

電子文書の情勢について (PDF, Web, EPUB など)

我々が「楽」に「まっとうな形式」の文書を作成・公開するには
あるいは, TeX 文書 (?) から EPUB を作るには

佐々木洋平

uwabami@gfd-dennou.org

摂南大学 理工学部 建築学科/基礎理工学機構

2020/10/23 @ ITPASS セミナー

ここ数年で電子書籍が一般に流通し出したものの、理工系の文書を電子書籍として組版するには課題が多い様に思える。一方で電子ジャーナルとして流通している論文等では Web 上で割と綺麗に組版されている。本講演では「電子書籍」や「電子ジャーナル」のデータ構造について紹介した後に、我々が取り組めそうな/取り込めそうな手法やその課題についてお話す (なにかとても素晴らしい提案や解決方法があるわけではないので雑談に終始するかもしれない)。

本日のお品書き

- ① 電子書籍と電子ジャーナル
- ② (我々の) 技術文書作成環境
- ③ 現状の対応策 (その試み) など.

censored

Amazon Kindle Paperwhite 防水機能搭載

電子書籍とは何か.

「電子書籍」

電子書籍（でんししょせき）とは、主にインターネット上で流通する電磁的に記録された読み物の総称である。

『電子書籍 - Wikipedia』

- ▶ 書籍データのデジタル化 (例: 辞書データの CD-ROM 化).
その後「電子ブック」がチラホラと…
 - 電子辞書については EPWING という規格があった。
⇒ 後継・派生規格はオープンになっていない。
 - これらは「専用端末」「専用ソフトウェア」での利用を想定していた。
- ▶ PDF の仕様は 1993 に無償公開, 2008 に ISO 32000-1 (標準化).
 - 印刷可能なデータの流通が始まる。
- ▶ インターネットによる「電子書籍」の流通開始.
 - 2007: Amazon 「Kindle」 発売, Kindle ストア開設。
 - 2010: Apple 「iPad」 発売 ← 「電子書籍元年」という(らしい).
 - 手軽な「端末」の登場 + サービスの提供。
 - なお, Kindle は書籍の販売ではない ことに注意。

「Kindle は書籍を販売していない」

オーウェル事件 (電子書籍販売停止と無断削除)

2009年7月17日: Amazon は再著作権を持たない出版社が販売していた電子書籍2点の販売を停止. これにあわせてユーザーがすでに購入した本も Kindle から無断で削除したのち料金を払い戻した.

- ▶ 本を「所有」するのではなく「書籍データを閲覧する権利」を契約
⇒ あくまで「貸本」業である.
- ▶ 本を「物理的」に所有するのは、どんどん贅沢になっていくのかも.

それはさておき
閑話休題

「電子書籍」ならではの特徴

手軽な端末の普及によって「紙の本」に縛られないデータの提供へ。

- ▶ 端末の画面サイズに依存する表示/しない表示.
 - 拡大縮小 (ズームイン・ズームアウト) ⇒ リフローへの対応.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam facilisis, diam id malesuada venenatis, lacus magna facilisis sapien, pretium varius elit velit sit amet elit. Ut commodo hendrerit lacus, sit amet pretium eros ornare sed. Cras placerat ex in diam ullamcorper, id tristique orci venenatis. Duis faucibus, lacus sit amet varius mattis, ex velit dignissim dui, id rutrum nunc tortor sit amet nunc. Suspendisse at sem vehicula, placerat urna eu, placerat est. Quisque quis quam eros. Maecenas sagittis tempus lectus, a feugiat purus finibus vitae. In et velit eros.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam facilisis, diam id malesuada venenatis, lacus magna facilisis sapien, pretium varius elit velit sit amet elit. Ut commodo hendrerit lacus, sit amet pretium eros ornare sed. Cras placerat ex in diam ullamcorper, id tristique orci venenatis. Duis faucibus, lacus sit amet varius mattis, ex velit dignissim dui, id rutrum nunc tortor sit amet nunc. Suspendisse at sem vehicula, placerat urna eu, placerat est. Quisque quis quam eros. Maecenas sagittis tempus lectus, a feugiat purus finibus vitae. In et velit eros.

リフロー型テキストの例 ~Wikipedia: 電子書籍~

- ▶ リッチコンテンツの可能性.
 - 画像⇒動画, インタラクティブツールの同梱.

(現在主流の) 電子書籍のフォーマット

① Portable Document Format(PDF): 拡張子 .pdf

- 電子書籍というより電子文書の規格
- 2008: ISO 32000-1 として標準化 (= PDF v1.7)
 - PDF 1.7 + Adobe Extension level.1~8...Adobe 独自拡張アリ
- 2017: PDF v2.0...まだあまり普及していない.

② Kindle File Format: 拡張子 .azw → .kfx

- 独自仕様
- HTML5 + CSS3(のサブセット) + DRM support.
- Kindle 端末, Kindle アプリケーションでの利用を想定

③ EPUB: 拡張子 .epub

- 現状での **標準フォーマット**
- 2007: IDPF(国際電子出版フォーラム) が規格策定
- 2011: EPUB3 公開
- EPUB ⊂ HTML(のサブセット) を Zip で固めた形式
 - EPUB3 から HTML5 + CSS ← **MathML** が使えるようになった.

電子ジャーナルとそのフォーマット

話は変わって電子ジャーナルについて.

- ▶ 歴史は良く知らない
 - テキストベースの検索システムは 1980 年台にあった模様.
- ▶ PDF の仕様が公開され普及
 - 「雑誌論文の PDF」がそのまま流通するように.
 - PDF v1.x の難点: 構造化データに乏しい (殆ど無い) こと.
- ▶ 一方で HTML ベースの論文公開の試みもなされてきた
 - 例: 「ながれマルチメディア」など
 - 近年では JATS (Journal Article Tag Suite) XML を用いて論文そのものを記述→ HTML として公開する雑誌も増えてきている.
 - XML であることの利点: 論理構造, 付加情報
 - 出力結果は文字データに付け加えて動画も提供できる.
- ▶ 「主流のフォーマット」は?
 - 出版社が保管しているデータは JATS XML になりつつある (気がする).
 - とはいえ JATS を手で書くのは非現実的 (ので, 出版社の人頑張ってる).

電子書籍と電子ジャーナル: まとめ?

- ▶ 手頃な端末とネットワークによって「電子書籍」が広まりつつある
 - 実態は「貸本業」だったり, (従来の)「買い切り」だったり.
 - DRM をどうするか, というのが相変わらず課題かも.
- ▶ 形式の変遷も落ち着いた感
 - PDF(固定形式) から「リフロー可能 + リッチコンテンツ」な形式へ.
 - だいたい「HTML5 + CSS3 (+ MathML) + 動画」といった形式.
 - 論文の Web 表示.
 - Kindle File Format.
 - EPUB.
- ▶ 変換ツールも増えつつある.
 - もはや「ツールでなんとかなる」のかもしれない.
 - 元原稿が $\text{T}_\text{E}\text{X}$ の場合は相変わらず辛い?
 - 詳細は後述.

なにか質問はありますか？
Have any Questions?

想定する技術文書の例: 「理論マニュアル」

▶ 例: 地球流体電磁気学倶楽部理論ノート

- 現在は pLaTeX 用の文書を作成
pLaTeX + dvipdfmx で PDF を作成
- dcnote, 過去には latex2html など

- dcnote (メンテナ不在?)

- PDF を画像化

- index 用の HTML 生成

- latex2html + 独自パッチ

- dennou style 対応 (?)

▶ 「教科書の電子化」をするには?

- pLaTeX ⇒ PDF で割と満足?

- TeX 組版による美しい文書. 流麗な数式.

- hyperlink 対応 (hyperref).

- リッチコンテンツはどうするか

- 動画入れたいよね.

磁気流体力学(MHD)の定式化

吉田茂生・竹広真一・佐々木洋平・林祥介

2002年7月5日

目次

1	電磁流体の基礎方程式	3
1.1	電磁場の式とその簡単な復習	3
1.2	流体力学の式	6
2	MHD 近似	8
2.1	光速に比べ遅い現象の記述: MHD と EHD	8
2.2	MHD 近似の下での電荷保存則	12
2.3	MHD 近似の下でのオームの法則	13
2.4	誘導方程式の導出	15
2.5	ローレンツ力とマクスウェル応力テンソル	17
2.6	磁場のエネルギーの式とジュール熱	18
3	MHD の基礎方程式	20
4	電磁場の境界条件	21
4.1	法線成分	21
4.2	接線成分	23
4.3	境界条件まとめ	25
5	回転系での MHD 方程式	27
6	エネルギー保存則	28
6.1	運動エネルギー保存則	28
6.2	磁場エネルギー保存則	29
6.3	内部エネルギー保存則	30
6.4	全エネルギー保存則	31

教科書の電子化, のために

何がしたいか, もしくはしたくないのか.

- ▶ 原稿はテキストで書きたい.
 - 版管理の手軽さ・容易さは失いたくない.
 - 特定の編集ソフトウェアには依存したくない.
- ▶ 数式は $\text{T}_\text{E}\text{X}$ で書きたい
 - できれば (これまで使っていた) マクロを使いたい.
 - 個人的には割と諦めて欲しい部分ではある
- ▶ 複数の形式に出力したい
 - PDF
 - HTML
 - EPUB (⇒変換したら Kindle Format にも変換可能?)
- ▶ 動画も同梱したい
 - 動画形式は選択の余地がありそう.

テキストベースの解決策

① T_EX 文書をそのまま使う

- EPUB3 では MathML が使えるが, Kindle では無理
⇒ SVG にしておくのが良いだろう.
- テキストフィルタ型: Pandoc, latex2html など
 - 拡張性はほぼ無い.
- T_EX エミュレート型: latexml, plastex, HeVeA など
 - 独自の T_EX マクロへの拡張が必須
- DVI ウェア型: TeX4ht など
 - DVI には文書構造が無い.

② 他の形式で原稿を書く

- 至極真っ当 (?) な解決策
- ではどの形式で書くか? HTML や XML を手書きするなんて無理ゲー.
- reST や Markdown は読み易いが…

テキストの「マークアップ」の設計 (思想)

- ▶ できる限り **暗黙** のマークアップ派: Markdown

Markdown

"...Markdown-formatted document should be **publishable as-is**, as plain text, **without looking like** it's been marked up with tags or formatting instructions."

<http://daringfireball.net/projects/markdown/>

- おかげさまで亜種が沢山存在していて…
 - 「あなたの Markdown はどの Markdown ですか?」状態.
- ▶ 概ね明示派: L^AT_EX
 - 開きタグ, 閉じタグ.
 - \$ や ^ などが若干暗黙的?
- ▶ 完全明示派: HTML, XML
 - 開きタグ, 閉じタグ.

(我々の) 技術文書作成環境: まとめ?

- ▶ 既存の文書に + α
 - 様々な形式で出力したい \Rightarrow 豊富な閲覧環境に対応したい.
 - 動画も付けたい
- ▶ テキストベースで原稿を作成したい.
 - できれば \TeX
 - 割とシンドイ (気がする)
 - latexml はまだ望みがありそうな気もするが.
 - \TeX が無理でも
なるべく「書き易く」「読み易い」マークアップで書きたい.

なにか質問はありますか？
Have any Questions?

- ▶ 最初に竹広さんから相談があったのは 2018/04/05
 - ① 数式 \LaTeX は捨てがたい. 他のマークアップ使いたくない
 - ② Web 上の画像や動画を取り込みたい.
 - ③ 実験プログラム, スクリプトもなんとかできれば.
 - ④ 長く持ちそうな形式でないと困る.
- ▶ 佐々木が勧めたのは Re:VIEW (レビュー)
 - ① 数式は \LaTeX
 - ② PDF, HTML, EPUB への出力可能
 - ③ プログラムやスクリプトは考える必要があるけれども.
 - ④ いざとなったら (中間出力の) \TeX を持っておけば良いだろう.

Re:VIEW とは?

Re:VIEW とは?

- ▶ 書籍作成にフォーカスした汎用マークアップの仕様およびその処理系
 - マークアップの仕様は RD の派生
 - 出力は HTML, \LaTeX , XML(for InDesign) が実装されている
 - ワンコマンドで...
 - EPUB 出力: review-epubmaker
 - PDF 出力: review-pdfmaker
- ▶ 開発体制と歴史
 - 2002: 青木峰郎氏が自身の執筆環境として開発
 - 2008: 武藤健志氏が開発に参加
 - 汎用的な書籍出版に活用できるように機能拡張
 - 現在も開発・保守・拡張中
 - その後 高橋征義氏, 角征典氏がコミッタとして参加
- ▶ ライセンス: GPL v2.1
- ▶ 利用実績が増えつつある.
 - 達人出版会の基本フォーマット, 技術系同人誌,
 - オライリー・ジャパン, インプレスでも原稿記法として採用

- ▶ 元々は竹広さんの作成されたファイル
- ▶ ソースの置き場所:

<https://dennou-k.gfd-dennou.org/member/uwabami/test/>

- Re:VIEW ファイル:
 - <https://dennou-k.gfd-dennou.org/member/uwabami/test/test.re>
- HTML 出力:
 - <https://dennou-k.gfd-dennou.org/member/uwabami/test/test.html>
- PDF 出力:
 - <https://dennou-k.gfd-dennou.org/member/uwabami/test/book.pdf>
- EPUB 出力:
 - <https://dennou-k.gfd-dennou.org/member/uwabami/test/book.epub>
 - 適当な EPUB 用のリーダーを利用して閲覧して下さい。

- ▶ ソースの見た目は割と良い, と思う.
- ▶ html,epub としての見栄えは CSS で調整.
- ▶ 数式は現在 PNG.
 - SVG に変更した方が良いかもしれない.
- ▶ (質問 1) 実際に使い始めるにはどうしたら良いか?
 - 「Re:VIEW クイックスタートガイド」を参照されたい.
 - T_EX 環境を導入した後に、gem で review を入れる.
- ▶ (質問 1) エディタのサポート状況は?
 - 発表者は Emacs を偏愛していたため知らなかったが他にも色々あった。
⇒ <https://github.com/kmuto/review/wiki> の Tools を参照.

- ▶ Wikipedia: 電子書籍
 - <https://ja.wikipedia.org/wiki/電子書籍>
- ▶ latexml
 - <https://dlnf.nist.gov/LaTeXML/>
- ▶ markdown
 - <https://daringfireball.net/projects/markdown/>
- ▶ Re:VIEW
 - <https://github.com/kmuto/review>