グループワーク① 「本文 PDF の作り方 －OCR ソフトウェアとアクロバット－」

＜ 近畿大学中央図書館 ： 伊豆田 幸司 ・ 青木 斐 ＞

本日の目的： 機関リポジトリで学術情報を公開するにあたり、電子データが存在しない紀要・雑誌等の資料について、スキャナと 2 種類のソフトを使用して、電子データを作成する方法を学ぶ

本文 PDF を作成する前に

原資料の媒体によって、作業方法は異なる。

1. 電子媒体

PDF ： そのまま利用
※透明テキストが付いていない場合は作成

Word や Power Point 等のファイル： 印刷コマンドで PDF に変換

2. 紙媒体･･･ 雑誌や図書の遡及が中心

裁断できる場合 ： 裁断後、オートフィーダでスキャン
同時に OCR(Optical Character Reader)も実施

裁断できない場合： 1 枚ずつスキャン→画像ファイルを作成後、一気に OCR

※OCR の主な目的とは？

・画像形式の PDF を検索可能な PDF へ変換する
・メタデータ入力のために本文をテキスト化する

設定を細かく変更し、修正を加えれば、OCR精度は向上する
⇒ データの容量や作業効率との兼ね合いが重要

〈今回の使用機器およびソフト〉
●スキャン
スキャナ ： EPSON ES-10000G
オートドキュメントフィーダ ： EPSON ESA3ADF2
※紙枚数 100 枚(80g/m²) 縦：A3～A5、横：A4～A6、両面対応

スキャナ ： FUJITSU FI-6130
※自動給紙方式 ADF（オートドキュメントフィーダ）
給紙枚数 50 枚 (80g/m²) 最大：縦 A４、最小：A8、両面対応

●OCR ソフト
Adobe Acrobat 9.0 Professional と e.Typist V.13.0 を併用

1： e.Typist を活用して、透明テキスト付きPDFを作成する

1. e.Typist を起動

2. [ファイル]→[スキャナの選択]→[スキャナドライブの選択]ウインドウで、スキャナを選択し、[OK]ボタンをクリック

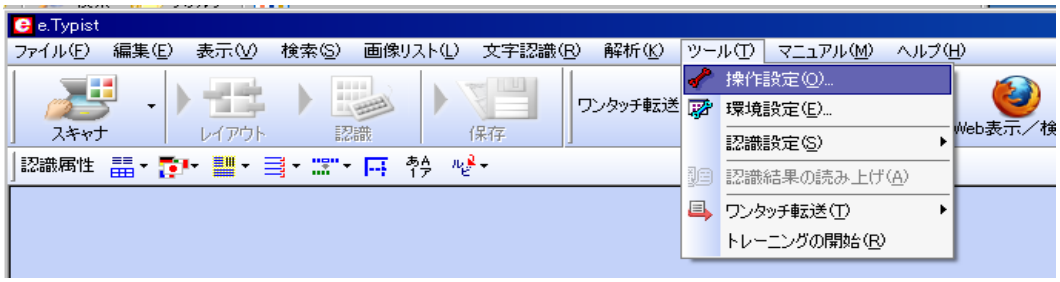
3. [スキャナ読込]モードを選択

※[ファイル読込]モードを選択すると、画像ファイルから透明テキスト付き PDF を作成できる

4. [文字認識]→[認識言語指定]→[日本語(欧文混合)]を選択
   ※言語に応じて、認識言語の設定を変更
   V. 13. 0 からの追加機能： アジア言語(簡体字・繁体字・ハングル)の透明テキスト付きPDFを作成できる

※混合認識時の欧文言語を選択

[ツール]→[操作設定]→[認識]→[混合認識時の欧文言語]を選択(複数選択も可能)→[OK]ボタンをクリック



混在認識時言語は、複数選択することもできるが、
その分、全体的な認識精度が落ちるので、選択は
慎重に行う。

日本語とアジア言語の混合認識はできない。

5. [ワンタッチ転送 Acrobat]→[高圧縮透明テキスト付きPDF]を選択

6. [Acrobat に転送]ボタンをクリック→スキャナの「TWAIN ドライバ」ウインドウが表示→原稿に応じて、スキャンの設定を変更

【基本設定(近畿大学の場合)】

解像度・・・基本 400dpi
サイズ・・・等倍
イメージタイプ(カラータイプ)・・・原稿に応じて、モノクロ、グレー(8bit)、カラー(24bit)を使い分ける

※カラー写真や、モノクロでもグラフ等の精密な画像が含まれる場合は、それぞれのイメージタイプで PDF を作成し、モノクロを基調として、後からカラーやグレー設定のページと差し替える

7. スキャンを実施

8. スキャン終了後、スキャナの「TWAIN ドライバ」ウインドウを閉じる→[レイアウト]→[認識]→Acrobat に転送
⇒ここからは自動

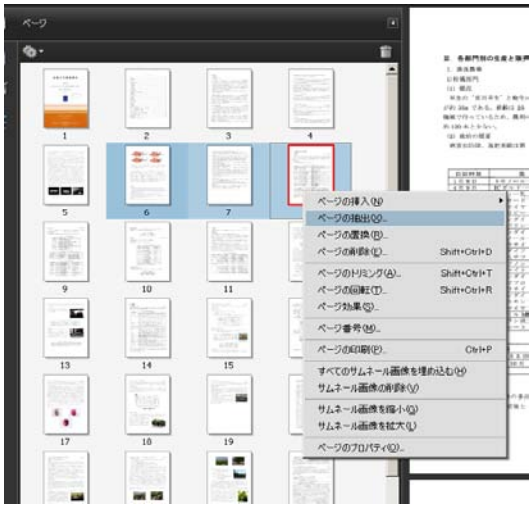
9. Acrobat のウインドウが表示→[ファイル]→[名前をつけて保存]

2: Adobe Acrobat を活用して、PDFを加工する

※ファイルを加工する前は、必ずバックアップを取る

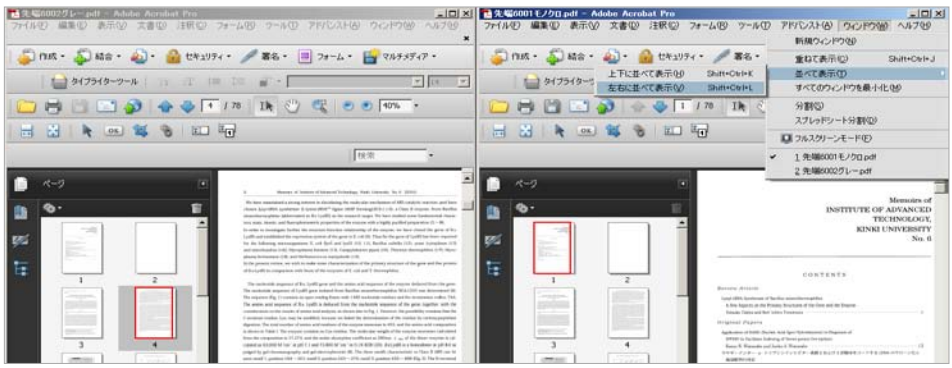
1. 加工したいPDFを開き、左のバー上で右クリック→[ページ]をクリックにサムネイルを表示

2. サムネイル上で右クリック→ 必要があれば、ページの挿入・抽出・置換・削除・回転等を実施



※イメージタイプが異なるページを差し替えて、モノクロベースの公開ＰＤＦを作成する場合

加工したいファイルをすべて開く→[ウインドウ]→[並べて表示]→[左右に並べて表示]をクリック
→差し替えたいページをドラッグ＆ドロップするとコピー＆ペーストできる



2. タイトルや作者等を(メモ帳等に)コピー＆ペーストして、一時保存

3. [ファイル]→[プロパティ]→概要にタイトルや作成者等を入力→ファイルを保存（メタデータ用にバックアップ）

4. PDF にセキュリティを設定する

開いているPDFを閉じる→[アドバンスト]→[文書処理]→[バッチ処理]→[シーケンスの編集]
→[コマンドの選択]ウインドウを表示して、各種設定を編集する

【セキュリティの基本設定（近畿大学の場合）】

セキュリティ方法：パスワードによるセキュリティ
文書のパスワード：いいえ（閲覧にパスワード入力は不要）
内容をコピー：許可しない
文書を変更：許可しない
印刷：許可
高品質で印刷：許可

バッチシーケンスウインドウで、[セキュリティ]を選択し、[シーケンスを実行]ボタンをクリック

シーケンスの実行確認ウインドウが表示されるので、[OK]ボタンをクリック

[処理するファイルを選択]ウインドウが表示されるので、PDF を選択

[フォルダの参照]ウインドウが表示されるので、保存先を指定し、[OK]ボタンをクリック

[権限パスワードの入力]ウインドウ→パスワードを入力し、[OK]ボタンをクリック

[権限パスワードの確認]ウインドウ→パスワードを入力し、[OK]ボタンをクリック

PDF 完成

【参考文献】

XooNIps 入力マニュアル（奈良大学仕様）
http://nijc.brain.riken.jp/xoonips/index.php?Documents

「近畿における機関リポジトリコミュニティ形成の支援」連続研修会　第3回事例報告　資料
「複合機でスキャン -貧乏暇無し、自力でのＰＤＦ化- 」森下 映理（奈良女子大学図書課電子情報係）
http://cont.library.osaka-u.ac.jp/kinki3/3-nara_wu.html

「機関リポジトリ登録論文のデジタル化について: 市販の OCR ソフトにて作成した透明テキストデータの調査」外崎　みゆき（関東学院大学図書館）
http://opac.kanto-gakuin.ac.jp/cgi-bin/retrieve/sr_detail.cgi?U_CHARSET=EUC-JP&CGILANG=japanese&SUNO=&HTMLFILE=sr_sform.html&SRC_BODY=1&ID=NI90000004&PID=NI90000004