML

Semistrukturierte Datenmodellierung

JSON/YAML-Abbildung

```
{
  "Instructors": [
    {
      "Instructor": {
        "instructorNumber": "p20621",
        "name": "Josef Altmann",
        "tel": 22610,
        "mobile": 22699,
        "email": "josef.altmann@fh-hagenberg.at"
      }
    },
    {
      "Instructor": {
        "instructorNumber": "p22080",
        "name": {
          "first": "Julian",
          "last": "Haslinger"
        },
        "email": "julian.haslinger@fh-hagenberg.at",
        "tel": 06641234567
      }
    }
  ]
}
```

JSON

```
--- #YAML-Mapping
Instructors:
- Instructor:
    instructorNumber: p20621
    name: Josef Altmann
    tel: 22610
    mobile: 22699
    email: josef.altmann@fh-hagenberg.at
- Instructor:
    instructorNumber: p22080
    name:
        first: Julian
        last: Haslinger
    email: julian.haslinger@fh-hagenberg.at
    tel: 06641234567
```

YAML

Semistrukturierte Datenmodellierung

- Unterschiedliche Datenmodelle für semistrukturierte Daten
 - Object Exchange Model (OEM)
 - eXtensible Markup Language (XML)
 - JavaScript Object Notation (JSON)
 - YAML Ain't Markup Language (YAML)
 - ...

- Unterschiede liegen in
 - der Syntax
 - den Beschreibungs- und Ausdrucksmöglichkeiten
 - der Abfrage- und Manipulationssprache
 - den Anwendungsbereichen

... but the goal is the same – modelling of semi-structured data!

Inhalt

- Strukturierte Daten
- Semistrukturierte Daten
- Datenmodelle für semistrukturierte Daten
- **Einführung in XML**

Das Phänomen XML

- "XML is the ASCII of the 21th century."
- "XML is the ASCII of the Web" (Tim Bray, Co-Editor von XML 1.0)
- "If I invent another programming language, its name will contain the letter **X**." (N. Wirth, Software Pioniere Konferenz, Bonn 2001)

Google-Suche 2019:

| | |
|--|-----------------|
| JavaScript | 3.240 Mio. |
| Database | 2.020 Mio. |
| Java | 744 Mio. |
| XML | 687 Mio. |
| SQL | 317 Mio. |
| JSON | 178 Mio. |
| C++ | 77 Mio. |
| YAML | 21 Mio. |
| "University of Applied Sciences Upper Austria" | 97 K |
| "Josef Altmann" | 14 K |

Motivation für XML

Von HTML zu XML

- HTML (HyperText Markup Language) ist die "Lingua Franca" zur Beschreibung von Hypertextdokumenten im Web
- Grundkonzept: Auszeichnungen ("Markup") in Form von "Tags" (start tag, end tag)
- Einschränkungen:
 - Beschränkte Anzahl vordefinierter Tags
 - Erweiterungen um (proprietäre) Tags
 - Tags beschreiben vorwiegend Layout-Aspekte
 - Strukturelle Tags fehlen
 - Suche im Web erschwert; nur einfache Anfragen möglich

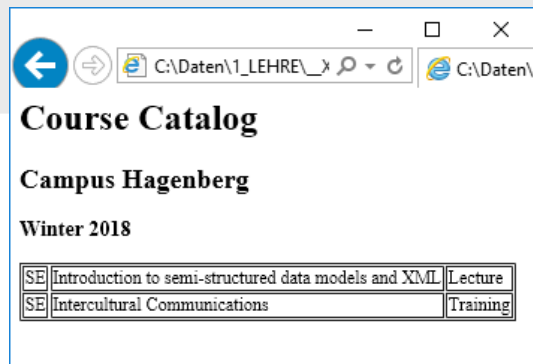
"The problem with HTML-WYSIWYG is that what you see is all you've got."
(Brian Kerningham, Entwickler von C)

Motivation für XML

Von HTML zu XML

HTML beschreibt das
Layout des Dokumentinhalts

```
<h1>Course Catalog</h1>
<h2>Campus Hagenberg</h2>
<h3>Winter 2018</h3>
<table border="1">
  <tr>
    <td>SE</td>
    <td>Introduction to ...</td>
  </tr>
  <tr>
    <td>SE</td>
    <td>Intercultural Communications</td>
  </tr>
  ...
</table>
```

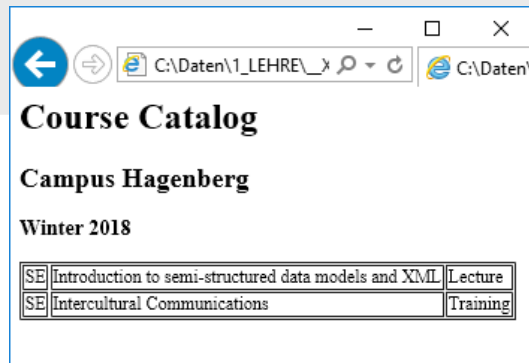


Motivation für XML

Von HTML zu XML

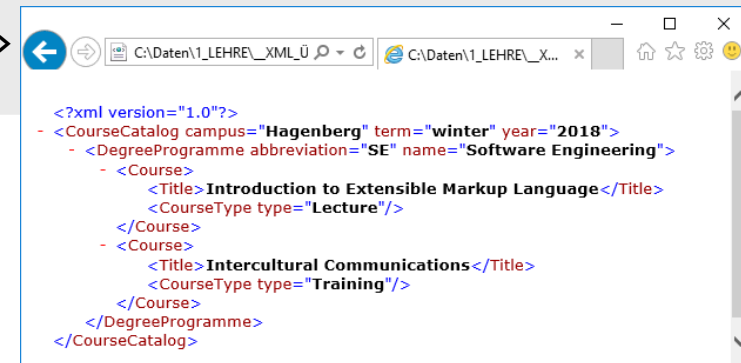
HTML beschreibt das
Layout des Dokumentinhalts

```
<h1>Course Catalog</h1>
<h2>Campus Hagenberg</h2>
<h3>Winter 2018</h3>
<table border="1">
  <tr>
    <td>SE</td>
    <td>Introduction to ...</td>
  </tr>
  <tr>
    <td>SE</td>
    <td>Intercultural Communications</td>
  </tr>
  ...
</table>
```



XML beschreibt die
Struktur u. Semantik des Dokumentinhalts

```
<CourseCatalog year="2018" term="winter" campus="Hagenberg">
  <DegreeProgramme name="Software Engineering" abbrev="SE">
    <Course>
      <Title>Introduction to ...</Title>
      <CourseType type="Lecture"/>
    </Course>
    <Course>
      <Title>Intercultural Communications</Title>
      <CourseType type="Training"/>
    </Course>
  </DegreeProgramme>
</CourseCatalog>
```



"XML will become the ASCII of the 21st century - basic, essential, unexciting."
(Tim Bray, Co-Editor of XML 1.0)

Was ist nun XML?

■ Generische Auszeichnungssprache (Markup-Sprache)

- textbasierte Sprache, die Dokumente mit zusätzlichen Tags („Markierungen“ = Strukturinformationen) versieht
- Kombination von Inhalt (Daten) und Informationen über den Inhalt (Metadaten) in einem Dokument
- keine Tags vorgegeben, beliebige Tags möglich
- Metasprache zur Definition von Sprachen

```
<CourseCatalog year="2018" term="winter" campus="Hagenberg">
  <DegreeProgramme name="Software Engineering" abbreviation="SE">
    <Course>
      <Title>Introduction to semi-structured data models and XML</Title>
      <CourseType type="Lecture"/>
    </Course>
    <Course>
      <Title>Intercultural Communications</Title>
      <CourseType type="Training"/>
    </Course>
  </DegreeProgramme>
</CourseCatalog>
```

Merkmale von XML

- **Layout-Unabhängigkeit**
 - Trennung Struktur u. Semantik des Inhalts von dessen Darstellung
- **Plattform- und Herstellerunabhängigkeit**
 - Lizenzfreie W3C-Standards
- **Erweiterbarkeit**
 - Tags und Attribute können neu definiert werden (Metasprache)
- **Strukturierbarkeit**
 - Tags können beliebig geschachtelt werden
- **Semistrukturiertheit**
 - Inhalt kann nicht-strukturierte Teile enthalten
 - Information trägt einen Teil der Struktur mit sich
- **Selbstbeschreibend**
 - ... für den Menschen: einfach zu lesen u. zu erstellen
 - ... für die Maschine: einfach zu generieren u. zu parsen
- **Validierbarkeit**
 - XML-Dokumente können ein Dokumentenmodell, d.h. eine formale Beschreibung ihres Vokabulars und ihrer Grammatik aufweisen und gegenüber diesem validiert werden.

Merkmal – Wohlgeformtheit

■ Wohlgeformtheit *

- Es existiert genau ein Wurzelement
- Jedes Start-Tag muss ein dazugehöriges End-Tag besitzen
- Tags dürfen einander nicht überschneiden
- Attributwerte müssen in Anführungszeichen stehen (paarweise "..." oder '...')
- Element- und Attributbezeichner müssen XML-Namen sein (Namenskonvention)
- XML ist case-sensitive (SE != se)
- Ein Element darf nicht zwei Attribute mit gleichem Namen besitzen
- Kommentare dürfen nicht innerhalb Tags stehen
- Reservierte Zeichen < und & dürfen nicht innerhalb von Elementinhalten oder Attributwerten auftreten (müssen ggf. maskiert werden)
- ... es gibt noch mehr

```
<CourseCatalog year="2018" term="winter" campus="Hagenberg">
  <DegreeProgramme name="Software Engineering" abbreviation="SE">
    <Course>
      <Title>Introduction to semi-structured data models and XML</Title>
      <CourseType type="Lecture"/>
    </Course>
    <Course>
      <Title>Intercultural Communications</Title>
      <CourseType type="Training"/>
    </Course>
  </DegreeProgramme>
</CourseCatalog>
```

* grundlegende syntaktische Korrektheit,
unabhängig vom anwendungsspezifischen Einsatz

Merkmal – Wohlgeformtheit

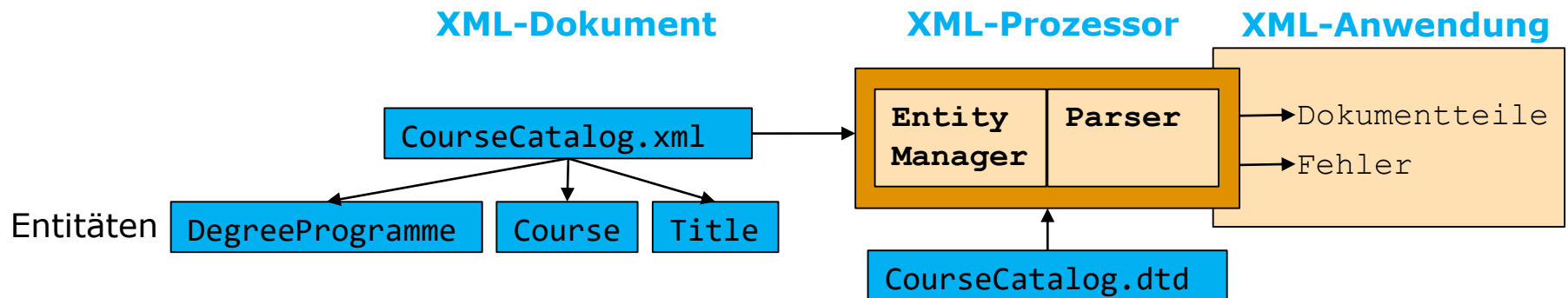
- Wohlgeformte Dokumente können zwar von jedem XML-Parser eingelesen werden, jedoch sagt die Wohlgeformtheit nichts darüber aus
 - welche Elementnamen überhaupt vorkommen dürfen,
 - in welcher Reihenfolge die Elemente im XML-Dokument erscheinen müssen,
 - welches Element Kindelement eines anderen sein darf
 - wie oft die Elemente im XML-Dokument erscheinen dürfen,
 - welche Attribute in bestimmten Elementen verwendet werden dürfen
 - ...

Merkmal – Gültigkeit

- Gültigkeit (Validität)
 - XML-Dokument ist wohlgeformt und
 - entspricht einem formalen Dokumentenmodell
- Formales Dokumentenmodell definiert
 - die ihr bekannten bzw. von ihr akzeptierten Elemente (**Vokabular**) sowie
 - die Dokumentenstruktur (**Grammatik**)
- Formales Dokumentenmodell kann mit Hilfe von Schemata definiert werden, z.B. mit
 - einer sog. **Document Type Declaration** (DTD, beschränkte Möglichkeiten) oder
 - einem **XML Schema** (aktuell)

Merkmal – Gültigkeit

- XML-Prozessoren lesen XML-Dokumente ein und
 - überprüfen entweder nur deren Wohlgeformtheit (nicht-validierende Prozessoren)
 - oder auch deren Validität (validierende Prozessoren)



- können in Anwendungen (z.B. Web-Browser, Textverarbeitung, DB-Server, Web-Server) eingebunden werden
- zerlegen ein XML-Dokument in seine Informationseinheiten und erstellen einen Baum

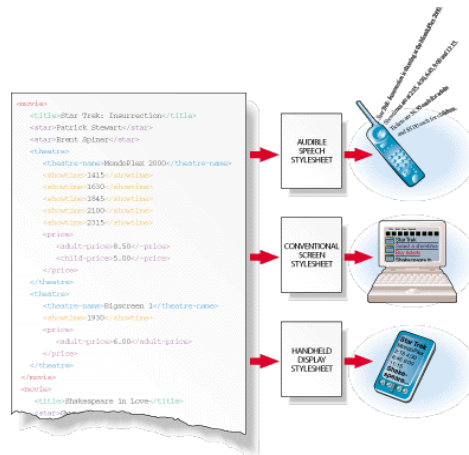
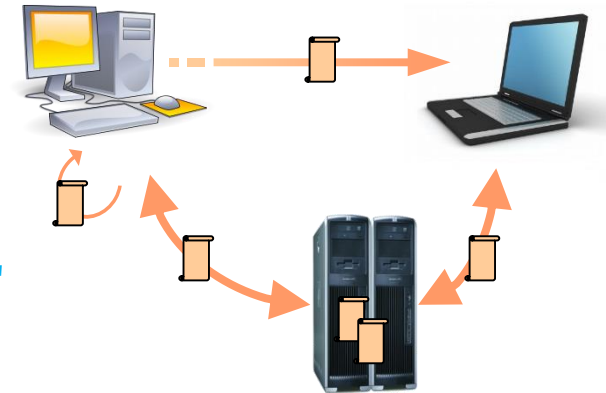
Anwendungsbereiche von XML

Wobei soll XML unterstützen?

■ Globale Sprache für den Datenaustausch

- über XML als reine Austauschnotation oder
- zusätzlich über gemeinsames Schema

"write once, read everywhere"

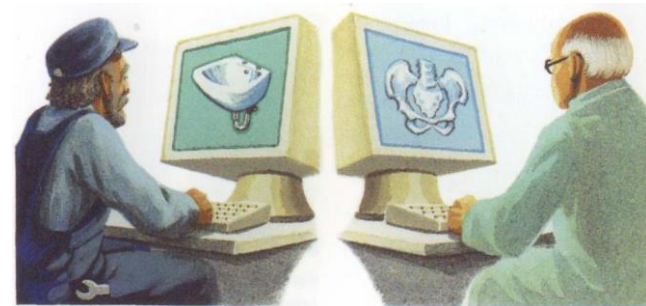


■ Multi-Delivery

- ein und derselbe Inhalt kann auf verschiedenen Geräten unterschiedlich präsentiert werden

■ Intelligente Suche

- statt einfacher Schlagwortsuche in HTML-Dokumenten, strukturbasierte Suche in XML-Dokumenten möglich



Anwendungsbereiche von XML

Branchen – "Vertikalisierung von XML"

■ Dokumentenmodelle für ...

- Reiseindustrie [openTravel](#)
- Personalwesen [XML-HR](#)
- Sprachapplikationen [VoiceXML](#)
- Vektorgraphiken [SVG](#)
- Mobile Applikationen [WML](#)
- Gesundheitswesen [HL7](#)
- Mathematik [MathML](#)
- Börsentransaktionen [FIXML](#)
- Finanzinformationen [OFX](#)
- eGovernment [eGovML](#)
- Chemie [CML](#)
- News [NewsML](#)
- Literatur [Gutenberg](#)
- Semantische Webs [RDF, OWL](#)
- Geodaten [GML](#)
[OpenStreetMap](#)
- ...

■ ... Electronic Commerce

- [xCBL](#): XML Common Business Library (Commerce One)
- [BizTalk](#): Microsoft
- [cXML](#): Commercial Extensible Mark-up Language
- [ebXML](#): OASIS + XML/EDI
- [FpML](#): Financial Products Markup Language
- [ebInterface](#): Rechnungsstandard

■ ... E-Learning

- [IMS](#) Learning Design, Beschreibung von E-Learning-Content: www.imsglobal.org/learningdesign/
- [SCORM](#) (Sharable Content Object Reference Model)

■ ... Büroanwendungen

- [ODF](#): OpenDocument
- [MS](#): Office Open XML

[en.wikipedia.org/wiki/List_of_XML_markup_languages]

Anwendungsbereiche von XML

XML Standardisierung – "Horizontalisierung von XML"

■ XML Namespaces

- Unterstützung global eindeutiger Element- und Attributnamen

■ XPath (XML Path Language)

- Pfadausdrücke zur Navigation in XML-Dokumenten

■ XQuery (XML Query Language)

- XML-Anfragesprache

■ XML Schema

- Sprache zur Beschreibung von XML-Schemata in XML

■ XSL (Extensible Stylesheet Language)

- **XSLT**: Transformation von XML-Dokumenten (deklarativ)
- **XSL-FO**: Formatierung von XML-Dokumenten (deklarativ)

■ DOM (Document Object Model)

- API für den prozeduralen Zugriff auf XML-Dokumente

■ ...

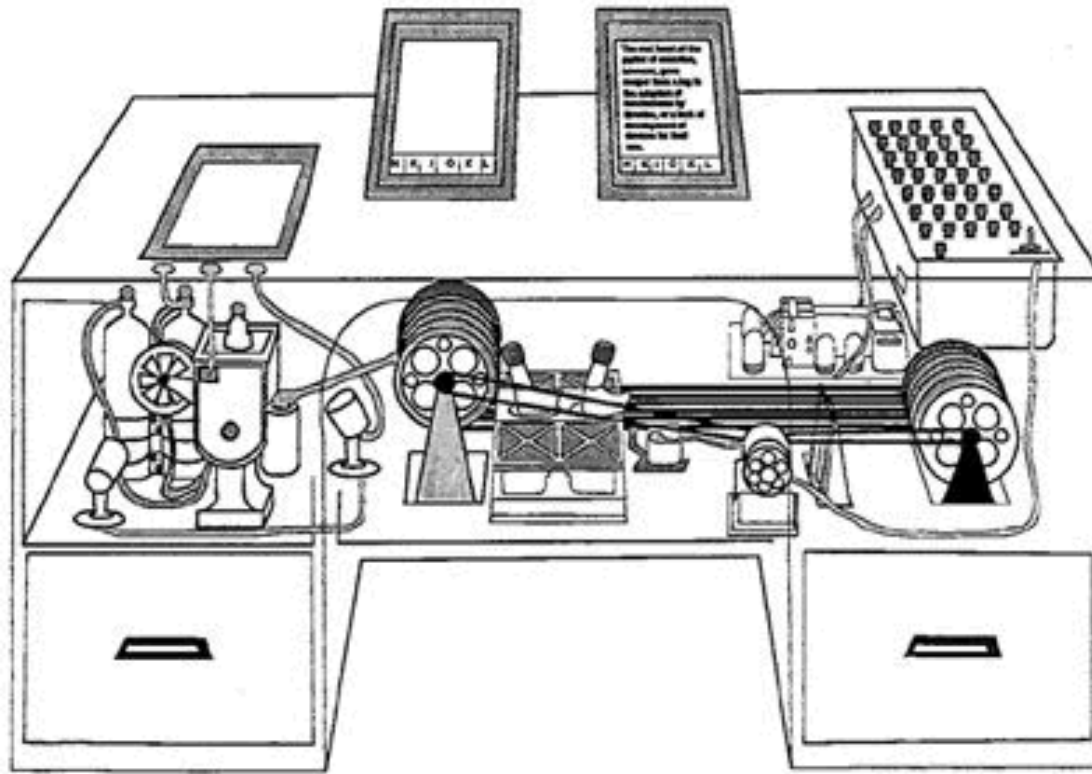
"It takes ten minutes to understand (base) XML, but then ten month to understand the new technologies hung around it."
(Peter Chen)

Geschichte von XML

- 1945: Hypertext
- 1969: GML
- 1986: SGML (ISO Standard)
- 1989: HTML (Tim Berners-Lee, CERN)
- 1994: W3C gegründet
- 1996: SGML Subset Arbeitsgruppe gegründet
- 1998: XML 1.0
- ...
- 2006: XML 1.1
- ... laufend neue Industriestandards

Geschichte von XML

Memex – Geburtsstunde von Dokumentenbeschreibungssprachen



Memex-Maschine (**M**emory **E**xender)



Bush Vannevar, "As we may think", 1945
[www.ps.uni-sb.de/~duchier/pub/vbush/vbush-all.shtml]

Geschichte von XML

"Markup" und "Generic Coding"

- Informationsgehalt eines Dokuments wird von seiner äußeren Form getrennt
- Buchdruck: "Editorial structure tags" (Ende der 60er)
⇒ "Generic Markup"
- "Markup": Begriff aus dem Verlagswesen
 - Nach der inhaltlichen Überprüfung eines Werkes erfolgt Bearbeitung durch einen Layouter (Hinzufügen von Seitenformat, Zeichensätze, ...)
 - Zuerst manuell, danach elektronisch unterstützt (Steuerzeichen, Makros)
- Konzept ist weiterhin aktuell
 - Textverarbeitungsprogramme arbeiten mit Formatvorlagen
 - Darstellung nicht über Steuerzeichen, sondern als bereits formatierter Text
- Generic Coding (GenCode-Konzept)
 - Wesentliche Idee von SGML
 - Struktur und logische Elemente eines Textes kennzeichnen
 - Die Kennzeichnungen (Markup) beschreiben dann die Art der gekennzeichneten Elemente ⇒ sie beschreiben diese genauer
 - Nicht format-orientiert

Geschichte von XML

Historische Entwicklung von Dokumentenbeschreibungssprachen

> 20 Jahre XML

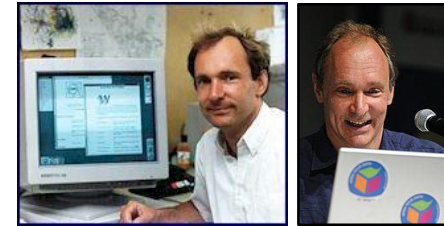
| | | |
|------|---|--|
| 2008 | XML 1.0, 5th Edition | www.w3.org/TR/2008/REC-xml-20081126/ |
| 2006 | XML 1.1, 2nd Edition | www.w3.org/TR/2006/REC-xml11-20060816/ |
| 2006 | XML 1.0, 4th Edition | www.w3.org/TR/2006/REC-xml-20060816/ |
| 2005 | XML 1.0, 3rd Edition | www.w3.org/TR/2004/REC-xml-20040204/ |
| 2004 | XML 1.1 | www.w3.org/TR/2004/REC-xml11-20040204/ |
| 2000 | XML 1.0, 2nd Edition | www.w3.org/TR/2000/REC-xml-20001006 |
| 1998 | XML 1.0 | www.w3.org/TR/1998/REC-xml-19980210 |
| 1996 | XML (eXtensible Markup Language) | XML Arbeitsgruppe (Bosak, Bray, Clark et al.) |
| 1994 | HTML-Abweichungen | Netscape, Microsoft |
| 1989 | HTML (Hypertext Markup Language) | CERN (Tim Berners-Lee) |
| 1986 | SGML (Standard Generalized Markup Language – ISO 8879) | ISO (International Standards Organization) |
| 1978 | Standardisierung (GML & GenCode) | ANSI (American National Standards Institute) |
| 1969 | GML (Generalized Markup Language) | IBM (Goldfarb, Mosher, Lorie) |
| 1967 | GenCode (Generic Coding) | Graphic Communications Association |

Organisation hinter XML

World Wide Web Consortium (W3C)



- World Wide Web Consortium [www.w3.org]
 - 1994 am MIT gegründet
 - Direktor: Tim Berners-Lee
- Industriekonsortium
 - ca. 480 Mitgliedsunternehmen und Forschungseinrichtungen (aktuelle Mitglieder, www.w3.org/Consortium/Member/List)
- Aufgaben
 - Entwicklung von einheitlichen Technologien, die den Fortschritt des Webs fördern und seine Interoperabilität sicherstellen
 - Keine Normierungsorganisation im klassischen Sinn
 - kann Einhaltung von Normen nicht auf rechtlichen Wege einklagen
 - definiert deshalb "lediglich" Empfehlungen (Recommendations)
 - Prozesse bzw. Reifegrade von Standards im "Consortium Process Document" beschrieben.
- Produkte
 - Recommendations (90%)
 - Software (10%)



Organisation hinter XML

World Wide Web Consortium (W3C)

[www.w3.org]

XML TECHNOLOGY

On this page → [technology topics](#) → [news](#) → [upcoming events and talks](#)

XML Technologies including XML, XML Namespaces, XML Schema, XSLT, Efficient XML Interchange (EXI), and other related standards.

XML Essentials

XML is shrouded by a set of essential technologies such as the infonet and namespaces. They address issues when using XML in specific applications contexts.

Efficient Interchange

XML standards are omnipresent in enterprise computing, and are part of the foundation of the Web. Because the standards are highly interoperable and affordable, people have wanted to use them in a wide variety of applications. However, in some settings (on devices with low memory or low bandwidth, or where performance is critical) experience has shown that a more efficient form of XML is required.

Schema

Formal descriptions of vocabularies create flexibility in authoring environments and quality control chains. W3C's XML Schema, SML, and data binding technologies provide the tools for quality control of XML data.

Security

Manipulating data with XML requires sometimes integrity, authentication and privacy. XML signature, encryption, and xkms can help create a secure environment for XML.

Transformation

Very frequently one wants to transform XML content into other formats (including other XML formats). XSLT and XPath are very powerful tools for creating different representations of XML content.

Query

XQuery (supported by XPath) is a query language for XML to extract data, similar to the role of SQL for databases, or SPARQL for the Semantic Web.

Components

The XML ecosystem is using additional tools to create a richer environment for using and manipulating XML documents. These components include style sheets, xlink xml id, xinclude, xpointer, xforms, xml fragments, and events.

Processing

A processing model defines what operations should be performed in what order on an XML document.

Internationalization

W3C has worked with the community on the internationalization of XML, for instance for specifying the language of XML content.

Publishing

XML grew out of the technical publication community. Use XSL-FO to publish even large or complex multilingual XML documents to HTML, PDF or other formats, include SVG diagrams and MathML formulas in the output.

STANDARDS

- Web Design and Applications
- Web Architecture
- Semantic Web
- XML Technology
- Web of Services
- Web of Devices
- Browsers and Authoring

W3C World Wide Web Consortium

Views: [desktop](#) [mobile](#) [print](#) W3C in Ihrer Nähe [Go](#)

[STANDARDS](#) [PARTICIPATE](#) [MEMBERSHIP](#) [ABOUT W3C](#)

TECHNICAL REPORTS

By date
By group

WEB AND INDUSTRY

- Automotive
- Publishing
- Entertainment (TV and Broadcasting)
- Web and Telecommunications
- Web of Data
- Web of Things
- Web Payments
- Web Security

WEB FOR ALL

- Accessibility
- Internationalization
- W3C A to Z

COMMUNITY AND BUSINESS GROUPS

TALKS AND APPEARANCES

EVENTS

W3C Welcomes IDPF as Organizations Officially Combine to Develop Roadmap for Future of Publishing
1 February 2017 | [Archive](#)

First Public Working Draft: TTML Profiles for Internet Media Subtitles and Captions 1.0.1 (IMSC1)
3 February 2017 | [Archive](#)

Two Notes Published by the Multimodal Interaction Working Group
3 February 2017 | [Archive](#)

Seven Notes Published by the Evaluation and Repairs Tool Working Group
3 February 2017 | [Archive](#)

CSS Snapshot 2017 Note Published
31 January 2017 | [Archive](#)

W3C Invites Implementations of Cooperative Scheduling of Background Tasks
31 January 2017 | [Archive](#)

Data on the Web Best Practices are now a W3C Recommendation
31 January 2017 | [Archive](#)

[More news...](#)

The World Wide Web Consortium (W3C) is an international community that develops open standards to ensure the long-term growth of the Web.

W3C operates under our Code of Ethics and Professional Conduct. Become a Friend of W3C to support the W3C mission and free developer tools.

W3C BLOG

Data on the Web? Here's How
31 January 2017 by Phil Archer

DPUB-ARIA 1.0 Is Released as a Candidate Recommendation
15 December 2016 by Tzviya Siegmán

Making smooth HTML5-based video playback a reality with Media Source Extensions
23 November 2016 by Philippe le Hegaret

JOBES

Open positions for a Web Accessibility Engineer (China), Web Standard

XML ...

- ... ist eine effiziente, flexible und einfache Metasprache, mit der Auszeichnungssprachen definiert werden können, d.h. eine Sprache mit der Sprachen definiert werden können
- ... trennt Inhalt, Struktur und Layout
- ... ist sprach- und plattformunabhängig
- ... ist ein textbasiertes Format (Unicode)
- ... wurde speziell für das Internet/Web entwickelt
- ... dient vorwiegend als Dateiaustauschformat
- ... bildet den Kern einer "Technologie-Familie", die Validierung, Abfrage und Transformation unterstützt
- ... ist ein W3C-(Industrie-)Standard und offen!

XML ...

- ... ist keine Programmiersprache
- ... ist kein Transportprotokoll
- ... ist keine Datenbank