

# 電子書籍と Web と XML の 組版技術

2011/10/22

むらかみ しんゆう  
村上 真雄



Twitter: [@MurakamiShinyu](https://twitter.com/MurakamiShinyu)  
[murakami@antenna.co.jp](mailto:murakami@antenna.co.jp)

## はじめに～自己紹介

- 1990 年ごろ、[テキスト整形ツール「XTR」](#)を趣味で開発しフリーソフトウェアとして公開。それから T<sub>E</sub>X のこと意識するようになる（勝手にライバル視）。
- それで、組版とマークアップ言語（SGML～XML/HTML）、スタイルシート言語（XSL/CSS）に関心を持つようになる。
- 1999 年、XML 自動組版ソフトの開発を、当時個人プログラマーとして関わりのあった[アンテナハウス\(株\)](#)に企画提案、Antenna House Formatter として製品化。
- [AH Formatter](#) は、多言語の大量の XML データからの自動組版などで威力を発揮して、けっこう世界で使われています。
- XSL および CSS の仕様とその実装に携わり、その関係で電子書籍フォーマット EPUB の日本語組版拡張に関わる。

# EPUB って何？

- リフロー型（ページが固定でない）電子書籍の世界標準仕様
- IDPF という標準化団体が策定
  - ◆ IDPF = International Digital Publishing Forum
  - ◆ 世界の出版社や ITC 企業を中心に構成される
- 文書内容は XHTML、レイアウトは CSS で指定（ウェブの標準技術を応用）、それに書籍の情報や目次データを合わせて ZIP 形式で圧縮したもの
  - ◆ W3C という標準化団体が、(X)HTML と CSS の仕様を策定
  - ◆ (X)HTML は 5 がこれからの標準。CSS は 2.1 が標準、それに拡張機能を追加するレベル 3（CSS3）の仕様がモジュール別に標準化が進む
- EPUB2.0 は、XHTML1.1 + CSS2 だった。縦書きが出来ないなど、日本語組版で問題があった。
- EPUB 仕様改訂 → EPUB 3.0 へ（HTML5 と CSS3 縦書き仕様など採用）

# EPUB3.0 はこうなった

- 最新の Web 標準技術を利用 : (X)HTML5 と CSS2.1 + CSS3 (一部)
- [CSS3 Writing Modes](#) : 縦書き、縦中横
  - ◆ まだドラフトのためプロパティ名に -epub- プレフィックス付きで採用 :  
-epub-writing-mode など
- [CSS3 Text](#) : 禁則、行揃え、圏点など
  - ◆ 同じく -epub- プレフィックス付きで採用 : -epub-text-emphasis など
- ルビ (HTML5 ruby)
- 段組 (CSS3 Multi-column)
- 外字は? まず Unicode です。Unicode に無い文字はフォント埋め込みで
- SVG (ベクターグラフィック)、MathML (数式)
- マルチメディア (音声、動画)
- DAISY との連携 : だれにでもアクセシブルな書籍・教科書へ

# EPUB3.0 で使えるようになった CSS3 の日本語組版機能の例

## □ [CSS3 Writing Modes](#) : 縦書き、縦中横

縦書きの指定 :

```
html { -epub-writing-mode: vertical-rl }
```

縦書きで欧文や数字を直立に (デフォルトは横倒し) :

```
html { -epub-text-orientation: upright }
```

縦中横の指定の例 :

```
span.縦中横 { -epub-text-combine: horizontal }
```

## □ [CSS3 Text](#) : 圏点、禁則、行揃えなど

圏点の指定の例 :

```
em.ゴマ点 { -epub-text-emphasis: sesame }
```

```
em.蛇の目 { -epub-text-emphasis: double-circle }
```

圏点  
ゴマ点  
蛇の目  
など

EPUB3.0  
で縦中横

たてちゅうようこ

強い禁則の指定：

```
html { -epub-line-break: strict }
```

弱い禁則の指定：

```
html { -epub-line-break: loose }
```

## □ ルビ (HTML5 ruby)

グループルビの例：

```
<ruby>秋桜<rt>コスモス</rt></ruby>
```

熟語ルビの例：

```
<ruby>秋<rt>あき</rt>桜<rt>ざくら</rt></ruby>
```

モノルビの例：

```
<ruby>秋<rt>あき</rt></ruby>の<ruby>桜<rt>さくら</rt></ruby>
```

秋あき  
のの  
桜さくら  
はは  
秋あき  
桜ざくら  
かか  
秋コスモス  
桜桜  
か。

# EPUB に限らず、Web 標準技術 (XML、HTML5、CSS) が「本」のフォーマットの主流に

□ EPUB3 はもちろん

□ [Amazon Kindle Format 8](#)

“KF8 replaces the Mobi format and adds over 150 new formatting capabilities including support for **HTML5 and CSS3**.”

□ [Opera Reader: Paging the Web](#)

“The new CSS features are described in the **CSS Generated Content for Paged Media** module.”

# XML 組版とは何か

## □ XML による構造化文書のマークアップとその組版指定

```
<document>
  <title>簡単XML入門</title>
  <author>村上 真雄</author>
  <chapter>
    <title>XMLとは何か</title>
    <section>
      <title>XMLで文書を構造化</title>
      <p>文書を論理的な構造で表すことができます。
        例えばこんな感じ。</p>
    </section>
  </chapter>
</document>
```

- ◆ XML の前身は SGML。HTML は SGML の応用だった。HTML を XML 仕様に直したものが XHTML。
- ◆ XML 文書を組版するための規格 XSL (Extensible Stylesheet Language)

# XSL とは : XSLT と XSL-FO

- XSL は「拡張可能なスタイルシート言語(Extensible Stylesheet Language)」で、XML 文書をレイアウトするためのもの。
- XSL は XML の変換を行う XSLT (XSL Transform)仕様とレイアウトを表現する XSL-FO (XSL Formatting Objects)仕様からなる。
- 元 XML 文書を XSLT を使って XSL-FO 形式に変換して、XSL-FO を組版する。  
XSL-FO を組版するソフトウェアとして次のものがある：
  - ◆ [Apache FOP](#) : オープンソフトの XSL-FO 組版エンジン
  - ◆ [RenderX XEP](#) : 米国 RenderX 社の製品
  - ◆ [AH Formatter](#) : わがアンテナハウスの製品
- XSLT は、XSL-FO への変換だけでなく使い道はいろいろ。一番使われているのは XML から HTML への変換。XML から T<sub>E</sub>X への変換にも使われているよう。

## XSL スタイルシートの例

以下はとても簡単な XSL スタイルシートの例なのだけど、難しそうに見えるところが難点です。こんな感じ：

```
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
                xmlns:fo="http://www.w3.org/1999/XSL/Format"
                version='1.0'>
  <xsl:template match="document">
    <fo:root>
      <fo:layout-master-set>
        <fo:simple-page-master master-name="all-pages">
          <fo:region-body margin="2cm"/>
        </fo:simple-page-master>
      </fo:layout-master-set>
      <fo:page-sequence master-reference="all-pages">
        <fo:flow flow-name="xsl-region-body">
          <xsl:apply-templates />
        </fo:flow>
      </fo:page-sequence>
    </fo:root>
  </xsl:template>
  <xsl:template match="document/title">
    <fo:block text-align="center" font-size="32pt">
```

```
<xsl:apply-templates />
</fo:block>
</xsl:template>
<xsl:template match="document/author">
  <fo:block text-align="end" font-size="20pt">
    <xsl:apply-templates />
  </fo:block>
</xsl:template>
<xsl:template match="chapter/title">
  <fo:block font-size="20pt">
    <xsl:apply-templates />
  </fo:block>
</xsl:template>
<xsl:template match="section/title">
  <fo:block font-size="12pt">
    <xsl:apply-templates />
  </fo:block>
</xsl:template>
<xsl:template match="p">
  <fo:block text-align="justify" text-indent="1em">
    <xsl:apply-templates />
  </fo:block>
</xsl:template>
</xsl:stylesheet>
```

## 代表的な XML 文書フォーマット

以下は、文書を書くのに使える XML 文書フォーマットの主なもの：

- DocBook：本のコンテンツ制作のための XML フォーマット
- DITA (Darwin Information Typing Architecture)：情報をトピックというモジュールで表現して「マップ」で組み立てる。国際的標準化団体 OASIS 標準
- DAISY XML、その次世代は [DAISY4 \(ZedAI\) = Z39.86 Authoring and Interchange Framework Specification](#)：ZedAI から EPUB3 に変換することができます。
- XHTML、最新は (X)HTML5（この資料も XHTML5 で書いてます）

# CSS 組版

以下は先ほどの簡単な XSL スタイルシートと同じことをする CSS スタイルシート

```
document, title, author, p { display: block }
document > title {
  text-align: center;
  font-size: 32pt;
}
document > author {
  text-align: end;
  font-size: 20pt;
}
chapter > title {
  font-size: 20pt;
}
section > title {
  font-size: 12pt;
}
p {
  text-align: justify;
  text-indent: 1em;
}
```

こんなふうに、CSS は(X)HTML だけでなく任意の XML のレイアウトにも使えます。

## Web ブラウザエンジンばかりではない CSS の実装

CSS Paged Media 仕様（CSS のページ媒体向け仕様）によって、CSS でページ組版が可能です。

以下は、CSS でページの組版（HTML や XML を組版して PDF などに出力）をするソフトウェアの例です：

- ❑ [Prince XML](#) : オーストラリアの YesLogic 社の製品
- ❑ [PDFreactor](#) : ドイツの RealObjects 社の製品
- ❑ [Copper PDF](#) : 日本の GNN 社の製品
- ❑ [AH Formatter](#) : わがアンテナハウスの製品。XSL と CSS の両方をサポート

EPUB など電子書籍は、画面上に端末上でページ組版するもので、Web ブラウザエンジン（WebKit など）がよく使われますが、上記のソフトウェアは、T<sub>E</sub>X と同様に紙の本の印刷に使えるレベルの組版をするものです。

# ■ 進化する CSS 仕様

CSS 仕様は、W3C で標準化が進められています。

- 日本語組版に大いに関わる CSS3 仕様：
  - ◆ [CSS3 Writing Modes](#)
  - ◆ [CSS3 Text](#)
  - ◆ [CSS3 Ruby](#)
- ページ組版に関係する仕様：
  - ◆ [CSS3 Multi-column Layout](#)
  - ◆ [CSS3 Paged Media](#)
  - ◆ [CSS3 Generated Content for Paged Media](#)
  - ◆ [CSS Exclusions and Shapes](#)
- ほかにいろいろな

## アンテナハウス AH Formatter のフロード拡張の例

左上に 2 段抜きで配置

```
-ah-float: multicol left top;
width: 3gr;
```

いろはにほへとち  
りぬるを、わかよ  
たれそ、つねなら  
む。うみのおくや  
まけふこえて、あ  
さきゆめみしゑひ

もせすん。いろは  
にほへとちりぬる  
を、わかよたれそ、

つねならむ。うみ  
のおくやまけふこ  
えて、あさきゆめ  
みしゑひもせすん。

いろはにほへとち  
りぬるを、わかよ  
たれそ、つねなら

む。うみのおくや  
まけふこえて、あ  
さきゆめみしゑひ  
もせすん。いろは  
にほへとちりぬる  
を、わかよたれそ、  
つねならむ。

右下に 3 段抜きで配置

```
-ah-float: multicol right bottom; width: 5gr;
```

詳しくは、[「AH Formatter V6 の紹介」](#)をご覧ください。

# 構造化する HTML5

## □ 文書構造を表す要素が充実

- ◆ article
- ◆ header
- ◆ footer
- ◆ hgroup
- ◆ section
- ◆ aside
- ◆ figure
- ◆ figcaption

## □ 既存の HTML 要素を、意味のあるものとして定義しなおした

- ◆ hr : 段落レベルのテーマの変わり目を表す (水平線ではない)
- ◆ small : 細目といった注釈を表わす (文字を小さくするためではない)

## ■ おわりに～参考情報

- epub café <http://www.epubcafe.jp/>
  - ◆ 総務省「電子出版環境整備事業」、「EPUB 日本語拡張仕様策定」のすべての成果物を公開。  
「[EPUB 日本語拡張 コンテンツ制作者向けチュートリアル](#)」など
- IDPF <http://idpf.org/> ・ [EPUB3.0 仕様](#)が公開されてます。
- W3C CSS current work & how to participate  
<http://www.w3.org/Style/CSS/current-work>
  - ◆ EPUB3 日本語組版に関する [CSS3 Text](#) や [Writing Modes](#) など、最新のドラフト仕様があります。

私も、ツイッター [@MurakamiShinyu](#) や「[CSS 組版ブログ](#)」で、情報発信しています。どうぞよろしくお願いします。