

デジタルドキュメント(9)

高久雅生

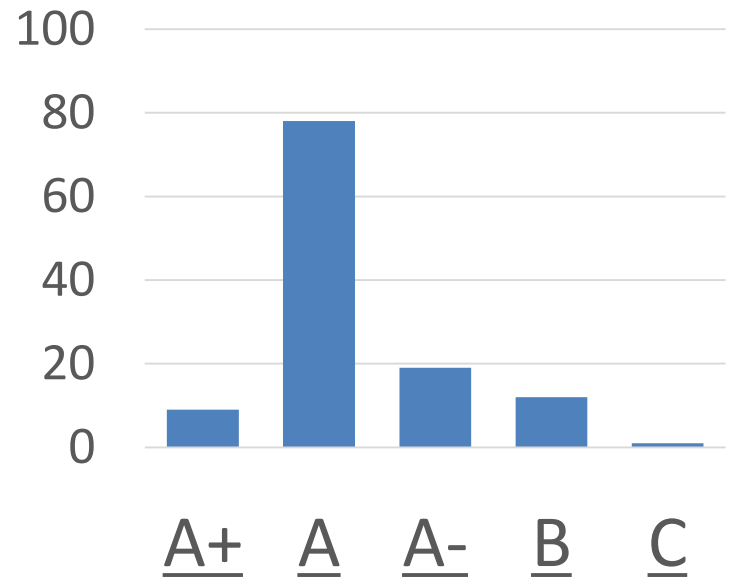
2013年6月13日(木)3・4時限

(再掲) 第3回レポート課題

- 電子書籍を一点選び、読んでみることに。
- 読んだ電子書籍を具体例に即して文章で説明すること。その際、以下の各項目に関する説明を加えること。
 - 閲覧のための前提条件および閲覧環境
 - 読んでみて気づいた点
 - コンテンツの配信元、配信形態
 - ドキュメントのフォーマット(ファイル形式)
 - 電子書籍としてのコンテンツの特徴
- 取り上げた電子書籍の書誌事項を必ず記載すること。
 - SIST-02準拠の形式を用いること

第3回レポート(講評)

- 全体によく書かれていた、という印象。
- 選択した電子書籍の種別
 - 小説 70
 - 一般書 23
 - コミック 16
 - 雑誌/新聞 6
 - 論文 3



第3回レポート(講評)(2)

～選択した電子書籍の提供プラットフォーム～

- 青空文庫 36
- Amazon Kindle 10
- iTunes Store 9
- (その他:出版社サイト) 9
- ニコニコ静画 5
- Yahooブックストア 4
- BinB Store 4
- Google Books 4
- Newsstand 3
- 国立国会図書館近代デジタルライブラリ 3
- Puboo 3
- Eエブリスタ 2
- BookLive 2
- honto 2
- Kinoppy 2
- 言い値書店 1
- ソク読み 1
- 漫画on Web 1
- DMM.com 1
- ビットウェイ 1
- J-STAGE 1
- BOOK Walker 1
- コミックシーモア 1
- Pixiv 1
- Digital Book Japan 1
- Project Gutenberg 1
- マガストア 1
- Amazonなか見検索 1
- iTunes U 1
- Sony Reader Store 1
- 任天堂DS 1
- CiNii 1
- Jコミ 1
- えあ草子 1
- Linedoc 1
- (その他:その他Webサイト) 1

(前回の復習 = ふりかえり)

- デジタルドキュメントとドキュメントフォーマット
- (半)構造化文書
- HTML
 - 文書例
 - 要素、タグ、属性
 - ハイパーリンク
 - 構造と見栄え
 - 歴史と標準化
- XML
 - 文書例
 - 整形式

前回の出席カードから(質疑、コメント)

- 前回スライド14枚目の下から2行目 が抜けていたように思う。
 - ありがとうございます。資料作成時の誤りでぬけていたようです。
 - Web掲載版は修正のうえ、差し替えてあります。
- 授業資料およびプレゼンテーション方法について
 - スライド一枚ごとの情報量がやや多すぎる気がします。
 - 1つのスライドに内容をつめすぎている。
 - 余白が少ない(天地左右、行間、時が太い、大きい)。
 - 文章の重要度(見出し・本文等)が分かりにくい(字のサイズ等の対比が小さい、色や区切り図形等を使用していない、囲み枠・下線などが効果的でない)。
 - 授業資料としては要らない部分(「出席票の提出」等)が毎回付いている。
 - 多数のご指摘ありがとうございます。可能な範囲で反映できるようにします。
- 最終レポート課題について
 - 最終レポートのような大きなレポートはあるのでしょうか。もしあるようなら、早めに告知していただくとありがたいです。
 - 期末レポートについて早めに詳細が知りたいです。他のレポート等あるため、時間に余裕を持って取り掛かりたいので、よろしくお願いします。
 - 第4回レポート課題の内容は来週にお知らせします。
 - 分量はこれまで通り、 \times 切にある程度の余裕を持たせる予定です。
- A-でもAを付けてほしい。
 - お約束できません。可能な範囲で、良いレポートを作成することに尽力ください。

前回の出席カードから(感想)

1. レポートの内容でなく、引用などの部分で厳しい減点をするのなら、授業中に引用の書き方などを紹介していただくと良いんじゃないかと思います。
2. レポートの回数はちょうど良いと思います。
3. レポートの難易度、回数はたとえ5回でも適切です。
4. レポートなどはいまのままでよいと思います。
5. レポートの回数も書き方についての指摘も適切だと思う。
6. レポート課題の量としては、ちょうど良いと思います。
7. 他の授業で一通りHTMLについては習ったが、今回の授業で自分がほとんど忘れてしまっていたことに気付いた。復習になって良かった。
8. HTMLは以前、別の広義で学んだので、良い復習の機会になりました。
9. HTMLを書いたことはあったが、frameタグというのは使ったことがなかった。
10. マークアップ言語は他の授業でもやっているのので、理解が深まってよかった。
11. 決まった参考文献の書き方はやはり見やすくきれいだと思った。
12. HTMLはマークアップ言語という授業で用いることがあるが、いろいろなことができて面白いと感じている。
13. この授業でHTMLが取り上げられるとは思わなかった。
14. 同じ効果を持つ文字修飾のタグが複数あるということが意外でした。
15. 「戻る」についての話が興味深かったです。

(番外編...)

- 結婚されていますか？もし、されていたらプロポーズの言葉を教えてください。本気で悩んでいます。

本日のお品書き

- デジタルドキュメントとドキュメントフォーマット
- LaTeX
 - 文書例
 - LaTeX文書の構成要素
 - 利用例 / 組版
- PDF
 - 歴史, PostScript
 - 構成要素
 - 文書フォーマットの構造

ドキュメントフォーマット (3)

LaTeX

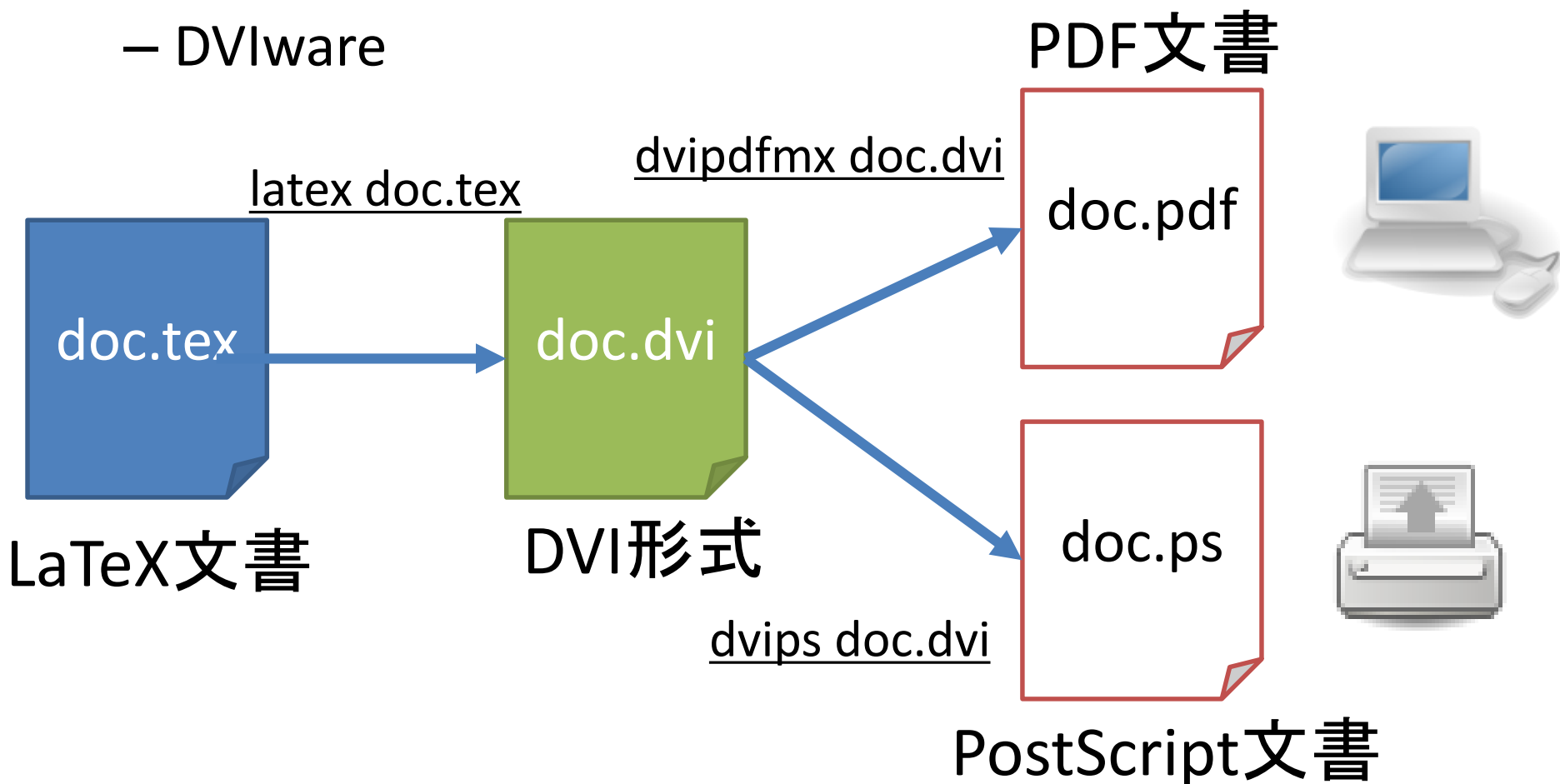
- テキストベースの文書記述言語
- 組版ツール
- (発音: ラテフ, ラテック, レイテック)
- 組版ツールTeX by Donald Knuth
 - 1978年から開発 ← 『The Art of Computer Programming』
 - 数式対応に優れている
 - 豊富な拡張機能(マクロ, フォント)
 - 一種のプログラミング言語
- LaTeXは、TeXを文書構造記述に拡張したもの
 - 日本語対応版: pLaTeX

LaTeX文書の構成要素

- メタデータ、文書本文
 - `\title{}`, `\author{}`, `\maketitle`
 - `\begin{document} ... \end{document}`
- 章節構造
 - `\section{}`, `\subsection{}`, `\subsubsection{}`
- 参照構造
 - `\label{}` \leftrightarrow `\ref{}`
- 文献参照
 - `\cite{}` \leftrightarrow `\bibitem{}`

LaTeX文書の組版(フォーマット変換)

- LaTeX → DVI → PostScript, PDF
 - DVI (Device Independent)
 - DVIware



PDF (Portable Document Format)

- Portable: 持ち運び可能な...
- Document Format: 文書形式
- コンピュータの機種や環境によらず、オリジナルのイメージをかなりの程度正確に再生できる。
- 歴史的には、PostScript (プリンタ用ベクタ描画言語) が源流として存在
- 印刷媒体 + コンピュータ上でのデータ交換
 - ディスプレイモニタ上での表示
 - 文書情報 (メタデータ)
 - ページ送り; ランダムアクセス
- 異なる環境でもレイアウトがほぼ一定

PDFの歴史

- 1993年 Adobe Systems社により開発
 - PDF version 1.0
 - Acrobat 1.0同時発売
- 1994年 Acrobat Reader無償配布
- 1995年 Netscape Navigator用プラグイン配布
 - Web普及への戦略
- 2008年 ISOによる標準化
- ※ただし、PDF以前の歴史が重要
 - PostScriptとDTP技術

PostScript

- PostScript
 - Adobe Systems (アドビ・システムズ)社が開発したページ記述言語
 - 描画をベクトル化することにより高い表現力と品質の向上を実現
- 1985年 PostScriptプリンター(Apple LaserWriter) およびイメージセッタ(Linotype社)の開発
- 1989年 日本語PostScriptプリンタの登場
- DTP(Desktop Publishing)の爆発的普及に貢献

PostScriptからPDFへ

- PDFはPostScriptを元に策定
- PostScriptのターゲットは印刷（紙）
- PDFはコンピュータ上でのデータ交換
 - ディスプレイでの表示
 - 文書情報（メタデータ）の扱い
 - ページ
 - ランダムアクセス

PDFの特徴

- 表示・印刷：異なる環境でもレイアウトがほぼ一定
- セキュリティ：設定可能
- ファイルサイズ：小
 - データ圧縮：テキスト、図画像
- 閲覧支援：しおり，リンク，コメント，注釈
- 検索：透明テキスト
- インタラクティブ：フォーム機能により入力欄を設定可能
- アクセシビリティ：音声化
- マルチメディア：対応

PDFの印刷と表示

- 表示用アプリケーション
 - 無償配布⇒OSに依存しない表示
 - MS Word, Excel等はOSが変わるとレイアウトも変わる
 - HTMLはウィンドウサイズにより左右される
- PDFの罨
 - Adobe Readerのバージョンが異なると正しく表示できない
 - 同一フォントを持っていないと文字化けやレイアウトが崩れる場合も
 - 解決方法:フォントの埋め込み
 - (課題)全体のファイルサイズが大きくなる。フォントのライセンス問題。

PDFのセキュリティ機能

- 閲覧パスワード
- 編集パスワード
 - 編集
 - 印刷
 - テキストと画像のコピーを制限
 - ※PDF言語仕様上は、単なる“紳士協定”であることに注意...
- 電子署名による作成・配布者の認証も可能

PDFファイルの作成

- PDFはPostScriptを元に策定された
- アドビシステムズのAcrobat
 - 印刷機能にAdobe PDFプリンタを追加
 - 印刷機能においてPDFファイルを作成する
 - 文書の編集は一般的に各種アプリケーションにまかせている
- Mac OS Xは各種アプリケーションから印刷でPDF化可能
- 最近では各種アプリケーションが直接PDF出力をサポート

PDFファイルの実際の作成

- Adobe製品⇒そのままPDF化可能
- Microsoft Office
 - プリンタとして出力
 - そのままPDF化可能 (MS Office 2007から)
- LaTeX
 - PDFLaTeX: LaTeXソースからそのままPDFへ
 - dvi2pdf: dvi形式からPDFへ
 - それ以外: dvi形式からPostScript形式にしてから、別途PDFに変換する
- 紙 ⇒ スキャン (+ OCR)
 - 有名な製品ScanSnap

PDFの課題

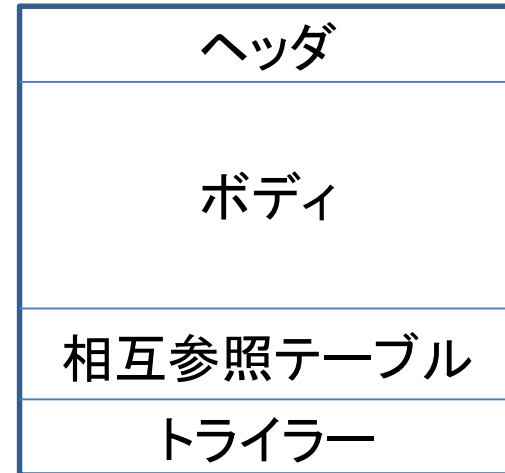
- 複雑な仕様
 - PDF 1.7 (ISO 32000-1, 2008) は750ページ以上(付録含む)
 - http://www.adobe.com/jp/devnet/pdf/pdf_reference.html
 - ISO標準化以前はAdobe一社による管理→デファクト標準の部分も一部残る
- ディスプレイでの閲覧支援とは言うものの...
- しおり, リンク, コメント, 注釈に関しては、作成時につける
 - Adobe Readerで閲覧しながら付けることはできない
 - 専用アプリケーションソフトウェアが必要

PDFの構造

- 一言で言えば、とても複雑
 - http://www.adobe.com/jp/devnet/pdf/pdf_reference.html から参照できる。
- ランダムアクセスを許すために、文書要素(オブジェクト)を分割して記述している
 - 逆に言えば、常に数百・数千ページあっても、簡単にページジャンプを実現できる。
 - 相互参照により、閲覧・編集ソフトウェアが文書構造を再構成しなおして表示する。
- オブジェクト(Object)
 - 記述のための要素。PDF文書の最も基本的な構成要素
 - 文字列、数値、配列、画像データ、フォント等。

PDFの構造 (2)

- ファイル構成 (PDF: 7.5 File Structure, p.38):
- ヘッダ (header)
%PDF-1.4
- ボディ (body)
...
- 相互参照テーブル (cross-reference table)
xref
0 45
0000000000 65535 f
0000003343 00000 n
0000003216 00000 n
...
- トレイラー (trailer)
trailer
<</Root 1 0 R/Info 2 0 R/Size 45>>
startxref
17538
%%EOF



PDFの構造 (3)

- 文書構造 (PDF: 7.7 Document Structure, p.70)
 - Objectsの階層関係
 - 参照関係で表現
 - トレイラー → ドキュメントカタログ → ページリスト → 各ページ (次スライド)
- コンテントストリーム (PDF: 7.9 Content Streams and Resources, p.81)
 - ページの見た目やグラフィカルな要素を記述する命令. 文書構造とは独立

1 0 obj

<</Pages 18 0 R/Type/Catalog>>

endobj

18 0 obj

<</Type/Pages/Count 1/Kids[3 0 R]/MediaBox[0 0
595.28 841.89]>>

endobj

3 0 obj

<</Resources 17 0 R/Type/Page/Parent 18 0
R/Contents[16 0 R]>>

endobj

(再掲)ドキュメントフォーマットの切り口 (1)

- テキスト (text) vs バイナリー (binary)
 - ビットデータ
 - 文字コードによる解釈
 - 外字
- フォーマットの指定・識別・判別
- シンプルコンテンツ vs 複合メディア
 - 埋め込みコンテンツ
 - ハイパーリンク
- メタデータ
 - 埋め込みメタデータ
 - 外部メタデータ記述
- 文書レイアウト
 - ページ概念
- 文書内の書式要素
 - 見栄え / スタイル
 - フォント
- ファイル vs ストリーム
 - データの保存・蓄積と配信

(再掲)ドキュメントフォーマットの切り口 (2)

- オープンフォーマット
 - 移植可能性 / ソフトウェア独立性
 - Free / proprietary
- 標準化
 - デファクト標準とデジュール標準 (“de facto” vs “de jure”)
- 文書フォーマットのバージョン
- フォーマット変換
 - テキスト → HTML
 - LaTeX → PDF
- 圧縮
 - 可逆 vs 非可逆
- セキュリティ
 - パスワード
 - 電子署名
- 長期保存
- デジタルフォレンジック

まとめ

- デジタルドキュメントとドキュメントフォーマット
- LaTeX
 - 文書例
 - LaTeX文書の構成要素
 - 利用例 / 組版
- PDF
 - 歴史, PostScript
 - 構成要素
 - 文書フォーマットの構造

次回予告

- 残テーマ
 - 他のドキュメントフォーマット
 - EPUB
 - Open Office XML, etc.
 - オフィス文書
 - 長期保存やデジタルライブラリーサービスとの関連
- 全体の振り返り
- そのほか
 - 授業評価アンケート
 - 第4回レポート課題: ✕切は 7/9 予定。分量は通常と同様。