

Technik digitaler Editionen mit XML/XSLT

Susanne Haaf

3. Münchner Digital Humanities Summerschool
11.-15. September 2017

XML: Regeln für die Wohlgeformtheit

XML: Regeln

- XML: Extensible Markup Language
- Ein XML Dokument
 - muss wohlgeformt sein hinsichtlich der grundlegenden Regeln für den XML Standard
 - muss valide sein in Bezug auf ein zugehöriges Schema (DTD, RNG, XSD, Schematron)

XML-Regeln

1. Ein XML Dokument beginnt immer mit einer **XML Deklaration**:

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

2. XML ist eine Baumstruktur, die mit einem Wurzelement beginnt (i.e. ein Element, das alle anderen XML-Strukturen des Dokuments umschließt)
3. XML Element-tags sehen so aus:
 - `<tagname>` (Starttag)
 - `</tagname>` (Endetag)
 - `<tagname/>` (leeres Element)

(Zu jedem Starttag gehört ein Endetag.)

This is an **example** containing *highlighted text*. →

This is an `<bold>example</bold>` containing `<italics>highlighted text</italics>`.

XML-Regeln

4. Elemente können ineinander verschachtelt sein ...

This is an **example** containing *highlighted text*. →

This is an **<bold>example** containing
<italics>highlighted text</italics></bold>.

5. ... aber überlappende Strukturen sind nicht erlaubt!

This is an **example** *containing highlighted text*. →

☹ This is an **<bold>example**
<italics>containing</bold> highlighted
text</italics>.

☺ This is an **<bold>example**
<italics>containing</italics></bold>
<italics>highlighted text</italics>.

XML-Regeln

6. Jedes Element kann Attribute und Werte enthalten, die es weiter spezifizieren. Attribut-Wert-Paare werden jeweils in das Starttag geschrieben:

```
<tagname attribute="value">...</tagname>
```

This is an **example** containing *highlighted text*. →

This is an **<hi type="bold">example containing
<hi type="italics">highlighted text</hi></hi>.**

7. Dasselbe Attribut kann nicht zweifach im selben Element verwendet werden...

☹ **<hi type="bold" type="italics">Example</hi>**

8. ... aber ein Attribut kann mehrere Werte erhalten.

☺ **<hi type="bold italics">Example</hi>**

XML-Regeln

9. XML erlaubt persönliche Kommentare, die nicht zum Primärtext gehören (und nicht geparkt werden).
Kommentare sehen so aus:

`<!-- comment -->`

10. Entitäten beginnen immer mit '&' und enden mit ';'.

e.g. Schaft-s: `ſ`

11. Es gibt reservierte Zeichen, die Teil der XML-Kernstruktur sind. Wenn sie zudem im Fließtext auftauchen, müssen sie speziell ausgezeichnet werden:

`<` `→` `<`; z.B.: 3 `<` 4

`>` `→` `>`; 4 `>` 3

`&` `→` `&`; Marks `&` Spencer