

Einführung in XML und XSLT

Am Beispiel von www.ccc.de

Lars Weiler <pylon@ccc.de>

Easter(H)egg — 19. April 2003



Inhalt

1. Definitionen und Hintergrund

- ▷ Extensible Markup Language (XML)
- ▷ Document Type Definition (DTD)
- ▷ Cascading Style Sheets (CSS)
- ▷ Extensible Stylesheet Language (XSL)
- ▷ XSL Translation (XSLT)
- ▷ Ressource Description Framework (RDF)
- ▷ Nützliche Tools

2. Funktionsweise von www.ccc.de

3. Selber machen!



Definitionen und Hintergrund



Extensible Markup Language (XML)

- ▷ 1998 vom *World Wide Web Consortium (W3C)* veröffentlicht
- ▷ Untermenge der *Standard Generalized Markup Language*
- ▷ Einfaches Format, leicht von Mensch und Maschine lesbar
- ▷ Die Einfachheit erlaubt eine große Vielfaltigkeit

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<vortrag id="XML-und-XSLT">
  <referent email="pylon@chaosdorf.de">Lars Weiler</referent>
  <titel>Definitionen und Hintergrund</titel>
  <untertitel>Extensible Markup Language (XML)</untertitel>
  <inhalt>Hier steht der Vortrag.</inhalt>
  <!-- Zwischendurch ein Kommentar -->
  <inhalt>Und noch mehr...</inhalt>
</vortrag>
```



Regeln

- ▷ Alles muss in Elemente eingeschlossen sein
- ▷ Elemente können einzeln stehen
- ▷ Elemente können Attribute enthalten
- ▷ Elemente können wiederum Elemente enthalten
- ▷ Keine überkreuzende Verschachtelung
- ▷ Alle Zeichen für Daten sind erlaubt – aber:
 - `<` → `<`
 - `>` → `>`
 - `&` → `&`
 - `"` → `"`
 - `'` → `'`



Document Type Definition (DTD)

- ▷ Gibt die Struktur einer XML-Datei vor
- ▷ Definiert Elemente, Attribute und Entities
- ▷ Schränkt den Bereich der erlaubten Zeichen ggf. ein

```
<!-- Vortrags-DTD -->
<!ELEMENT vortrag      (referent+, titel, undertitel?, inhalt*)>
<!ATTLIST vortrag      id      ID      #REQUIRED>
<!ELEMENT referent     (#PCDATA)>
<!ATTLIST referent     email   CDATA   #IMPLIED>
<!ELEMENT titel        (#PCDATA)>
<!ELEMENT undertitel   (#PCDATA)>
<!ELEMENT inhalt       (#PCDATA)>
```

? : Element ist optional

+ : Element darf mehrmals vorkommen, aber mindestens ein mal

* : Element darf kein, ein oder mehrmals vorkommen



Der Browser zeigt aber nur das XML in einer Baumstruktur an!

Wie kriegt man jetzt eine schöne Ausgabe hin?

Lösung: CSS und XSL/XSLT

CSS: Cascading Style Sheets

XSL: Extensible Stylesheet Language

XSLT: XSL Transformations



Cascading Style Sheets (CSS)

- ▷ Definiert Formateigenschaften von Elementen
- ▷ Ermöglicht die Formatierung für unterschiedliche Ausgabegeräte
- ▷ Erlaubt zentrale Formate

```
* {display: block;}
vortrag {background-color: yellow; color: blue;}
referent {font-size: 0.8em; padding: 0.5em 0 0.5em 0;}
titel {
    font-size: 2em;
    padding: 2em 0 2em 0;
    text-align: center;
}
untertitel {font-size: 1.5em; padding: 0.2em 0 0.5em 0;}
inhalt {padding: 0 3pc 0.4em 3pc; color: green;}
```



Extensible Stylesheet Language (XSL)

- ▷ Ebenso eine Auszeichnungssprache
- ▷ Wesentlich mächtiger als CSS
- ▷ XSL wird mit Hilfe von XML definiert (keine eigene Syntax)
- ▷ Erlaubt z.B. auch Bedingungen, Aufzählungen, u.v.m.

```
<!-- vorherige Elemente ausgelassen -->
<xsl:template match="titel">
  <div style="font-size: 2em; padding: 2em 0 2em 0;
           text-align: center;">
    <xsl:value-of select="." />
  </div>
</xsl:template>
<!-- weitere Elemente ausgelassen -->
```



XSL Transformation (XSLT)

- ▷ Browser unterstützen noch kein XSL, von daher muss die Umwandlung (Transformation) in HTML serverseitig geschehen
- ▷ XSLT wandelt XML in ein durch XSL definiertes Format um
- ▷ Transformation aus einem *Quellbaum* in einen *Ergebnisbaum*

Quellbaum:

Element: vortrag

Attribut: id

Wert: XML-und-XSLT

Element: referent

Attribut email

Wert: pylon@chaosdorf.de

Wert: Lars Weiler

...



Resource Description Framework (RDF)

- ▷ Enthält Metadaten zu Dokumenten
- ▷ Geeignet für Katalogsysteme, News-Listen, u.v.m.
- ▷ Besteht aus XML

```
<?xml version="1.0"?>
<RDF xmlns="http://w3.org/TR/1999/PR-rdf-syntax-19990105#"
      xmlns:DC="http://purl.org/DC#">
  <description about="vortrag.xml">
    <DC:Title>XML und XSLT</DC:Title>
    <DC:Creator>Lars Weiler</DC:Creator>
    <DC>Date>2003-03-11</DC>Date>
    <DC:Subject>XML und XSLT verhelfen zum semantischen
      Web -- am Beispiel von www.ccc.de</DC:Subject>
  </description>
</RDF>
```



Nützliche Tools

XML Syntax Checker: `xmllint`

```
$ xmllint --valid --noout vortrag3.xml
vortrag3.xml:3: validity error: element vortrag does not
  carry attribute id
</vortrag>
      ^
```

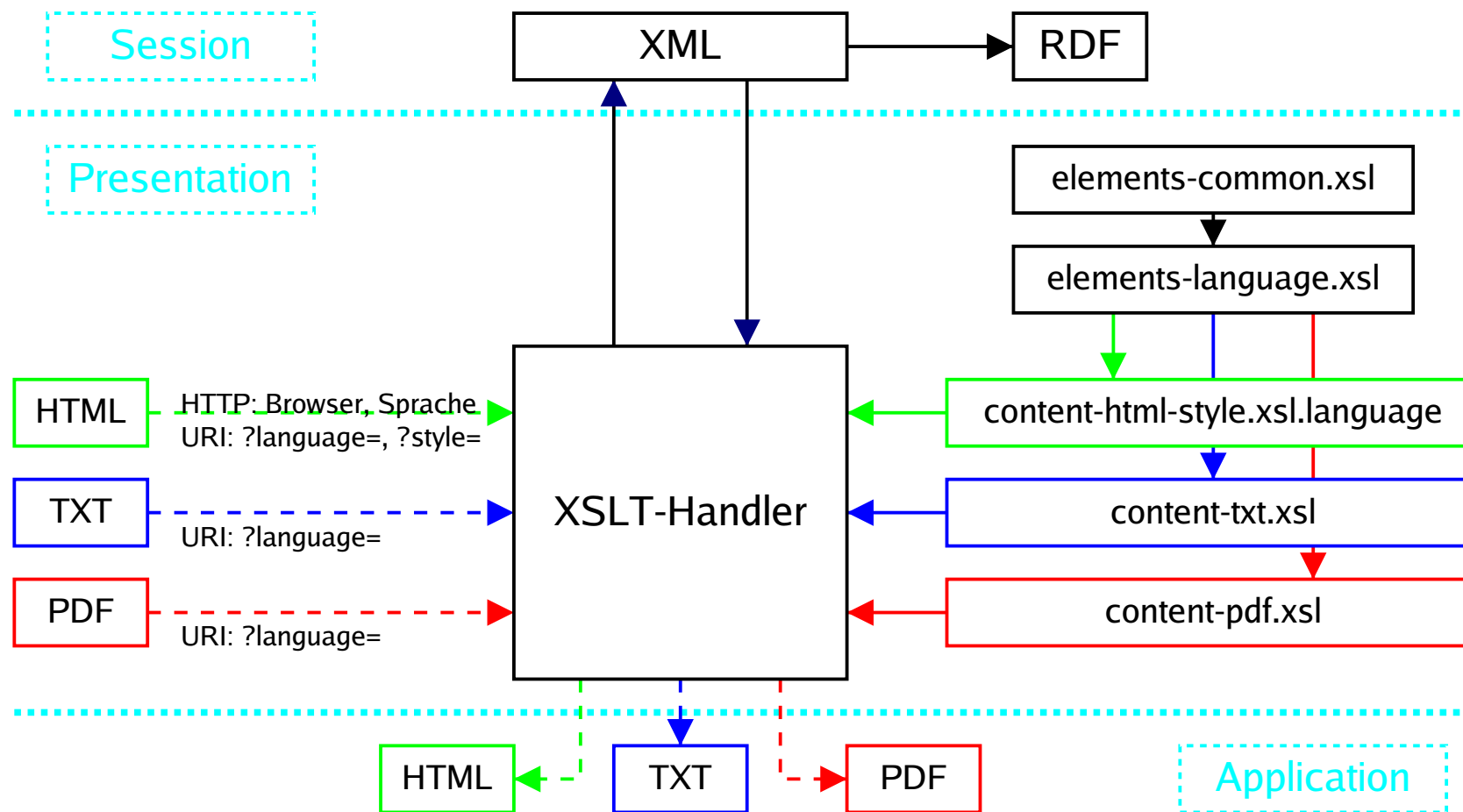
XSLT Processor: `xsltproc`



```
$ xsltproc vortrag.xsl vortrag2.xml
<html>
<head>
<meta http-equiv="Content-Type" content="text/html;
  charset=iso-8859-1">
<title>Definitionen und Hintergrund</title>
<style type="text/css">
  body {background-color: yellow; color: blue;}
</style>
</head>
<body>
  <p><a href="mailto:pylon@chaosdorf.de">Lars Weiler</a></p>
  <h1 style="text-align:center;">Definitionen und
    Hintergrund</h1>
  <h2>Extensible Markup Language (XML)</h2>
  <p style="color:green;">Hier steht der Vortrag.</p>
  <p style="color:green;">Und noch mehr...</p>
</body>
</html>
```



Funktionsweise von www.ccc.de





Weitere Informationen

Diese Folien im Netz:

http://www.chaosdorf.de/~pylon/files/XML_und_XSLT.pdf

- ▷ World Wide Web Consortium
- ▷ Stefan Münz, *SELFHTML*
- ▷ Stefan Mintert, *Informationen zur Extensible Markup Language (XML)*
- ▷ Kuno Dünhölter, *Das Web automatisieren mit XML*
- ▷ ZVON.org, *The Guide to the XML Galaxy*
- ▷ Chaosradio 79, *Das Semantische Web*