

Textkodierung mit XML

Summer School "Digitale Edition" 2017
Christiane Fritze

#ideschool17



Überblick

- XML-Grundlagen: Was ist XML und wie geht das?
- Das XML-Dokument und seine Regeln
- XML schreiben - erste Fingerübung
- XML – und wie weiter?
- Zusammenfassung



Text kodieren mit XML - warum

- Mit XML wird der Text „aufgeschlossen“ - für die Maschine
- XML explifiziert Implizites oder Interpretiertes
- Mit XML wird dem Text Information hinzugefügt
- XML vereinfacht die (maschinelle) Nachnutzung des Textes in anderen Kontexten

Beispiel

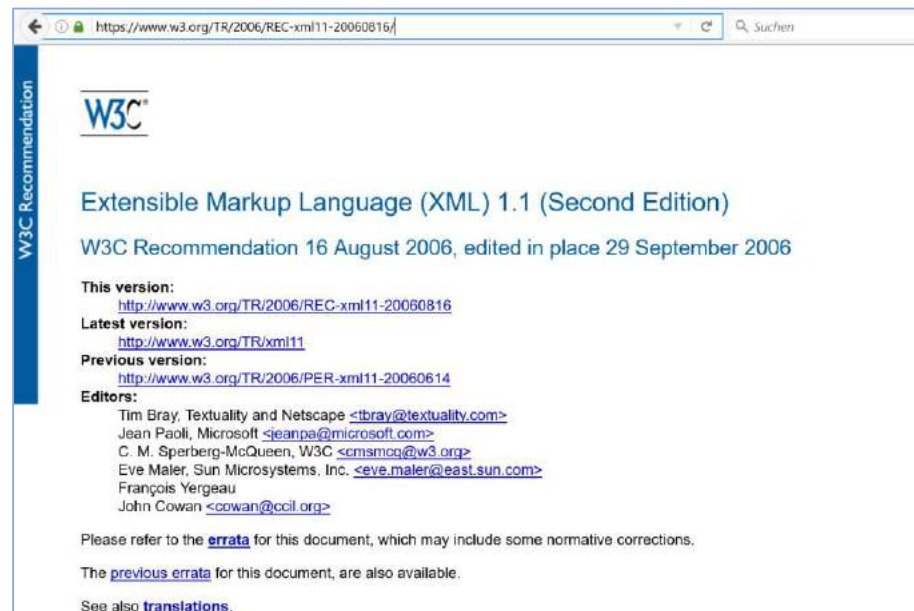


Arndt, Johann: Von wahrem Christenthumb. Bd. 1. Magdeburg, 1610. [Faksimile 55]

```
<fw place="bottom" type="sig">  
  <hi rendition="#fr">Erfter Theil D</hi>  
</fw>  
<fw place="bottom" type="catch">  
  <hi rendition="#fr">erwecket</hi>  
</fw>  
<lb/>
```

XML Grundlagen

- XML heißt e**X**tensible **M**arkup **L**anguage. → erweiterbar
- XML ist ein internationaler W3C-Standard.
 - 10.2.1998 erste Empfehlung
 - 16.8.2006: 4. Edition ist der aktuell gültige Standard XML 1.1.



The screenshot shows a web browser window with the URL <https://www.w3.org/TR/2006/REC-xml11-20060816/>. The page content includes the W3C logo, the title "Extensible Markup Language (XML) 1.1 (Second Edition)", and the date "W3C Recommendation 16 August 2006, edited in place 29 September 2006". It also lists the current, latest, and previous versions with their respective URLs, and the names and contact information of the editors: Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, François Yergeau, and John Cowan. At the bottom, it mentions errata and translations.

Was ist XML?

- ... ein weit verbreiteter Standard für die **Beschreibung** und den **Austausch** von Daten.
- ... **trennt Struktur und Darstellung** voneinander, Inhalt und Präsentation.
- ... XML-Dokumente werden nach einem **Dokumentenmodell** entwickelt.
 - Dokumentenmodell als XML-Schema
- ... **menschen- und maschinenlesbar**

```
<element attributname="attribut">  
    Mein Text steht hier.  
</element>
```

Was ist XML?

- XML ist eine universelle Metasprache.
- Einzelne Vokabulare
 - Z.B. xHTML, TEI, EAD, METS/MODS, SVG etc.
- Verschiedene Nutzungen, z.B.
 - Webpräsentation
 - Visualisierungen
 - Druckausgaben
- Menschen- und maschinenlesbar
- einfach

```
<element attributname="attribut">  
    Mein Text steht hier.  
</element>
```

Was ist XML?

- XML wird von einer breiten Softwarepalette unterstützt.
- XML hat eine große internationale Nutzer- und Entwickler-gemeinde.
- XML ist die Grundlage vieler Anwendungsstandards („XML is everywhere“).
- XML umfasst eine ganze Familie von begleitenden Standards.

```
<element attributname="attribut">  
    Mein Text steht hier.  
</element>
```

Überblick

- XML-Grundlagen: Was ist XML und wie geht das?
- Das XML-Dokument und seine Regeln
- XML schreiben - erste Fingerübung.
- XML – und wie weiter?
- Zusammenfassung



XML-Regeln

- Alles ist Text, Zeichendaten
- Markup bedeutet, Text mit Auszeichnungen zu versehen ...
- Dafür gibt es (Auszeichnungs-)Elemente.
 - Öffnendes Tag ... Text ... Schließendes Tag
- Elemente können Attribute haben.
- Elemente enthalten Elemente, Text, beides - oder nichts.

```
<elementname>Elementinhalt</elementname>
```

```
<elementname  
attributname="attributwert">Elementinhalt</elementname>
```

```
<WasSollDas></WasSollDas> = <WasSollDas/>
```



XML-Regeln

- Alle Elemente müssen richtig geschachtelt sein.
- Es gibt ein – und nur ein - Wurzelement.



XML-Regeln: Elemente

- Elementnamen müssen mit einem **Buchstaben**, **Unterstrich** oder **Doppelpunkt** beginnen.
 - Ergo: Sie dürfen nicht mit Zahlen beginnen.
- Elementnamen unterscheiden Groß- und Kleinschreibung.
 - `<Postkarte>` ≠ `<postkarte>`
- Elementnamen können Buchstaben, Zahlen, Bindestriche, Punkte oder Unterstriche, Umlaute und Akzente enthalten.
- Elementnamen dürfen nicht mit `xml` beginnen.
- Elementnamen können beliebig lang sein.
- Die Verwendung von `<` `>` `&` `'` und `"` ist nicht erlaubt.

XML-Regeln: Attribute

- Ein Element darf beliebig viele Attribute tragen.
- Ein Element darf nicht zweimal das gleiche Attribut haben.
- Attributwerte müssen in Anführungszeichen stehen.

```
color="red"
```

- Attributwerte dürfen alle möglichen Zeichen enthalten

XML-Regeln: Dokumentaufbau

- XML-Deklaration

```
<?xml version="1.0" encoding="UTF-8"?>
```

- Processing instructions

```
<?xml-stylesheet type="text/xsl" href="„transformieren.xsl“?>
```

- „eigentlichen“ Elemente, schön geschachtelt

```
<Postkarte><Anrede>Lieber  
<Name>Paul</Name></Anrede>  
</Postkarte>
```

- Kommentare

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

XML-Regeln: Entitäten

- Was ist mit bereits durch XML besetzten Zeichen?
 - Dafür gibt es spezielle Entitäten:

< wird zu `<`;

> wird zu `>`;

& wird zu `&`;

“ wird zu `"`;

‘ wird zu `'`;

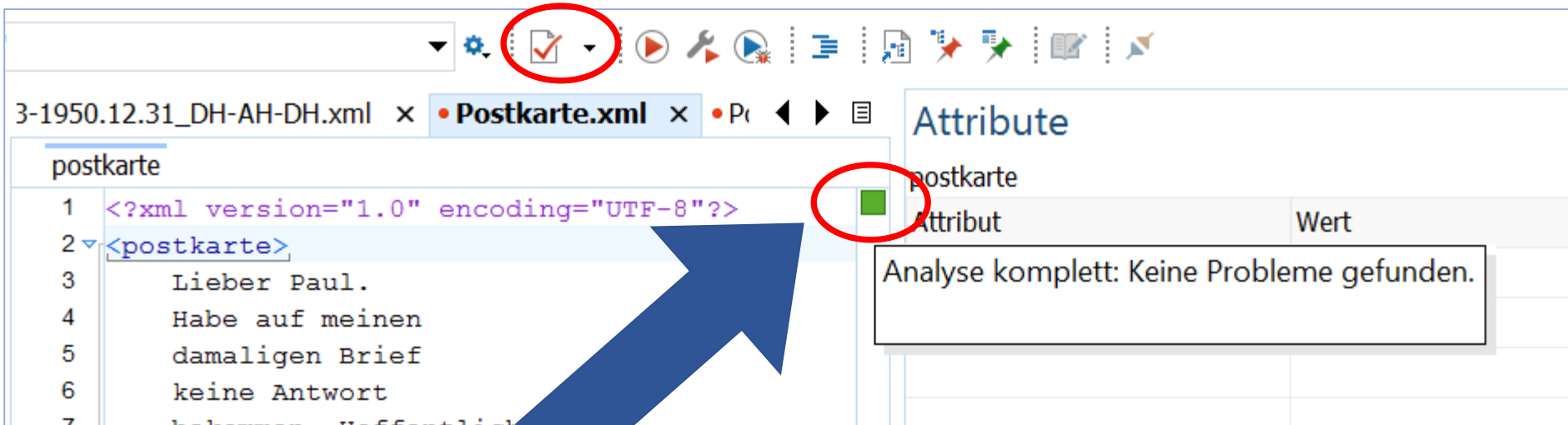
Alternativ können sie als CDATA-Sektionen markiert werden:

```
<![CDATA[ Inhalte mit <spitzen> Klammern ]]>
```

XML-Prüfung

- Die Prüfung erledigt ein Parser.
- Prüfung auf Wohlgeformtheit.
 - Ein XML-Dokument ist **wohlgeformt**, wenn es den Regeln des XML-Standards entspricht.
- Prüfung auf Validität/Gültigkeit.
 - Ein XML-Dokument ist **valide**, wenn es wohlgeformt ist und der Grammatik des XML-Schemas entspricht.

XML-Prüfung im Oxygen-Editor



The screenshot shows the Oxygen XML Editor interface. The main editor displays the XML content of a file named 'Postkarte.xml'. The XML structure is as follows:

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <postkarte>
3   Lieber Paul.
4   Habe auf meinen
5   damaligen Brief
6   keine Antwort
7   bekommen. Hoffentlich
```

The 'Attributes' panel on the right shows a table with the following structure:

Attribut	Wert

A message box in the foreground displays the text: 'Analyse komplett: Keine Probleme gefunden.'

Überblick

- XML-Grundlagen: Was ist XML und wie geht das?
- Das XML-Dokument und seine Regeln
- XML schreiben - erste Fingerübung.
- XML – und wie weiter?
- Zusammenfassung



XML schreiben: Hilfsmittel

- einfache Textprogramme (MS Editor, TextPad, BBedit)
- XML-Editoren
 - erleichtern die Arbeit mit XML, z.B.
 - XML Notepad <http://www.microsoft.com/en-us/download/details.aspx?id=7973>
 - XMLSpy <http://www.altova.com/de/xmlspy.html>
 - oXygen <http://oxygenxml.com/>
 - jEdit <http://www.jedit.org/>
- In meiner Lieblings-Textverarbeitung und anschließender Konversion mittels Makros, Skripte oder Programmen

XML selber schreiben

- *Ein* Beispiel: eine Postkarte



Graz, Stadtpfarrkirche und Herrengasse, Verlag Josef Kienreich, Graz, gelaufen von Graz nach Wien, ca. 1906–1908, mit Autotypie kombinierter Lichtdruck, Graz Museum



Rückseite Graz, Stadtpfarrkirche und Herrengasse

Warum Oxygen

- plattformunabhängig
- als Eclipse plug-in erhältlich
- subversion client für kollaboratives Arbeiten
- TEI Unterstützung
- Unterstützung aller Schemasprachen
- Syntaxvervollständigung
- Tool-tip Dokumentation
- XSLT und FOP Unterstützung für Transformationen nach HTML, PDF...



Integrierte Dokumentation

```
<div type="entry" xml:id="e1950-02-13">  
  <dateline><handShift new=""/>Montag, <date when-iso="1950-02-13">13.  
    Feber 1950</date>:  
  <p>Zeitig auf, wenig intere  
    bei <lb/>mir angestellt.</p>
```

indicates a handNote element describing the hand concerned.

nkauf, Ordnu

Integrierte Dokumentation

```
<div type="entry" xml:id="e1950-02-13">
```

```
<dateline><handShift new="" />Montag, <date when-iso="1950-02-13">13.
```

```
Feber 1950</date>:</
```

```
<p>Zeitig auf, wenig interes
```

```
bei <lb/>mir angestellt.
```

```
<p>Um 10 Uhr auf die Uni gef
```

```
<p>Das Wetter war lind, wech
```

```
<p>Zur Ebertprüfung angemeld
```

```
<lb/>jetzt!/, Platz ange
```

```
Formalitäten. <lb/>Nm. m
```

```
<lb/>Seiten Ebert gelernt
```

```
◇ #e1950-02-24
```

```
◇ #e1950-02-25
```

```
◇ #e1950-02-26
```

```
◇ #e1950-02-27
```

```
◇ #e1950-02-28
```

```
◇ #handwritten
```

```
◇ #stenographed
```

```
◇ #typewritten
```

Vorschau:

Von Andreas Okopenko mit einer Schreibmaschine geschriebener Text....

grosse Prüfung für lange Zei
Dekanat um die üblichen
sten sehcs<!-- wie Eigenkorre
<lb/>schon durchs Ueberschrei

medium

rend

rendition

a Attrib

Eleme

exercise3-complete.xml [D:\data\work\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml] - <oxygen/> XML Editor

File Edit Find Project Perspective Options Tools Document Window Help

XPath 2.0

Project

- sample.xpr
 - css
 - debugger
 - fo
 - import
 - jsp
 - nvdI
 - relaxng
 - schematron
 - svg
 - wsdl
 - xquery
 - dita
 - docbook
 - ooxml
 - tei
 - xhtml
 - personal.schema.css

Outline

- TEI "http://www.tei-c.org/ns/1.0"
 - teiHeader
 - text
 - body
 - div "verse" PROGRESS.
 - head PROGRESS.
 - p ["Giving evidence recently before a Select Committee of the House of Commons, Miss C. E. Collet, of the Home Office, said the commercial laundry was killing the small hand laundry."--Evening News.]
 - lg
 - The little crafts! How soon they die!
 - In cottage doors no shuttle clicks;
 - The hand-loom has been ousted by
 - A large concern with lots more sticks.
 - lg
 - The throb of pistons beats around;
 - Great chimneys rise on Thames's banks;
 - The same phenomena are found
 - In Sheffield. (Yorks) and Oldham (Lancs).
 - lg
 - 'Tis so with what w
 - 'Tis this that chief
 - That not with tend

```

1 <?xml version="1.0" encoding="UTF-8" standalone="no" type="text/xml"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>Demonstration exercise for Punch</title>
7       </titleStmt>
8     <publicationStmt>
9       <p>Unpublished</p>
10    </publicationStmt>
11   <sourceDesc>
12     <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
13   </sourceDesc>
14 </fileDesc>
15 </teiHeader>
16 <text>
17   <body>
18     <div type="verse">
19       <head> PROGRESS.</head>
20       <p>["Giving evidence recently before a Select Committee of the House of Commons,
21         Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
22         the small hand laundry."--Evening News.]</p>
23       <lg>
24         <l>The little crafts! How soon they die!</l>
25         <l>In cottage doors no shuttle clicks;</l>
26         <l>The hand-loom has been ousted by</l>
27         <l>A large concern with lots more sticks.</l>
28       </lg>
29       <lg>
30         <l>The throb of pistons beats around;</l>
31         <l>Great chimneys rise on Thames's banks;</l>
32         <l>The same phenomena are found</l>
33         <l>In Sheffield. (Yorks) and Oldham (Lancs).</l>
34       </lg>
35       <lg>
36         <l>'Tis so with what w</l>
37         <l>'Tis this that chief</l>
38         <l>That not with tend</l>

```

So sieht das Dokument im Oxygen aus.

Text Grid Author

D:\data\work\esu\vorbilder_fremd\oxford_2009\exercise3-complete.xml | U+003C | 1:1

Project

- sample.xpr
 - css
 - debugger
 - fo
 - import
 - jsp
 - nvdI
 - relaxng
 - schematron
 - svg
 - wsdl
 - xquery
 - dita
 - docbook
 - ooxml
 - tei
 - xhtml
 - personal-schema.css

```

1 <?xml version="1.0" encoding="UTF-8" standalone="no" ?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3 <teiHeader>
4 <fileDesc>
5 <titleStm>
6 <title>Demonstration exercise for Punch</title>
7 </titleStm>
8 <publicationStm>
9 <p>Unpublished</p>
10 </publicationStm>
11 <sourceDesc>
12 <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
13 </sourceDesc>
14 </fileDesc>
15 </teiHeader>
16 <text>
17 <body>
18 <div type="verse">
19 <head> PROGRESS.</head>
20 <p>["Giving evidence recently before a Select Committee of the House of Commons,
21 Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
22 the small hand laundry."--Evening News.]</p>
23 <lg>
24 <l>The little crafts! How soon they die!</l>
25 <l>In cottage doors no shuttle clicks;</l>
26 <l>The hand-loom has been ousted by</l>
27 <l>A large concern with lots more sticks.</l>
28 </lg>
29 <lg>
30 <l>The throb of pistons beats around;</l>
31 <l>Great chimneys rise on Thames's banks;</l>
32 <l>The same phenomena are found</l>
33 <l>In Sheffield. (Yorks) and Oldham (Lancs).</l>
34 </lg>
35 <lg>
36 <l>No longer now the housewife makes</l>
37 <l>Her rare preserves, for what's the good?</l>

```

Prüfung auf Wohlgeformtheit

Outline

- TEI "http://www.tei-c.org/ns/1.0"
 - teiHeader
 - text
 - body
 - div "verse" PROGRESS.
 - head PROGRESS.
 - p ["Giving evidence r...
 - lg The little crafts! H...
 - lg The throb of pistor...
 - lg No longer now the...
 - lg 'Tis so with what w...
 - lg 'Tis this that chief...
 - lg That not with tend...

Project

- sample.xpr
 - css
 - debugger
 - fo
 - import
 - jsp
 - nvdI
 - relaxng
 - schematron
 - svg
 - wSDL
 - xquery
 - dita
 - docbook
 - ooxml
 - tei
 - xhtml
 - personal-schema.css

```
1 <?xml version="1.0" encoding="UTF-8" standalone="no" type="text/xml"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3 <teiHeader>
4 <fileDesc>
5 <titleStm>
6 <title>Demonstration exercise for Punch</title>
7 </titleStm>
8 <publicationStm>
9 <p>Unpublished</p>
10 </publicationStm>
11 <sourceDesc>
12 <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
13 </sourceDesc>
14 </fileDesc>
15 </teiHeader>
16 <text>
17 <body>
18 <div type="verse">
19 <head> PROGRESS.</head>
20 <p>["Giving evidence recently before a Select Committee of the House of Commons,
21 Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
22 the small hand laundry."--Evening News.]</p>
23 <lg>
24 <l>The little crafts! How soon they die!</l>
25 <l>In cottage doors no shuttle clicks;</l>
26 <l>The hand-loom has been ousted by</l>
27 <l>A large concern with lots more sticks.</l>
28 </lg>
29 <lg>
30 <l> The throb of pistons beats around;</l>
31 <l>Great chimneys rise on Thames's banks;</l>
32 <l>The same phenomena are found</l>
33 <l>In Sheffield. (Yorks) and Oldham (Lancs).</l>
34 </lg>
35 <lg>
36 <l>No longer now the housewife makes</l>
37 <l>Her rare preserves, for what's the good?</l>
```

Prüfung auf Gültigkeit

Outline

- TEI "http://www.tei-c.org/ns/1.0"
 - teiHeader
 - text
 - body
 - div "verse" PROGRESS.
 - head PROGRESS.
 - p ["Giving evidence r
 - lg The little crafts! H
 - lg The throb of pistor
 - lg No longer now the
 - lg 'Tis so with what w
 - lg 'Tis this that chief
 - lg That not with tend

Project

- sample.xpr
 - css
 - debugger
 - fo
 - import
 - jsp
 - nvdI
 - relaxng
 - schematron
 - svg
 - wSDL
 - xquery
 - dita
 - docbook
 - ooxml
 - tei
 - xhtml
 - personal-schema.css

```
1 <?xml version="1.0" encoding="UTF-8" standalone="no" type="text/xml"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3 <teiHeader>
4 <fileDesc>
5 <title>Demonstration exercise for Punch</title>
6 </title>
7 <publication>
8 <p>Unpublished</p>
9 </publication>
10 <sourceDesc>
11 <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
12 </sourceDesc>
13 </fileDesc>
14 </teiHeader>
15 <text>
16 <body>
17 <div type="verse">
18 <head>PROGRESS.</head>
19 <p>["Giving evidence recently before a Select Committee of the House of Commons,
20 Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
21 the small hand laundry."--Evening News.]</p>
22 <lg>
23 <l>The little crafts! How soon they die!</l>
24 <l>In cottage doors no shuttle clicks;</l>
25 <l>The hand-loom has been ousted by</l>
26 <l>A large concern with lots more sticks.</l>
27 </lg>
28 <lg>
29 <l>The throb of pistons beats around;</l>
30 <l>Great chimneys rise on Thames's banks;</l>
31 <l>The same phenomena are found</l>
32 <l>In Sheffield. (Yorks) and Oldham (Lancs).</l>
33 </lg>
34 <lg>
35 <l>'Tis so with what we see</l>
36 <l>'Tis this that chiefly grieves</l>
37 <l>That not with tenderness</l>
38 </lg>
39 </body>
40 </text>
41 </TEI>
```

Dokument ist valide

Outline

- TEI "http://www.tei-c.org/ns/1.0"
 - teiHeader
 - text
 - body
 - div "verse" PROGRESS.
 - head PROGRESS.
 - p ["Giving evidence r
 - lg The little crafts! H
 - lg The throb of pistor
 - lg No longer now the
 - lg 'Tis so with what w
 - lg 'Tis this that chief
 - lg That not with tend

Project

- sample.xpr
 - css
 - debugger
 - fo
 - import
 - jsp
 - nvd1
 - relaxng
 - schematron
 - svg
 - wsdl
 - xquery
 - dita
 - docbook
 - ooxml
 - tei
 - xhtml

```

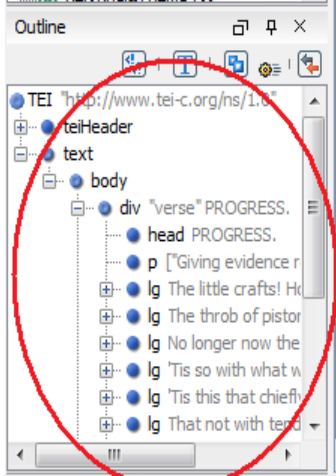
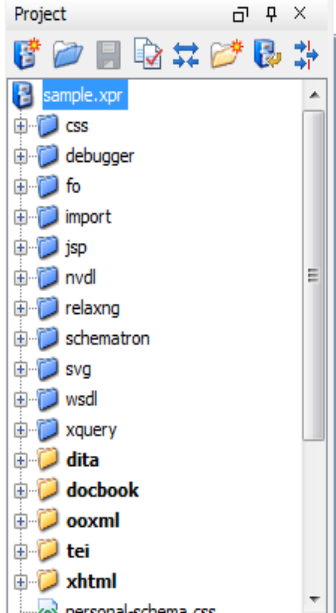
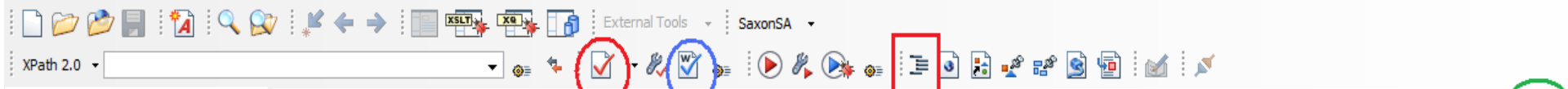
1 <?xml version="1.0" encoding="UTF-8" standalone="no" type="text/xml"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStm>
6         <title>Demonstration exercise for Punch</title>
7       </titleStm>
8     <publicationStm>
9       <p>Unpublished</p>
10    </publicationStm>
11    <sourceDesc>
12      <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
13    </sourceDesc>
14  </fileDesc>
15 </teiHeader>
16 <text>
17   <body>
18     <div type="verse">
19       <head> PROGRESS.</head>
20       <p>["Giving evidence recently before a Select Committee of the House of Commons,
21         Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
22         the small hand laundry."--Evening News.]</p>
23       <lg>
24         <l>The little crafts! How soon they die!</l>
25         <l>In cottage doors no shuttle clicks;</l>
26         <l>The hand-loom has been ousted by</l>
27         <l>A large concern with lots more sticks.</l>
28       </lg>
29       <lg>
30         <l>The throb of pistons beats around;</l>
31         <l>Great chimneys rise on Thames's banks;</l>
32         <l>The same phenomena are found</l>
33         <l>In Sheffield. (Yorks) and Oldham (Lancs).</l>
34       </lg>
35       <lg>
36         <l>No longer now the housewife makes</l>
37         <l>Her rare preserves, for what's the good?</l>

```

Dokumentenstruktur in Kurzform

Outline

- TEI "http://www.tei-c.org/ns/1.0"
 - teiHeader
 - fileDesc
 - titleStm
 - title
 - publicationStm
 - p
 - sourceDesc
 - p
 - text
 - body
 - div "verse"
 - head
 - PROGRESS.
 - p ["Giving evidence r..."]
 - lg
 - The little crafts! H...
 - The throb of pistor...
 - No longer now the...
 - 'Tis so with what w...
 - 'Tis this that chief...
 - That not with tend...



```

1 <?xml version="1.0" encoding="UTF-8" standalone="no" type="text/xml"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3 <teiHeader>
4 <fileDesc>
5 <titleStm>
6 <title>Demonstration exercise for Punch</title>
7 </titleStm>
8 <publicationStm>
9 <p>Unpublished</p>
10 </publicationStm>
11 <sourceDesc>
12 <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
13 </sourceDesc>
14 </fileDesc>
15 </teiHeader>
16 <text>
17 <body>
18 <div type="verse">
19 <head> PROGRESS.</head>
20 <p>[Giving evidence recently before a Select Committee of the House of Commons,
21 Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
22 the small hand laundry.]</p>
23 <lg>
24 <l>The little crafts! How soon they die! </l>
25 <l>In cottage doors no shuttle clicks; </l>
26 <l>The hand-loom has been ousted by </l>
27 <l>A large concern with lots more sticks.</l>
28 </lg>
29 <lg>
30 <l> The throb of pistons beats around; </l>
31 <l>Great chimneys rise on Thames's banks; </l>
32 <l>The same phenomena are found </l>
33 <l>In Sheffield. (Yorks) and Oldham (Lancs).</l>
34 </lg>
35 <lg>
36 <l>No longer now the housewife makes </l>
37 <l>Her rare preserves, for what's the good? </l>

```

Dokument hübsch machen [STR + Y + Shift]

Project

- sample.xpr
 - css
 - debugger
 - fo
 - import
 - jsp
 - nvdI
 - relaxng
 - schematron
 - svg
 - wsdl
 - xquery
 - dita
 - docbook
 - ooxml
 - tei
 - xhtml
 - personal-schema.css

```
1 <?xml version="1.0" encoding="UTF-8" standalone="no" ?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStm
6         <title>Demonstration exercise for Punch</title>
7       </titleStm>
8       <publicationStm
9         <p>Unpublished</p>
10      </publicationStm>
11      <sourceDesc>
12        <p>Material from the Gutenberg text of Punch for 1 July 1914</p>
13      </sourceDesc>
14    </fileDesc>
15  </teiHeader>
16  <text>
17    <body>
18      <div type="verse">
19        <head> PROGRESS.</head>
20        <p>["Giving evidence recently
21          Miss C. E. Collet, of the Home Office, said the commercial laundry was killing
22          the small hand laundry."—Evening News.]</p>
23        <lg>
24          <l>The little crafts! How soon they die! </l>
25          <l>In cottage doors no shuttle clicks; </l>
26          <l>The hand-loom has been ousted by </l>
27          <l>A large concern with lots more sticks.</l>
28        </lg>
29        <lg>
30          <l> The throb of pistons beats around; </l>
31          <l>Great chimneys rise on Thames's banks; </l>
32          <l>The same phenomena are found </l>
33          <l>In Sheffield. (Yorks) and Oldham (Lancs).</l>
34        </lg>
35        <lg>
36          <l>No longer now the housewife makes </l>
37          <l>Her rare preserves, for what's the good? </l>
```

Gerade störende Zeilen einklappen und wieder ausklappen

Outline

- TEI "http://www.tei-c.org/ns/1.0"
 - teiHeader
 - text
 - body
 - div "verse" PROGRESS.
 - head PROGRESS.
 - p ["Giving evidence recently
 - lg The little crafts! How soon they die! </l>
 - lg The throb of pistons beats around; </l>
 - lg No longer now the housewife makes </l>
 - lg 'Tis so with what w
 - lg 'Tis this that chief
 - lg That not with tend

XML schreiben – im XML-Editor



- Starten Sie das Programm Oxygen.
- Legen Sie eine neue XML-Datei an.
 - Strg+N | Menü Datei → Neue Datei
- Transkribieren Sie die Postkarte.
- Aufgabe:
 - Versuchen Sie, Ihr Wissen über das Dokument festzuhalten
 - Textbestand?
 - Strukturen?
 - Zusätzliches Wissen?
 - Welche Informationen werden gebraucht, um später eine Edition daraus zu machen?
- Lassen Sie sich helfen: rote Unterkringelungen?



XML selber schreiben

The screenshot shows the XML Editor (oXygen) interface. The main editor window displays the following XML code:

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <postkarte>
3   Lieber Paul.
4   Habe auf meinen
5   damaligen Brief
6   keine Antwort
7   bekommen. Hoffentlich
8   Geht es dir gut.
9   Fröhliche Weihnachten
10  wünscht Dier Dein Bruder
11  KarlFried. Busson Graz
12  Kaiser Franz Joseph Kaserne. Einjährig Freiwilliger.
13 </postkarte>
```

The left sidebar shows a project tree with the following structure:

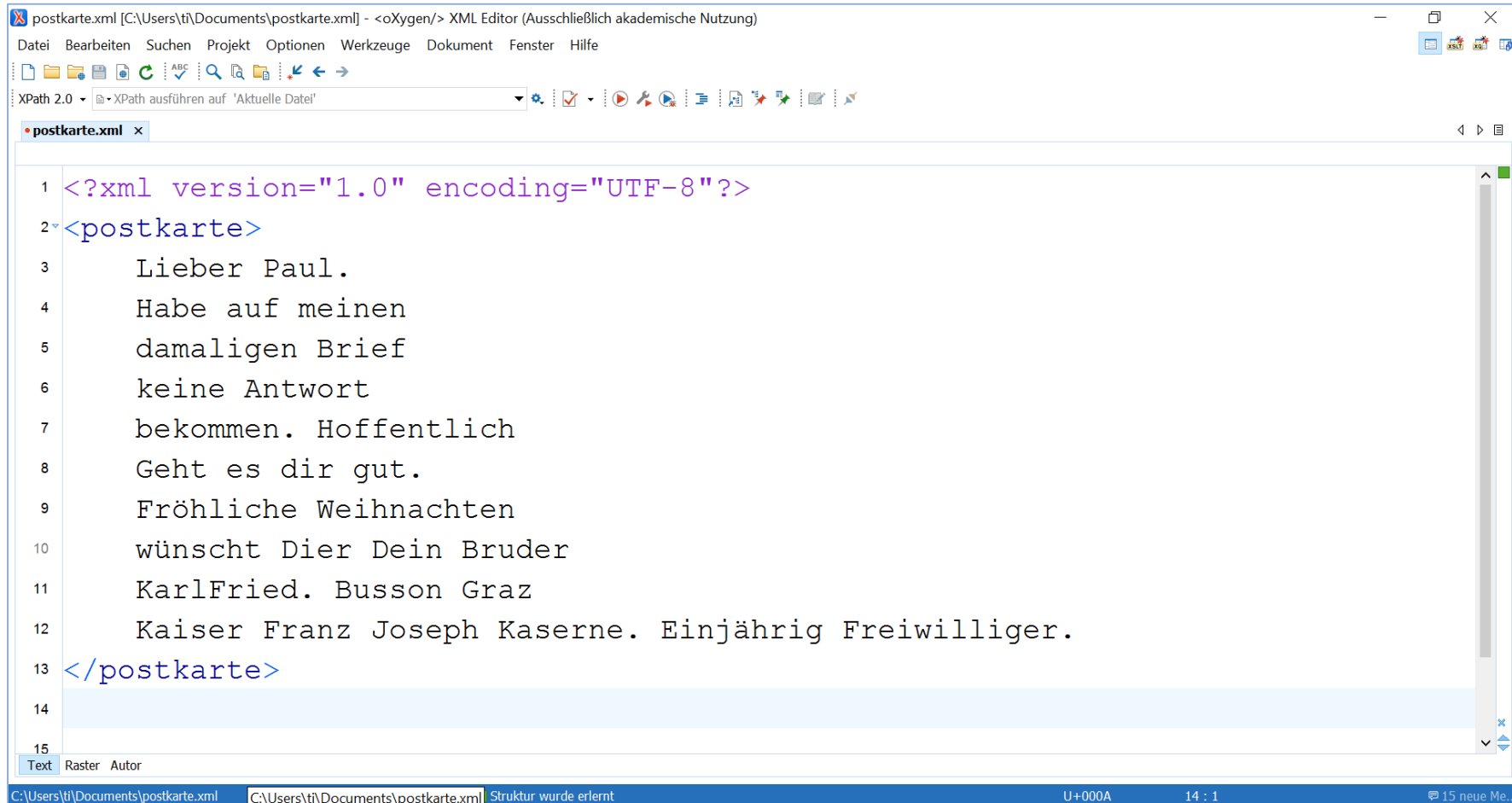
- sample.xpr
 - sample.xpr
 - css
 - debugger
 - epub
 - fo
 - import
 - json
 - jsp

The right sidebar contains several panels:

- Attribute**: A table with columns 'Attribut' and 'Wert'.
- Transformation...**: A panel for transformation settings.
- Zuordnu... Umwandlung**: A panel for mapping and conversion settings.

The status bar at the bottom indicates the file path 'C:\Users\...\Documents\postkarte.xml', the status 'Struktur wurde erlernt', the encoding 'U+000A', the page number '15 : 1', and the date 'Sonntag, 4. September 2016'.

XML selber schreiben



The screenshot shows the XML Editor interface with a file named 'postkarte.xml' open. The editor displays the following XML code:

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <postkarte>
3     Lieber Paul.
4     Habe auf meinen
5     damaligen Brief
6     keine Antwort
7     bekommen. Hoffentlich
8     Geht es dir gut.
9     Fröhliche Weihnachten
10    wünscht Dier Dein Bruder
11    KarlFried. Busson Graz
12    Kaiser Franz Joseph Kaserne. Einjährig Freiwilliger.
13 </postkarte>
```

The status bar at the bottom indicates the current position in the document: 'U+000A 14 : 1' and '15 neue Me.'.

Ihre Ideen zur Kodierung

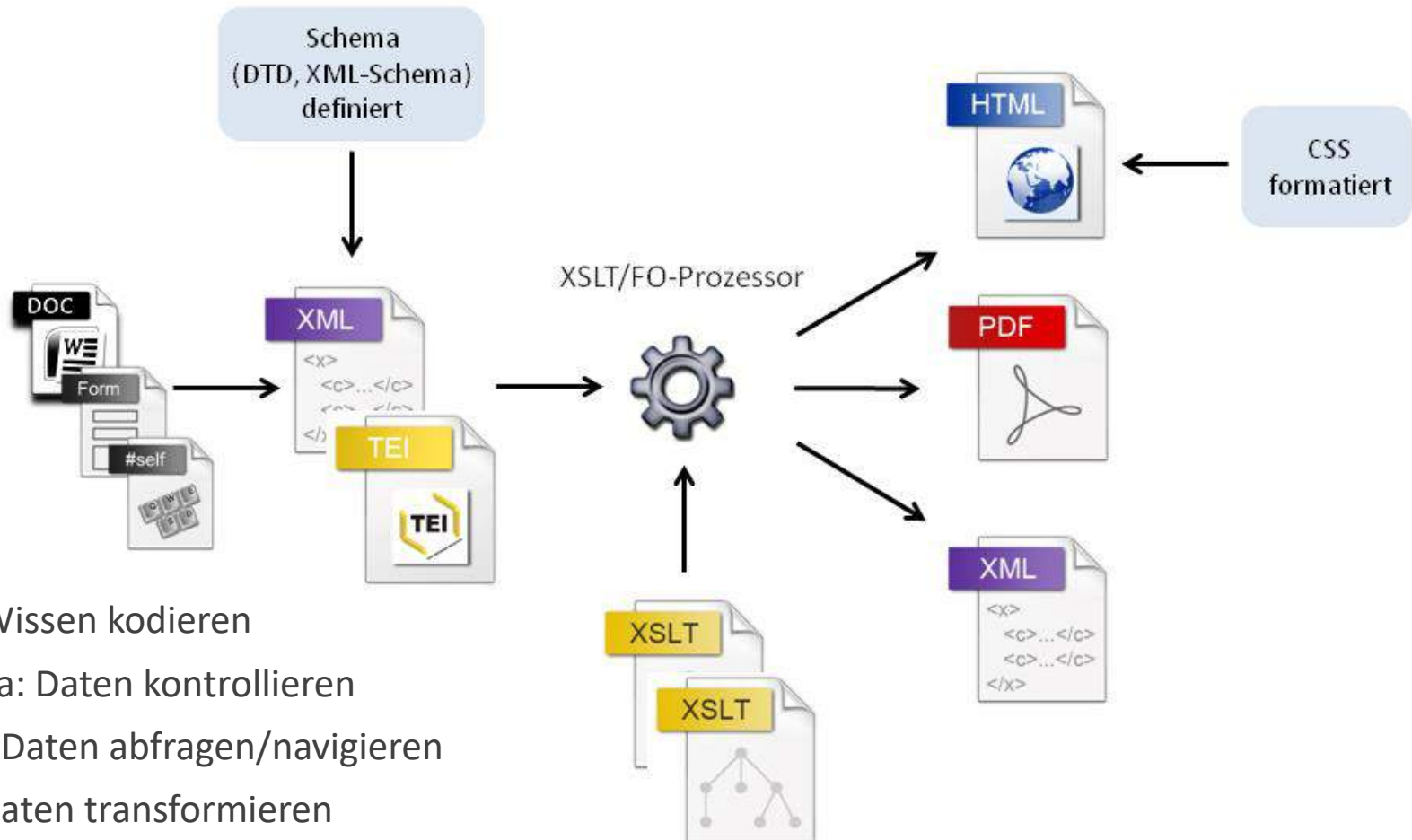
- Was haben Sie ausgezeichnet?
- Element oder Attribut?

Überblick

- XML-Grundlagen: Was ist XML und wie geht das?
- Das XML-Dokument und seine Regeln
- XML schreiben - erste Fingerübung.
- XML – und wie weiter?
- Zusammenfassung



XML-basierte Dokumentverarbeitung



XML: Wissen kodieren

Schema: Daten kontrollieren

XPath: Daten abfragen/navigieren

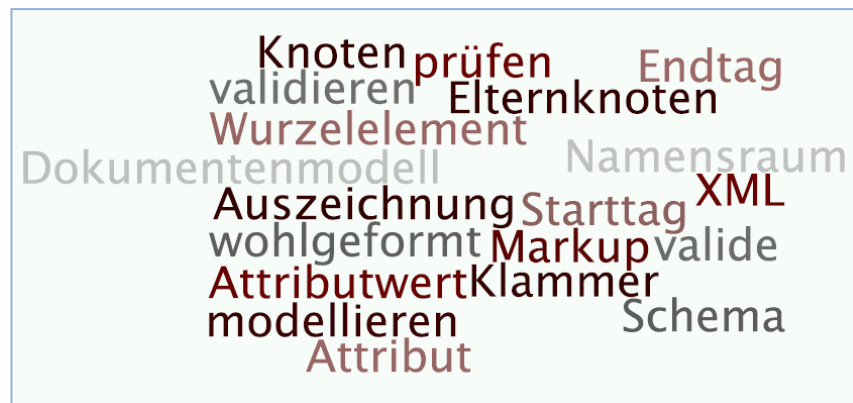
XSLT: Daten transformieren

XML-Familie

- XML ... beschreibt strukturierte Daten
- XPath ... erlaubt die Navigation in XML-Daten
- XML Schema ... beschreibt ein striktes Datenmodell
- XSL ... eXtensible Style Language
 - XSLT ... transformiert XML-Dokumente
 - XSL-FO ... beschreibt eine formatierte Ausgabe (z.B. für den Druck)
- XQuery ... ist eine XML-Datenbankabfragesprache
- XInclude ... lädt XML-Dokumente ineinander
- XLink ... beschreibt komplexe Links
- XPointer ... ermöglicht komplexe Referenzierung
- XForms ... beschreibt Eingabeformulare

Überblick

- XML-Grundlagen: Was ist XML und wie geht das?
- Das XML-Dokument und seine Regeln
- XML schreiben - erste Fingerübung.
- XML – und wie weiter?
- Zusammenfassung



Kurz: XML ist...

- XML ist einfach.
- XML ist wie HTML, nur anders:
 - XML ist erweiterbar.
 - XML muss wohlgeformt sein.
 - XML kann validiert werden.
- besonders gut geeignet, um Texte zu strukturieren.

Alles über ein XML-Dokument

- Ein XML-Dokument kann enthalten:
 - Elemente, eventuell mit Attributen
 - Verarbeitungsanweisungen
 - Kommentare
 - Entitätsreferenzen
- Ein XML-Dokument muss wohlgeformt sein und kann validiert werden.
- XML Attributwerte müssen in " " stehen.
- XML-Dokumente sind als lineare Zeichenketten kodiert.
- XML-Dokumente beginnen mit einer besonderen Verarbeitungsanweisung.

Anhang

- Literatur
- Gedanken zu XML
- Benutzung des Oxygen-Editors



Literatur zu XML

- Vonhoegen, Helmut, *Einstieg in XML. Grundlagen, Praxis, Referenz*, Galileo Press, Bonn 2009⁵.
- St. Laurent, Simon/Fitzgerald, Michael, *XML. kurz & gut*, O'Reilly, Köln 2006³.
- W3 Schools Tutorial: <http://www.w3schools.com/xml/> (XML),
- W3C Spezifikation: <http://www.w3.org/TR/xml/>

XML - was die Leute sagen...

- XML ist ein Akronym
- XML steht für eXtensible Markup Language
- XML ist "plain text"
- XML ist einfach
- XML ist wie HTML, nur anders
- XML ist eine Auszeichnungssprache
- XML sagt: "invent your own tags"
- XML ist eine Metasprache
- XML ist eine Grammatik
- XML ist eine Datenstruktur
(ein Datenformat? Ein Datenmodell?)
- XML beschreibt strukturierte Informationen
- XML ist ein Baum
- XML ist ein serialisierter Graph
- XML ist eine hierarchisierte Sequenz
- XML ist ein (Daten-?)Standard
- XML ist ein Standard des W3C
- XML ist eine Untermenge von SGML
- XML ist plattformunabhängig
- XML unterstützt
- XML wird von einer breiten Softwarepalette ist zukunftssicher
- XML ist flexibel
- XML kann ein Austauschformat sein
- XML ist eine Notation
- XML ist dokumentnah
- XML ist die Grundlage von Online-Ressourcen
- xHTML ist XML
- XML ist mächtig, weil es abstrahiert und weil es viele "Freunde" hat
- XML "is everywhere"