

# 資料の電子化マニュアル

## 目次

1.電子化の流れ.....	4
1-1.資料の分別.....	4
1-2.電子化作業.....	4
1-3.チェック作業.....	4
1.データ内容チェック.....	4
2.データ形式・ファイル名チェック.....	4
3.PDF 分割、統合.....	5
2.電子化データの作成について.....	6
3.「透明テキスト付 PDF」について.....	7
4.OCR について.....	8
5.スキャナについて.....	9
5-1.フラットベッドスキャナ.....	9
5-2.ODF（オートドキュメントフィーダ）スキャナ.....	9
5-3.ブックスキャナ.....	9
5-4.スタンドスキャナー.....	9
6. OCR ソフトについて.....	10
7.資料の電子化の方法.....	11
7-1.裁断.....	11
7-2.スキャン.....	11
7-2-1.スキャン設定.....	11
7-2-2.ファイルの保存.....	11
7-2-3.ファイルの分割・結合.....	11
8.e.Typist の使い方.....	13
8-1. 操作設定.....	13
8-2.連続自動処理の設定.....	13
8-3.認識の設定.....	14
8-4.ファイル読込の設定.....	14
8-5.保存先の設定.....	14
8-6.連続自動処理の開始.....	15
8-7.加工前ファイル選択.....	15
8-8.ページ選択.....	15
8-9.実行.....	16
8-10.修正.....	16
8-11.保存.....	17

8-12.確認.....	17
8-13.文字認識 .....	17
9.AcrobatDC の使い方.....	19
9-1.acrobatDC の起動 .....	19
9-2.ファイルを選択.....	19
9-3.ファイルの追加.....	20
9-4.フォルダを指定.....	20
9-5.出力先指定 .....	21
9-6.テキスト認識の設定 .....	22
9-7.フォルダの確認.....	23
9-8.加工後フォルダ確認 .....	24
10.フォルダ構造 .....	25
10-1.加工前 .....	25
10-2.加工後 .....	25
10-3.アップ用 .....	26
11.参考資料.....	27

# 1.電子化の流れ

## 1-1.資料の分別

電子化出来ているものと、出来ていないものに分ける。

### 【電子化出来ているもの】

パソコンで見たり聞いたりできるデジタル化されたもの（テキスト、画像ファイル、動画、音声）、紙媒体の電子化データ（画像）には透明テキストが付加されているもの。

### 【電子化出来ていないもの】

パソコンで見られないアナログなもの（紙媒体、フィルム、テープなど）。

※紙媒体以外は業者に相談。

## 1-2.電子化作業

電子化出来ていないもの（紙媒体）を電子化する。

1.裁断

2.スキャン（スキャナの種類の説明）

3.OCR で透明テキスト付与（OCR ソフトの種類の説明、使い方）

詳しくは、「7.資料の電子化の方法」にて。

## 1-3.チェック作業

公開できる内容か確認する。

### 1.データ内容チェック

プライバシーを侵害していないか確認する。

引用などで、論文掲載に対しては許諾を取っていても、リポジトリ公開に対して許諾を取っていない場合の再許諾。

→問題があると思ったら著者に確認し、再度許諾申請する。

### 2.データ形式・ファイル名チェック

#### 【データ形式】

テキスト、画像、動画。ほとんどのファイルフォーマットに対応している。

登録時に登録出来なかったらリポジトリ部へ連絡する。

#### 【ファイル名作成ルール】

1.ファイル名は半角英数字に統一する。

2.病院名（機関コード）＋内容（雑誌論文や紀要など）＋番号

※頭5ケタは必ず機関コードにする。

3.内容や番号は各機関で規則性をもって決めればOK。

例)

八尾市立病院年報 1 号 P90-100

27031nenpo01090.pdf

八尾市立病院年報 25 号 P2-21

27031nenpo25002.pdf

巻・号・頁の桁合わせ（ボリュームに合わせて二ケタにするか三ケタにするなど）は各機関で統一する。

ゼロ埋めしないとソートを掛けた時にファイル順番が変わる為。

ハイフン「-」やアンダースコア「\_」で区切ると分かりやすい。

ファイル名に推奨する文字は以下の通り。

半角英字：「a」～「z」

半角数字：「0」～「9」

ハイフン：「-」

アンダースコア：「\_」

※全角は使わない（文字化けする可能性があるため）。

※半角大文字英字「A」～「Z」や半角スペース「 」や半角カッコ「(」「)」や半角中点「・」やピリオド「.」やカンマ「,」や、「\$」「|」「¥」などの特殊記号などは使わない方がよい（OS の環境によって使えない場合がある）。

### 3.PDF 分割、統合

紀要など複数論文データを 1 ファイルでもらった場合は PDF ファイルを分割、統合する。

また一論文でもファイルの容量が大きい場合は分割する。

大きいファイルも登録することは可能だが、ダウンロードに時間がかかるため、ユーザーに優しい設定で登録する。（最大でも 200M 以下にする。）

ファイルの大きさは動画で 512M を超える場合はリポジトリ部に相談する。

## 2.電子化データの作成について

リポジトリにデータを登録するのに、紙媒体のままでは登録出来ない。

そのため、スキャナで取り込んで「PDF」化する。

(スキャンしただけでは、ただの画像のため、文字として認識出来ず、検索出来ない)

OCR ソフトを使って、「透明テキスト付 PDF」化する。

### 3. 「透明テキスト付 PDF」について

PDF には電子媒体経由の PDF と、紙媒体経由の PDF がある。

電子媒体経由の PDF→Word やエクセルなどのソフトから作成した PDF。文字データを有する。

紙媒体の PDF→ただの画像。そのため OCR 機能を使って文字データを付加する。

ただの画像に文字を付加したもの＝「透明テキスト付 PDF」

## 4.OCR について

OCR ; Optical Character Recognition

光学式文字読取

OCR とは、スキャナなどで入力された画像情報の中から、文字の形状に基づいて文字を識別し、コンピュータ上で扱える文字データへと変換する仕組みのことである。

OCR の機能を備えた装置やソフトウェアも、同じく OCR と呼ばれる。この場合の OCR は Optical Character Reader の略とされる。

( Weblio 辞書より)



## 5.スキャナについて

スキャナとは、印刷物や写真などを画像情報として読み取り、データとしてパソコンに保存するための装置。

- 1.フラットベッドスキャナ
  - 2.ODF オートドキュメントフィーダスキャナ
  - 3.ブックスキャナ
  - 4.スタンドスキャナー
- などがある。

### 5-1.フラットベッドスキャナ

原稿をガラス台に固定し、下から光を当てて読取装置を動かして画像を読み取るスキャナ。安価だが、1枚ずつ取る手間がかかる。

### 5-2.ODF（オートドキュメントフィーダ）スキャナ

自動読み取り装置付きスキャナ。紙をセットすると自動的に取り込むことができる。両面同時にスキャン出来て便利。

### 5-3.ブックスキャナ

読み取り面が角まであり、裁断せずに読み込むことができる。裁断出来ない本などに。

### 5-4.スタンドスキャナー

スキャナの下に本を広げた状態で読み取るため、裁断せずに読み込むことができる。裁断出来ない本などに。

## 6. OCR ソフトについて

OCR ソフトとは、スキャナで読み取った文字を解析し、文字データとして認識して入力するためのソフトウェアのことである。

e.Typist v.15.0

Adobe Acrobat DC

読み取り革命 v.15

などがある。

## 7.資料の電子化の方法

### 7-1.裁断

裁断機で裁断する

手動の場合は、刃が垂直に下りてくる垂直裁断型が適している。

### 7-2.スキャン

#### 7-2-1.スキャン設定

【スキャンの設定】 ScanSnap での設定例

ファイル形式：PDF

解像度：400dpi

枠消去：上下左右各 1 cm程度

連続読み取り：ON

カラーモード

カラー写真：カラーモード

白黒写真：グレースケール

#### 7-2-2.ファイルの保存

【ファイルの保存】

加工前（OCR化前）、加工後（OCR化後）のフォルダを作って格納するところをあらかじめ決めておく。

ファイル名は任意でわかりやすいように設定する。

（フォルダさえ分ければファイル名は同じでもよい。ファイル名の後ろに a や b をつけてもよい。アップロードするファイルではないので全角を使用してもよい。）

#### 7-2-3.ファイルの分割・結合

【ファイルの分割・結合】

スキャンしたデータを分割したい。

（別の論文にまたがっている時）

バラバラにスキャンしたデータを結合したい。

そんな時は、分割・結合ソフトを使う。

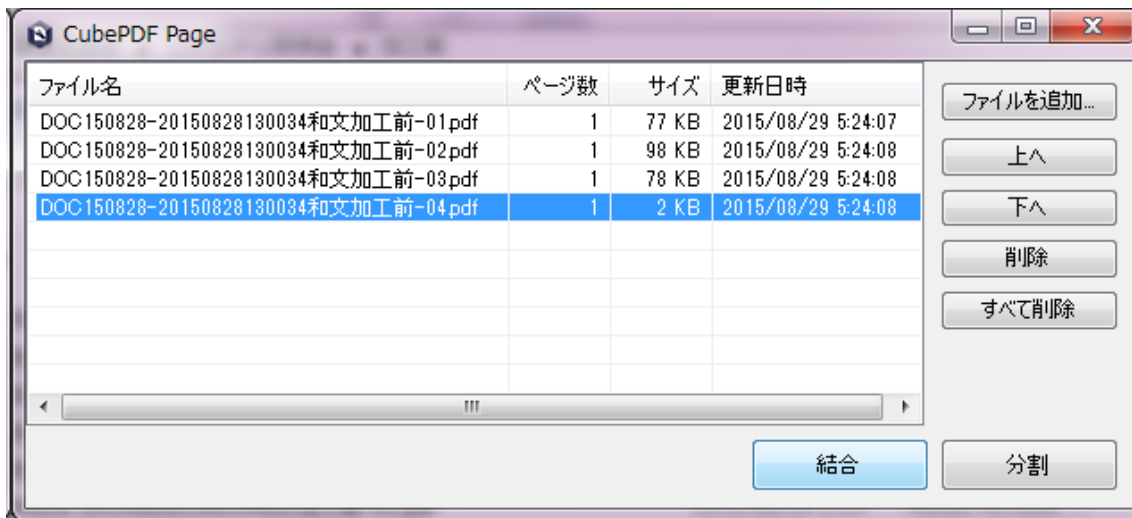
CubePDF Page…フリーソフト（使いやすい）

#### 7-2-4. CubePDF Page の使い方

【使い方】

ファイルをドラッグして

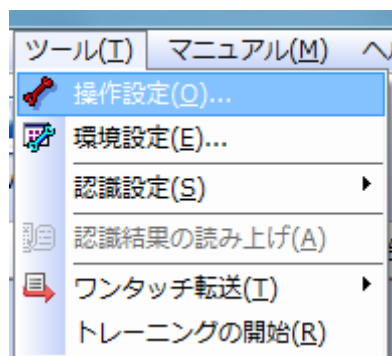
「分割」を押すとファイルが分割され、「結合」を押すと結合される。



## 8.e. Typist の使い方

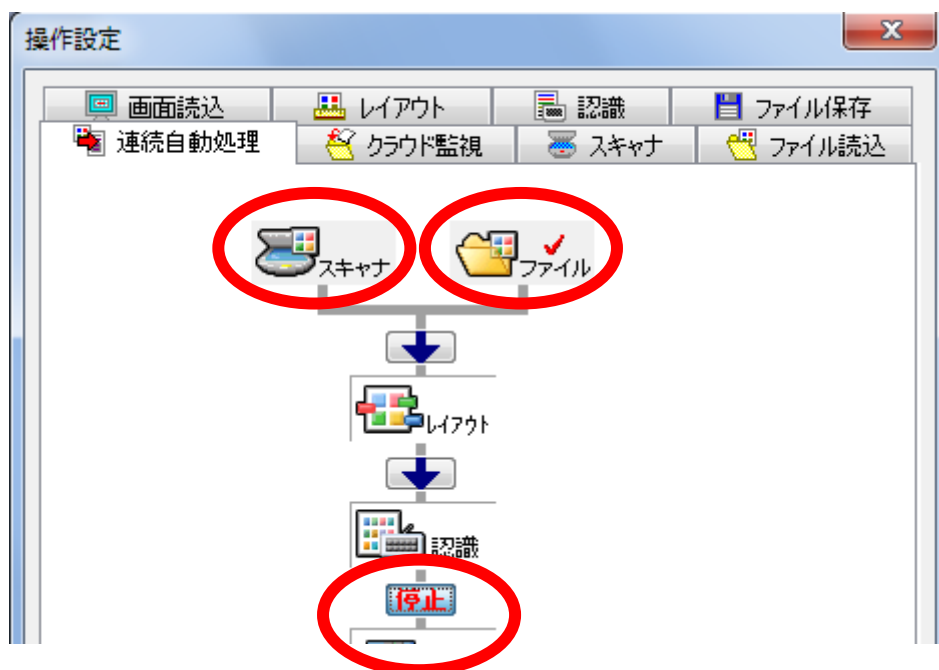
### 8-1. 操作設定

ツール→【操作設定】を開く。



### 8-2. 連続自動処理の設定

「スキャナ」※1 か「ファイル」※2 にチェックを入れる。認識のあとで「停止」※3 にする。



※1. 「スキャナ」…お持ちのスキャナが TWEIN 対応の場合はスキャナから直接取り込んで処理できる。（各スキャナの説明書を確認）

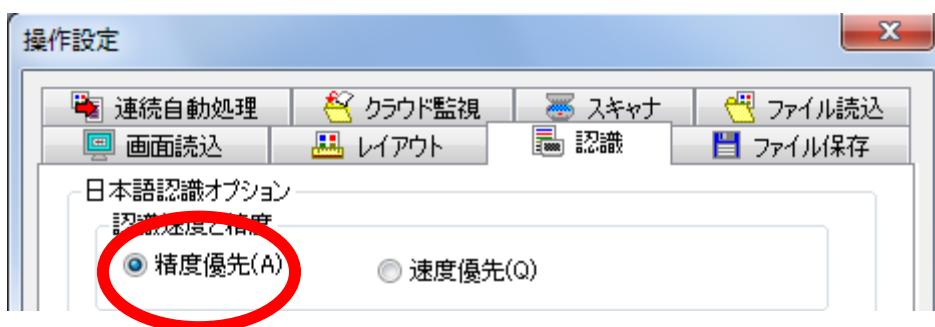
ScanSnap S1300 は TWEIN 非対応。スキャナから直接取り込みできない。

※2. 「ファイル」… TWEIN 非対応の場合。もしくは既画像ファイルがある場合はファイルを選択。

※3. 正しく取り込めているか確認する為、認識のあとで「停止」にする。  
TWEIN…画像入力装置用の技術的な規格のこと

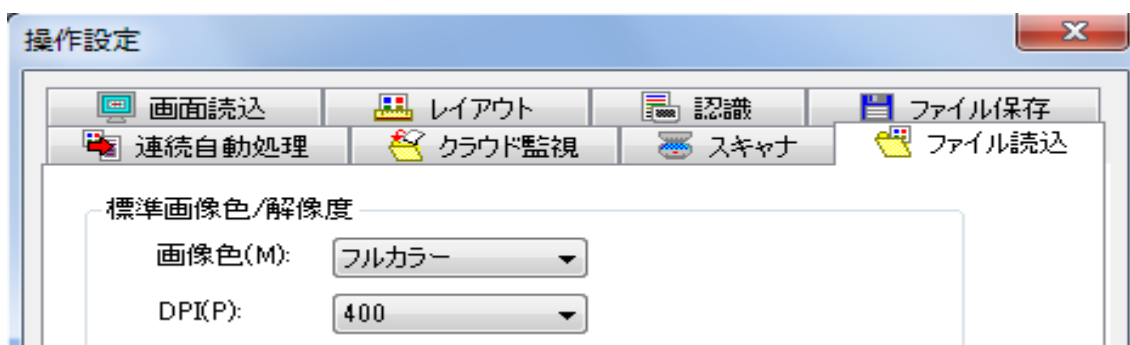
### 8-3. 認識の設定

認識速度と精度を「精度優先」で設定。



### 8-4. ファイル読込の設定

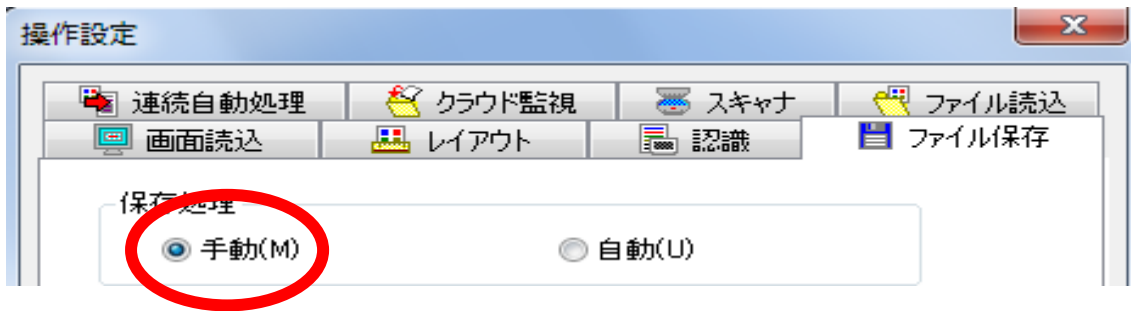
ファイル読込の画像色を  
カラー混在の場合… 「フルカラー」  
白黒のみの場合… 「白黒」  
解像度… 「400」



### 8-5. 保存先の設定

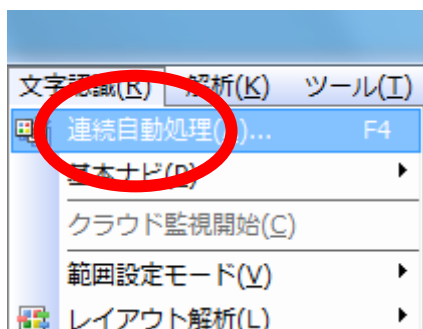
保存先を「手動」で設定。

単純なテキストファイルなら「自動」でも問題なく認識できる。  
最初は「手動」でやり方を覚える。



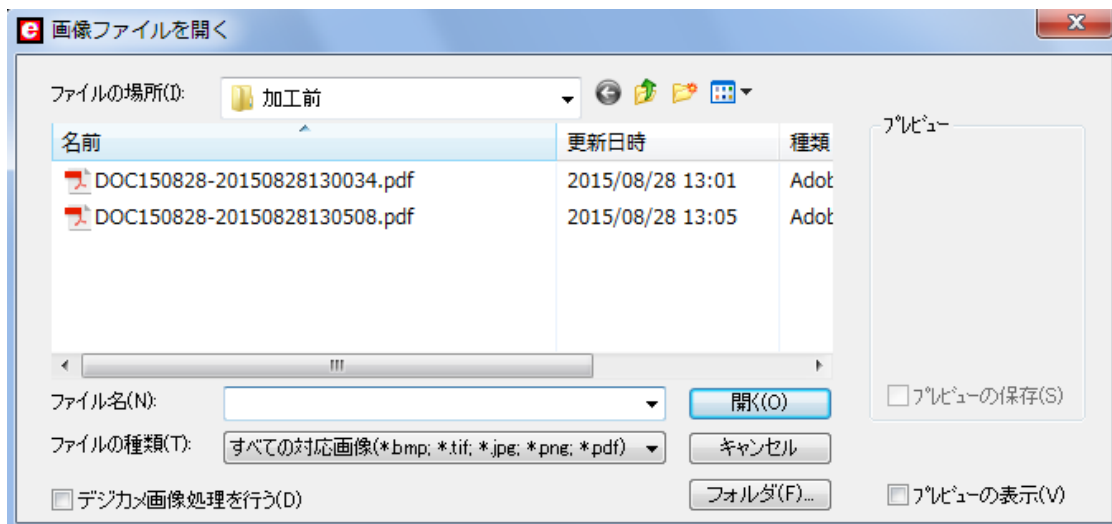
### 8-6.連続自動処理の開始

メニュー【文字認識】から「連続自動処理」を選択する。



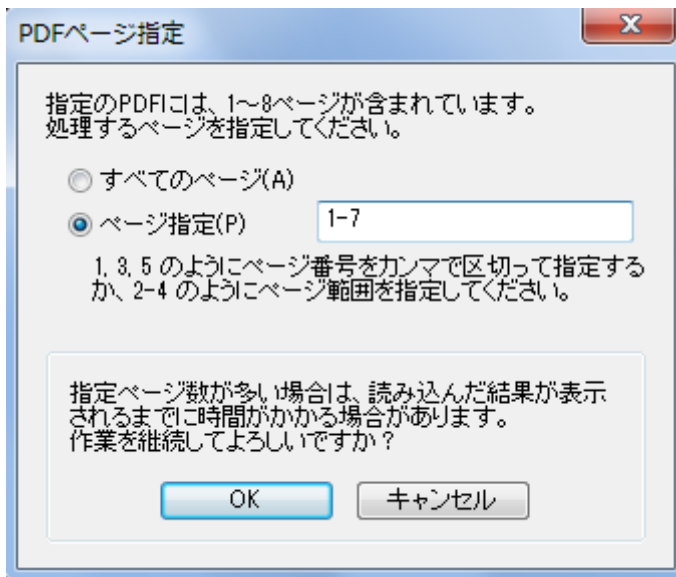
### 8-7.加工前ファイル選択

ファイルを選択して「開く」



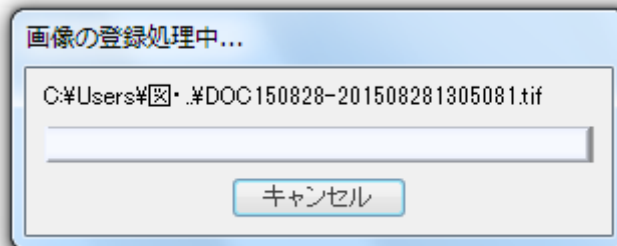
### 8-8.ページ選択

ページ指定する（前後に別の論文が入っている場合は、ページ指定して省く）。



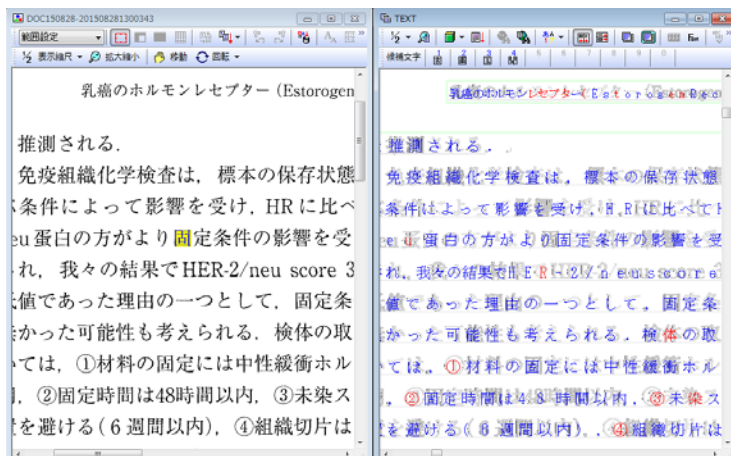
### 8-9.実行

自動で処理する。



### 8-10.修正

認識後に画面が表示されるので、画像とテキストを確認して修正する。





## 8-11.保存

修正が終わったら保存する。

ファイル名は自館で分かりやすい名前を付ける。

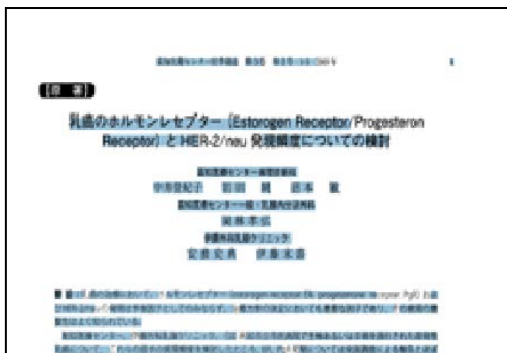
ファイルの種類は「高圧縮 PDF 画像（透明テキスト付）」を選ぶ。

ファイル名(N):	1881-5022_3_2_1-3_after.pdf	保存(S)
ファイルの種類(T):	高圧縮PDF画像 (透明テキスト付き) (*.pdf)	キャンセル

## 8-12.確認

Adobe ReaderなどでPDFファイルを開く。

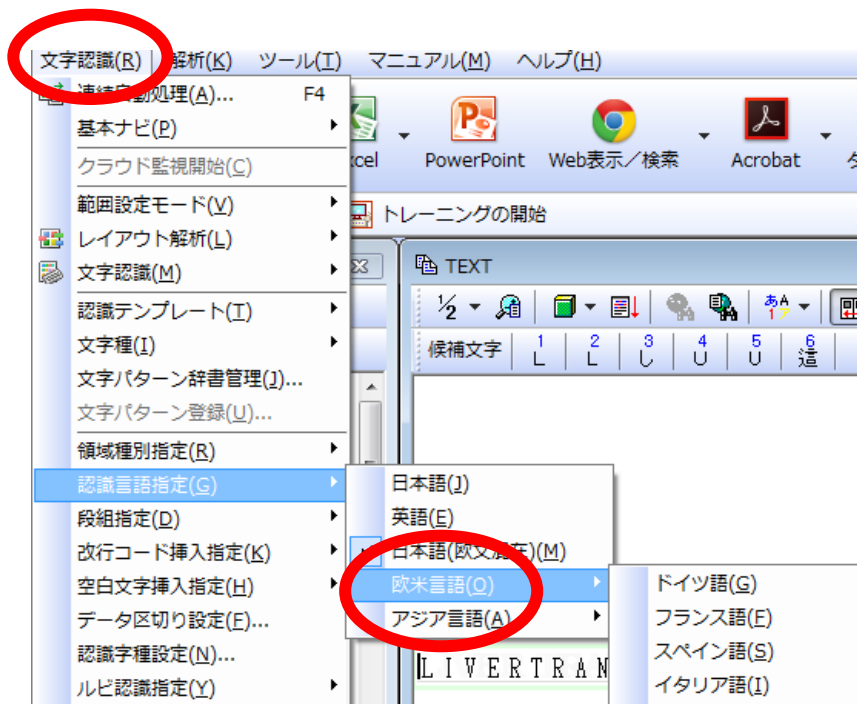
PDFファイルに透明テキストがついている（青色に反転する）ことを確認する。



## 8-13.文字認識

日本語、英語以外を使うときは、文字認識で使う言語を選ぶ。

和文論文も英文論文も、特に設定を分けることなく同じ方法でOCRできる。

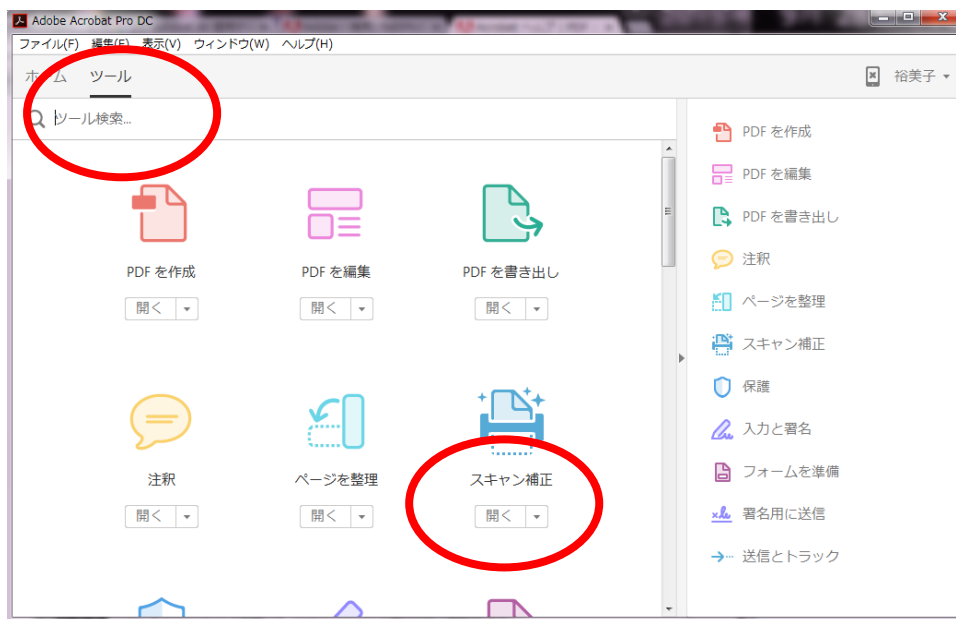


## 9.AcrobatDC の使い方

### 9-1.acrobatDC の起動

ツールをクリック

スキャン補正の「開く」をクリックしてさらにプルダウンメニューの「開く」をクリック



### 9-2.ファイルを選択

単数の時…「ファイルを選択」をクリック

複数の時…「または複数のファイル内のテキストを認識」をクリック



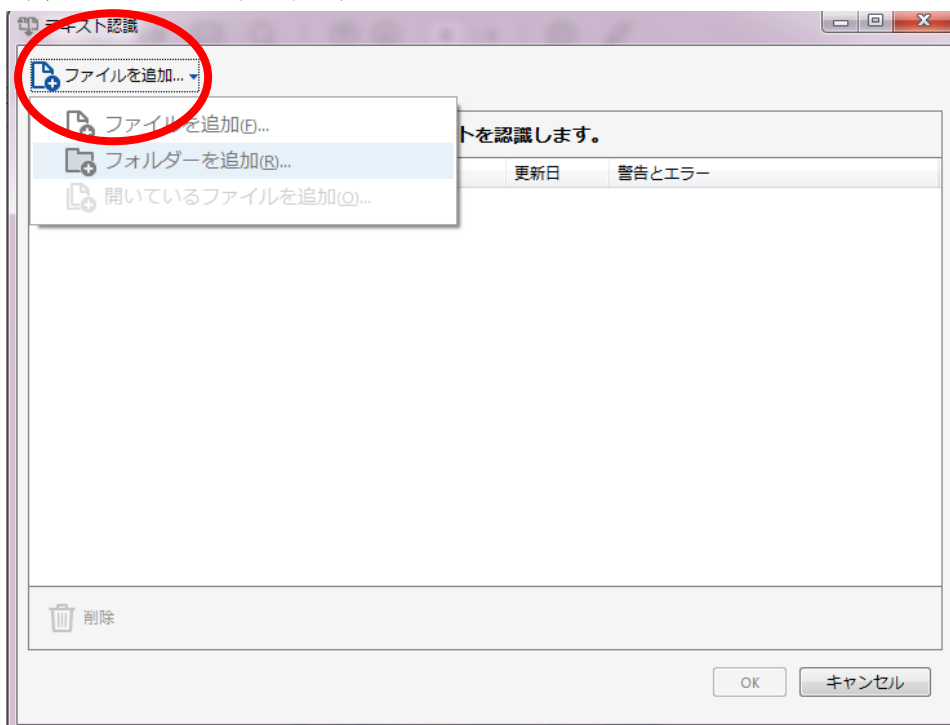
### 9-3. ファイルの追加

ファイルを追加をクリック

ファイルを追加する場合…「ファイルを追加」をクリック

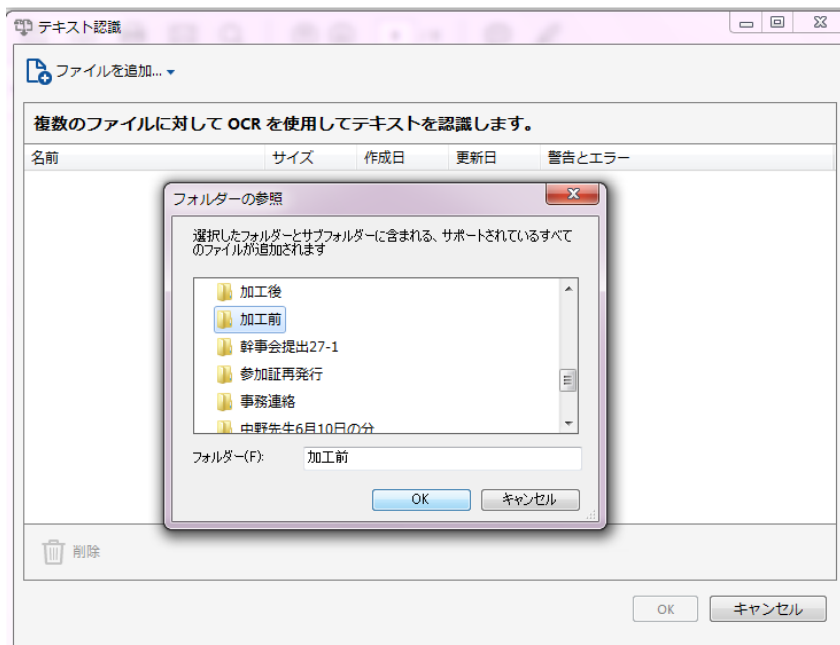
フォルダを追加する場合…「フォルダを追加」をクリック

(今回はフォルダ指定する)

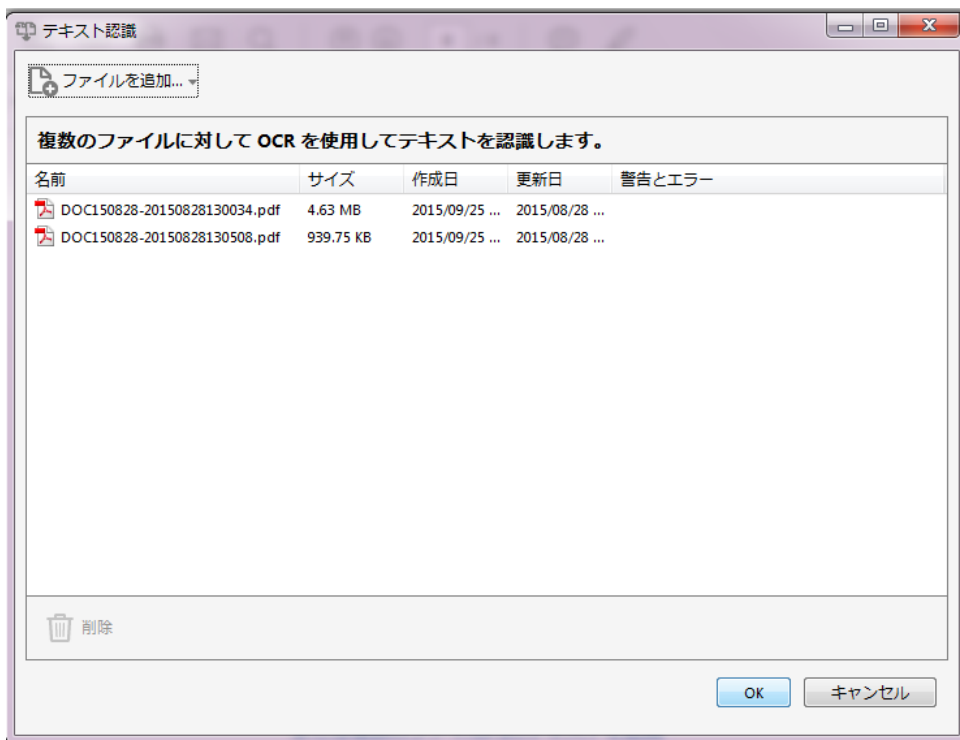


### 9-4. フォルダを指定

今回は加工前というフォルダを選択する。



これでよければ、OK を押す。



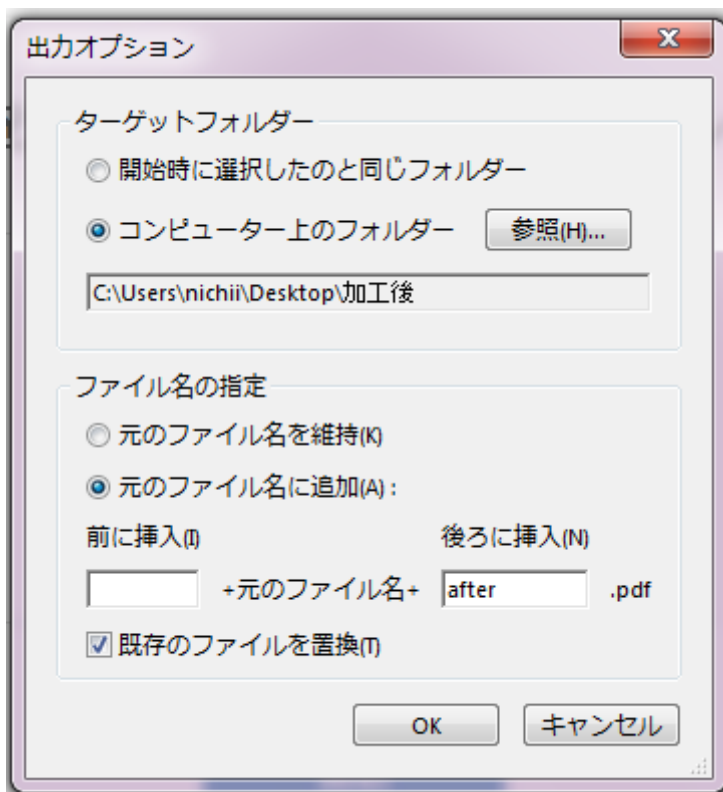
### 9-5.出力先指定

ターゲットフォルダーをデスクトップの「加工後」にする。

ファイル名を「元のファイル名に追加」で「after」を挿入する。

(各機関の分かりやすい規則を作ってファイル名を付ける)

OK を押す。



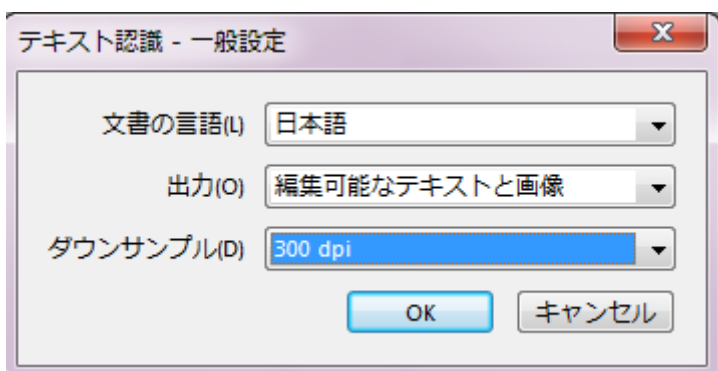
### 9-6. テキスト認識の設定

言語…読み込む論文の言語に設定

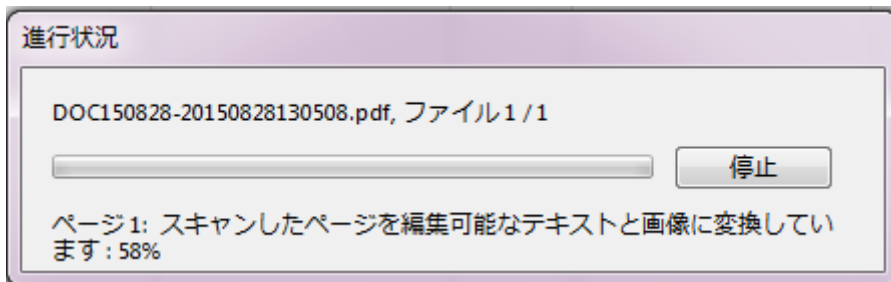
出力…「編集可能なテキストと画像」を選ぶ。

ダウンサンプル…300dpi を設定する。

(一度の認識で日本語なら日本語の論文ばかり選ぶ)



◆変換中



◆終了

終わるとまた、最初の画面に戻る。

Acrobat を終了する。

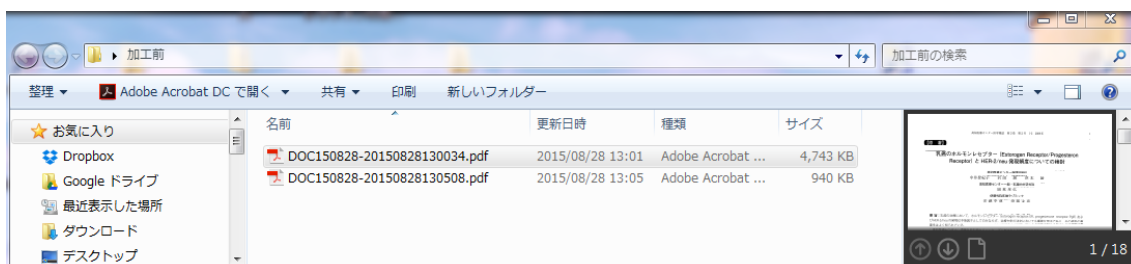


### 9-7.フォルダの確認

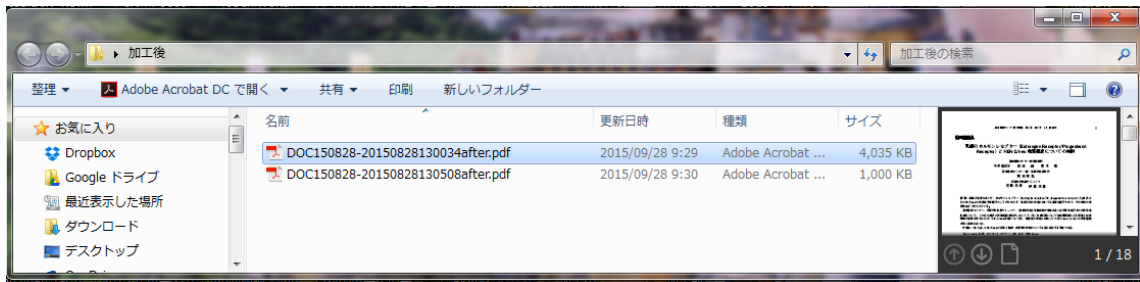
加工後に変換されたファイルが入っているか確認する。

「加工後フォルダ」に、after を付けた PDF ファイルが保存されていれば OK。

「加工前フォルダ」



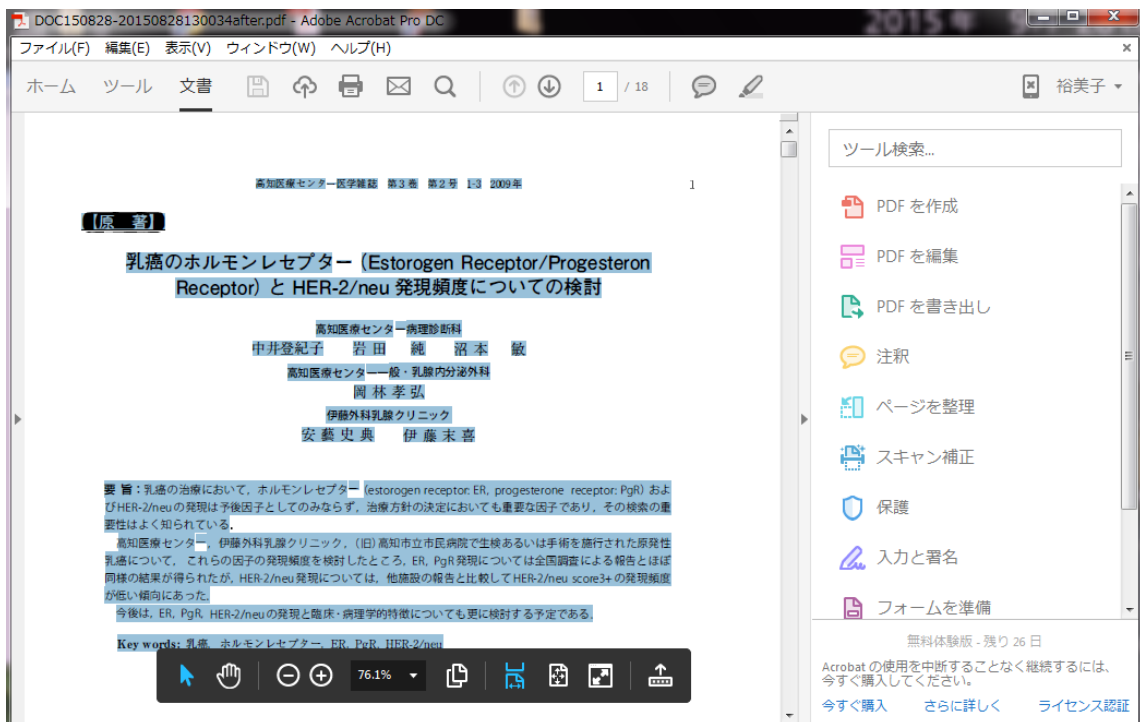
「加工後フォルダ」



## 9-8.加工後フォルダ確認

テキストが付加されていたら OK。

(テキストを選択すると薄青く反転する)





## 10.フォルダ構造

一番上のフォルダを「〇〇病院リポジトリ」とし、その中に「紀要」「年報」などのフォルダを作成し、さらにその中に巻、号フォルダを作成する。

このフォルダ構造はあくまで例である。

### 10-1.加工前

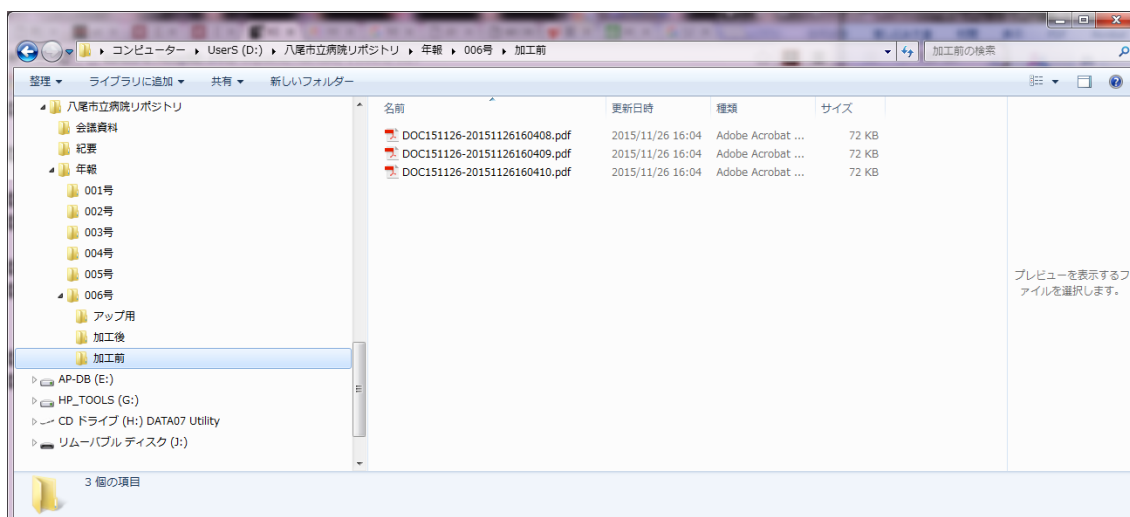
スキャンして保存したファイルをそのままの名前で加工前に入れる。

フラットベッドの場合は1頁ずつとるので、ファイルも頁分作成される。

オートフィーダの場合は10頁ずつ作成するので数ファイル。

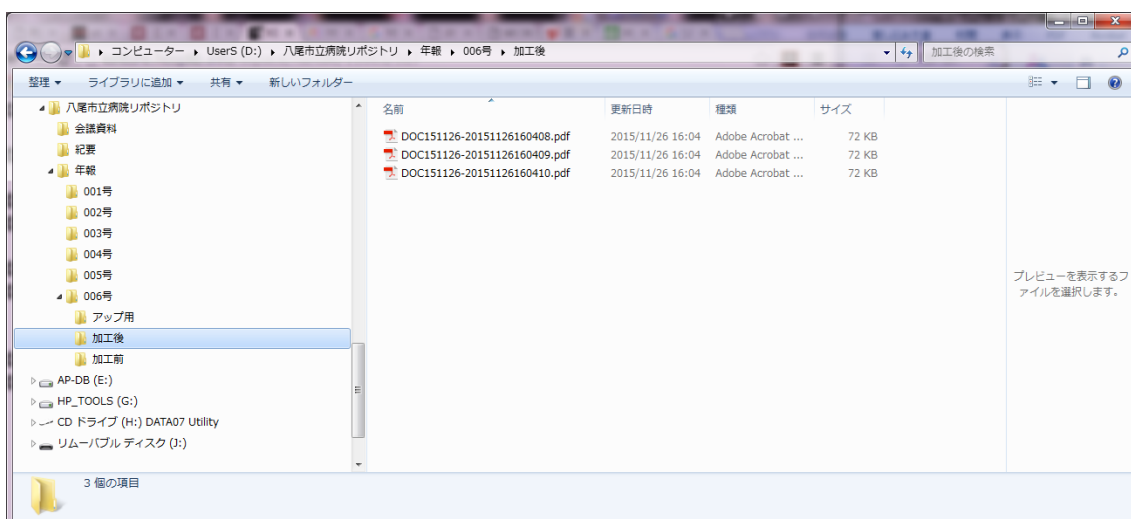
コピー機などでスキャンすると一冊まるごとで1ファイルになる。

(表示例は東芝のコピー機でスキャンしたファイル名)



### 10-2.加工後

加工前のファイルをOCRソフトで開き、テキストを付与した後、加工後のフォルダに保存する。別フォルダなので同じ名前でも大丈夫だが、分かりにくい場合はファイルの後ろに「a (after)」や「後 (加工後)」などつけてもよい。



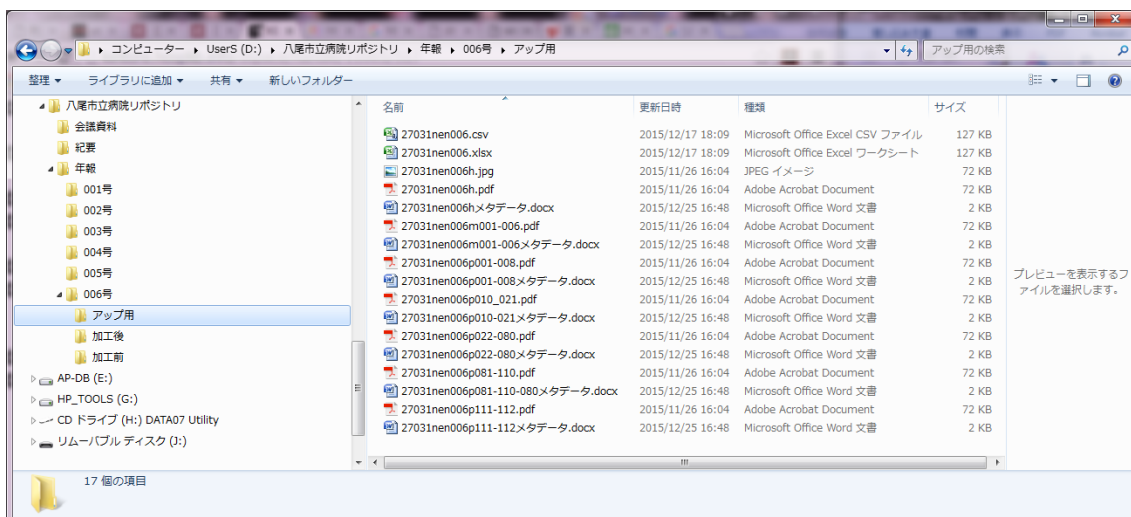
### 10-3. アップ用

加工後のファイルを論文ごとに保存する。

ファイルの分割・結合は「7-2-3.ファイルの分割・結合」を参照。

保存ファイルはアップ用のファイル規則に従う（頭5ケタが機関コード）

メタデータは1ファイルずつでもひとまとめでもよい。



## 11. 参考資料

「資料の電子化（スキャン、OCR）」

関西福祉大学 西本 朱美 氏

近畿病院図書室協議会第 134 回研修会 「資料の電子化（スキャン、OCR）」配布資料

「紙資料の電子化（スキャン・PDF 化・OCR 処理）」

森下 映理 氏

[https://ir-suishin.repo.nii.ac.jp/?action=pages\\_view\\_main&active\\_action=repository\\_view\\_main\\_item\\_detail&item\\_id=96&item\\_no=1&page\\_id=25&block\\_id=42](https://ir-suishin.repo.nii.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=96&item_no=1&page_id=25&block_id=42)

「機関リポジトリと著作権 Q&A」

黒澤 節男 氏

<http://ir.lib.hiroshima-u.ac.jp/ja/00023065>