

Coding the Sources

Digitales Edieren in den Geisteswissenschaften

Online Summerschool und Workshop

Lübeck/Online, 17.08. – 27.08.2020



Forschungsstelle
für die Geschichte
der Hanse und des Ostseeraums



XML-Einführung

XML und oXygen Vertiefung

Roman Bleier
roman.bleier@uni-graz.at
Lübeck/Online, 18.08.2020



Forschungsstelle
für die Geschichte
der Hanse und des Ostseeraums



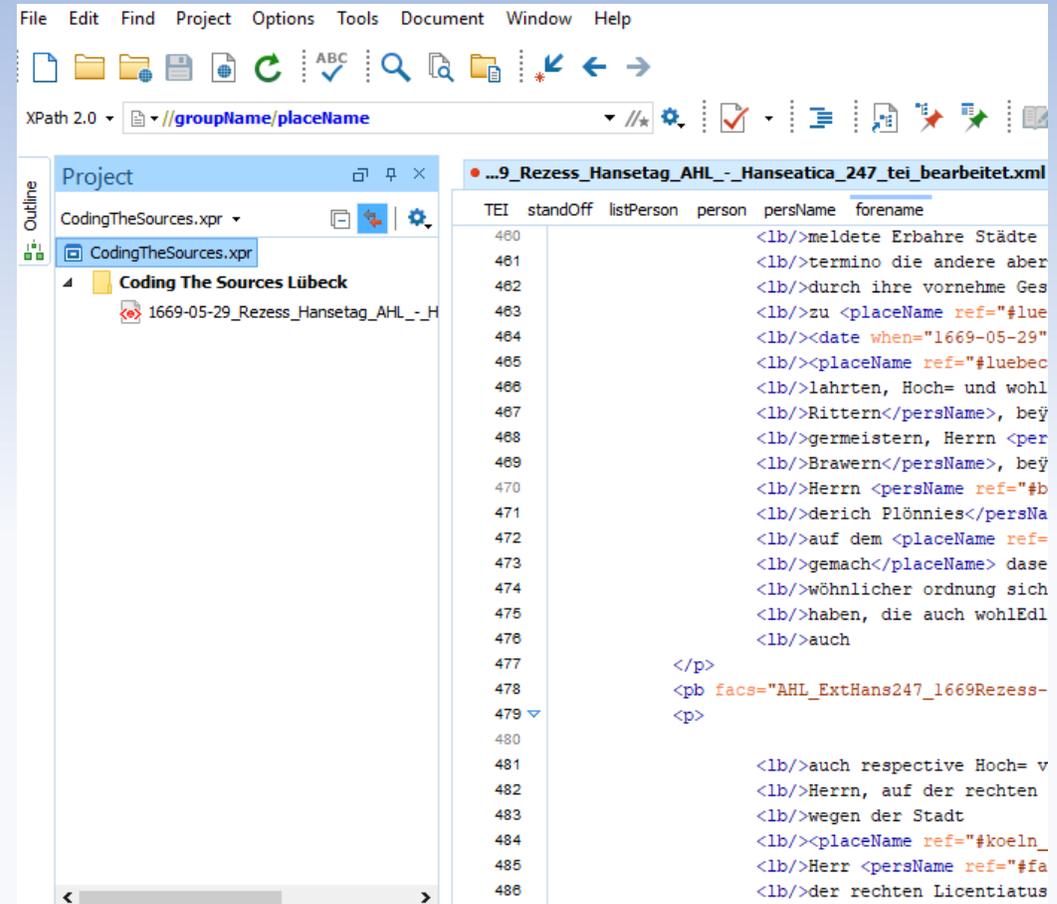
Übersicht

- oXygen Project View
- Wohlgeformtes vs valides XML
- RELAX NG

Danke an meine KollegInnen des IDE, deren Folien ich als Basis für meinen Vortrag verwendet habe.

oxygen Project View

- Organisation und Management von XML Dateien
- Zu einem Projekt gehörende Dateien können virtuell zusammengeführt werden



Wohlgeformtes vs valides XML

XML Dokumente müssen wohlgeformt und valide sein !!!

- Wohlgeformt: das XML entspricht den Regeln für XML Dokumente (Syntax)
 - Ein Wurzelement
 - Element- und Attributnamen entsprechen den festgesetzten Regeln
 - Elemente dürfen nicht überlappen
 - Anfangs- und Endtag müssen übereinstimmen
 - Groß- und Kleinschreibung

Wohlgeformtes vs valides XML

XML Dokumente müssen wohlgeformt und valide sein !!!

- Valide: dem XML ist ein DTD oder XML Schema zugeordnet gegen das es validiert werden kann
 - Ein Schema beschreibt die Struktur eines XML Dokuments
 - Die Namen für Elemente und Attribute werden definiert
 - Inhaltsmodelle für Elemente werden beschrieben: kein Inhalt (leeres Element), Text, gemischter Inhalt (Text und Elemente), nur Elemente
 - Attribute und Attributwerte werden definiert

RELAX NG (REgular LAnguage for XML Next Generation)

RELAX NG

```
<Adressbuch>  
  <Eintrag>  
    <Name>Sherlock Holmes</Name>  
    <Adresse>221B Baker Street London</Adresse>  
  </Eintrag>  
  <Eintrag>  
    <Name>Bruce Wayne</Name>  
    <Adresse>Wayne Manor 1007 Mountain Drive  
    Gotham</Adresse>  
  </Eintrag>  
</Adressbuch>
```

RELAX NG

- In dem Beispiel werden definiert:
- Das Wurzelement <Adressbuch>
- Das Kindelement <Eintrag>
- Die Kindelemente (von Eintrag) <Name> und <Adresse>
- Die Inhaltsmodelle (Elemente oder Text)

```
3 <grammar xmlns="http://relaxng.org/ns/structure/1.0">
4   <start>
5     <element name="Adressbuch">
6       <oneOrMore>
7         <element name="Eintrag">
8           <element name="Name">
9             <text></text>
10          </element>
11         <element name="Adresse">
12           <text></text>
13        </element>
14       </element>
15      </oneOrMore>
16     </element>
17   </start>
18 </grammar>
19
```

RELAX NG

RELAX NG Syntax:

- Wurzelement `<grammar>`
- RNG Namensraum `@xmlns`
- Element `<start>` Beginn der Regeln
- Definition des Wurzelements und Kindelemente mit `<element>` und `@name`
- Die Inhaltsmodelle (`<element>` oder `<text>`)

```
3 <grammar xmlns="http://relaxng.org/ns/structure/1.0">
4   <start>
5     <element name="Adressbuch">
6       <oneOrMore>
7         <element name="Eintrag">
8           <element name="Name">
9             <text></text>
10          </element>
11         <element name="Adresse">
12           <text></text>
13        </element>
14       </element>
15      </oneOrMore>
16     </element>
17   </start>
18 </grammar>
19
```

RELAX NG

Definition von Inhaltsmodellen:

1. Einfache Inhalte (nur Text oder nur Elemente)

```
<!-- Einfache Inhaltsmodelle (nur Text) -->
<Name>Lübeckischen Sýndico</Name>

<!-- Einfache Inhaltsmodelle (nur Elemente) -->
<Person>
  <Name>
    <Vorname>Heinrich</Vorname>
    <Nachname>Hüneken</Nachname>
    <Titel>Doktor beider Rechte</Titel>
  </Name>
</Person>
```

2. Gemischte Inhalte

```
<!--Gemischte Inhaltsmodelle (Elemente und Text)-->

<Absatz>...Hanseatici die propositio
  dem <Name>Lübeckischen Sýndico</Name>
  von dem <Name>Löblichen Directorio</Name>...
</Absatz>
```

3. Leere Elemente

```
<!-- Leere Inhaltsmodelle (leere Elemente) -->

<Zeilenumbruch />
```

Wohlgeformtes vs valides XML

Einige RELAX NG Elemente:

- Einfache Inhalte
 - <text>
- Gemischte Inhalte
 - <interleave>
 - <zeroOrMore>
 - <oneOrMore>
- Leere Elemente
 - <empty>

```
<element name="Name"><text></text></element>
```

```
<element name="Name">  
  <interleave>  
    <text/>  
    <zeroOrMore>  
      <ref name="Zeilenumbruch"/>  
    </zeroOrMore>  
  </interleave>  
</element>
```

```
<oneOrMore>  
  <element name="Seitenumbruch">  
    <empty/>  
  </element>  
</oneOrMore>
```

Wohlgeformtes vs valides XML

RELAX NG Definition von Attributen:

- <attribute> mit @name
- Man kann Werte definieren mit <choice> und <value>
- <optional>

```
<element name="Name">
  <optional>
    <attribute name="Typ">
      <choice>
        <value>Person</value>
        <value>Ort</value>
        <value>Organisation</value>
      </choice>
    </attribute>
  </optional>
  <interleave>
    <text/>
    <zeroOrMore>
      <ref name="Zeilenumbruch"/>
    </zeroOrMore>
  </interleave>
</element>
```

Übung für den Nachmittag

- Im Ordner der Einheit auf GDrive: XML Datei mit Transkription, RNG-Datei mit einem Basisschema
- Das Schema in die Transkriptionsdatei einbinden
- Zeichnet die Transkriptionsdatei nach den Schemaregeln aus
 - Seiten-, Zeilenumbrüche, Absätze, Namen mit @Typ, Datum
- Valideriung muss erfolgreich sein!

Für Fortgeschrittene:

- Fügt ein Element <Titel> und <Signatur> dem Schema hinzu
- Fügt dem Element <Datum> ein Attribut @normalisiert hinzu

Weiterführende Literatur:

W3C Schools. XML Tutorial: <https://www.w3schools.com/xml/>
RELAX NG Buch. <http://books.xmlschemata.org/relaxng/page2.html>

Fragen?

roman.bleier@uni-graz.at



Forschungsstelle
für die Geschichte
der Hanse und des Ostseeraums

