

Einführung in XML

Übersicht: Einführung in XML

- Die Historie von XML
- XML - die standardisierte Datenfreiheit
- XML, DTDs und Validierung
- XSL oder CSS - Formatierung für XML-Elemente
- Das W3-Konsortium und XML
- XML und der Softwaremarkt
- Die Aufgaben von XML-Parsern
- XML und die Browser
- Wichtige XML-Standardsprachen



Markierungssprachen

- Spezielle Markierungssprachen (Markup Languages), z.B. RTF, HTML, ...
 - ◆ Für bestimmte Anwendungen/Geräte entworfen
 - ◆ Menge an Markierungselementen vorgegeben und nicht beliebig erweiterbar
 - ◆ Der Schwerpunkt liegt meist auf der Präsentation
 - ◆ Wenig portabel, da meist nur von bestimmten Anwendungen interpretierbar



Markierungssprachen

- Allgemeine Markierungssprachen, z.B. SGML, XML
 - ◆ Markierung beschreibt Struktur bzw. Bedeutung der einzelnen Elemente
 - ◆ Legt nicht die Interpretation (Verwendung) und die Präsentation fest
 - ◆ Metasprachen, d.h. mit ihrer Hilfe werden geeignete Markierungen definiert

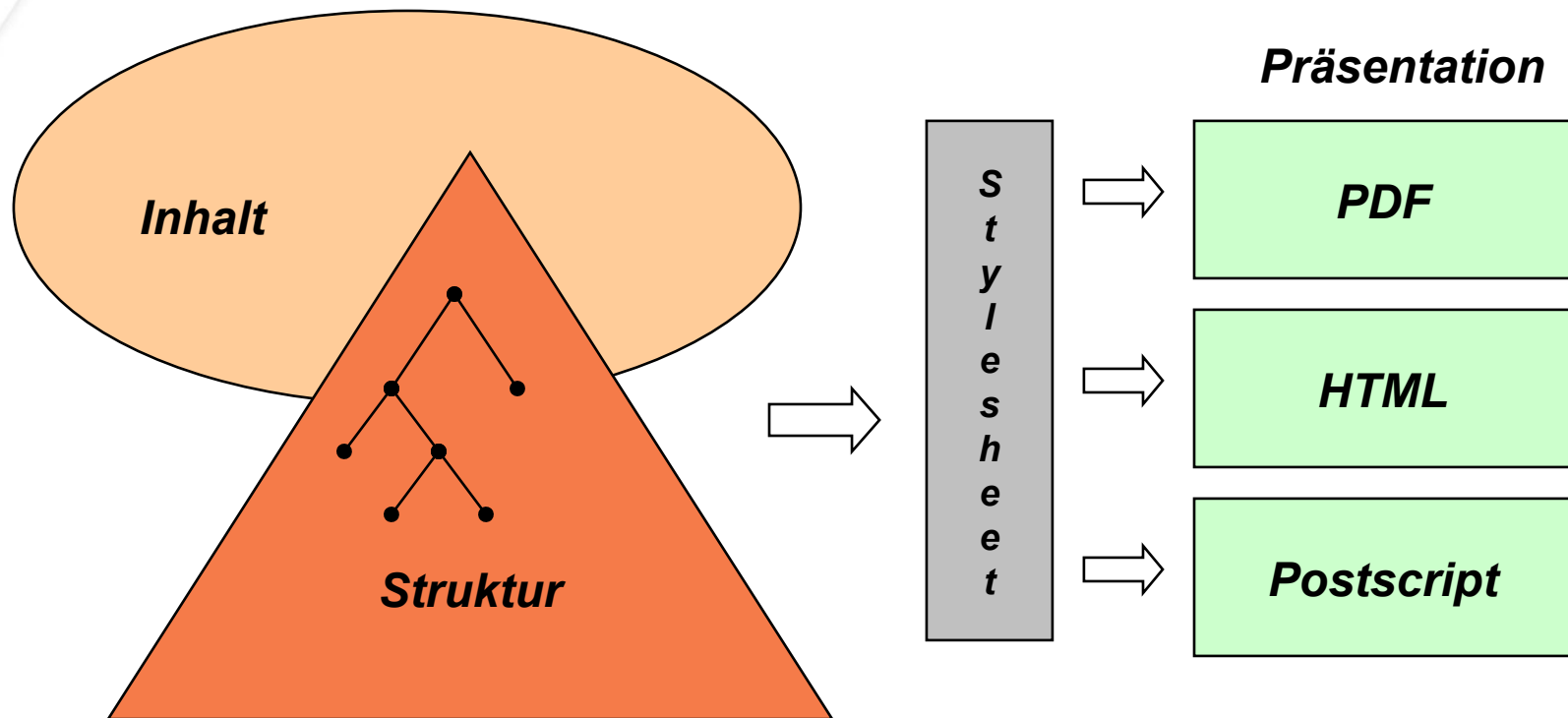


Basis-Technologie XML

- XML ist heute der Oberbegriff für eine ganze Familie von Technologien und Standards zum Strukturieren von Informationen, Formatieren von Dokumenten bis hin zur Datenfilterung
- XML (eXtensible Markup Language) ist im engeren Sinne ein Regelwerk, um Auszeichnungssprachen (Markup Languages) zu erstellen
- Grundgedanke: Trennung von Inhalt, Struktur und Präsentation
- Entwurfsziele:
 - ◆ SGML-kompatibel
 - ◆ möglichst universell einsetzbar
 - ◆ XML-Dokumente sollen leicht erstellbar und lesbar sein



Inhalt – Struktur - Darstellung



Die Historie von XML

- 1969 GML (Generalized Markup Language) wird bei IBM entwickelt
- 1986 Festlegung des ISO Standard 8879 für SGML
- 1989 Entwicklung von HTML als SGML Applikation am CERN (Organisation européenne pour la recherche nucléaire)
- 1996 Beginn der Arbeit an einer vereinfachten Version von SGML vom W3-Konsortium (www.w3.org)
- 1998 Standard für XML 1.0 wird festgelegt
- 2001 Vorstellung von Schema
- 2004 Standard für XML 1.1 wird festgelegt



Änderungen in der Version 1.1

- XML 1.0 unterstützt Unicode bis Version 2.0.
- XML 1.1 muss Unicode bis Version 4.0 unterstützen.
 - ◆ Bei XML 1.0 ist alles verboten, was nicht erlaubt ist. Bei XML 1.1 ist das genau umgekehrt.
 - ◆ Es werden zusätzlich Zeichen für Zeilenumbruch und den Beginn des Zeichens freigegeben.
 - ◆ Es wurden spezielle Steuerzeichen eingeführt, die immer als Entitäten angegeben werden müssen.

Das W3-Konsortium und XML

- Das World Wide Web Consortium (W3C)
(<http://www.w3c.org>) legt Standards (im Sinne von Einsatzempfehlungen) für XML-Anwendungen fest. Dazu zählen
 - ◆ XML
 - ◆ PNG
 - ◆ SVG
 - ◆ HTML ...
- Das W3-Konsortium ist ein offenes, herstellerunabhängiges Gremium und wurde 1994 von Tim Berners-Lee (CERN) und dem MIT gegründet.
 - ◆ Die mehr als 400 Mitglieder sind Vertreter aus Wirtschaft und Lehre.



Ein Blick auf eine XML-Datei

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

```
<produkte>
```

```
  <artikel>
```

```
    <einkaufspreis>15,00</einkaufspreis>
```

```
    <verkaufspreis>59,00</verkaufspreis>
```

```
    <beschreibung>
```

Rechtwinkliger Besen für die Sauberkeit in
allen Ecken. Passend auf jeden Besenstiel.

```
  </beschreibung>
```

```
  </artikel>
```

```
</produkte>
```



Grundbegriffe

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

`<produkte>` Wurzelement

Start-Tag `<artikel>` Inhalt
└──┘
`<beschreibung> Besen </beschreibung>` End-Tag

`</artikel>` Attribut
└──┘
`<artikel preis="123" >Tanne</artikel>`

`<artikel/>` Leeres-Element-Tag

`</produkte>`

Wohlgeformtheit

- XML-Dokumente müssen **wohlgeformt** sein (**well-formed**).
- Das gilt, wenn es unter anderem
 - ◆ genau ein Wurzelement gibt,
 - ◆ Start- und Ende-Tags unter Beachtung von Groß/Kleinschreibung übereinstimmen,
 - ◆ die Elemente korrekt verschachtelt sind,
 - ◆ Attributwerte stehen stets in Anführungszeichen.

`<artikel><absatz>...</absatz></artikel>`



`<artikel><absatz>...</artikel></absatz>`



Validierung

- Zusätzlich lassen sich Regeln für die Struktur definieren.
 - ◆ Diese Regeln werden unabhängig von den Daten definiert.
 - ◆ Das Testen dieser Regeln nennt sich **Validierung**.
- XML-Dokumente, die der Strukturbeschreibung entsprechen, sind **gültig** oder **valid**.

XML-Dokument mit zugehöriger DTD

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

```
<!DOCTYPE produkte [  
  <!ELEMENT produkte (artikel*)>  
  <!ELEMENT artikel (einkaufspreis, verkaufspreis, beschreibung)>  
  <!ELEMENT einkaufspreis (#PCDATA)>  
  <!ELEMENT verkaufspreis (#PCDATA)>  
  <!ELEMENT beschreibung (#PCDATA)>  

```

```
<produkte>  
  <artikel>  
    <einkaufspreis>15,00</einkaufspreis>  
    <verkaufspreis>59,00</verkaufspreis>  
    <beschreibung>  
      Rechtwinkliger Besen für die Sauberkeit in allen Ecken.  
      Passend auf jeden Besenstiel.  
    </beschreibung>  
  </artikel>  
</produkte>
```



Grammatiken

- Zur Validierung bieten sich an:
 - ◆ Die DTD (Document Type Definition) wurde von SGML übernommen und ist Bestandteil des XML-Standards.
 - ◆ Eine detaillierte Form der Strukturbeschreibung ist mit XML-Schema möglich.
 - ◆ Weitere Möglichkeiten sind RELAX NG, Schematron, ...

Kurzcharakterisierung von XML

- XML dient zur Strukturierung der verschiedensten darzustellenden Daten.
 - ◆ XML ist ein systemunabhängiges Datenformat.
- XML-Dokumente bestehen aus menschenlesbarem Text und sind nicht binär.
- XML ist eine Familie von Techniken.
 - ◆ XML ist eingebettet in anderen Techniken, wie Abfragesprachen.
- XML ist modular.
 - ◆ Eigene Datenformate für spezielle Anwendungen können sehr leicht definiert und verarbeitet werden.



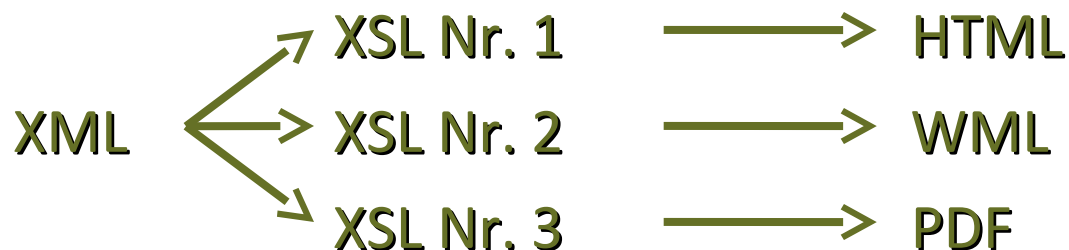
Aufgaben von XML-Parsern

- Der XML-Standard legt zwar das Datenformat fest, beschreibt aber weniger wie Computer-Programme diese Dokumente verarbeiten.
- Ein XML-Prozessor (auch Parser) liest XML-Dokumente ein und erlaubt den Zugriff auf Inhalt und Struktur.
- Dabei überprüft er, ob
 - ◆ die Datenstruktur nach den Regeln von XML strukturiert ist (Wohlgeformtheit)
 - ◆ die Datenstruktur nach einer Strukturbeschreibung gültig ist (z. B. einer DTD).



Darstellung mit CSS und XSL

- XML beschreibt lediglich die Daten. Damit ist aber keine Visualisierung verbunden.
- Zur Darstellung bieten sich an:
 - ◆ CSS ermöglicht eine einfache Darstellung der XML-Daten in einem Webbrowser.
 - ◆ XSL wird verwendet, um XML in andere Textformate umzuwandeln.

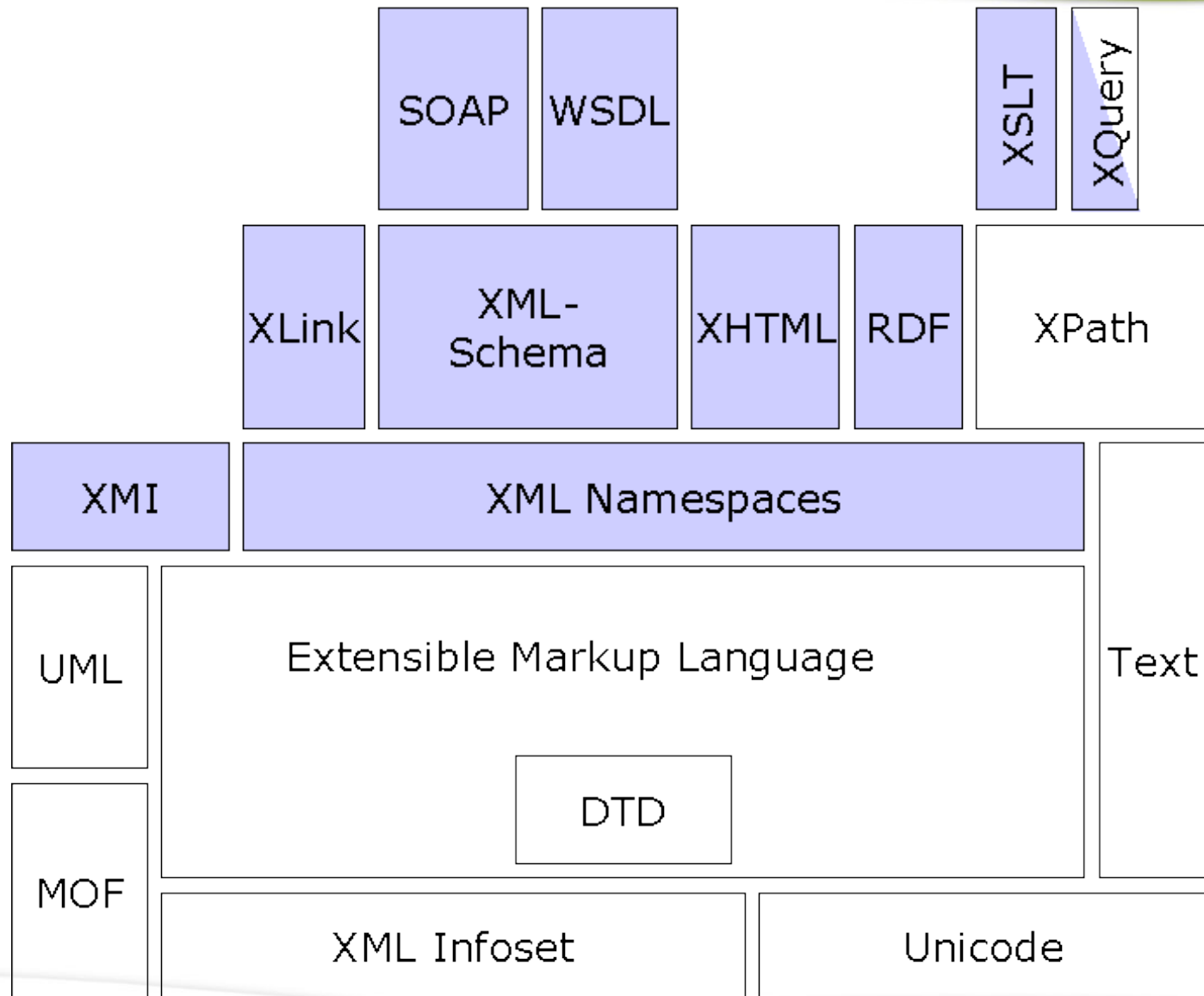


XML und die Browser

- Die modernen Browser wie Internet Explorer oder Firefox stellen XML-Dokumente hierarchisch da.
- Falls Visualisierungsinformationen (CSS) mit dem XML-Dokument verbunden sind, wenden die Browser diese Formatierungsanweisungen für die Visualisierung an.
- Durch XSL können XML-Daten vom Server oder vom Client in HTML umgewandelt werden.



Die XML-Sprachfamilie



Wichtige XML-Standardsprachen

- Eine Menge von XML-Elementen, die eine festgelegte Bedeutung haben, werden als XML-Applikation bezeichnet.
 - ◆ SVG: Vektorgrafik
 - ◆ MathML: Mathematische Formeln
 - ◆ SMIL: Multimedia-Inhalte
 - ◆ RDF: Metadaten zur Beschreibung von Dokumenten
 - ◆ WML: Beschreibungssprache für Inhalte von WAP-Seiten
 - ◆ GML: Austauschformat für Geodaten
 - ◆ ...



Zusammenfassung

- XML ist ein universelles Format für die Beschreibung von Daten.
- Die Regeln für erlaubte Elemente und Attribute wird unabhängig von den XML-Daten beschrieben.
- Durch CSS können XML-Daten in einem Webbrowser dargestellt werden.
- Mit XSL können XML-Daten in andere Formate umgewandelt werden.
- Das W3-Konsortium legt die Basisstandards fest.

