

# EPUB

## – Anatomie eines elektronischen Buchformates

---

Dr. Thomas Meinike  
Hochschule Merseburg (FH)

~~04.11.2010 – Wiesbaden~~

---

# EPUB?

## → Einstieg (1)

### • EPUB

- Kurzform für **E**lectronic **P**ublication, Version 2.0 wurde 2007 vom Gremium IDPF (International Digital Publishing Forum) als Standard für elektronische Bücher (E-Books) verabschiedet
- Aktuell liegt die Revision 2.0.1 (09/2010) vor
- EPUB verwendet etablierte Technologien wie XML, XHTML, CSS sowie MIME-Typen und kann somit in existierende Publikationsprozesse integriert werden
- Es sind mittlerweile zahlreiche E-Ink-Lesegeräte verfügbar, welche zunehmend Konkurrenz durch Tablets erhalten
- Im Folgenden wird die EPUB-Architektur vorgestellt und auf praktische Aspekte der Umsetzung von E-Books eingegangen

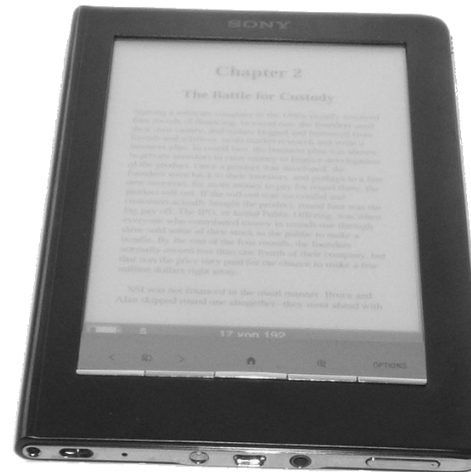
→ Einstieg (2)

## • E-Ink-Lesegeräte

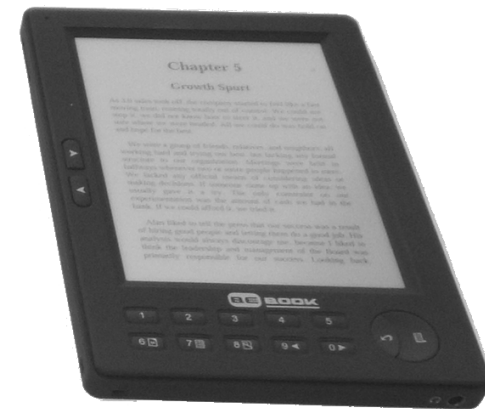
➤ Im Rahmen eines Projektes untersucht:



iRex iLiad



SONY PRS-600



BeBook Mini

→ Einstieg (3)

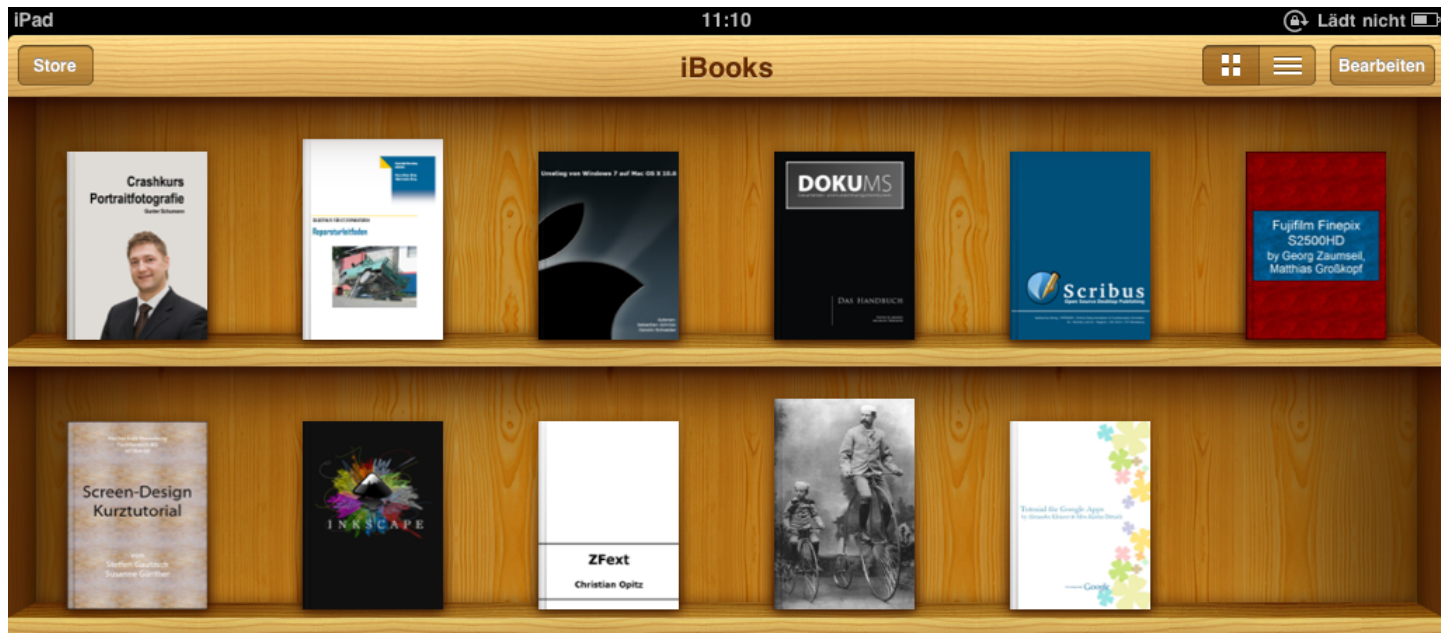
## • **Eigenschaften von E-Ink-Lesegeräten**

- Geringer Energiebedarf (Strom fließt nur beim Umblättern zum Neurendern der Seiten, keine Hintergrundbeleuchtung)
- Akkuladung hält somit Tage bis Wochen
- Üblich sind (noch) s/w-Displays mit 16 Graustufen
- Typische Auflösung um 160 dpi mit 600 x 800 Pixeln
- Anzeige von Text sowie von Raster- und Vektorgrafiken
- Darstellbar sind verschiedene Formate, nicht jeder Reader kennt alle
- Hauptsächlich werden EPUB, Mobipocket und PDF unterstützt, seltener DOC, PPT, ...

→ Einstieg (4)

## • Konkurrenz zu E-Ink-Lesegeräten → iPad

- Kostenlose Anwendung iBooks ermöglicht Darstellung von E-Books im EPUB-Format auf iPad und iPhone
- Höhere Auflösung (iPad: 768 x 1024 Pixel), Farbdisplay, Beleuchtung
- Gekaufte oder selbst erstellte Bücher werden mit iTunes übertragen

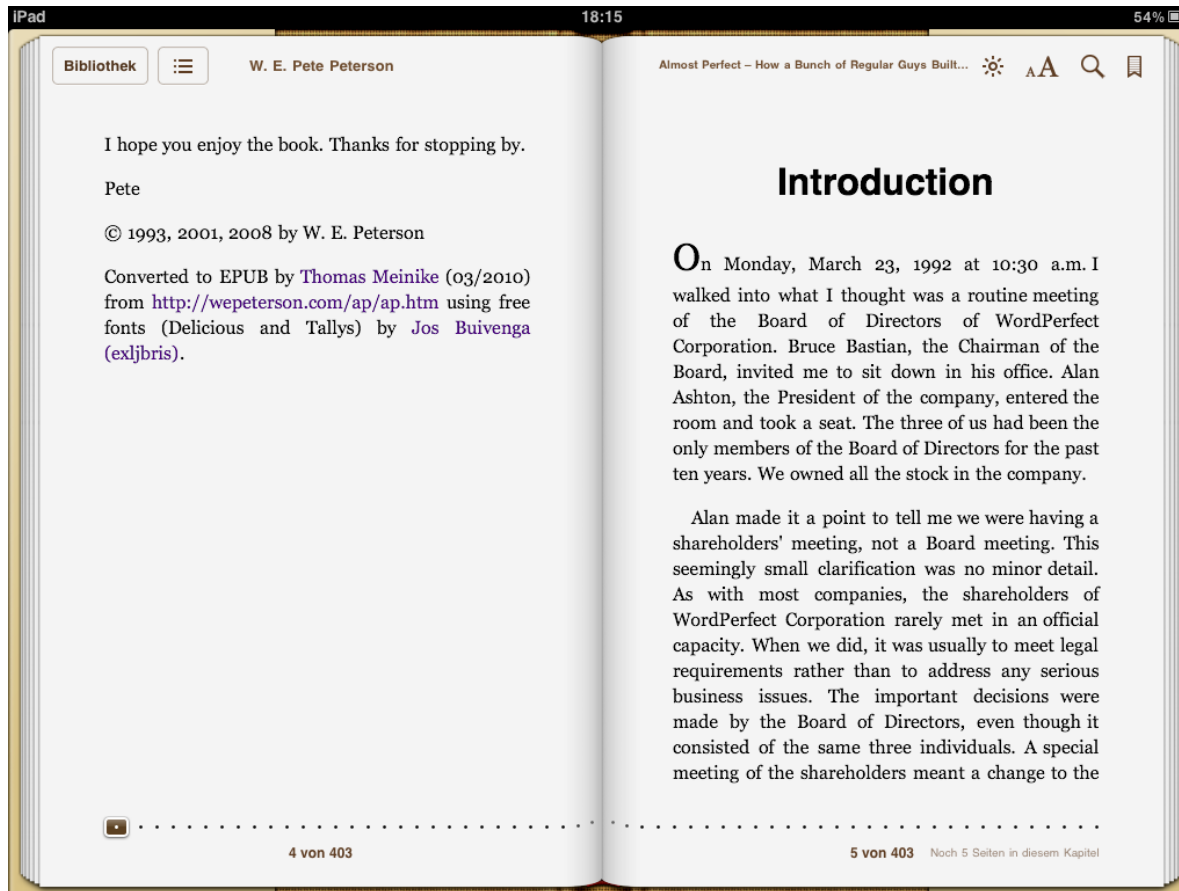


Arbeiten aus dem Masterstudiengang TRW (2010)

→ Einstieg (5)

# • Konkurrenz zu E-Ink-Lesegeräten → iPad

## ➤ Buchansicht in iBooks

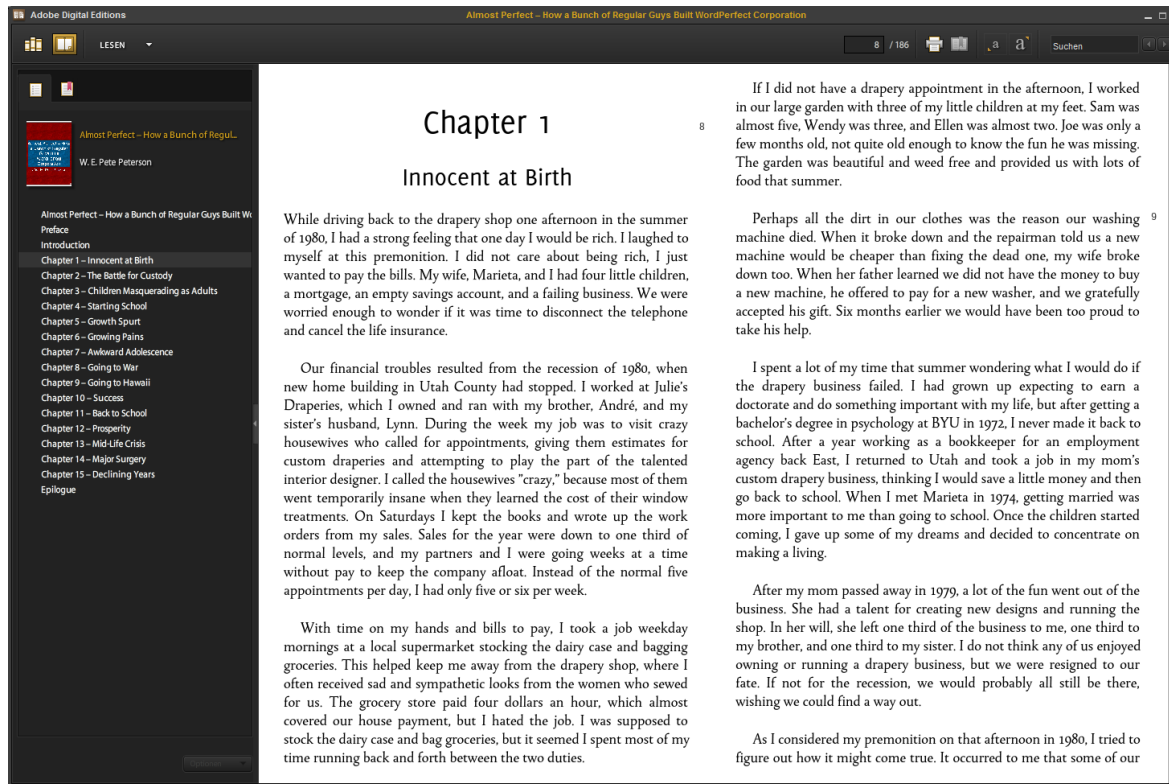


E-Book-Umsetzung  
des Autors (2010)

## → Einstieg (6)

# • Lesen ohne Hardware

## ➤ Adobe Digital Editions



## ➤ Firefox-Addon EPUBReader und weitere Software

# EPUB intern

→ EPUB-Aufbau (1)

## • **Beteiligte Spezifikationen (2.0.1)**

### ➤ **OCF (Open Container Format)**

- Definiert Verzeichnisstruktur und ZIP-Dateiformat

### ➤ **OPF (Open Packaging Format)**

- Beschreibt die erforderlichen bzw. optionalen Metadaten, die Leseabfolge und den Aufbau des Inhaltsverzeichnisses

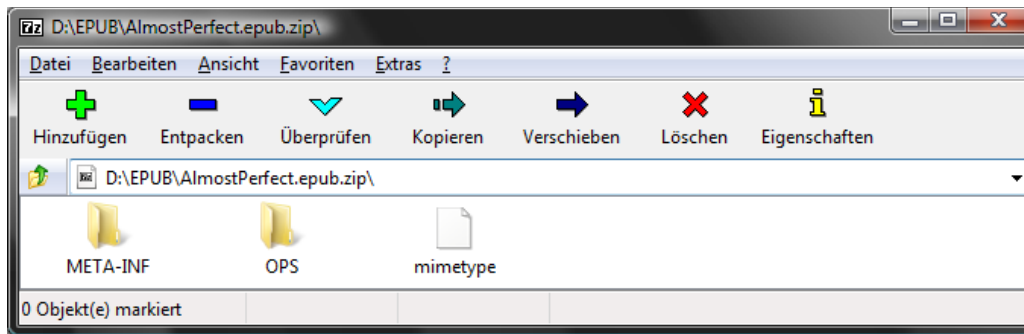
### ➤ **OPS (Open Publication Structure)**

- Legt inhaltliches Vokabular fest (u. a. XHTML / DTBook, CSS, Bildformate, Schriftarten)

## → EPUB-Aufbau (2)

### • Archivstruktur

- EPUB-Dateien sind ZIP-gepackte Archive, mit einem Werkzeug wie 7-Zip leicht einzusehen:



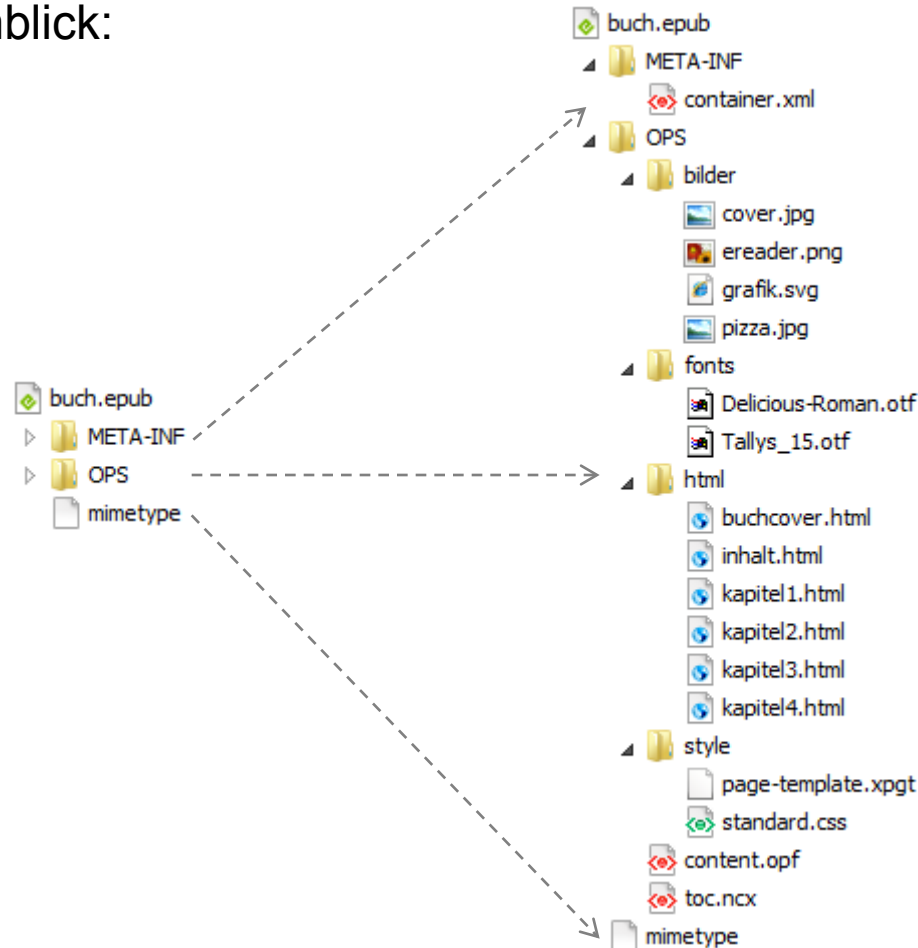
- Im Archiv sind die Unterverzeichnisse **META-INF** und OPS enthalten sowie die unkomprimierte (!) Datei **mimetype** (enthält als einzige Zeile: **application/epub+zip**)
- META-INF enthält **container.xml**, OPS die eigentlichen Inhalte mit formaler Beschreibung (content.opf) sowie die Navigation (toc.ncx)
- Alle Inhalte sind als Unicode zu speichern (UTF-8/UTF-16)

[Hinweis: Pflichtbezeichnungen sind rot markiert, alle anderen sind frei wählbar]

→ EPUB-Aufbau (3)

## • Archivstruktur

➤ Detaillierter Einblick:



## → EPUB-Aufbau (4)

### • META-INF

#### ➤ container.xml:

```
<?xml version="1.0" encoding="UTF-8"?>
<container version="1.0" xmlns="urn:oasis:names:tc:opendocument:xmlns:container">
  <rootfiles>
    <rootfile full-path="OPS/content.opf"
              media-type="application/oebps-package+xml"/>

    <rootfile full-path="PDF/book.pdf"
              media-type="application/pdf"/><!-- ggf. alternative PDF-Version -->
  </rootfiles>
</container>
```

- Attribut full-path verweist auf content.opf (bzw. book.pdf)
- Im Verzeichnis META-INF können weitere optionale Dateien für Signaturen, Verschlüsselung und digitales Rechteverwaltung liegen (signatures.xml, encryption.xml, rights.xml)

## → EPUB-Aufbau (5)

### • OPS

#### ➤ content.opf → package:

- Beschreibt die Paketstruktur
- Metadaten [dc:\* (Dublin Core) / meta]
- Auflistung der Navigations-, Inhalts- und Zusatzdateien des Buches
- Steuerinformationen und optionale Angaben zur Bedeutung der Einträge

```
<?xml version="1.0" encoding="UTF-8"?>
<package xmlns="http://www.idpf.org/2007/opf"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  unique-identifier="BookId" version="2.0">
  <metadata><!-- ... --></metadata>
  <manifest><!-- ... --></manifest>
  <spine toc="toc"><!-- ... --></spine>
  <tours><!-- ... --></tours><!-- in 2.0.1 deprecated -->
  <guide><!-- ... --></guide>
</package>
```

## → EPUB-Aufbau (6)

### • OPS

- content.opf → package → metadata mit Kindelementen (dc-Prefix):
  - Pflichtangaben: title, language, identifier

```
<metadata>
```

```
<dc:title>Titel</dc:title>
```

```
<dc:language>de</dc:language>
```

```
<dc:identifier id="BookId">http://epub.example.net</dc:identifier>
```

```
</metadata>
```

- Optionale Angaben: contributor | coverage | creator | date | description | format | publisher | relation | rights | source | subject | type
- dc:identifier-Inhalt: URI bei eigenen Werken ohne ISBN, ISBN bei Verlagsveröffentlichungen: urn:isbn:xxxxxxxxxxxxxx
- id-Wert (hier **BookId**) muss mit unique-identifier von <package ...> übereinstimmen

## → EPUB-Aufbau (7)

### • OPS

- content.opf → package → manifest:
  - Listet die physischen Inhalte, ihre IDs und Medientypen auf

```
<manifest>
```

```
<item id="toc" href="toc.ncx" media-type="application/x-dtbnx+xml"/>
```

```
<item id="CoverDesign" href="CoverDesign.jpg" media-type="image/jpeg"/>
```

```
<item id="CoverPage" href="CoverPage.html" media-type="application/xhtml+xml"/>
```

```
<item id="style" href="style.css" media-type="text/css"/>
```

```
<item id="epub.embedded.font1" href="fonts/font.otf" media-type="font/opentype"/>
```

```
<item id="kap1" href="kapitel1.html" media-type="application/xhtml+xml"/>
```

```
<!-- weitere item-Elemente ... -->
```

```
</manifest>
```

- href-Werte (insbesondere Unterverzeichnisse wie `fonts`) werden relativ zum OPS-Verzeichnis angegeben
- Eintrag `toc` bezieht sich auf die separate Navigationsstruktur
- `item` kann das Attribut `fallback="IDRef"` für Alternativinhalt erhalten

## → EPUB-Aufbau (8)

### • OPS

#### ➤ content.opf → package → spine:

- Gibt die anzuzeigenden Archivteile vor und legt ihre lineare Abfolge beim Blättern von oben nach unten fest

```
<spine toc="toc">  
  <itemref idref="CoverPage" linear="no"/>  
  <itemref idref="kap1"/>  
  <itemref idref="kap2"/>  
  <itemref idref="kap3"/>  
</spine>
```

- Wert toc bezieht sich auf manifest-Eintrag toc.ncx
- linear="no" schließt Teile von der Leseabfolge aus
- idref-Werte korrespondieren mit item/@id-Werten im manifest-Element

## → EPUB-Aufbau (9)

### • OPS

- content.opf → package → tours:
  - Optional und verworfen in EPUB 2.0.1, ermöglicht eine Vorschau auf besonders interessante Stellen im Buch

```
<tours>
```

```
<tour id="tourID" title="Vorschau">
```

```
<site href="kapitel1.html#fragment_id" title="..."/>
```

```
<site href="kapitel2.html#fragment_id" title="..."/>
```

```
</tour>
```

```
</tours>
```

- Mehrere tour-Kindelemente sind möglich

## → EPUB-Aufbau (10)

### • OPS

#### ➤ content.opf → package → guide:

- Optional, vermittelt die Bedeutung der einzelnen Abschnitte

```
<guide>
```

```
<reference type="cover" title="Titelseite" href="CoverPage.html"/>
```

```
<reference type="text" title="Kapitel 1" href="kapitel1.html"/>
```

```
<reference type="text" title="Kapitel 2" href="kapitel2.html"/>
```

```
<reference type="text" title="Kapitel 3" href="kapitel3.html"/>
```

```
</guide>
```

- Mögliche type-Attributwerte für das Element reference:  
acknowledgements | bibliography | colophon | copyright-page | cover |  
dedication | epigraph | foreword | glossary | index | loi (list of illustrations) |  
lot (list of tables) | notes | preface | text | title-page | toc (table of contents)

## → EPUB-Aufbau (11)

### • OPS

#### ➤ toc.ncx:

- TOC = Table of Contents, Navigation und enthält weitere Metadaten zum Buch  
– entstammt dem DAISY-Standard (Digital Accessible Information System → daisy.org)

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

```
<!DOCTYPE ncx PUBLIC "-//NISO//DTD ncx 2005-1//EN"  
"http://www.daisy.org/z3986/2005/ncx-2005-1.dtd">
```

```
<ncx xmlns="http://www.daisy.org/z3986/2005/ncx/" xml:lang="de" version="2005-1">
```

```
<head>
```

```
<meta name="dtb:uid" content="http://epub.example.net"/>
```

```
<meta name="dtb:depth" content="1"/>
```

```
<meta name="dtb:totalPageCount" content="0"/>
```

```
<meta name="dtb:maxPageNumber" content="0"/>
```

```
</head>
```

```
<docTitle><text>Titel</text></docTitle>
```

```
...
```

- Buch-ID (uid) wie zuvor definiert, depth = Verschachtelungstiefe der Navigation (≥ 1), Page-Daten: 0 (da keine festen Seitenzahlen), docTitle = Buchtitel

## → EPUB-Aufbau (12)

### • OPS

#### ➤ toc.ncx:

- navMap enthält navPoint-Elemente (weiter verschachtelbar)

```
<navMap>
  <navPoint id="navPoint-1" playOrder="1">
    <navLabel><text>Inhaltsverzeichnis</text></navLabel>
    <content src="inhalt.html"/>
  </navPoint>
  <navPoint id="navPoint-2" playOrder="2">
    <navLabel><text>Kapitel 1</text></navLabel>
    <content src="kapitel1.html"/>
  </navPoint>
  <!-- weitere navPoint-Elemente ... -->
</navMap>
</ncx>
```

- navLabel = Text im Inhaltsverzeichnis, content = referenziertes XHTML-Dokument (ggf. kapitel.html#fragment), playOrder = Reihenfolge

## → EPUB-Aufbau (13)

### • Inhalte, Formatierung, Fonts

#### ➤ XHTML 1.1 [DTBook, Custom XML]

- Alle wesentlichen Elemente (hx, p, em, strong, a, img, object, dl, ol, ul, table, pre, div, span, style, link, ...)
- Bildformate: GIF, JPEG, PNG, SVG
- Script-Code soll nicht ausgeführt werden, Inhalte müssen zugänglich sein

#### ➤ CSS2-Unterstützung

- Einschließlich externer Schriftarten (OpenType / TrueType) über @font-face:

```
@font-face {  
  font-family: "Delicious";  
  font-style: normal;  
  font-weight: normal;  
  src: url("fonts/Delicious-Roman.otf") format("opentype");  
}
```

```
h1,h2 { font-family: "Delicious", sans-serif; }
```

- Möglichst relative Angaben verwenden (em)
- Bildgrößen an typische Auflösungen anpassen

## → EPUB-Aufbau (14)

### • Tests

#### ➤ EpubCheck

- Testet Integrität und Standardkonformität
- Aufruf: `java -jar "X:\Pfad_zu\epubcheck-1.0.5.jar" buch.epub`

#### ➤ EpubPreflight

- Prüft auf leere oder zu große Inhalts- und Bilddateien (> 300 KB bzw. 10 MB)
- Aufruf: `java -jar "X:\Pfad_zu\epubpreflight-0.1.0.jar" buch.epub`

```
... EPUB (buch.epub) wurde erzeugt.
```

```
* EPUB-Tests ...
```

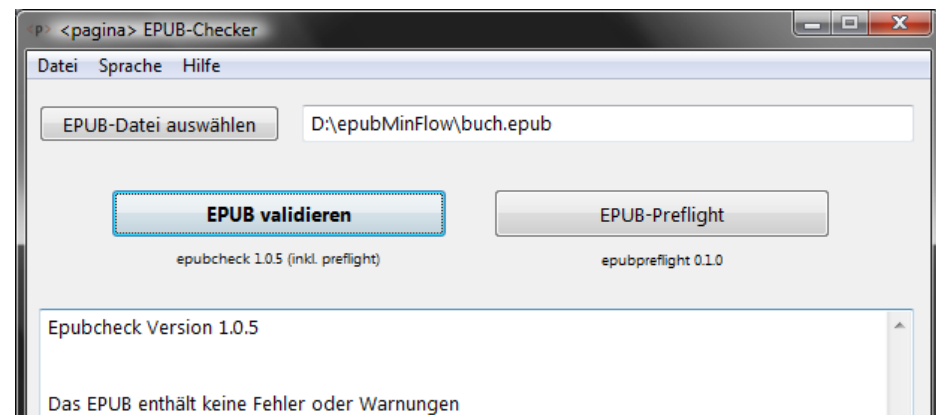
```
Epubcheck Version 1.0.5
```

```
No errors or warnings detected
```

```
EpubPreflight Version 0.1.0
```

```
No errors or warnings detected
```

#### ➤ <pagina> EPUB-Checker (Testwerkzeuge in GUI integriert)



## → EPUB-Aufbau (15)

### • Tests

- Valides Arbeiten und Testen vermeidet Fehlermeldungen:

The screenshot shows a web browser interface with a top navigation bar containing 'Bibliothek', a menu icon, and icons for settings, font size, search, and a bookmark. Below the navigation bar, there are four distinct error messages, each enclosed in a light red rectangular box. Each message follows a similar structure: a bold heading 'This page contains the following errors:', a line of text describing the error (e.g., 'error on line 5 at column 11: Encoding error'), a bold heading 'Below is a rendering of the page up to the first error.', and a preview of the page content up to that point.

**This page contains the following errors:**  
error on line 5 at column 11: **Encoding error**  
**Below is a rendering of the page up to the first error.**

*This page contains the following errors:*  
error on line 15 at column 33: **StartTag: invalid element name**  
*Below is a rendering of the page up to the first error.*

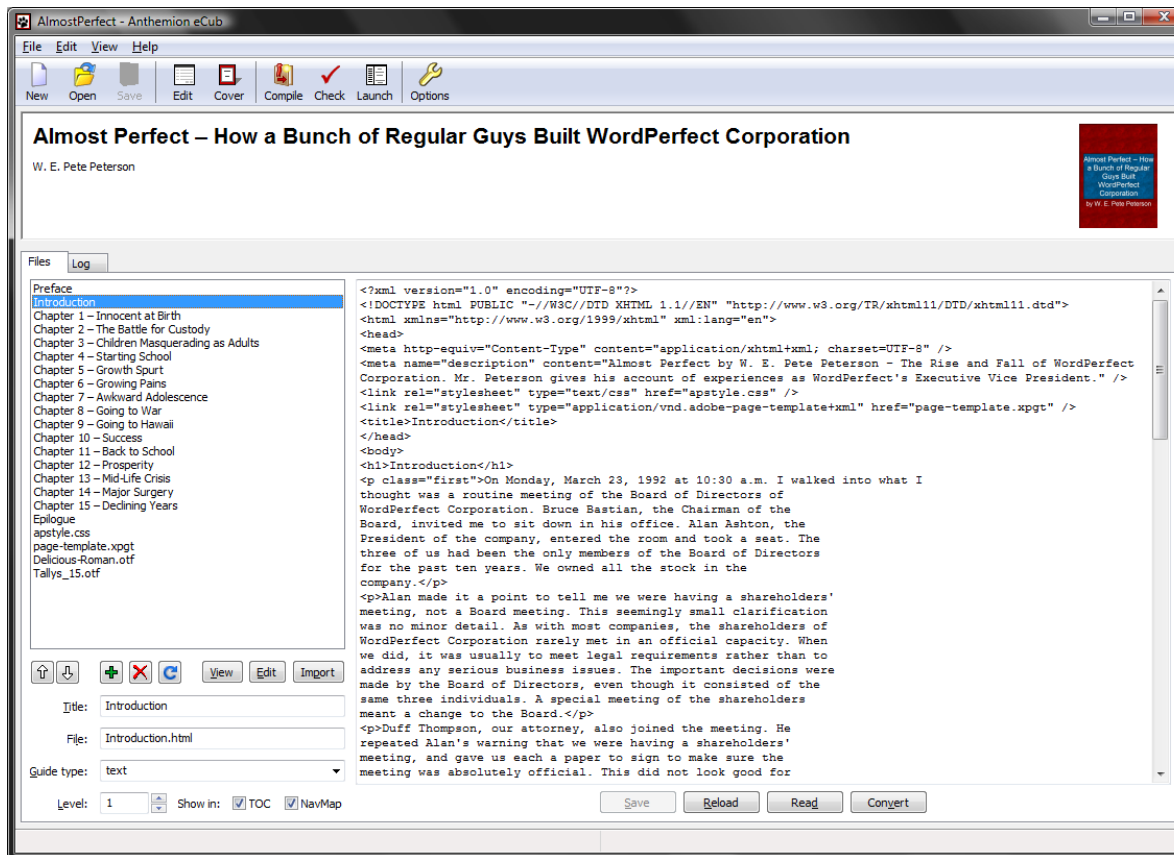
**This page contains the following errors:**  
error on line 23 at column 70: **Opening and ending tag mismatch: div line 0 and p**  
**Below is a rendering of the page up to the first error.**

*This page contains the following errors:*  
error on line 19 at column 12: **attributes construct error**  
*Below is a rendering of the page up to the first error.*

## → EPUB-Erstellung (1)

### • Software

- Export aus InDesign ab CS3, RoboHelp 8, Help & Manual 5, ...
- Separate Programme, z. B. eCub:

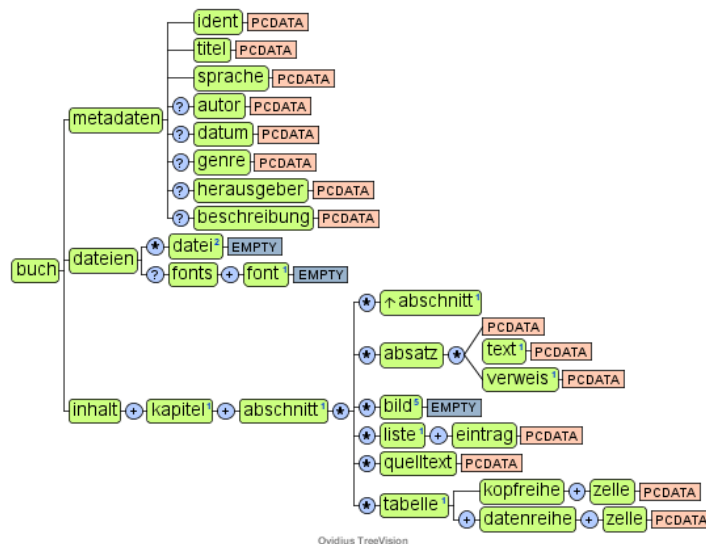


- Komfortable Entwicklungsumgebung
- XHTML + CSS hinzufügen / bearbeiten
- Struktur / Navigation aufbauen
- Metadaten-Editor
- Cover-Designer
- EPUB-Tests integrierbar
- EPUB-Generierung auf Knopfdruck
- Frei verfügbar (Linux, MacOS, Windows)

## → EPUB-Erstellung (2)

### • XSLT-Prozess epubMinFlow

- XML-Struktur (buch.xml) wird mit XSLT-2.0-Stylesheet (epub.xsl) über einen Batchlauf in komplettes E-Book überführt



#### Schritte:

- Verzeichnisse vorbereiten (CSS, Bilder, Fonts, HTML)
- buch.xml transformieren → container.xml, content.opf, toc.ncx, XHTML-Kapiteldateien (mittels xsl:result-document)
- ZIP-Komprimierung
- EPUB-Tests

Ergebnisstruktur zeigt Folie 12

```
C:\Windows\system32\cmd.exe
D:\epubMinFlow-go
EPUB-Build-Prozess gestartet ...

* Vorbereitung ...

style\page-template.xpgt
style\standard.css
  2 Datei(en) kopiert.
bilder\cover.jpg
bilder\ereader.png
bilder\grafik.svg
bilder\pizza.jpg
  4 Datei(en) kopiert.
fonts\Delicious-Roman.otf
fonts\Tallys_15.otf
  2 Datei(en) kopiert.

* XSLT-Prozess ...

wrote D:\epubMinFlow\epub\META-INF/container.xml
wrote D:\epubMinFlow\epub\OPS/content.opf
wrote D:\epubMinFlow\epub\OPS/toc.ncx
wrote D:\epubMinFlow\epub\OPS/html/buchcover.html
wrote D:\epubMinFlow\epub\OPS/html/inhalt.html
wrote D:\epubMinFlow\epub\OPS/html/kapitel1.html
wrote D:\epubMinFlow\epub\OPS/html/kapitel2.html
wrote D:\epubMinFlow\epub\OPS/html/kapitel3.html
wrote D:\epubMinFlow\epub\OPS/html/kapitel4.html

* Kompression ...

adding: mimetype (stored 0%)
adding: META-INF/container.xml (deflated 31%)
adding: OPS/bilder/cover.jpg (deflated 3%)
adding: OPS/bilder/ereader.png (deflated 0%)
adding: OPS/bilder/grafik.svg (deflated 63%)
adding: OPS/bilder/pizza.jpg (deflated 0%)
adding: OPS/content.opf (deflated 70%)
adding: OPS/fonts/Delicious-Roman.otf (deflated 43%)
adding: OPS/fonts/Tallys_15.otf (deflated 18%)
adding: OPS/html/inhalt.html (deflated 38%)
adding: OPS/html/inhalt.html (deflated 52%)
adding: OPS/html/kapitel1.html (deflated 46%)
adding: OPS/html/kapitel2.html (deflated 47%)
adding: OPS/html/kapitel3.html (deflated 41%)
adding: OPS/html/kapitel4.html (deflated 56%)
adding: OPS/style/page-template.xpgt (deflated 79%)
adding: OPS/style/standard.css (deflated 56%)
adding: OPS/toc.ncx (deflated 76%)

... EPUB (buch.epub) wurde erzeugt.

* EPUB-Tests ...

Epubcheck Version 1.0.5
No errors or warnings detected

EpubPreflight Version 0.1.0
No errors or warnings detected

D:\epubMinFlow>
```

# EPUB!



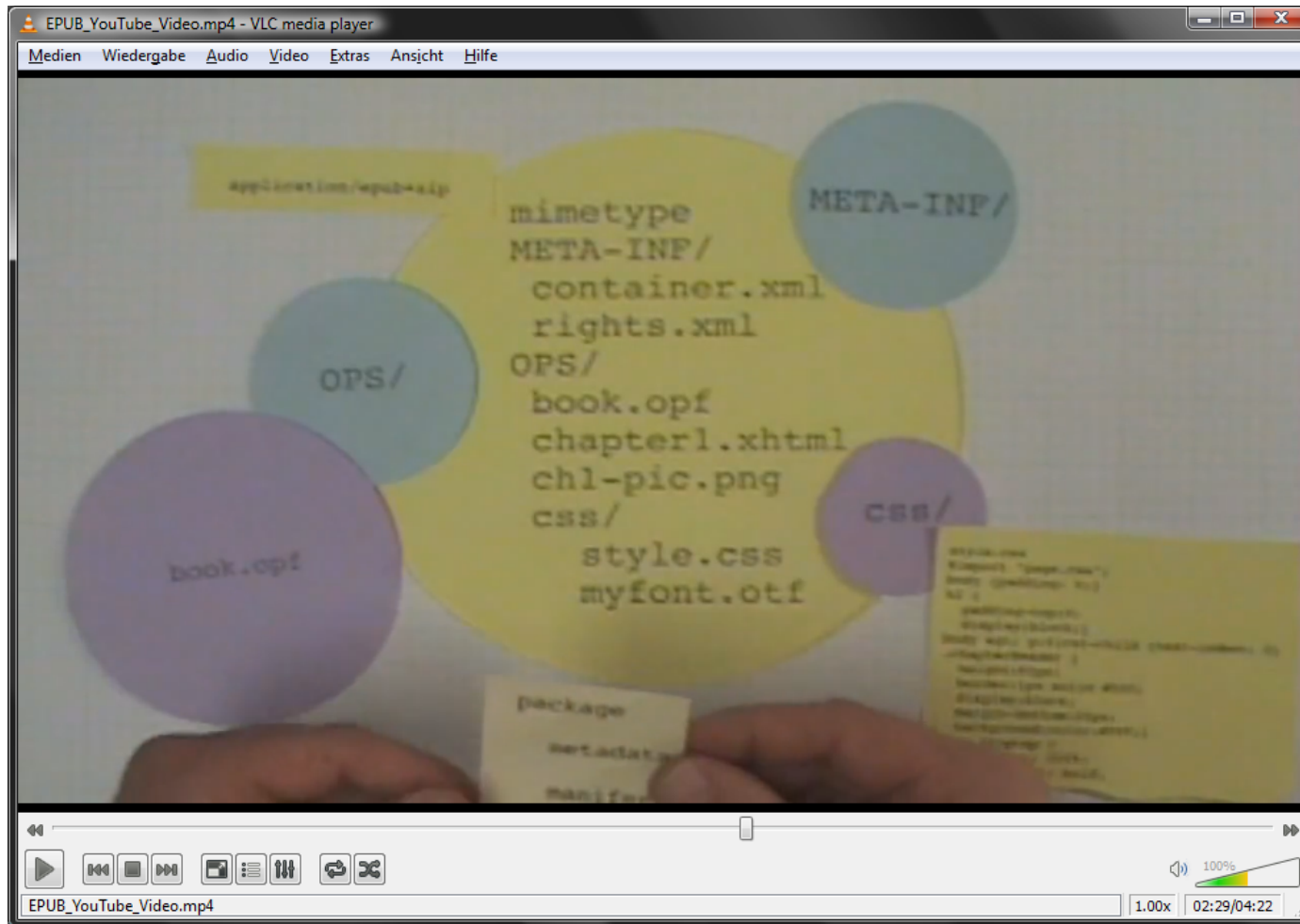
## → Zusammenfassung und Ausblick (1)

### • EPUB ...

- ... ist gegenwärtig das populärste und breit unterstützte E-Book-Format (Kindle-Reader unterstützen es bisher nicht → Konvertierung nach Mobipocket möglich)
- ... lässt sich weitgehend problemlos erzeugen und verwenden
- ... erfährt Weiterentwicklung in Richtung HTML5 (audio- / video-Element)  
Neue Geräte wie Tablets (iPad) bieten bereits mehr Möglichkeiten als die E-Ink-Lesegeräte
- IDPF-Arbeitsgruppe diskutiert als Nachfolger EPUB 2.1 / 3.0 (Mai 2011?)
- Autoren-Tipp: Buch von E. Castro gibt hilfreiche Tipps u. a. zur Nutzung von InDesign und Word zur EPUB-Produktion (u. a. Anpassung der Stylesheets)
- Entwickler-Tipp: <oXygen/> 12 unterstützt direkte EPUB-Bearbeitung

→ Zusammenfassung und Ausblick (2)

- **EPUB in < 5 Minuten erklärt**



## → Referenzen

- ADE: <http://adobe.com/de/products/digitaleditions>
- Castro, E.: EPUB Straight to the Point; Peachpit Press 2010
- eCub / Jutoh: <http://juliansmart.com>
- EpubCheck / EpubPreflight: <http://code.google.com/p/epubcheck>  
Eigenständige Anwendung EPUB-Checker: <http://pagina-online.de/software/epub-checker>
- EPUB Logo: [http://idpf.org/EPUBlogo/epublogocontest\\_winner.htm](http://idpf.org/EPUBlogo/epublogocontest_winner.htm)
- International Digital Publishing Forum (IDPF): EPUB-Spezifikationen; <http://idpf.org/specs.html>
- Kemp, J.: Video – iPad eBook format overview of the .epub file; <http://youtu.be/vvGrFZdSDww>
- Meinike, T.: Einfach publizieren und benutzen – EPUB-Format in Theorie und Praxis; Entwickler Magazin 4.2010, S. 99–106 (weitere Quellen in diesem Artikel)  
[http://www.iks.hs-merseburg.de/~meinike/PDF/Meinike\\_Einfach\\_publizieren\\_und\\_benutzen\\_EM\\_0410.pdf](http://www.iks.hs-merseburg.de/~meinike/PDF/Meinike_Einfach_publizieren_und_benutzen_EM_0410.pdf)
- Meinike, T.: Projekt epubMinFlow – Ein minimaler Workflow zur automatisierten Umsetzung von E-Books im EPUB-Format; <http://datenverdrahten.de/epubMinFlow>
- <oXygen/> XML Editor 12: <http://oxygenxml.com>

## → Kontakt

- E-Mail: [thomas.meinike@hs-merseburg.de](mailto:thomas.meinike@hs-merseburg.de)
- WWW: <http://www.iks.hs-merseburg.de/~meinike>