

# PDF活用のための基礎知識

2007年6月

アンテナハウス株式会社  
<http://www.antenna.co.jp>

小林 徳滋  
[koba@antenna.co.jp](mailto:koba@antenna.co.jp)

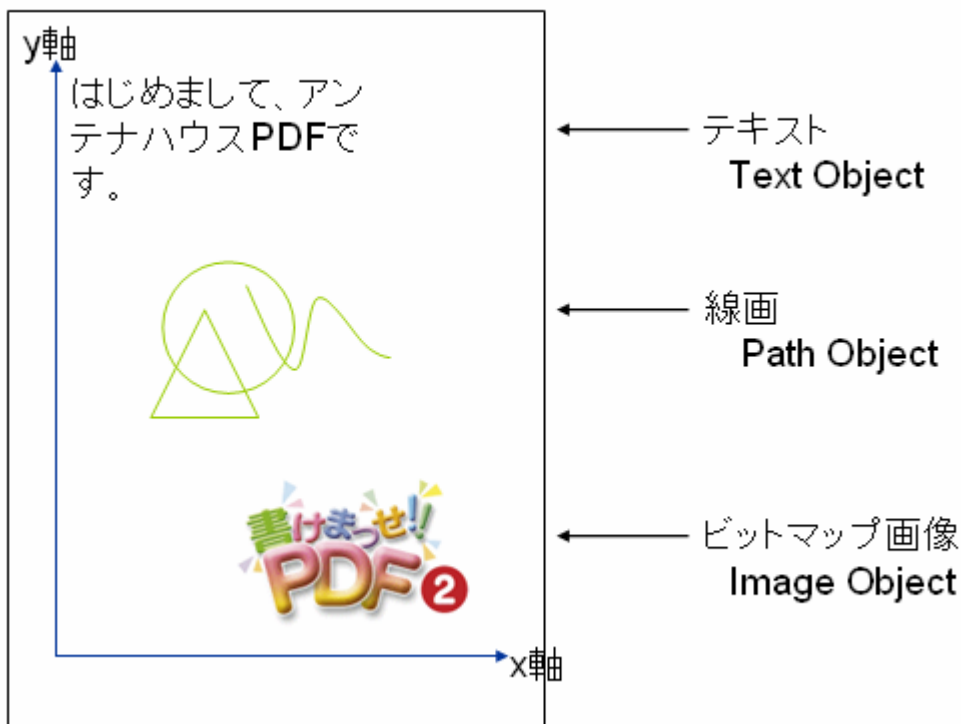
# 今日のお話の内容

1. PDFとは
2. PDFの作成方法、その進化
3. PDFで文字表示する仕組みとフォント埋め込み
4. PDFの構造とリニアライズ
5. PDFの編集
  - 本文編集（タッチアップ、上書き、ページの入れ替え）
  - しおりの作成と編集
  - 注釈・リンクの編集
6. PDFの再利用
7. 出版・印刷におけるPDF利用
8. PDFの将来

# PDFは電子の紙

- 紙への表記をデジタル形式で実現したもの

PDFの内容表示用の主なオブジェクト



線画オブジェクトは2次元座標系の上に数学的な直線・曲線（パス）として表現されることが出来ます。そうしたパスに線幅指定、色指定したり、パスで囲む領域を塗り潰したりすることで、図形が表現されます。

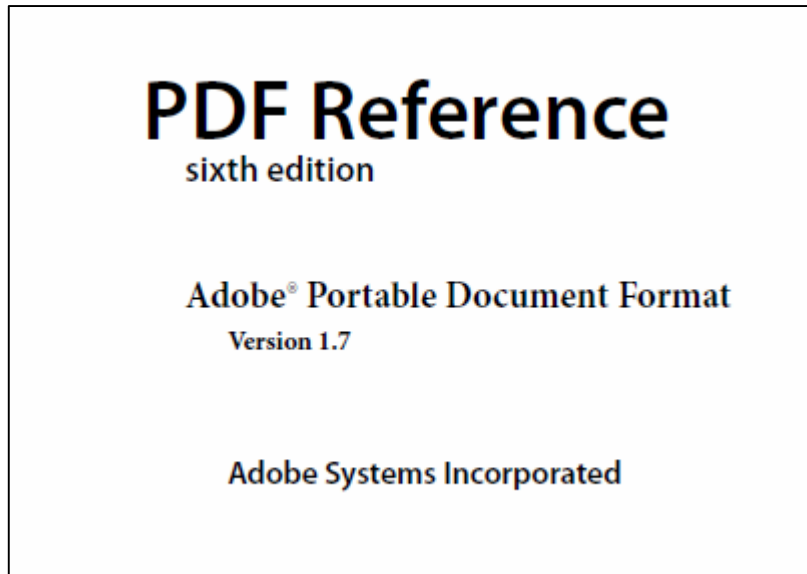
文字はビットマップとしてドットの塗り潰しで表すか、あるいは、文字の輪郭(アウトライン)を曲線で表して囲まれた部分を塗り潰すアウトラインフォントの方式で表現します。

PDFファイルは、オブジェクトを規定するデジタルデータの塊

# 誰でもPDFソフトを開発できる

- ファイル形式「**P**ortable **D**ocument **F**ormat」の頭文字

PDF仕様を規定する仕様書(PDF Reference)が公開されています。



「この本はPDFのファイル形式についての説明を提供し、第一に、PDFファイルを直接生成するPDF Producerアプリケーションの開発者のためのものである。また、開発者が既存のPDFファイルを読んで、内容を解釈し、変更するPDF Consumerアプリケーションを記述するのに十分な情報を含んでいる。」

1993年 発表

2007年1月29日

ISOの標準にするため仕様をAIIIMIに譲渡し、国際標準化作業中

- Acrobat: AdobeのPDF作成・加工ソフトの名前
- Adobe Reader: PDFの表示ソフト名

# 相互運用が重要課題

## PDFの生成ソフト

Acrobat / Distiller

Antenna House PDF Driver

XSL Formatter (XML⇒PDF)

## PDFの消費ソフト

Adobe Reader

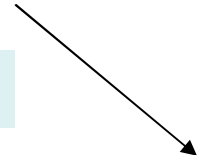
PDF RIP

PDFから商業印刷

リッチテキストPDF 3

いきなりPDF to Data 2

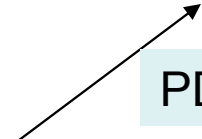
PDFを生成



配布・流通

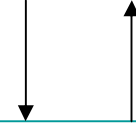
PDFファイル

PDFを表示・活用



PDFを編集・加工

PDFの編集ソフト



Acrobat

いきなりPDF Edit 2

書けまっせ!!PDF

いきなりPDF Professional 3

# 起源はデジタル印刷技術

## デスクトップ・パブリッシング

PageMaker → Quark → InDesign

## ページプリンタ

(1ページ全体を描画してからプリント)

LaserWriter (Apple)

## ページ記述言語

PostScript

1997年PostScript 3

PDFはPostScriptを超えた

PDF ワークフローへ

DTP 3種の神器

1993年PDFの誕生  
(Acrobat Distiller)

PDFの進化

PostScriptからPDFに変換

DTPソフトはPDFを直接入出力

1980年代後半

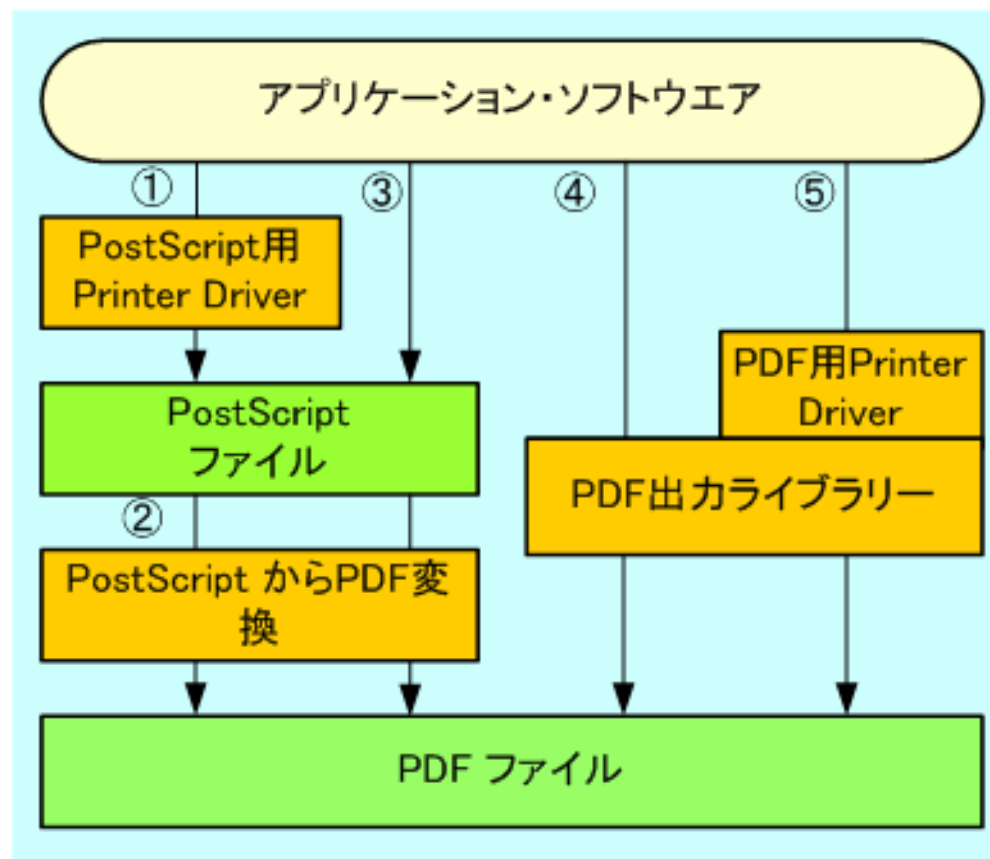
1990年代後半

2000年代後半

# 例外として紙のPDF化も

- 紙のPDF化は印刷起源のPDFとは別のもの
  - 拡大・縮小に弱い
- 紙をスキャナーで読み取り画像化、  
⇒その画像をPDF化したもの
- PDFにすることで、複数ページの管理、メタデータをつけて管理が簡単に
- 透明テキスト付きPDF: OCRで認識したテキストを上配置し検索可能に

# デジタル生まれPDFの作成方法



- ① Word, 一太郎などアプリケーションから PostScript プリンタ・ドライバを使って PostScript ファイルを作る。
- ② PostScript ファイルを PDF ファイルに変換する。PostScript ファイルはまた、Webなどで流通している PostScript ファイルを使うこともある。
- ③ PostScript プリンタ・ドライバを使わずに、PostScript を作成し、PostScript を PDF ファイルに変換する。
- ④ Word, 一太郎などアプリケーションから、プリンタ・ドライバを通さずに PDF を作成する
- ⑤ アプリケーションの「印刷」メニューで見える PDF 作成用のプリンタ・ドライバを使って PDF ファイルを作成する。

①～③はPostScriptを経由する

④～⑤はPDFの命令を直接作成する



# PDF作成ソフトを分類すると

印刷機能を使う＝プリンタ・ドライバ方式

①

Acrobat

PrimoPDF

Jaws PDF Creator

Postscriptを經由する

Acrobat Distiller

GhostScript

②

Antenna House PDF Driver

⑤

DocuCom PDF Driver

SkyCom クセロPDF

Postscriptを經由しない

PDFLib

iText

④

印刷機能を使わない＝API提供

参考資料: [さまざまなPDF作成技術の概観](http://www.antenna.co.jp/PDF/reference/CreatePDF.htm)

<http://www.antenna.co.jp/PDF/reference/CreatePDF.htm>

# 主要なPDF作成用ドライバ

## • PDF作成用ドライバの機能比較表

製品名	Antenna House PDF Driver	Acrobat 8 Professional	SkyPDF Pro	クセロPDF2
バージョン	3.2.0	8.0.0	2.53	2.00β3
メーカー	アンテナハウス	アドビシステムズ社	スカイコム	クセロ
Windows Vista への正式対応	○	×	×	○(但し、β版)
2007 Office system上でのアドインによるPDF作成	○	×	×	×
JIS X 0213:2004 の 11,233文字を全て正しく扱えるか?	○	○	×(※1)	×(※2)
AES 暗号化方式	○	○	×	×
Web 表示用に最適化(リニアライズ)	○	○	×	×
透かし(ウォーターマーク)	○	○	○	×
JPEG2000 圧縮	○	○	○	×

○ Acrobat 8.1で Vista, Office2007対応

2007年4月24日現在の最新版を使用。各社とも製品が改善されていることがありますので、最新情報はお確かめください。

※1、Unicode の補助面に割り当てられている 303文字の検索ができない。

※2、Unicode の補助面に割り当てられている 303文字、Unicode にコードポイントのない 25文字をPDF に出力できない。

詳細情報: <http://www.antenna.co.jp/pdv/matrix.htm>

# PDF作成方法の変化

- PostScript (PS) 経由方式 ⇒ PDF Lib方式へ
  - PDF Lib方式の方が高速
  - PSでできないことが増えている
  - PS技術は過去形
- アプリケーションの「PDFで保存」(Save As PDF)
  - Office 2007
  - OpenOffice.org (Sun StarSuite)
  - DTP (Illustrator, InDesign)

# サーバサイドPDFの普及

- Webサーバでの高速PDF作成
  - 国税庁 確定申告書等作成コーナーなど
  - 航空券の申込み(E-commerce)など
- 帳票レベルではかなり普及している
- 文書のサーバサイドPDF化の課題
  - 高速高精度組版エンジンの重要性
    - XSL-FO (XSL Formatter)

PostScriptはWebサーバ上でのダイナミックなPDF生成には、遅すぎて使えない

# 文字コードとグリフ

- 文字コードとは言語情報交換・処理のために作成した抽象的な文字集合の要素番号。
- グリフ(字形)とは文字の表記形。
  - アウトラインフォントでは曲線を表すパラメータとして実現される。
- 文字コードとグリフは一対一対応ではない。

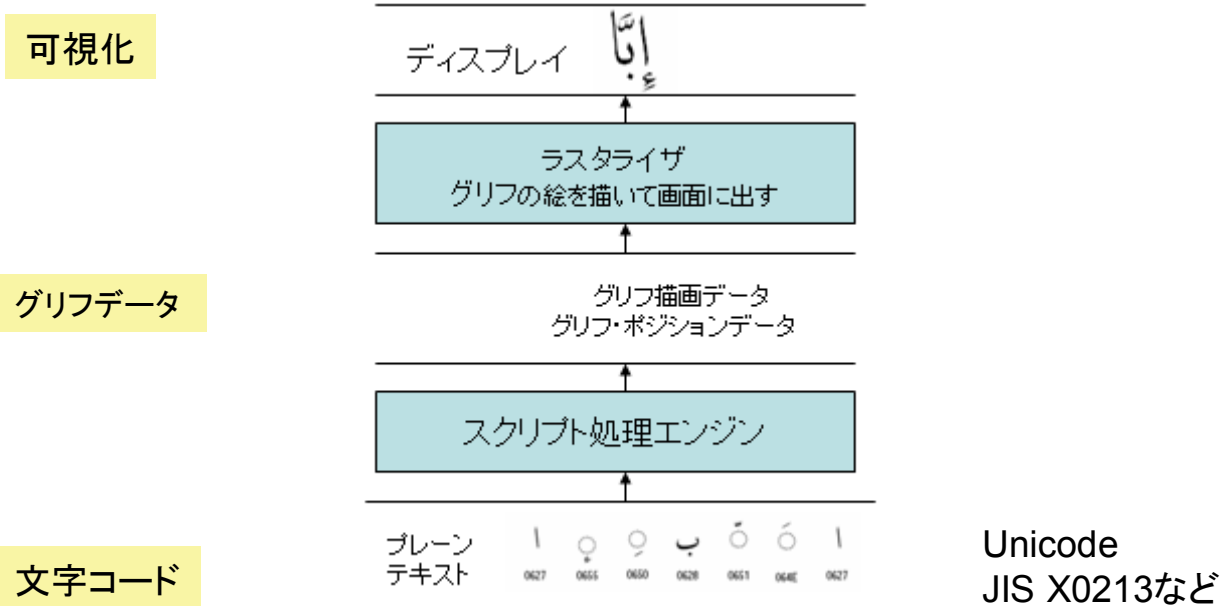
多対多対応

詳細は

コンピュータによるテキスト表記とPDFのフォント埋め込みについて  
<http://www.antenna.co.jp/pdf/reference/FontEmbedding.htm>

# 文字コードの可視化

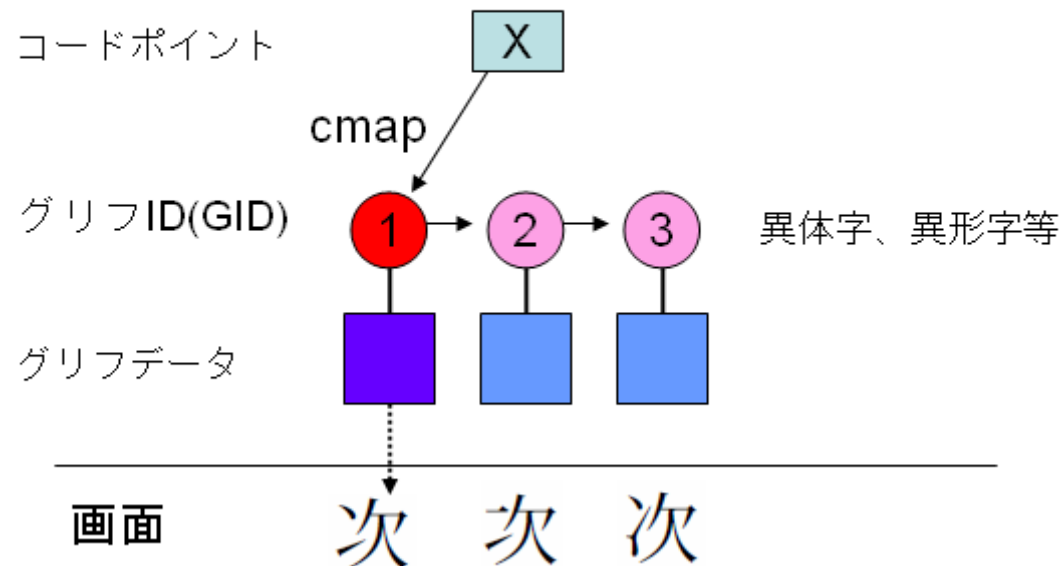
- 文字コードの並びは、フォントのグリフデータを用いて可視化される。



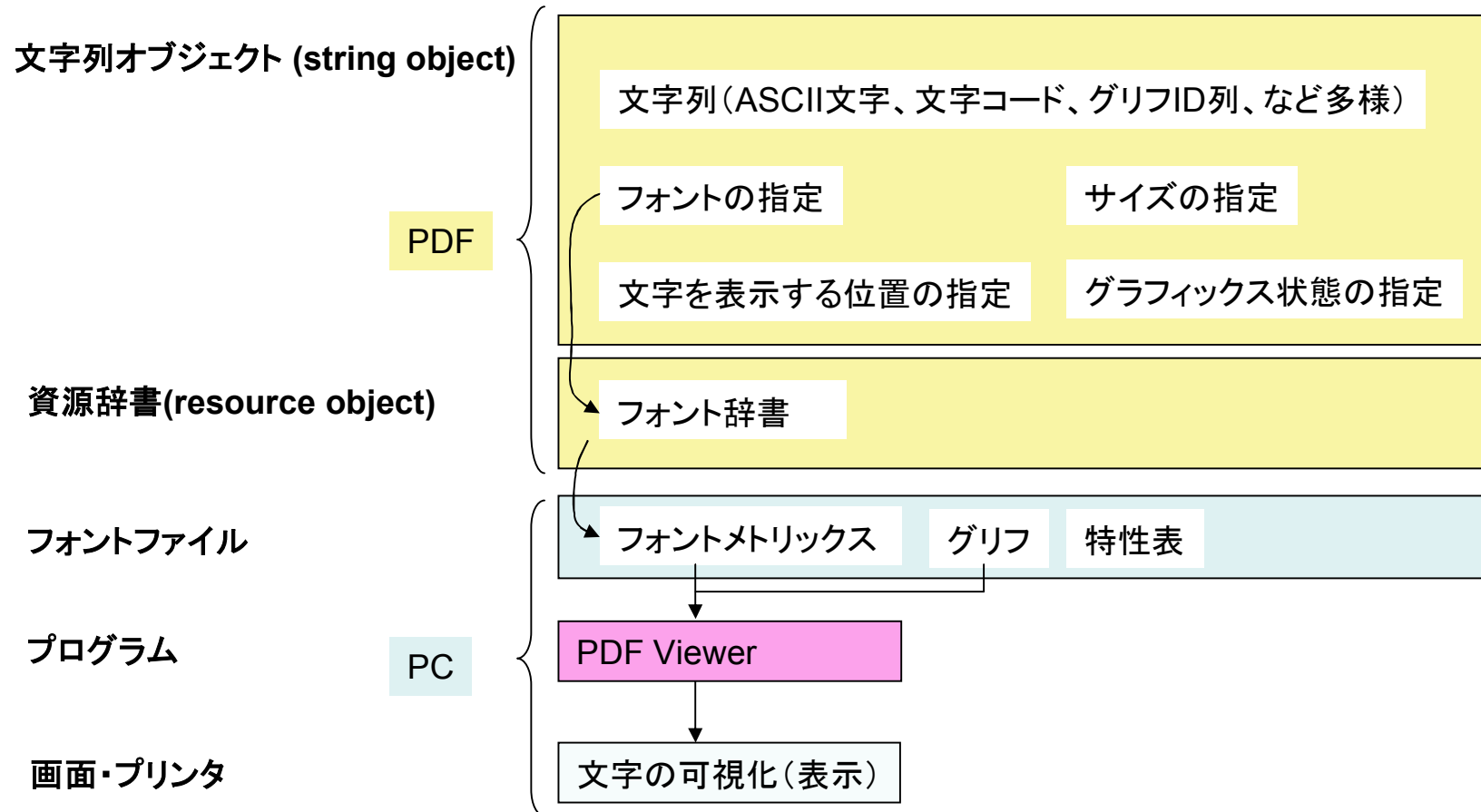
# コードポイントとグリフ

- 漢字では文字コード上のひとつのコードポイントに多数の字体＝異体字がありえる。

コードポイントとグリフデータ



# PDFでの文字表示





# フォント埋め込み

- フォントを埋め込まないPDFの表示
  - 受信相手のシステム上のフォントを使う。
  - Adobe Readerは必ずしもそうになっていない。
  - 特に海外へ送るときは注意が必要。
    - 例) 日本語の文字が少しでも入っていると英語の Adobe Readerではまったく表示できない。  
Windows\Fontsにフォントがあってもだめ。
- フォントを埋め込んだPDF
  - PDFにフォントのサブセットが添付される。全文字ではなく使っている文字だけが原則。

# フォントを埋め込んだPDF

文字列オブジェクト (string object)

PDF

グリフID列(原則として)

フォントの指定

サイズの指定

文字を表示する位置の指定

グラフィックス状態の指定

資源辞書(resource object)

フォント辞書

フォントファイル

フォントメトリックス

グリフ

特性表

プログラム

PC

PDF Viewer

画面・プリンタ

文字の可視化(表示)

フォントを埋め込むとフォントファイルから必要な情報がPDFに取り込まれる

# フォント埋め込みと著作権

- フォント(プログラム)は著作権保護の対象
- フォント埋め込みは、フォントのアウトラインデータをPDFに複製する処理
- 許可なくフォントを埋め込むと著作権侵害
- 多くのフォントは埋め込みが許可されている
- 埋め込み許可フラグ
  - TrueType、OpenTypeはフォント内にフラグ有り
  - PDF作成ソフトでプログラムでフラグを見て埋め込み許可を判定できる

# PDFファイル構造はWeb以前

作成直後のPDF

ヘッダ 例) %PDF-1.5
本体
相互参照表
トレイラ

PDFであることを識別するための情報

PDFの本体情報

PDFの本体にランダムアクセスするための情報

PDFファイルは最後にファイルサイズ、カタログ情報、暗号辞書などが登録されている

標準ではトレイラが最後にあるため、Adobe Readerなどの利用アプリケーションは通常、PDFファイルの一番後ろから読まねばならない。

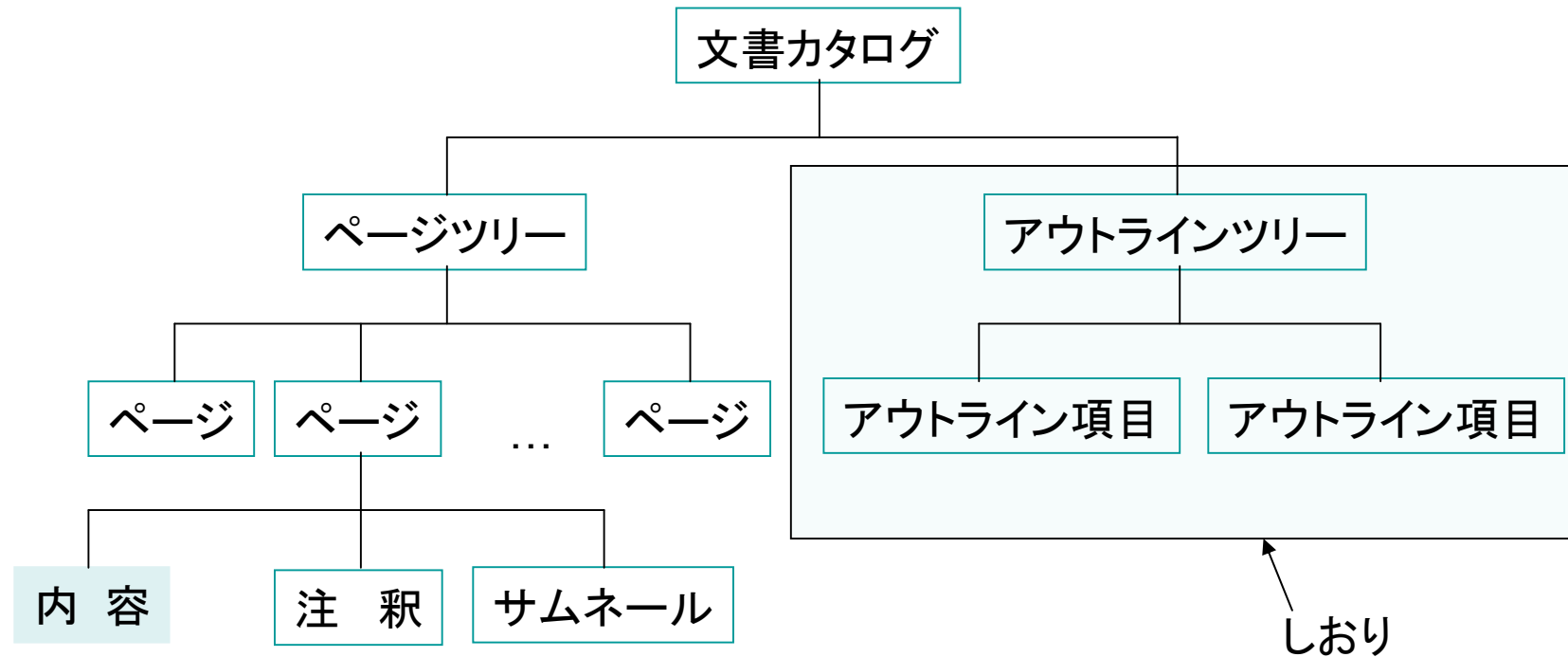
このため、ファイルの容量が大きい（ページ数の多い）PDFをWeb経由で表示しようとすると、全部ダウンロードするまで、画面には、内容がまったく表示できません。

# リニアライズ(Web最適化)

- Web経由でPDFをできるだけ早く表示するためのPDFのオプション仕様
- リニアライズはPDFファイルの内容を並び替える処理。
- リニアライズされたPDFなら、指定ページの内容を直ちにに取り出して表示できます。

参考: PDF Reference Appendix F

# PDF本体の構造(抜粋)



- ①PDFの内容は頁単位になっている(ワープロ文書(例:Word)とは違う)
- ②しおりの情報は、アウトラインツリーとして別管理になっている
- ③注釈は1頁毎に管理されていて、かつ、頁の内容とは別管理になっている

# PDFの編集

PDFはレイアウト済みの情報を記録したもののなのでレイアウトの大幅な変更は無理。

1. 本文の削除・挿入＝タッチアップ
2. ページに上書き・追加記入する
3. ページ単位の削除、結合、入れ替え
4. しおりの付加
5. 注釈(Annotation)、リンク、添付ファイル
6. フィールドへのデータ挿入
7. 電子署名

# PDFのページ内容

- PDFには1頁毎にページの内容を描画するための情報が保存されています。
- 文字は、行の一部である文字列単位で位置を指定して配置済み
- 線画を描画する線の命令など(ラインアート)
- 画像(イメージ画像など)



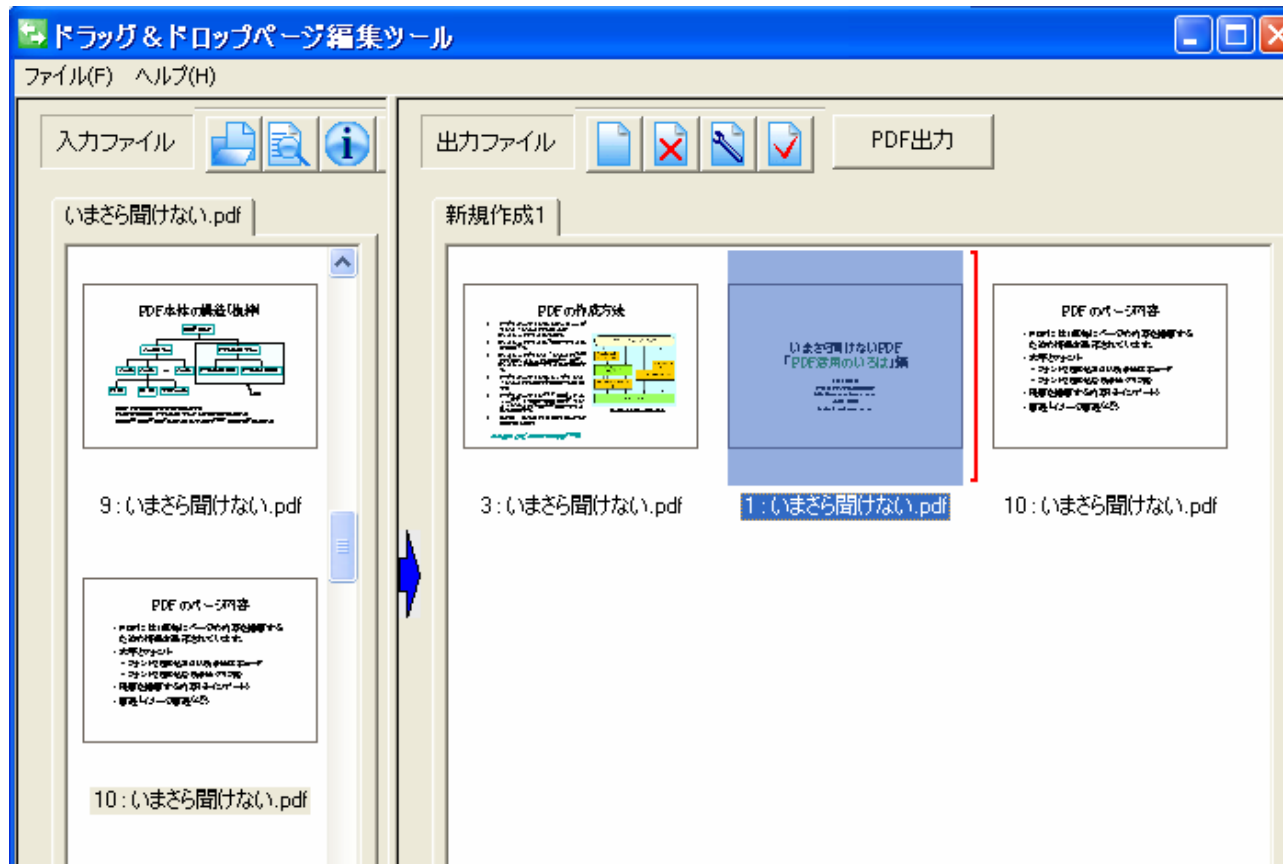
Q PDFに書き込む事は可能なんですか？  
どんなソフトがあれば書き込みできるのでしょうか？

PDFへの書き込みにはいろいろなレベルがあります。

1. PDFの本文の加筆・訂正（タッチアップ）  
「Adobe Acrobat」,「いきなりPDF Edit!」
2. PDFの空き欄（申請書など）に書き込み  
「書けませPDF」,「やさしくPDFへ文字入力」  
— 注釈やコメントを追加する  
「いきなりPDF Professional 3」

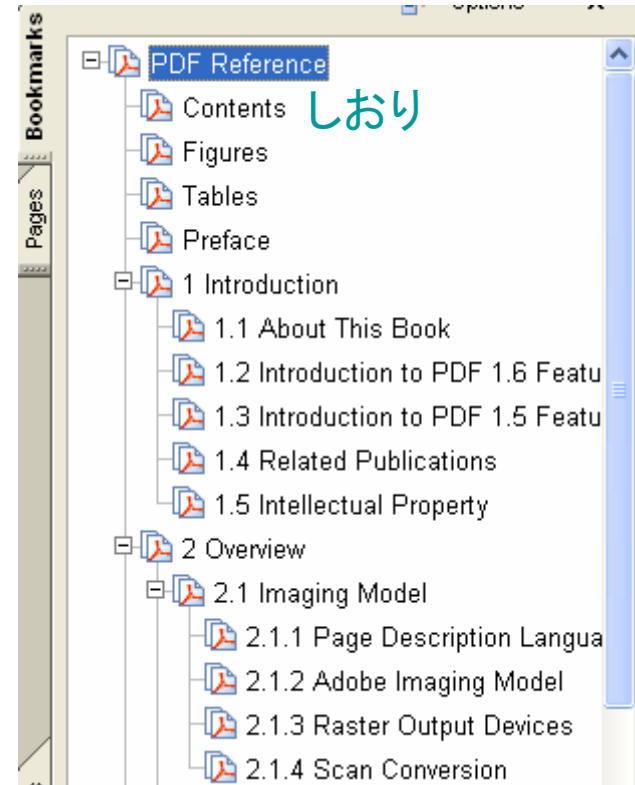
# PDFページ編集の例

- ページ単位で結合したり、分割したりできます。



# PDFのしおり

- しおり (Bookmark、アウトライン項目)
- しおりを階層化したツリーがアウトライン
  - 文書構造を表示する目次になる
- ドキュメント・カタログにて本体のページとは別に管理される



アウトライン

# PDF作成時のしおりの作成

しおりは便利なのに普及度は低い

- PDFを「印刷」メニューから作成する方法では、しおりを作成できない
  - 「印刷」メニューでは、紙への印刷動作を想定
- マクロを使ってOfficeと対話しながらPDFを作成時に「しおり」も出力する方式が必要
  - Antenna House PDF Driver : [Officeアドイン](#)
  - [Adobe PDF Maker](#)

# しおりの追加・削除・更新

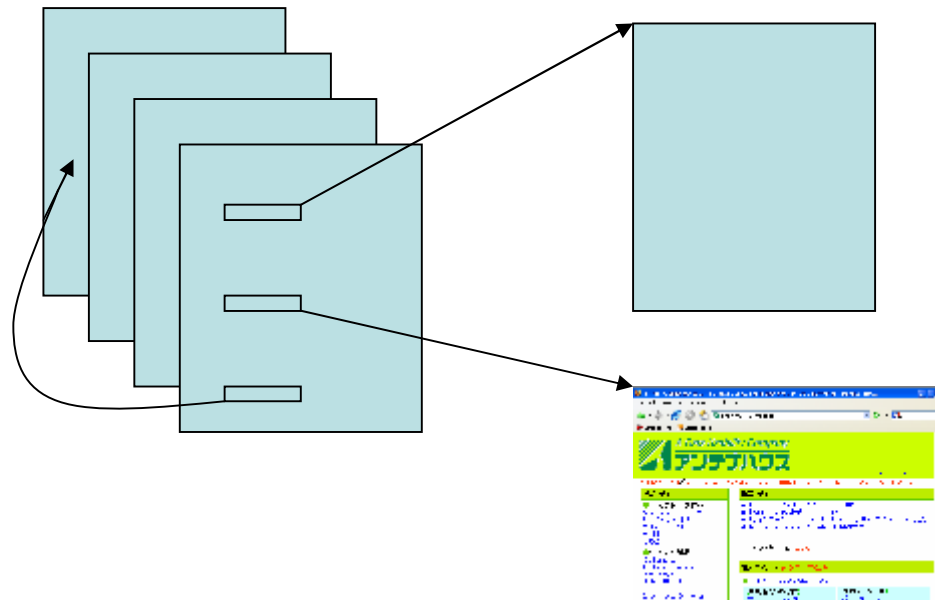
- しおりはPDFのページ内容とは別に管理されています。
  - ページの内容と別に、追加、更新、削除が可
- しおり作成デモ
  - 「アウトライナー」でしおり編集のデモ

# 注釈(コメント)など

- 注釈(コメント注釈など)は付加情報なので本文と別に管理されています。
- PDF 1.6の注釈の種類(Text annotation、Link annotation、Free text annotation、Line annotation。。。)
  - 「PDF Reference」 p.581
- 注釈編集ソフト
  - 「いきなりPDF Professional 3」

# PDFのリンク編集

- PDFではリンクは注釈の中の一種類  
– PDFの頁内容と別に追加、更新、削除可  
– 頁の内容には影響を与えない



# PDFデータの再利用

- PDFページ内容のデータ形式は、表示・印刷以外での利用を想定していない。
- 表示上繋がっているテキストは、PDFの内部で繋がっていないこともある。
  - 音声読み上げで、正しい順序で読めないことがある。
- (PDF生成ソフトによっては)PDFで表示される文字をテキストに取り出せないことがある。
  - PDFのテキストをコピーしたり、検索したりできない。
  - PDF内フォントにToUnicodeCMAPが設定されていないため。



# PDFデータ再利用ツール

- PDFからワープロへの変換
- 実現方法は次の二通りに分かれる。
  1. PDFを画像化してOCRする。OCR結果からワープロ文書、表計算に変換する
    - [いきなりPDF to Data2](#) (パナソニックソリューション+ソースネクスト)
    - [速攻！PDF to Data2](#) (SkyCOM+クロスランゲージ)
  2. PDFを解読して文字の配置、書式から文書構造を推定していく
    - [リッチテキストPDF3](#) (アンテナハウス)
    - その他海外ソフトの日本語版

# 出版、印刷におけるPDFの活用

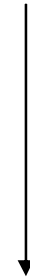
- 制作と印刷の中間段階として
  - ポストPostScriptとしての利用
- PDFを紙の代替メディアとするコンテンツ配信
  - 有価証券報告書、IR情報(無償)
  - JIS規格書(表示のみ無償、印刷可PDFは有償)
  - O' Reilly Safari Books (PDF有償販売)



# PDFによるコンテンツ配信

- デジタルデータはコピーしても情報劣化なし
  - オリジナルと複製は全く同等
  - なにも制限しないとコピーし放題になる
- セキュリティの設定
  - パスワードセキュリティ
  - 電子証明書による暗号化
  - 簡易的DRM: KeyRing
  - Digital Rights Management (DRM)

低コスト、緩い制限



高コスト、厳格な制限

# PDFによる有償コンテンツ配信

- 広告媒体価値を考えるならば、コピーされてもサーキュレーションの多い方が良いが広告料金にコピー数を反映する方法は？
- コンテンツの対価を得るためには、コピー制限を掛けるのは必須

# PDFのファイルサイズ

- ファイルサイズを決定する要因
  - 文字: フォント埋め込みの有無
  - 画像のサイズ
  - オブジェクト圧縮
- 画像ファイルは、影響が大きいので、印刷用の高品質画像とWeb配布用のサイズ最適化ファイルの2種類を用意して、目的に応じて使い分けるべき

# PDFの競合技術

- 競合製品または技術
  - DocuWorks (ゼロックス)
  - XPS (XML Paper Specification: マイクロソフト)
  - XAMLによる配信 (.NET、マイクロソフト)
  - XSL-FOによる配信 (W3C)
  - CSS3による配信 (W3C)
  - Flash (アドビ)

XAML, XSL-FO, CSS3は、HTML (ブラウザ) のレイアウトを高度化する技術

# PDF 2.0

- ISO標準のPDFに大きな期待
  - アドビの試行錯誤からの脱却が必要
- 紙に文字を表記する如く簡単に取り扱いたい
  - 基盤技術を構築し直す必要がある
  - 過去の遺産を引きずり過ぎ
  - 現PDFは超複雑
  - 重すぎ
- マルチメディアは？

Q PDFにflashを入れられるようですが、もう少し詳しく知りたい。

A

- FlashのサンプルA

- Flashを用意して、AcrobatでPDFを開き、ムービーツールを使用して、Flashを挿入する矩形領域を指定します。その後で開かれるファイル選択ダイアログでFlashのファイルを指定します。

- FlashのサンプルB

- Internet Explorer でFlash が含まれているサイトを開き、Acrobat のアドインで PDFを作成することができます。



Q PDFにformを設定して利用する方法を、もう少し詳しく知りたい。

A

• Formのサンプル

- Acrobat 7の「高度な編集」ツールバーを使用して、PDF上に配置することができます。
- Adobe Reader では チェックボックスのチェック、テキストボックスへの入力などはできますが、PDF 保存はできません。その状態で印刷することはできます。また、入力したデータだけを保存したり、メールで送信することができます。
- この機能を使用して、アンケートをPDFで作成して配布、回答だけを返信してもらうということができます。
- Sample-Form.pdf: Adobe Readerでフォームへ入力し、「電子メールで送信」をクリックすると入力データを送信したり、ファイル保存することができます。
- Sample-Form.pdf: Acrobat で開き、上記の保存データを「アドバンス」メニューのフォームから「データをフォームに取り込む」を使用することで、入力されたデータを表示することができます。(Acrobat 7でないとは取り込めない)

## Q 通常のPDFとPDF/X系のファイルの構造の違いについて

- A**
- PDF/Xは、PDFにより印刷用のデータ交換を行うための仕様です。このためPDFで使える機能に制限を設けています。
  - 通常のPDFとPDF/X系のファイルには構造の違いはありません。
  - 主な制限は、
    - 色空間の制限: PDF/X-1aは、CMYK、スポットカラーなどに限定
    - 圧縮の制限: LZW、JBIG2など使用禁止
    - 暗号化の制限
    - 透明の使用制限などです。
  - アンテナハウスPDF資料室にまとめてあります。

<http://www.antenna.co.jp/PDF/reference/>

- ◆ [PDF/Xについて \(1\) PDF/X-1a](#) – 2006/07/26
- ◆ [PDF/Xについて \(2\) PDF/X-3](#) – 2006/08/06