

# TiF Manager マニュアル

2023年7月

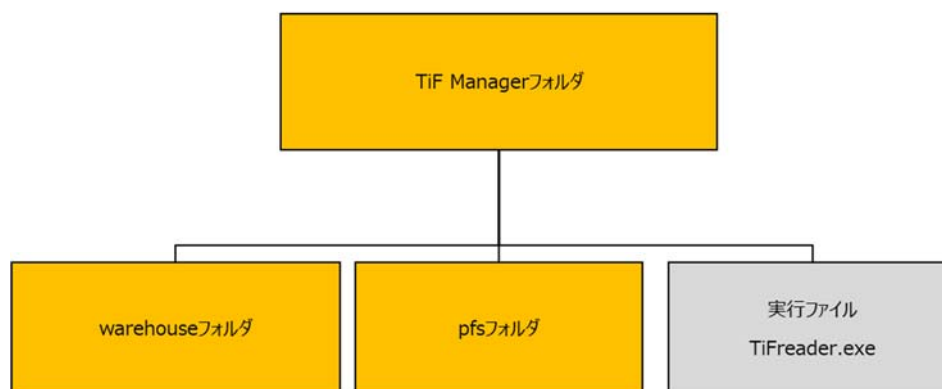


## 1. TiF Manager が稼働するためのシステム要件

- 64ビットの Windows（32ビットの Windows は不可、Microsoft 社がサポート中の Windows に限るので Windows8 以降）がインストールされたパソコン（インターネット接続不要）
- ご利用中の Windows で正常に稼働する Microsoft Excel
- 1GB 程度の記憶領域空き容量（パソコンのハードディスクでもネットワーク内フォルダでも可）

### 【インストール方法】

ダウンロードした zip ファイル「TiF Manager.zip」をフォルダごと解凍し、TiF Manager フォルダをお好みのフォルダに設置してください。TiF Manager フォルダの配下のフォルダ・ファイルが下記のような配置になっていれば問題なく稼働します。





目 的	次の商品に関する貿易業、売買業、問屋業、代理業並びに仲立業 イ)鉄、非鉄金属及びこれらの原料、製品並びに鉱産物 ロ)石炭、石油、天然ガスその他の燃料並びにこれらの副製品 ハ)各種機械器具(計量器、医療用具を含む)、製造設備、通信設備、公害防止設備等の設備、車輛、自動車、船舶、宇宙機器、航空機並びにこれらの部品 ニ)各種化学製品、塩、肥料、高圧ガス、火薬、薬品(医薬品、医薬部外品、動物用医薬品、麻薬、毒物、劇物を含む)、放射性同位元素、化粧品並びにこれらの原料 ホ)食糧、砂糖、油脂、飼料及びこれらの原料、畜類、農畜水産物、加工食品、酒類その他の食料、飲料 ヘ)各種繊維製品及びその原料 ト)各種毛皮製品及びその原毛皮 チ)木材、セメントその他の建築資材並びに住宅設備関連機器 リ)ゴム、パルプ、紙及びこれらの製品、たばこ、雑貨類 ス)工業用水、飲料水 前号に関連する探鉱、開発、製造加工業並びに廃棄・再生処理業
発行可能株式数	100000株
発行可能株式の割合	153.39 mm

何のために計測するかを説明しておきますと、小枠の短辺の指定をしなければ、傾いている PDF などでおかしな枠認識をしたり、カタカナの口(ろ)や漢字の口(くち)を枠認識したりすることで、膨大な TiF に対して無意味な OCR を実行してしまうからです。大枠の長辺は A4 タテ限定なので 210mm 以上となることはないのですが、210mm にしておく読み取りたい部分全体を、枠を無視して、重複で OCR してしまうことが希にあります。

## (2) PDF の TiF Manager フォルダへのコピー

インストールした TiF Manager フォルダの直下に読み込ませたい PDF ファイルを保存してください。

TiF Manager の直下には warehouse フォルダと pfs フォルダがありますが、これらのフォルダに保存しても、エラーとなるだけなのでご注意ください。

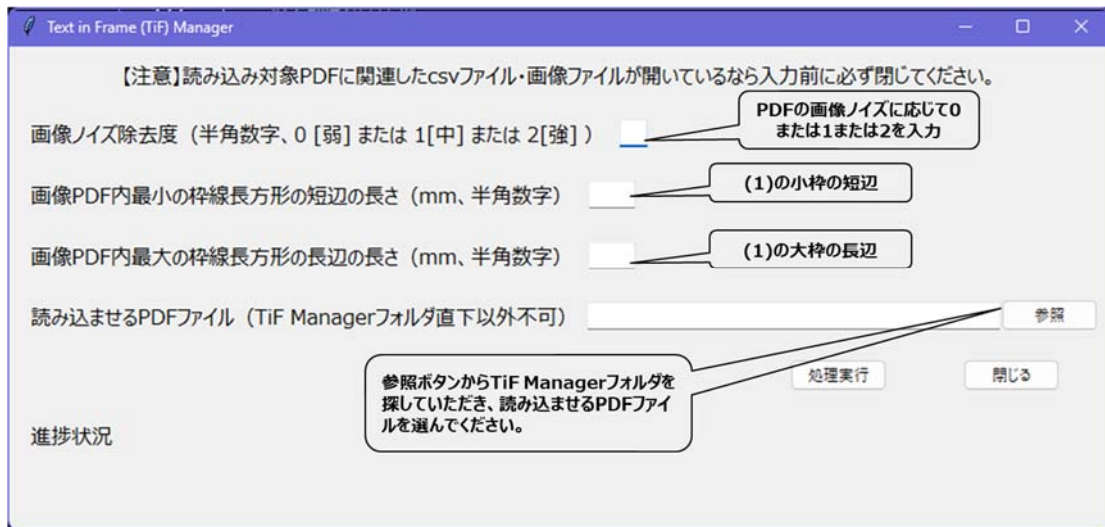
特に pfs フォルダは OCR エンジンなどの重要なプログラムファイルが保存してありますので、誤動作の原因となりますので、極力開かないようにしてください。

## (3) 実行ファイル(TiF reader.exe)のクリックと条件入力

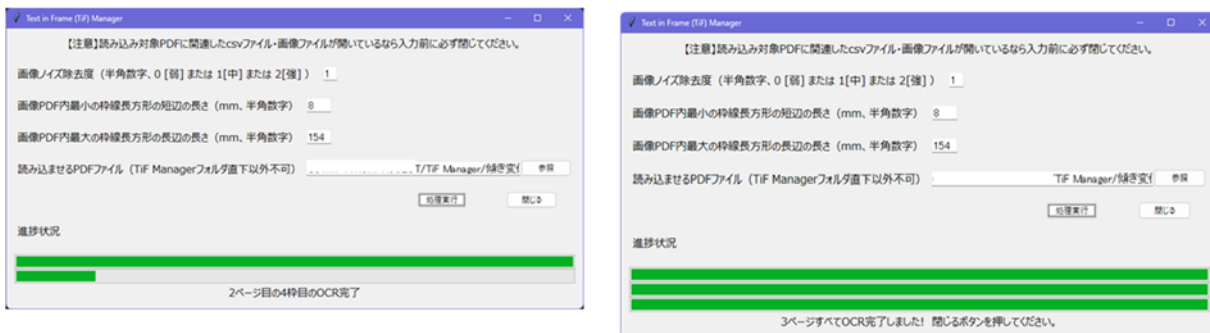
TiF Manager フォルダ配下の TiF reader.exe というクリックすると、ターミナルウィンドウが開き、続いて下記のようなウィンドウが立ち上がります。一時フォルダに pfs フォルダに含まれない追加のプログラムファイルを展開する必要があり、起動までに時間がかかります。

【注意】に表示していますが、一度実行して ocr\_kekka.csv や画像ファイルを開いて内容を確認して、条件を修正してもう一度やり直すといった場合に閉じ忘れが目立ちます。このウィンドウが開いた後でも遅くはありませんので、必ず閉じてから処理実行ボタンをクリックしてください。

下記吹き出しの要領で条件を入力いただきますが、「画像ノイズ除去度」の項目について説明しておきます。背景画像やスキヤナー読み取り紙面上の折り目や汚れを画像ノイズと呼びます。PDF の背景が真っ白で、枠線や文字が黒で鮮明な場合は、画像ノイズは少ないと考えられるため、[弱]に該当する 0 を入力してください。画像ノイズの程度によって[中]の 1 や[強]の 2 を入力していただきますが、ノイズ除去を強くすればするほど、黒い枠線・文字がかすれてしまい、枠や文字を誤認しやすくなってしまいます。「画像ノイズ除去度」を 1 にするか 2 にするか悩まれる場合は、テスト実行していただき、XXX フォルダ (XXX は PDF ファイル名) 直下に排出される枠線 ID 付の画像ファイル (XXX\_(ページ番号) .png) で画像ノイズの除去結果は確認できますので、確認後調整して実行してください。



すべて入力したら処理実行ボタンを押下します。OCRする文字数と枠数に比例して、処理時間はかかります。(3ページのすかし画像のあるPDFの場合の処理中の画面表示例)



傾いたままの紙資料のスキャナー保存で作成されている場合、枠が重複したり、正しく読み取れていなかったりすることがあります。この場合読み取り枠数が無駄に多くなり、処理時間が長くなります。最大の枠線長方形の長辺（大枠の長辺）を実寸より大胆に小さくする（3分の2～2分の1くらいにする）と重複がなくなったり、不適切な枠認識が解決したりしますので、お試しください。

入力内容不備があったり、CSV ファイルを開いたまま実行していたりすると処理を中止し、ターミナルウインドウにエラーメッセージが表示されます。

(4) warehouse フォルダ内に読み込ませた PDF ファイル名のフォルダが作成されます。

その中に下図のように 1 つの csv ファイル、読み込ませた PDF に枠 id を付した画像ファイル、TiF で細切れにした画像ファイルのフォルダ pieces フォルダが作成されます。

warehouse フォルダ直下の PDF ファイル名のフォルダは、まるごと切り取って任意のフォルダに移動することが可能です。

