

デジタルドキュメント (8)

高久雅生

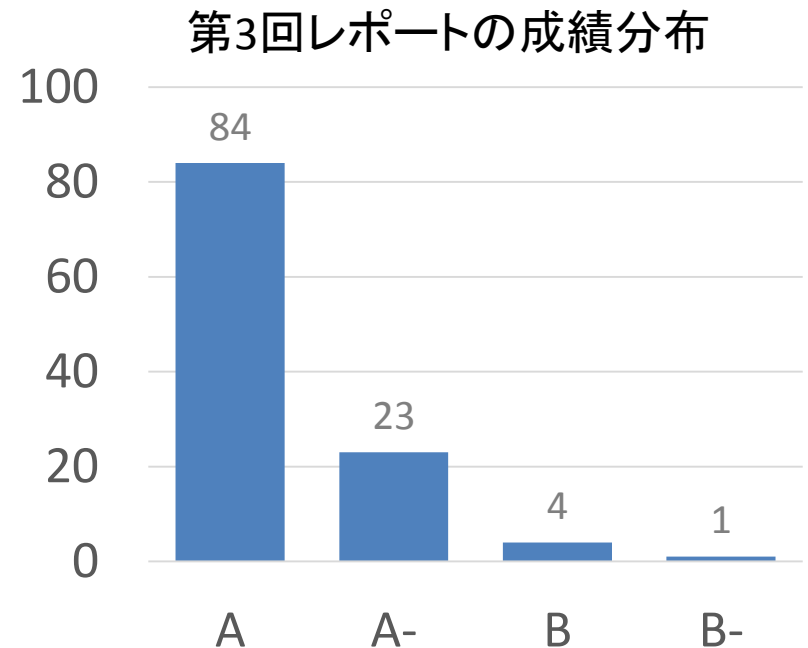
2014年6月5日(木)3・4時限

本日のお品書き

- (第3回レポートの返却・講評)
- (前回の振り返り)
- (前回講義での質問・要望への回答：2件)
- ドキュメントフォーマットの例
 - LaTeX
 - PDF
 - オフィス文書
- (第4回レポート課題)

第3回レポート講評

- 成績は全体にとっても良かった。
 - 内容的にもよく書けている人が多い印象を受けた。
- 選択した電子書籍の種別（昨年とほぼ同様）
 - 小説 68
 - 書籍 23
 - コミック 17
 - 論文 2
 - 絵本 1



第3回レポート講評 (2)

- 電子書籍プラットフォーム, 提供元
 - 青空文庫 36
 - Kindle 14
 - iBooks 9
 - Kobo 7
 - BinB 6
 - (その他: 出版社サイト) 4
 - ニコニコ静画 3
 - eBookJapan 3
 - (その他: iOSアプリ) 3
 - Honto 2
 - 小説家になろう 2
- Google Books
- パピレス
- とんりのヤングジャンプ
- i文庫S
- BookLive
- Jコミ
- GA-BON
- 漫画 on Web
- Book Walker
- ブックパス
- パプー
- マンガゲット
- 言い値書店
- マンガボックス
- Comixology
- 小説を読もう
- Newsstand
- eBOOKs
- Yahoo!ブックストア
- ガンガンONLINE
- Kinoppy

(前回の振り返り)

- ドキュメントフォーマットとは？
- ドキュメントフォーマットの切り口
- フォーマットの識別、判別
- プレインテキストのフォーマット
 - テキスト / 文字コード
- テキストフォーマットの例
 - Eメールのフォーマット
 - 軽量テキストフォーマット

質問・要望へのお返事 (1)

- いまいちこの授業での良いレポートというのがつかめないため、優秀なレポートがあったら公開して頂けるとありがたい。後学のためにもよろしく願います。

次回以降、検討します。なお、文例等については、第6回に紹介したレポートの作成法・作文技術の参考文献も確認するようにしてください。

質問・要望へのお返事 (2)

- 期日通りに提出してBより、遅れ提出でAの方が評価が低いということによいでしょうか。

Yes.

採点基準をあわせるために、遅れ提出の採点であっても「A」等をつけていますが、教員側での採点表には提出日・遅れ有無も記録してありますので、最終評定の際にはその分の減点が反映されます。

ドキュメントフォーマット (2)

LaTeX

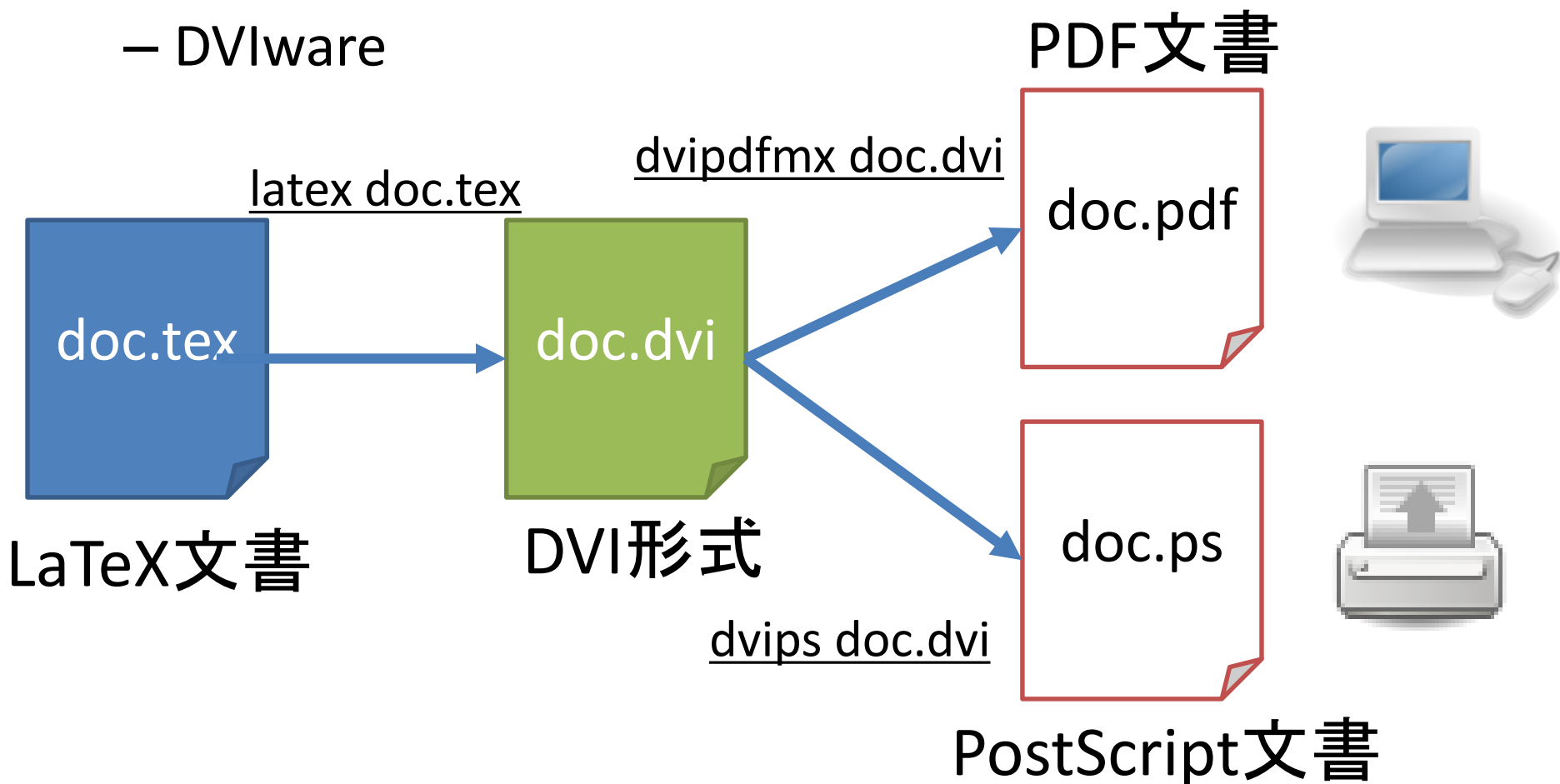
- テキストベースの文書記述言語
- 組版ツール
- (発音: ラテフ, ラテック, レイテック)
- 組版ツールTeX by Donald Knuth
 - 1978年から開発 ← 『The Art of Computer Programming』
 - 数式対応に優れている
 - 豊富な拡張機能(マクロ, フォント)
 - 一種のプログラミング言語
- LaTeXは、TeXを文書構造記述に拡張したもの
 - 日本語対応版: pLaTeX

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



LaTeXの実用例

- 理論計算機科学者による開発
 - 計算機科学、数学、物理学等の学術文献(論文)での利用が多い
- 例:
 - ACM論文執筆用テンプレート:
<http://www.acm.org/sigs/publications/proceedings-templates>
 - Springer論文執筆用テンプレート:
<http://www.springer.com/authors/book+authors?SGWID=0-154102-12-970131-0>
- 前回の

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
 - dvi_{ps}pdfmx: 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

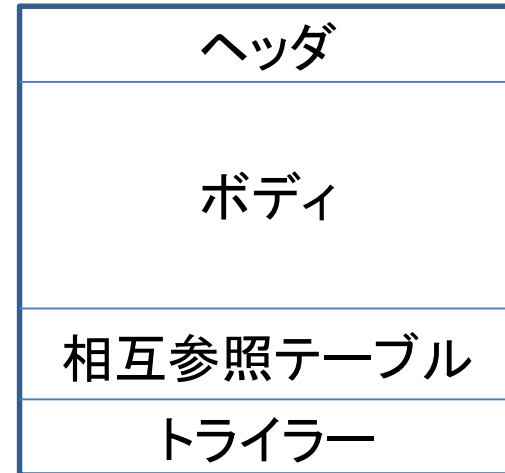
- 複雑な仕様
 - PDF 1.7 (ISO 32000-1, 2008) は750ページ以上(付録含む)
 - http://www.adobe.com/jp/devnet/pdf/pdf_reference.html
 - ISO標準化以前はAdobe一社による管理→デファクト標準の部分も一部残る
- ディスプレイでの閲覧支援とは言うものの...
- しおり, リンク, コメント, 注釈に関しては、作成時につける
 - Adobe Readerで閲覧しながら付けることはできない
 - 専用アプリケーションソフトウェアが必要

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

オフィス文書のフォーマット

オフィス文書の具体例

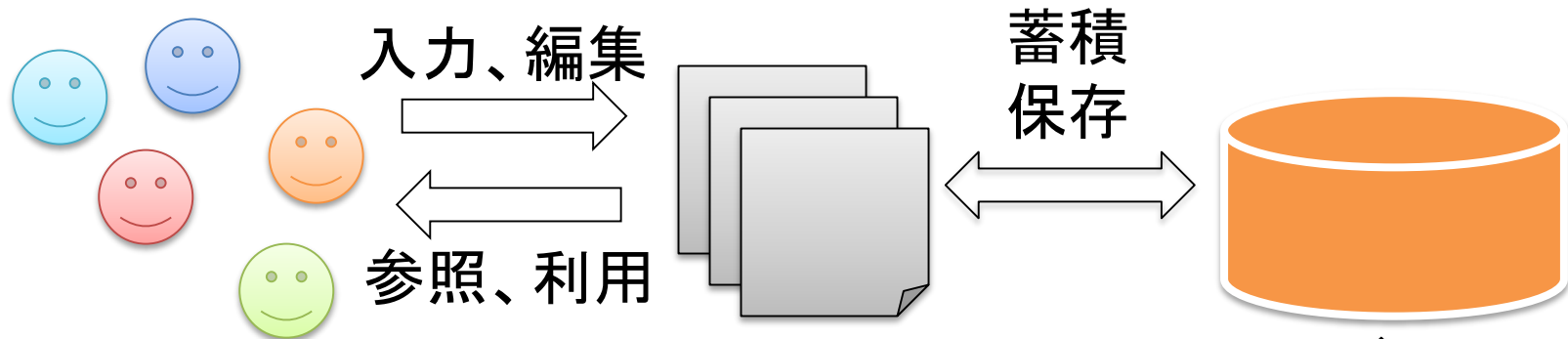
- 会議資料
- 議事録
- プレゼンテーション資料
- 広報資料
- 案内、お知らせ
- 会計書類：見積書、発注書、納品書、請求書
- 出張願、出張依頼、出張復命書
- 報告書
- 企画書
- 説明書、マニュアル
- 業務メモ
- 経歴書、業績書
- 連絡文書
- レターカバー

オフィス文書とデジタルドキュメント

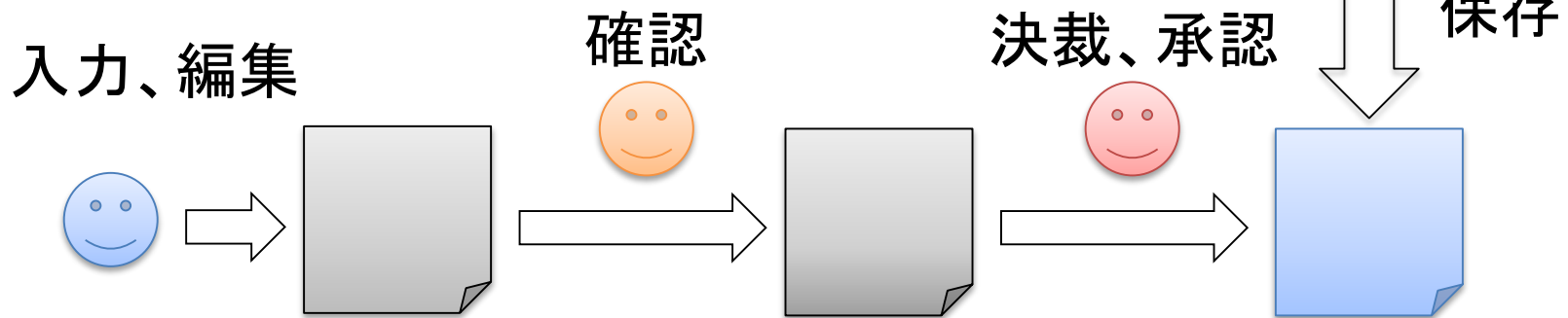
- オフィス文書はデジタルドキュメントの王道
- 文書のサイクルが出現
 - 作成
 - 編集
 - 承認
 - 管理
 - 配信・提供
 - 保存管理
- 関わるひとが多いことも特徴
 - ドキュメントの作成者
 - 共同入力者・編集者
 - 承認・決裁者
 - 利用者
 - 管理者
 - 2次的な利用者
- 紙の文書との併用も

オフィス文書に関する処理の種類

1. 文書共有、保存



2. 文書入力、決裁



オフィス文書の特徴

- 複数人による情報共有が前提とされている
 - 他者による引継ぎや検査
- 情報の保存
 - 日付(保存期間)、担当者など
- 定型形式
- 紙文書としての処理フロー
- 版管理機能

統合オフィスソフトウェアの事例

(オフィススイート; Office suite)

名称	文書	表計算	プレゼンテーション	画像処理
Microsoft Office	Word	Excel	Powerpoint	
LibreOffice	Writer	Calc	Impress	Draw
Google Docs	Docs	Spreadsheet, Form	Presentation	
iWorks	Pages	Numbers	Keynote	
JUST Suite	一太郎	三四郎		花子
Lotus (IBM)	WordPro	1-2-3		

オフィスソフトウェアが取り扱う ドキュメントフォーマット

- それぞれの文書ソフトウェア毎に独自フォーマットが用いられることが多い。
- 1990年代後半からは、Windowsとともに、Microsoft Office製品が市場シェアを伸ばし、それらのソフトウェアがオフィス文書コンテンツの代表格となってきた。
- 2000年代以降、オープンフォーマットの重要性が高まり、標準化された文書ソフトウェアが伸張している。
 - OpenOffice.org (2000-2011)
 - →ドキュメントフォーマット ODF (Open Document Format)
- また、別の動きとしてオンライン上での文書ソフトウェアの提供の動きもある。
 - Google Docs

Microsoft Officeのドキュメントフォーマット

- 長年の間、Microsoft Officeは独自形式(バイナリデータ)を用いていた
 - Microsoft Officeの初期バージョン以来、MS Office 2004まで
 - MS Office 2004以前の環境におけるドキュメントフォーマットは一社独占による文書形式 (proprietary)
 - 拡張子: .doc, .xls, .ppt
- 仕様の詳細が明らかでなかったため、MS社以外によるドキュメント利用が困難
 - 別のOS・機種における閲覧可能性が低い
 - 検索エンジンによるテキスト検索などでの応用
 - 社内文書の管理、検索においても問題
- 公的部門における調達への準拠(グローバル化への対応)

オフィス文書における

ドキュメントフォーマットの標準化 (ODF)

- 2000年にOpenOffice.orgがオープンソースソフトウェアとして公開されると、MS Officeの代替としての期待が高まった。同時に、データ交換のためにも、オープンフォーマットへの期待、ドキュメントフォーマットの互換性への期待も高まった。
- 2002-2005年にかけて議論され標準化された。
- Open Document Format (ODF)
 - 国際標準ISO/IEC 26300:2006

オフィス文書における ドキュメントフォーマットの標準化 (OOXML)

- ODFの動きに呼応して、Microsoft側も、Office 2007リリースにあわせ、新しいドキュメントフォーマットを提案、標準化
- Office Open XML (OpenXML, OOXML) の提案
- 2006年：ヨーロッパの標準化団体ECMAにおける標準化
 - ECMA-376
- 2008年：国際標準化機構ISOにおける標準化
 - ISO/IEC 29500

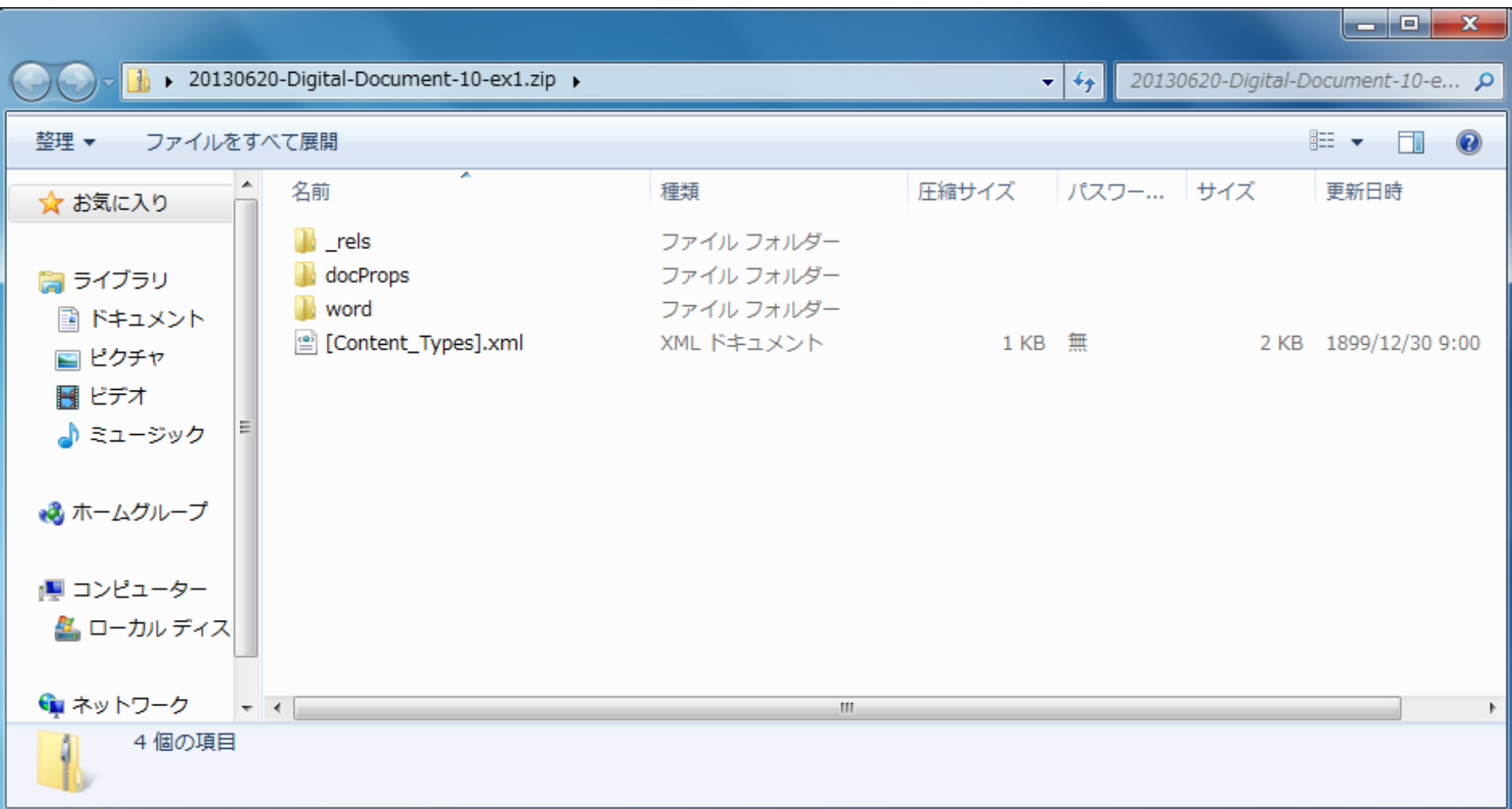
OOXML と ODF

- 国際標準化に向けた競争の激化
- 両者のドキュメントフォーマットは基本的には、かなりの部分で類似のもの
 - (しいて言えば、OOXMLの方がより複雑: Microsoft Officeが持つ機能全体に対応)
- 3つの主要な文書種別をカバー
 - ワードプロセッシング、スプレッドシート、プレゼンテーション
 - それぞれ Word/Writer, Excel/Calc, Powerpoint/Impress に対応
 - ODFフォーマットの拡張子: .odt, .ods, .odp
 - OOXML新フォーマットの拡張子: .docx, .xlsx, .pptx
 - WordPressingML, SpreadsheetML, PresenationML
- 複合オブジェクトを許す汎用形式
 - パッケージ形式 (Zip)
 - XML (コンテンツ+メタデータ)
 - マルチメディアファイル

OOXMLにおけるパッケージ形式

- Zip形式によるアーカイブ化＋圧縮
 - XML形式以外のファイル(画像等)の埋め込み、展開
 - テキスト形式であるXML文書の圧縮によるファイルサイズ縮小に貢献
- 任意のWord文書等の拡張子を .zip に変更してZipファイルとして展開、閲覧してみる
 - [Content_Types].xml
 - _rels/.rels
 - *word/document.xml* , ...
 - *docProps/core.xml* , ...

OOXMLパッケージの展開例



OOXMLパッケージの展開例

_rels/.rels

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<Relationships
xmlns="http://schemas.openxmlformats.org/package/2006/relationships
">
  <Relationship Id="rId3"
Type="http://schemas.openxmlformats.org/officeDocument/2006/relati
onships/extended-properties" Target="docProps/app.xml"/>
  <Relationship Id="rId2"
Type="http://schemas.openxmlformats.org/package/2006/relationships/
metadata/core-properties" Target="docProps/core.xml"/>
  <Relationship Id="rId1"
Type="http://schemas.openxmlformats.org/officeDocument/2006/relati
onships/officeDocument" Target="word/document.xml"/>
</Relationships>
```


OOXMLパッケージの展開例

word/document.xml

```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <w:document mc:Ignorable="w14 w15 wp14"
  xmlns:wps="http://schemas.microsoft.com/office/word/2010/wordprocessingShape"
  xmlns:wne="http://schemas.microsoft.com/office/word/2006/wordml"
  xmlns:wpi="http://schemas.microsoft.com/office/word/2010/wordprocessingInk"
  xmlns:wpg="http://schemas.microsoft.com/office/word/2010/wordprocessingGroup"
  xmlns:w15="http://schemas.microsoft.com/office/word/2012/wordml"
  xmlns:w14="http://schemas.microsoft.com/office/word/2010/wordml"
  xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main"
  xmlns:w10="urn:schemas-microsoft-com:office:word"
  xmlns:wp="http://schemas.openxmlformats.org/drawingml/2006/wordprocessingDrawing"
  xmlns:wp14="http://schemas.microsoft.com/office/word/2010/wordprocessingDrawing"
  xmlns:v="urn:schemas-microsoft-com:vml"
  xmlns:m="http://schemas.openxmlformats.org/officeDocument/2006/math"
  xmlns:r="http://schemas.openxmlformats.org/officeDocument/2006/relationships"
  xmlns:o="urn:schemas-microsoft-com:office:office"
  xmlns:mc="http://schemas.openxmlformats.org/markup-compatibility/2006"
  xmlns:wpc="http://schemas.microsoft.com/office/word/2010/wordprocessingCanvas">
  - <w:body>
    - <w:p w:rsidP="00A770D1" w:rsidRDefault="00A770D1" w:rsidRPr="00A770D1"
      w:rsidR="000C7B86">
      - <w:pPr>
        <w:pStyle w:val="1"/>
      </w:pPr>
      - <w:r w:rsidRPr="00A770D1">
        <w:t>デジタルドキュメント文書の例</w:t>
      </w:r>
    </w:p>
    - <w:p w:rsidP="00A770D1" w:rsidRDefault="00A770D1" w:rsidR="00A770D1">
```

ドキュメントフォーマットの切り口（再掲）(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

ドキュメントフォーマットの切り口（再掲）(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
 - プレインテキスト → 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

本日のまとめ

- 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
- オフィス文書とデジタルドキュメント
 - ワークフロー
 - 処理モデル
- OOXMLとODF
 - 標準化(独占規格から国際標準へ)
 - パッケージ形式による複合オブジェクトの埋め込み
 - 文書内容にあわせた文書要素

第4回レポート課題

- 以下の13種類のドキュメントフォーマットから、1つを取り上げて、デジタルドキュメントとしての具体例を挙げながら、そのフォーマットがどのような領域分野で用いられているか、閲覧・作成環境とその特徴、国際標準との関連について文章で説明してください
- 対象とするドキュメントフォーマット
 - HTML, SGML, XML
 - LaTeX, PDF, PostScript
 - EPub, Amazon Kindle, XMDF
 - Markdown
 - OOXML, ODF, RTF
- なお、取り上げたフォーマットに関する参考文献を一つ以上、必ず記載すること(SIST-02に準拠)

第4回レポート課題 (2)

- A4用紙:1ページ以上2ページ以内にまとめる
 - 2ページにわたる場合は裏面に記載のこと
- 課題番号(第4回レポート課題)、提出年月日、学籍番号、所属、氏名を提出用紙の一番上に必ず記入する
- 提出〆切:2014年6月17日(火)15:00
 - 春日エリア学群学務前のレポート提出BOXへ