

## Regeln zum Erstellen eines Inhaltsverzeichnisses

Wolfgang Schmidle

2010-03-16

Im folgenden werden Regeln beschrieben, wie aus einem XML-Text ein Inhaltsverzeichnis ("TOC") erstellt werden soll. Dabei wird vorausgesetzt, dass der Text gültig ist nach dem MPDL-Schema. Der Text hat eine logische Struktur, die mit <div> markiert wird. Aus diesen <div> wird das TOC erstellt. Wenn der Text keine <div> enthält, kann auch kein TOC erzeugt werden.

### Liste der div-Typen

Die Liste basiert auf den div-Typen in Benedetti, der hier als Modelltext fungiert. Außerdem enthält sie die Typen volume, index, multiflow, parallel. Die Liste kann jederzeit erweitert werden. Zu jedem div-Type ist angegeben, ob und wie es im TOC erscheinen soll.

Ein Stern \* bedeutet, dass der div-Typ selbst gedruckt werden sollte.

Beispiel: <div type="volume"> sollte im TOC als "Volume" erscheinen (mit großem "V").

#### 1. volume

Text: \*

CSS: fett, Leerzeile vorher

Numerierung: die Volumes haben eine getrennte Numerierung. Verwende das Attribut @n.

Untereinträge werden eingerückt: nein

#### 2. front, body, back

Text: \*

CSS: fett, Leerzeilen vorher und nachher

Numerierung: keine

Untereinträge werden eingerückt: nein

#### 3. toc, cover, title, errata

Text: \*

CSS: normal

Numerierung: keine

Untereinträge werden eingerückt: ja

#### 4. dedication, preface, index, appendix

Text: \*, und zusätzlich alle <head>, falls es welche gibt; return nach jeder head-Gruppe (siehe Erläuterung 1)

CSS: normal

Numerierung: keine

Untereinträge werden eingerückt: ja

#### 5. chapter, section, math:\* (zum Beispiel math:theorem), letter

Text: alle <head>; return nach jeder head-Gruppe. Sonst "[no heading]"

CSS: normal

Numerierung: getrennte Numerierung für diese Gruppe, in der heutzutage üblichen Form: 5, 5.1, 5.1.3 etc. (jeweils ohne Punkt am Ende)

Untereinträge werden eingerückt: ja

#### 6. multiflow, parallel, float, sowie unbekannte div-Typen und deren Untereinträge (kommen nicht in das TOC)

## Erläuterungen

1. Unmittelbar aufeinanderfolgende <head> bilden eine "head-Gruppe". (Einzelne <head> sind eine head-Gruppe für sich. Zusätzliche head-Gruppen in einem <div> entstehen durch die editorische Entscheidung, direkt vor der head-Gruppe kein <div> einzuführen.) Jede head-Gruppe sollte ihren eigenen link haben. Beispiel:

```
<div type="section">
  <head style="it">Motum rectum ... per fe <lb/>quicquid Ariftoteli uifum fit.</head>
  <head>CAP. XXV.</head>
  <p> ... </p>
  <pb/>
  <head>ALITER IDEM.</head>
```

...

wird zu

|      |   |           |
|------|---|-----------|
| 4.25 | <i>Motum rectum ... per fe quicquid Ariftoteli uifum fit.</i> CAP. XXV. | Page: 196 |
|      | ALITER IDEM.  | Page: 197 |

Anderes Beispiel:

|      |   |           |
|------|---|-----------|
| 2.11 | CAP. XI.  | Page: 143 |
|      | ALITER IDEM.  | Page: 144 |
| 2.12 | <i>JACOBO ... MEDIOLANENSI Serenißimi Ducis ... peritißimo.</i> CAP. VII. | Page: 145 |

Beachte, dass <head style="it"> wieder in kursiv umgesetzt wird. Die zusätzlichen head-Gruppen sollten, falls möglich, bündig mit der ersten head-Gruppe sein.

2. Ein Beispiel für ein <div> ohne <head> ist:

```
<div type="section" level="3" n="14">
  <div type="letter" level="4" n="1">
    <head>DE MOTV MOLAE ...</head>
    <head style="it">Illust. Ioanni Paulo ...</head>
    ...
  </div>
  <div type="letter" level="4" n="2">
    <head style="it">De rèuolutione rota putealis ...</head>
    <head>AD EVNDEM.</head>
    ...
  </div>
  <div type="letter" level="4" n="3">
    <head style="it">De machina ...</head>
    <head>AD EVNDEM.</head>
    ...
  </div>
</div>
```

Im TOC:

|                   |  |           |
|-------------------|--|-----------|
| 6.14 [no heading] |  |           |
| 6.14.1            | DE MOTV MOLAE ... <i>Illust. Ioanni Paulo ...</i>  | Page: 297 |
| 6.14.2            | <i>De rèuolutione rota putealis ...</i> AD EVNDEM. | Page: 298 |
| 6.14.3            | <i>De machina ...</i> AD EVNDEM.                   | Page: 299 |

Wenn auf ein <div> der Gruppe 5 unmittelbar kein <head> und kein <div>, sondern ein <p> folgt, schreibe nicht "[no heading]", sondern die ersten Wörter des Absatzes, am Ende "...", in eckigen Klammern.

Beispiel: 4.25 [Wir sind doch nunmehr ganz, ja mehr denn ganz ...]

Dieser Fall kommt allerdings in Benedetti nicht vor.

3. Zu "Untereinträge": Der Einrückungs-Level eines Eintrags entspricht der Anzahl seiner einrückenden Ancestors. Beispiel:

```
<div type="volume" level="1">
  <div type="body" level="2">
    <div type="chapter" level="3">
      <div type="section" level="4">
        <div type="section" level="5">
          <div type="letter" level="6">
```

Hier hat "letter" zwei nicht-einrückende ancestors, nämlich volume und body, sowie drei einrückende ancestors, nämlich chapter und zweimal section. Der Einrückungs-Level ist also 3.

4. Es könnte für jeden Dokumenttyp eigene Regeln für die Erstellung des TOC geben. Bisher gibt es `<text type="book">` wie bei Benedetti und `<text type="multivolume">`. Hier sehe ich keinen Grund für verschiedene Regeln. Das könnte aber zum Beispiel für `<text type="letter">` anders sein. Solange wir solche Dokumenttypen nicht in unserem Text-Korpus haben, müssen wir das noch nicht klären.

### **Zur Möglichkeit von expliziten Hinweisen zur TOC-Darstellung im XML-Text**

Mit den oben beschriebenen Regeln ist klar, dass im XML-Text keine Hinweise für die Erstellung des TOC benötigt werden. Man sollte solche Hinweise sicher auch nicht verpflichtend machen, denn die dadurch nötige Mehrarbeit erhöht die Hemmschwelle, unser Schema zu verwenden.

Frage ist nun, ob man sie wenigstens optional erlauben soll. Dabei sollte man allerdings bedenken, dass fast kein Text solche Hinweise enthalten wird, weil das Eintragen dieser Information in den Text viel Arbeit ist und das automatische Ergebnis in fast allen Fällen gut genug sein wird. Alles, was automatisiert in das XML eingetragen werden kann, kann genausogut vom Anzeigesystem übernommen werden.

Dafür spricht, dass eine bessere Feinsteuerung des TOC möglich wäre. Dagegen spricht, dass die Information konzeptionell nicht in den XML-Text gehört. Der Text muss nicht einmal wissen, dass in unserem System aus den `<div>` ein TOC erstellt wird.