



# Textauszeichnung mit XML

Ulrike Henny  
&  
Patrick Sahle



# Fahrplan

- XML: wieso – weshalb – warum?
- XML als Markup
- XML essentiell
- XML schreiben – am Beispiel
- XML schreiben – im Editor
- XML, die Regeln
- XML, die Philosophie

# XML, wieso, weshalb, warum

- XML = eXtensible Markup Language
- Standard des W3C seit 1998
- Metasprache zur Beschreibung (komplexer) strukturierter Daten
- Die Grundlage von vielen Dingen
  - Einzelne Datenformate
    - Z.B. XHTML, TEI, EAD, METS/MODS, SVG etc.
  - Datenaustausch
  - Verschiedene Nutzungen
    - Webpräsentation
    - Visualisierungen
    - Druckausgaben
    - Etc.
- Menschen- und maschinenlesbar
- einfach



# XML essentiell

- Alles ist Text, Zeichendaten
- Markup bedeutet, Text mit Auszeichnungen zu versehen ...
- Es gibt (Auszeichnungs-)Elemente
  - Öffnendes Tag ... Text ... Schließendes Tag
  - `<elementname>Elementinhalt</elementname>`
- Elemente können Attribute haben
  - `<elementname attributname="attributwert">Elementinhalt</elementname>`
- Elemente enthalten Elemente, Text, beides - oder nichts
- Elemente können geschachtelt sein; es gibt *ein* Wurzelement



# XML, ein Beispiel

```
<?xml version="1.0" encoding="UTF-8"?>  
  <notiz>  
    nicht anfangen zu labern,  
    <person normnummer="pnd:135774314">Patrick</person>  
  </notiz>
```

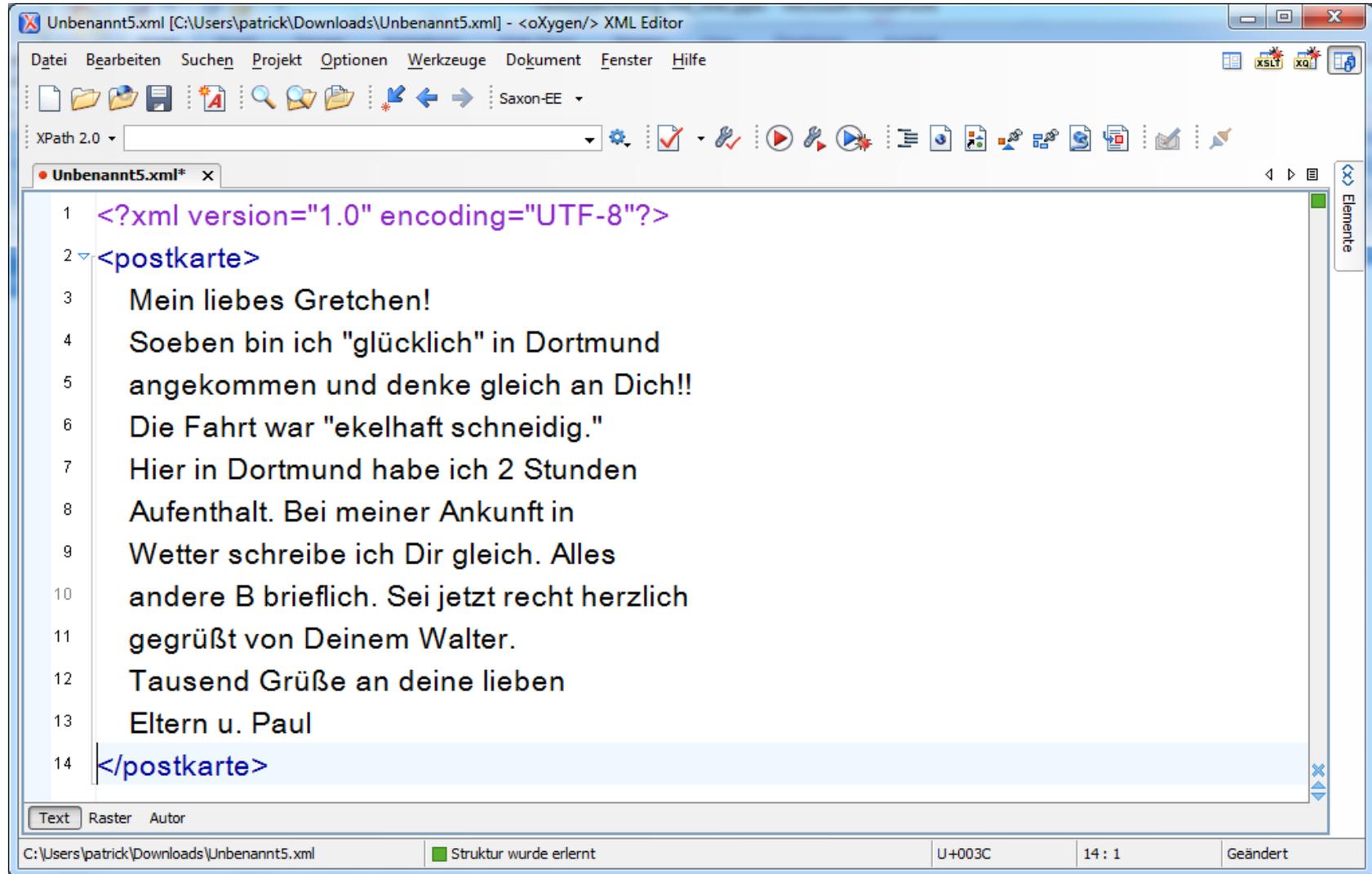
# XML selber schreiben – am Beispiel

- ein Beispiel: eine Postkarte
- andere Teilnehmerbeispiele?



# XML schreiben – im Editor

- ein* Editor: oXygen
- Eine XML-Datei anlegen
  - Strg+N | Menü Datei.Neue Datei | 
- Ein Wurzelement; Transkription ggf. Per copy&paste
- 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?



The screenshot shows the XML Editor interface with the following content:

```

1  <?xml version="1.0" encoding="UTF-8"?>
2  <postkarte>
3      Mein liebes Gretchen!
4      Soeben bin ich "glücklich" in Dortmund
5      angekommen und denke gleich an Dich!!
6      Die Fahrt war "ekelhaft schneidig."
7      Hier in Dortmund habe ich 2 Stunden
8      Aufenthalt. Bei meiner Ankunft in
9      Wetter schreibe ich Dir gleich. Alles
10     andere B brieflich. Sei jetzt recht herzlich
11     begrüßt von Deinem Walter.
12     Tausend Grüße an deine lieben
13     Eltern u. Paul
14 </postkarte>
  
```

The status bar at the bottom indicates: C:\Users\patrick\Downloads\Unbenannt5.xml, Struktur wurde erlernt, U+003C, 14 : 1, Geändert.



# XML, Regeln

- Zeichen*
- Aufbau einer XML-Datei*
- Schachtelung*
- Syntax / Regeln für Namen*
- Attribute*
- Wohlgeformtheit / Validität*



# XML, Regeln, Zeichen

- Alle Zeichen (z.B. aus Unicode) sind erlaubt ...*
- außer ...*
  - < → &lt;*
  - > → &gt;*
  - & → &amp;*
- „whitespace“ ... spielt keine Rolle*

# XML, Regeln, Aufbau einer XML-Datei

- *Bestandteile einer XML-Datei*
  - *XML-Deklaration:* `<?xml version="1.0" encoding="UTF-8"?>`
  - *Processing instructions:* `<?xml-stylesheet type="text/xsl" href="abc.xsl"?>`
  - *Elemente, geschachtelt:* `<root> ... </root>`
  
- *Kommentare*
  - `<!-- hier steht ein Kommentar -->`



# XML, Regeln, Syntax, Elemente

- *Elementnamen unterscheiden Groß- und Kleinschreibung*
  - *<Postkarte> ≠ <postkarte>*
- *Elementname dürfen nicht mit Zahlen anfangen*
- *Elementnamen dürfen nicht mit „xml“ anfangen*
  
- *Elemente können leer sein: <zeilenumbruch/>*
- *Auch leere Elemente können Attribute haben*  
*<zeilenumbruch nummer="24"/>*



# XML, Regeln, Attribute

- ein Element darf nicht zweimal das gleiche Attribut haben*
- Attributwerte müssen in Anführungszeichen stehen*
- Attributwerte dürfen alle möglichen Zeichen enthalten*

# XML, Begriffe

- *„wohlgeformt“*
  - *... nennt man ein XML-Dokument, das die Regeln von XML einhält*
  - *... alle Syntaxregeln eingehalten, alle Elemente geschlossen, keine Überlappungen*
  - *... wohlgeformte Dokumente können geparkt und verarbeitet werden*
  
- *„gültig“ / „valide“*
  - *...nennt man ein XML-Dokument, das den Regeln eines assoziierten Schemas entspricht*

# XML, Philosophie

- Eigene Erfahrungen?
- Unsere Beispiellösung
  
- Explizit-Machen von Wissen
- Elemente kreieren
- Text vs. Markup
- Beschreibung vs. Deutung
- „Metadaten“ / Head-Body
- Elemente vs. Attribute
- Textzentrierung vs. Dokumentorientierung
- Schachtelung, Hierarchie, primäre Hierarchie, leere Elemente
- XML als Metasprache vs. Konkrete Sprachen
- Lokale Dialekte vs. Gemeinsame Sprachen



# 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 wird von einer breiten Softwarepalette unterstützt
- XML 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"



# XML, weiterführende Hinweise

- Ein Tutorial: <http://www.w3schools.com/xml/default.asp>