



紙資料を正しくペーパーレス化した上で効率的に管理するシステム

# TiF Manager (ティフマネージャー)

---

誤読がゼロにはならないOCR結果の誤読発見・修正事務をデジタル化にふさわしくBPRします。

## 集計表やレポートなどの一切の資料作成事務（情報加工）の生産性は、テキストデータのコピーで高まります。

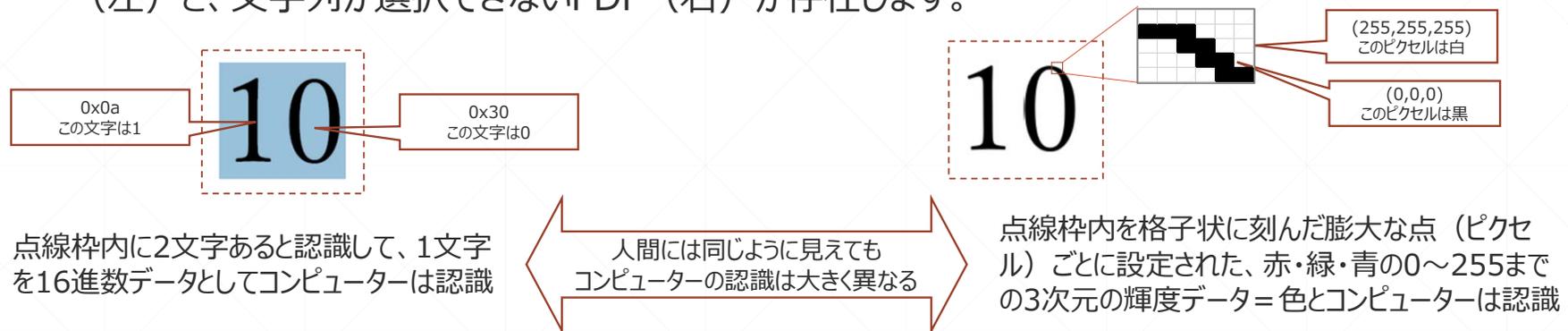
- 「目」でパソコン画面に表示したテキストを読み、読み手が正しく理解できるように「頭」を使ってテキストを加工して、「手」を使ってキーボードとマウスを動かして、ペーパーレス化時代の資料は作成されます。
- 情報加工に関するDX（デジタルトランスフォーメーション）は、人間の目・頭・手の負担をどこまで減らせるかへの挑戦と解釈することも可能です。

これまでは数値や短い文字列のテキストの加工は、EXCEL関数のような加工ツールでの「頭の利用削減」にとどまっていたが、Chat GPTのような生成AIは長い文字列の高度な加工を可能としたことで、今後は情報加工においてほとんど「頭」を利用しなくて済むようになるかもしれません。

- 特に「目」と「手」の利用削減は徹底して進み、テキストのコピーと貼り付け（マウスでの手動またはプログラムによる自動化分を含めたコピー）操作だけで済むようになっていて、短時間で大量の情報加工を可能としています。
- しかし、コピープロテクトもされていないテキストとして画面上には表示されながらも、コピーを拒絶する情報が存在し、目視とキーボードによる転記を情報加工の際に今なお求め続けています。
- 生成AI活用で「頭」を利用しなくてよい時代になっても、資料作成事務担当者の稼働時間のほとんどは、目と手を酷使して、この厄介なテキスト情報との格闘に費やされる可能性が高いのではないのでしょうか。

# 目と手を酷使することを求める困ったテキスト情報とは、テキストデータではない文字の画像データです。

- 他の資料にPDF資料からコピーしようとするときに文字列をマウスで色つき部分のように選択できるPDF（左）と、文字列が選択できないPDF（右）が存在します。



- 左を「テキストデータ」、右を「テキストデータもどきの画像データ」と呼ぶことにしますと、テキストデータからなる資料Xは他の資料Yにテキストデータとして転記（コピー）できますし、テキストデータもどきの画像データへのテキストデータ変換（いわゆるスクリーンショット）も含めた各種演算操作さえも可能です。

テキストデータもどきの画像データからなる資料は、ペーパーレス化したとはいえ、テキストデータのようなコピーはできないため、紙資料と同様に目視で画像を人間が読み取り、人間が画像を解釈し、テキストデータ化した上で他の資料にコピーするしかないのである。

## 銀行でのペーパーレス化は多くの資料でコピーがやりやすい形になってはいますが…

- 計数関連資料は数桁の数字からなる大量のテキストデータを扱いますが、PDFではなく基幹系システムからCSVファイルで抽出できるようになっており、EXCEL等がテキストビューアーとして機能して、集計処理などもやりやすく、効率的な管理ができるようになっています。
- 他方、数千以上の文字からなる長文の文書（例えば稟議や各種報告書など）は、PDFとする際にも段落や改行や箇条書きのフォーマットを維持したまま（画像ファイルではなく）テキストデータとしてPDF化されていることが一般的で、他の資料に苦も無くコピーできます。
  - ただし、長文の文書のある一文を抽出するといった場合、キーワードでなく、その文を直接検索するような利便性は提供されていません。
- テキストもどきの画像データは、数字ほど短いわけでもなく、長文よりは短い文字が計数関連資料のように多数の枠に仕切られた資料に集中します。
  - 登記事項証明書や勘定科目内訳書やアンケート結果など紙で提出された外部資料を、銀行内部でスキャナー処理で作成したPDFファイルが該当します。
- こうしたテキストもどきの画像データを人間の目や頭や手を煩わせずテキストデータ化し、（＝スクリーンショットの逆操作）コピーも可能とするための機能がOCR（光学的文字認識）です。

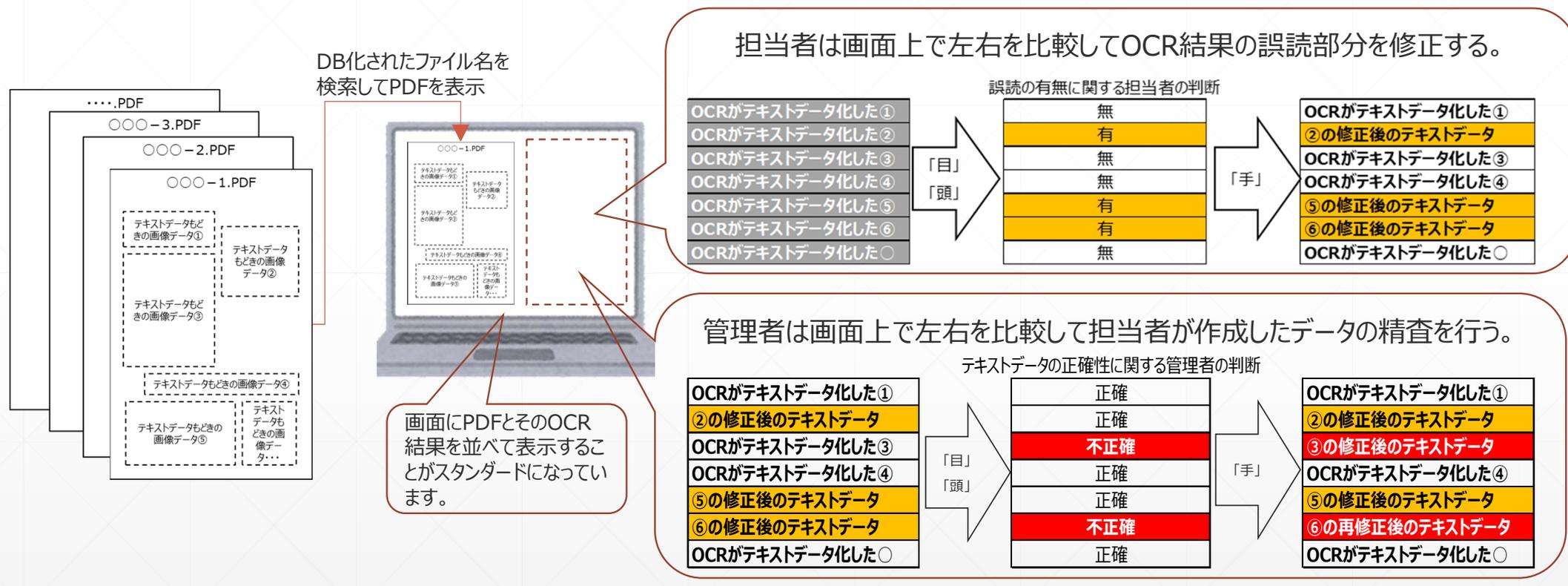
## 無限に存在する手書き文字画像をビッグデータととらえ、AIを活用することで、OCR技術も飛躍的に発達してきてはいますが・・・

- テキストデータもどきの画像データを、人間の「目」と「頭」に代わってテキストデータ化するOCRですが、基本は文字の画像データにおけるピクセル配置と該当するテキストデータの難易度の高い対応関係をコンピューターに覚えさせることで実現しています。
- 画数の多い漢字や画像ノイズの影響で、OCRでの誤読は頻繁に発生するので、人間の「目」と「頭」の助けをまだまだ借り続けているのが実態です。
- AIを活用したAI-OCRでは、文字の画像データと該当するテキストデータを対応を、使用する度に学習させていくことで、手書き文字の読み取りにさえ対応し、誤読は確実に減っていきますが、それでも誤読がなくなることはありません。
- ちょっとした誤読が経済上の損失につながりかねない銀行では、多かれ少なかれ誤読がある以上はAI-OCRであってもコンピューター任せにすること絶対にはなく、複数の人間の「目」と「頭」での再確認を介在させます。

OCRでの誤読がゼロとまらない限り、銀行ではテキストデータもどきの画像データをテキストデータ化する人間とテキストデータ化の正確性を確認する人間が完全に不要となることは、未来永劫ありません。

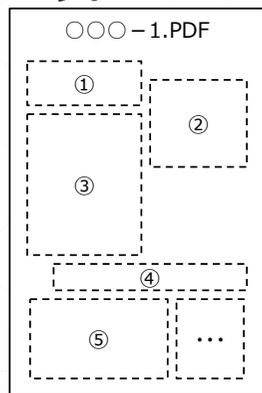
# OCR機能を装備した最先端の高価な画像文書管理システムはこんな形になっています。

- PDF文書上のすべての文字を人間が「手」で転記せずに済むようになっただけで、「目」と「頭」の酷使を減らす工夫の余地は、まだまだあります。



# TiF Managerは、保存した画像PDF文書上の情報を、銀行の文書管理担当者・管理者の「目」と「頭」にやさしい形に変換します。

- PDF内に枠線があれば枠線を認識して枠線毎にバラバラにして画像の子データベースを作り、画像毎に認識方法が異なる3種類（ナーシャ・アシュレー・ホア）のOCR結果をデータベースにセットします。



自動でバラバラにして画像化

これをText in Frame(TiF)と呼ぶことにします。

1つの画像に対し3種類のOCRを実行

OCR1がテキストデータ化した①	OCR2がテキストデータ化した①	OCR3がテキストデータ化した①
OCR1がテキストデータ化した②	OCR2がテキストデータ化した②	OCR3がテキストデータ化した②
OCR1がテキストデータ化した③	OCR2がテキストデータ化した③	OCR3がテキストデータ化した③
OCR1がテキストデータ化した④	OCR2がテキストデータ化した④	OCR3がテキストデータ化した④
OCR1がテキストデータ化した⑤	OCR2がテキストデータ化した⑤	OCR3がテキストデータ化した⑤
OCR1がテキストデータ化した⑥	OCR2がテキストデータ化した⑥	OCR3がテキストデータ化した⑥
OCR1がテキストデータ化した○	OCR2がテキストデータ化した○	OCR3がテキストデータ化した○

専門用語が並ぶことも多く、どこに重要な情報があるのかもわからない担当者にとっては文書内容確認に時間もかかり、「目」も「頭」も酷使することになります。

誤読がどんどん少なくなるという優秀なAI-OCRでは、読み取り結果に自信満々な様子で読み取り結果は、1つだけしか示されません。

TiF Managerは大量の文字を含むPDFの枠線毎に枠線IDを割り振り、文書を箇条書きのように短文化します。

TiF Managerは誤読相関の低いOCRで3通りの読み取り結果を取得でき、下手な鉄砲も数撃ちゃあたる的発想でテキストデータ化候補を3つ提示します。

# PDFの画像ノイズ（すかし画像や光学的処理から生じる汚れ）や傾きも自動で修正します。そして修正後のPDFの枠線毎にIDを割り振ります。

履歴事項全部証明書	
神奈川県横浜市中区濱町2丁目9番6号古井ビル4階 満帆商事株式会社	
会社法人番号	0101-69-5858
商号	満帆商事株式会社
本店	東京都港区港北1丁目5番2号
	神奈川県横浜市中区濱町2丁目9番6号古井ビル4階 令和2年1月31日移転 令和2年2月5日登記
公告する方法	官報に記載する方法により行う。
会社設立の年月日	平成6年8月8日
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、除薬、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
発行済株式の総数並びに種類及び株	発行済株式の総数 100000株
株式の譲渡制限に関する規定	当会社の発行する株式の譲渡による取得については、株主総会の承認を受けなければならない。ただし、当会社の株主に譲渡する場合には、承認したものとす。
役員に関する事項	取締役 早坂 光 宗 平成30年6月30日辞任 平成30年7月3日登記
	取締役 井上 洸 二 平成30年7月1日就任 平成30年7月3日登記
登記記録に関する事項	埼玉県蕨市塚越北町3丁目9番13号 代表取締役 井上 洸 二
	設立 平成25年3月31日登記

登記事項証明書を全く知らない人に対して「7」をテキストデータ化してくださいという指示が可能！

XXXというpdfファイル名であれば、XXX\_(ページ番号) という名前の画像ファイルpngファイルで、そのページの枠idをばらばらにした画像にどう割り振ったかがわかるファイルを生成します。

履歴事項全部証明書	
神奈川県横浜市中区濱町2丁目9番6号古井ビル4階 満帆商事株式会社	
会社法人番号	1 0101-69-5858
商号	3 満帆商事株式会社
本店	5 東京都港区港北1丁目5番2号
	7 神奈川県横浜市中区濱町2丁目9番6号古井ビル4階 令和2年1月31日移転 令和2年2月5日登記
公告する方法	9 官報に記載する方法により行う。
会社設立の年月日	11 平成6年8月8日
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、除薬、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	15 100000株
発行済株式の総数並びに種類及び株	17 発行済株式の総数 100000株
株式の譲渡制限に関する規定	19 当会社の発行する株式の譲渡による取得については、株主総会の承認を受けなければならない。ただし、当会社の株主に譲渡する場合には、承認したものとす。
役員に関する事項	取締役 早坂 光 宗 平成30年6月30日辞任 平成30年7月3日登記
	取締役 井上 洸 二 平成30年7月1日就任 平成30年7月3日登記
登記記録に関する事項	埼玉県蕨市塚越北町3丁目9番13号 代表取締役 井上 洸 二
	設立 平成25年3月31日登記

# 別途下記のようなCSVファイル「XXX\_OCR\_kekka.csv」をデータベースとして生成しますので、Excelで開いてOCR結果を確認できます。

ページ	ページ内の枠のid	読み取り画像リンク	ナーシャ読み取り	アシュレー読み取り	ホア読み取り	確定
1	1	.¥pieces¥満帆商事_page1-1.png	0101-69-5858	0101-69-5858	0101-69-5858	
1	2	.¥pieces¥満帆商事_page1-2.png	会社法人番号	会社法人番号	還 会社法人番号	
1	3	.¥pieces¥満帆商事_page1-3.png	満山商事株式会社	満山商事株式会社	満山商事株式会社	
1	4	.¥pieces¥満帆商事_page1-4.png		商号	還	
1	5	.¥pieces¥満帆商事_page1-5.png	2号	、東京都港区港北1丁目6番2号 呈	2号 握	
		神奈川県横浜市中区濱町2丁目9番6号古井ビル4階 .png	令和2年1月31日移転 令和2年2月5日鬮記	令和2年1月31日移転 令和2年2月5日合記	令和2年1月31日移転 -----ド----- ニ-----m----- ニ----- 令和2年?鬮記	
1	7	.¥pieces¥満帆商事_page1-7.png	神奈川県横浜市』 区演町2丁目9番6号古才 EEビル4階	神奈川横浜市中区演町2 丁目9番6号古井ビル4階	神奈川横浜市中区演町2丁 目9番6号古井ビル4階	神奈川横浜市中区演町 2丁目9番6号古井ビル4 階
1	8	.¥pieces¥満帆商事_page1-8.png	本店	本店 	本店	
1	9	.¥pieces¥満帆商事_page1-9.png	官報に記載する方法により行 う。	官報に記載する方法により行 う。	官報に記載する方法により行 う。	
1	10	.¥pieces¥満帆商事_page1-10.png	公告する方法	公告する方法	症 公告する方法	
1	11	.¥pieces¥満帆商事_page1-11.png	— 平成6年8朋 8日		—   平成6年8朋 8日	

“濱”のホアが正解

神奈川県横浜市中区濱町2丁目9番6号古井ビル4階

枠id7のリンクをクリックすればTiF画像をWindowsの画像ビューアーで表示

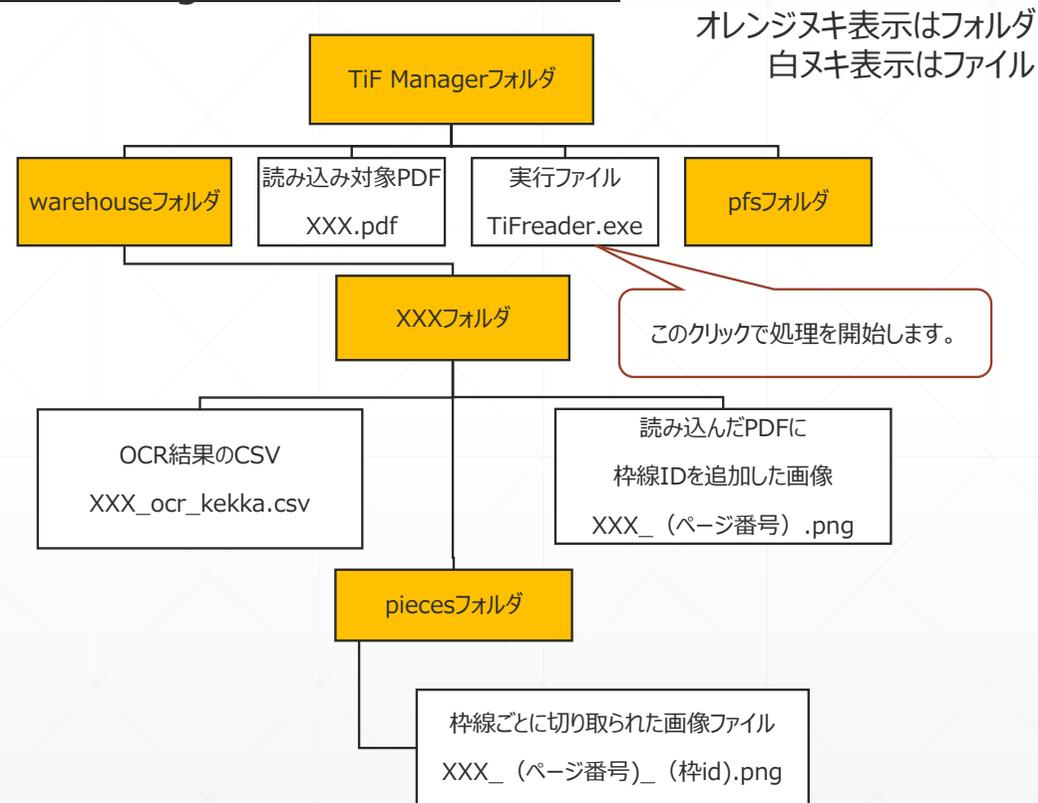
アシュレーとホアの  
違いはここ

ホア読み取り結果をコピーして、先頭行の“\_ \_”をとって確定！  
最小限のキーボード使用でテキストデータを確定できる。

# TiF ManagerはWindowsのフォルダ内に見える形で展開されており、各ファイルはTiF Managerフォルダ内の所定の位置に自動的に生成されます。

- 読み込み対象PDF (XXX.pdf)はご面倒でもTiF Managerフォルダの直下に配置する必要があります。
- XXX\_ocr\_kekka.csv、XXX\_ (ページ番号) .png、枠線毎に細切れにした画像を保管するpiecesフォルダは、XXXフォルダという自動作成されるフォルダに格納されます。
  - 予めXXXフォルダが存在している場合は削除してフォルダ構造を再作成します。
- XXXフォルダをまるごと切り取って、別の場所で保管することも可能です。
- pfsフォルダはプログラムファイルを格納してありますので、書き換えたり移動したりしないようお願いいたします。

## TiF Managerのフォルダ・ファイル構成



## 様々なTiFの枠配置とPDFに残る画像ノイズに対応するため、読み込むPDFに対して目視での紙面確認は必要となります。

- 誤読の原因でもある、登記事項証明書にあるようなすかし背景画像の除去や画像の傾き補正はTiF Manager内で自動で行っています。
- すかしやウォーターマークがあったり、紙資料に複数回コピーが行われていたり、紙に折れ目がついていたりすると画像ノイズが増えます。
  - ノイズ除去をすると、OCRでの誤読が増えるのですが、ノイズ除去しないと枠線さえも誤認してしまうため、どれくらい画像ノイズを除去するかを読み込むPDFの画像をご確認いただき、ノイズ除去度合いを0（弱）または1（中）または2（強）で下の①で指定していただきます。
- また複雑な枠線構成への対応や不要なデータ読み込み防止のために、読み込むPDFファイルの指定とは別途テキストデータ抽出を必要とする最大の枠長方形の短辺の長さ（mm）と最小の枠長方形の長辺の長さをミリ単位で下の②と③で指定してもらいます。

Text in Frame (TiF) Manager

【注意】読み込み対象PDFに関連したcsvファイル・画像ファイルが開いているなら入力前に必ず閉じてください。

画像ノイズ除去度（半角数字、0 [弱] または 1 [中] または 2 [強]）  ①

画像PDF内最小の枠線長方形の短辺の長さ（mm、半角数字）  ②

画像PDF内最大の枠線長方形の長辺の長さ（mm、半角数字）  ③

読み込ませるPDFファイル（TiF Managerフォルダ直下以外不可）  参照

処理実行 閉じる

進捗状況

詳細はユーザーズマニュアルを  
ご参照ください。



## 誤読はあるものの3種類のOCR結果を得た状態でPDFの持つすべての情報が予め保管されることのBPR上の意義は大きいのではないのでしょうか。

- それぞれ読み取り方が異なるOCRエンジンのため、人間の目で見ても読めないようなTiFでない限り、誤読の相関は低くなっている（＝誤読の分散が効く）、AI-OCRとは別の方法で誤読率を引き下げる効果があります。
- パート派遣の行職員（今や確保難や賃金上昇という逆風まであります！）を使ったとしても、「OCR結果の誤読部分を修正する」コストの割高感、機密保持を構わなければ子どもでもできそうな作業ゆえ、払拭できないままと思いますが、ペーパーレス化としてPDFのまま文書を保管する限り、永遠に悩み続けることとなります。
- TiF Managerを使うことは、3人のバーチャル担当者ががんばってテキストデータ化することに等しく、リアルな担当者を使わずとも、管理者の精査だけでテキストデータもどきの画像データをテキストデータ化を完結できるのではないかと考えています。
  - やめることのできない重いテキストデータ化作業を、極力軽量化できる形式に画像PDF上の情報をTiF Managerは変換します。

こわれるほどたたいて渡らなければ安心できない石橋にも、テクノロジーの進歩にふさわしい効率的な渡り方があるものです。

## 手書き文字対応はほぼ絶望的ですが、作成資料の読み手がOCRより確実に誤読しにくい人間である限り、こんなTiF Manager活用方法もあります。

- セミナー会場で参加者から回収した大量の紙アンケートのうち、手書きの感想・意見だけをまとめた資料を急ぎで作成しなければならないといった情報加工の場面は多いと思います。
- 同じフォーマットであって、枠のIDは何番（例では12）かははっきりしますので、CSVの読み取り画像リンクをクリックしてpiecesフォルダに収められた当該枠の画像をコピーして、右下のように並べれば、アンケート用紙PDFを1つずつ開いてスナップショットをコピーしなくても済みます。
  - 画像のコピペも面倒ならば、ocr\_kekka.csvを読み込み、枠idを12に絞り込んで読み取り結果列を削除したEXCELファイルにpiecesフォルダを添えたものでも対応できます。

ページ	ページ内の枠のid	読み取り例
1	12	〇〇というオの計明会 なヶ、 もそう 由に る友和光いそただエそい口。
2	12	ビジネス3も とりのげしそいす 人泉 ITかりで、フうき5人4か。正
3	12	いう4お 世言になっております>商商事の車務です。家 内のあ。た〇の〇tいう商品に2いて、刺社お得意、災 か5の9 1き合いがちあり行なので、近々マ 上ぐさせて ください。
4	12	胡林光光の多い地あでのカーボンクレジット沙用のポテ ンシャルがよく妊解できました。

読み取り例を提出しても怒られるだけなので・・・

〇〇という商品の説明会 なる、  
もう事判に おんえいたがこたい。

ビジネス誌でとりあげられている内容  
ばかりで、つまらなかつた。

いつもお世話になっております>商商事の車務です。ご案内のあつた〇  
〇という商品について、弊社お得意先からの引き合いが別行なの  
で、近々ご相談させていただきます。

森林資源の多い地元でのカーボンレ  
ジット活用のポテンシャルがよく理解で  
きました。

## TiF Managerご利用にあたっての注意点

- あくまで銀行内部での利用を想定したTiF Managerであり、銀行勤務役職員個人や銀行以外での使用許諾は想定しておりません。
- その分情報漏えい対策を完備した銀行内のパソコンであれば、制限なくコピーして使用することが認められています。
  - ただし、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は不要）でのみ動作します。CSVビューアーとしてそのWindowsパソコンで正常に稼働するMicrosoft Excelは必須となります。
- OCRエンジンの機械学習によるスペック改善なども今後見込まれますので、TiF Managerは適宜バージョンアップサポートさせていただきます。
- TiF Managerの特定PDFを対象にしたOCRの読み取り最適化や出力CSV（他データベースとの連携など）に関する各種設定の変更や取引先提供用途での開発は、有償にはなりませんが、カスタマイズも承りますので、トライアル版使用後にご相談ください。

## 2023年上期中、無償利用可能なトライアル版をモジュール工房ウェブサイトでダウンロード提供中です。

- トライアル版の使用期限はダウンロード時期に拘わらず一律2023年9月30日までとさせていただきます。
- モジュール工房のTiF Managerページからzipファイルでダウンロードできます。
- 解凍パスワードが設定されていますので、[yatsushiro@yatsushiro.co.jp](mailto:yatsushiro@yatsushiro.co.jp)宛てにメールで下記条件を満たしたメールでお問い合わせください。

- ✓ whoisで確認できる銀行ドメインの正規のメールアドレスから送られたメールである
- ✓ 所属部署・氏名がメール本文に記載されている
- ✓ zipファイルのダウンロード日時が記載されている

- 解凍パスワードはyatsushiro@yatsushiro.co.jpから、お問合せのあったメールアドレスに返信でお伝えいたします。