

VP index[®] Series

PROFESSIONAL SOFTWARE

表題欄OCR处理 (属性抽出)

1、概要



大量の社内図面を電子化するにあたり、図面固有の属性入力とファイル名指定の作業がボトルネックとなり、思うように電子化が先へ進まないといったケースをよく耳にします。

通常、データベース上の各フィールドに図面ごとの属性を入力しておくことで、大量図面の保管・管理・検索が行えるようになります。例えばデータベース登録する図面ごとに表題欄の図面番号/図面名称/設計者等の属性情報をキーボードから手入力するとなると非常に重労働となり、入力ミスも発生する可能性があります。またスキャニング時のファイル名も連続してスキャニングを行うと連番や日付等のユニークな名前が自動的に付いてしまうため、後で手作業で図面番号等にリネームする必要があります。こうしたボトルネックとなる属性抽出、ファイル名付け作業を大幅に軽減させるために開発されたソフトウェアが「VPindexシリーズ」です。



日々、大量の図面や帳票を電子化してるけど、ファイル名付けや属性入力作業をすべて手入力しているので手間と時間がかかりすぎる、もっと合理化できる道具はないかな？



良い解決方法があります！VPindexを利用することで、今まで手入力していたファイル名付けや属性入力作業を飛躍的に合理化できます。

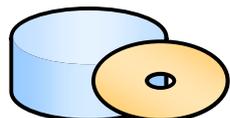
2、VPindexによるインデックス処理の流れ



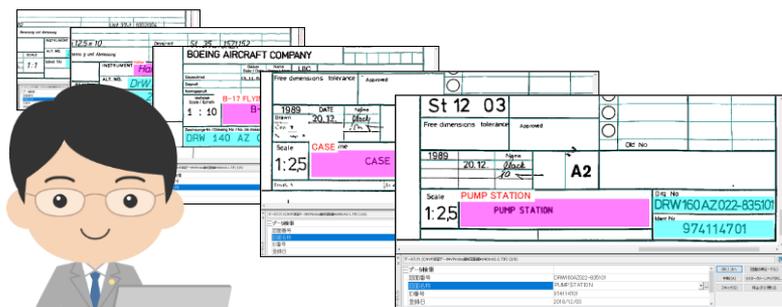
スキャンング → 自動クリーンアップ ⇒ 表題欄OCR処理 ⇒ データベース出力 → 各種管理システム



スキャナー
[ラスタ]



外部データ
[ラスタ]



管理DB

各種データベース



自動リネーム

img-123123001.tif
img-123123002.tif
img-123123003.tif
img-123123004.tif
img-123123005.tif



DRW160AZ022-835_PUMP STATION.tif
DRW160AZ022-836_CASE.tif
DRW140AZ005-710_B-17 FLYING.tif
DRW260AZ028-220_Peilstab.tif
DRW260AZ020-620_Halter.tif

スキャン直後

自動リネーム後

各種DB出力 (CSV出力例)

	A	B	C	D	E
1	ファイル名	図面番号	図面名称	ID番号	登録日
2	D:\SCAN\DRW160AZ022-835_PUMP STATION.tif	DRW160AZ022-835	PUMP STATION	974114801	2018/12/3
3	D:\SCAN\DRW160AZ022-836_CASE.tif	DRW160AZ022-836	CASE	974114701	2018/12/3
4	D:\SCAN\DRW140AZ005-710_B-17 FLYING.tif	DRW140AZ005-710	B-17 FLYING	974129001	2018/12/3
5	D:\SCAN\DRW260AZ028-220_Peilstab.tif	DRW260AZ028-220	Peilstab	974125801	2018/12/3
6	D:\SCAN\DRW260AZ020-620_Halter.tif	DRW260AZ020-620	Halter	274224	2018/12/3

3、VPindexの特徴

① 表題欄形状の罫線パターン認識

通常、社内には複数の表題欄（例えばA4用、A3用、A1用など）が存在します。識別したい複数の表題欄の罫線形状をパターン登録することで、インデックス処理時に自動的に合致する表題欄を識別し、指定したセル内をOCR処理します。

② ファイル名の自動付与

インデックス処理にて抽出した図面番号や図面名称等の情報をファイル名に自動付与します。柔軟にファイル名の定義ができるため、社内のルールに沿ったファイル名にすることができます。

③ 属性値によるフォルダ仕分け

抽出した属性値をそのままフォルダ名に割り当てることができますので、属性値を流用したフォルダ仕分けが行えます。例えば図面の標題欄から抽出した図面名や図面作成日などでフォルダ仕分けが行えます。

④ 各種データベース形式による属性出力

インデックス処理にて取得した属性情報はCSVファイル、属性付きXDWファイルをはじめとして各種データベース形式で出力することができます。また帳票形式に準じたCSV構造にカスタマイズ（※ VPindexのみの機能）することもできます。

⑤ OCRの認識率アップ

OCR処理する各セルごとに、OCRする文字種（日本語、英数字、数字のみ、特定文字）をセットすることができます。これにより認識率の飛躍的な向上が期待できます。また特定の文字列が使用される項目はリスト登録機能を使用すると便利です。

⑥ ラスタークリーンアップ

インデックス処理中に図面のクリーンアップ（自動ゴミ除去、自動傾き補正、自動回転等）をバッチ処理で行うことができます。これにより品質の向上した状態の最終図面を保管および管理することが可能となります。

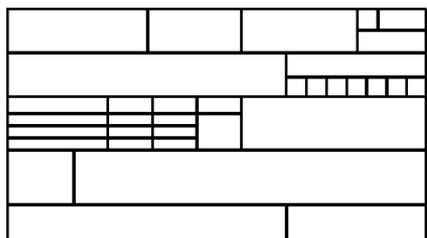
3-1、表題欄形状の罫線パターン認識

■複数パターンの表題欄を自動識別！

通常、社内には複数の表題欄（例えばA4用、A3用、A1用など）が存在します。識別したい複数の表題欄のフレーム形状をパターン登録することで、インデックス処理時に自動的に合致する表題欄を識別し、指定したセル内をOCR処理します。
 ※インデックス処理（表題欄抽出&OCR処理）が行えるフォーマット形式はXDW,PDF,TIF,JPEG等です。

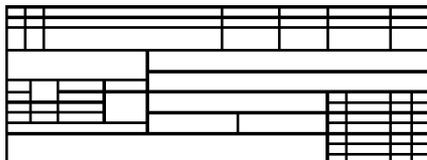
例：パターン1

Hersteller / Material / Matière		Gewicht / Weight / Poids		Teilenummer / Part No. / No. de pièce	
MANUFACTURER: BOEING AIRCRAFT COMPANY		35000 POUNDS LOADS		Kategorie: Modell No. / No. mod.	
Gezeichnet / Drawn		Name / Name / Nom		Documentation / Description / Documentation	
30.05.1992		LBC			
Verstärkung / Reinforcement / Renforcement		Bedeutung / Description / Description			
1 : 10		B-17 FLYING FORTRESS			
Zeichnungs-Nr. / Drawing No. / No. de dessin		Ident. Nr. / Ident. No. / No. d'ident.			
DRW 140 AZ 005 - 211.002		974125801			



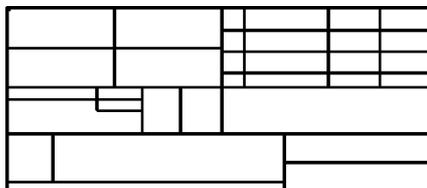
例：パターン2

Teil / Part		Abmessung / Dimension / Dimension		Gezeichnet / Drawn	
Halter		20.12.1989		ST 35	
INSTRUMENT		Name / Name / Nom		Date / Date / Date	
Halter		DrW260AZ020-620.208		15/11/52	
Skizze / Drawing		Skizzen-Nr. / Drawing No. / No. de dessin		Skizzen-Nr. / Drawing No. / No. de dessin	
274 224					

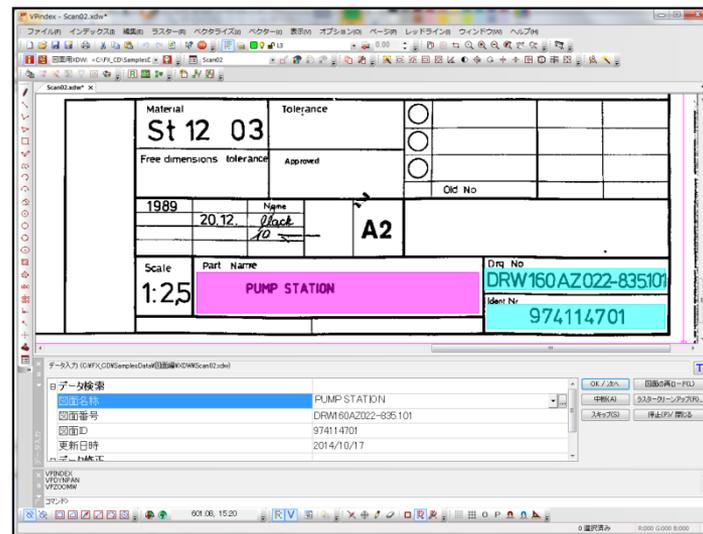


例：パターン3

Material		Tolerance			
St 12 03					
Free dimensions / Tolerance		Approved			
1989		20.12.		Name / Name / Nom	
				A2	
Scale		Part Name		Dwg No	
1:25		CASE		DRW 160 AZ 022 835.101 - 101	
				Ident. No.	
				974114801	



例えば、インデックス処理時に処理対象の図面がパターン1~パターン3の中のどれに合致するかを自動判別します。下図の例ではパターン3が自動的に使用されます。

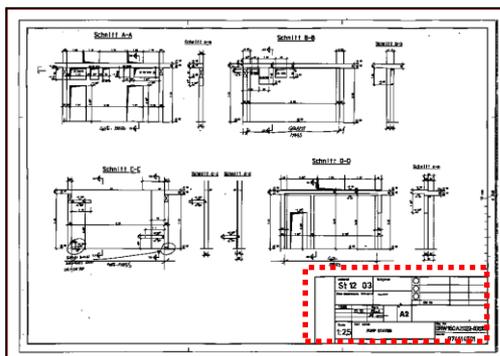


3-2、ファイル名の自動付与

■大量にあるスキャンデータのファイル名を手入力でリネームするのは大変！

紙図面の表題欄をOCR処理し、図面番号や図面名称等の情報をファイル名に自動付与します。
※インデックス処理（表題欄抽出&OCR処理）が行えるフォーマット形式はXDW,PDF,TIF,JPEG等です。

表題欄をOCR処理



Material	St 12 03	Tolerance			
Free dimensions tolerance		Approved			
1989	20.12.	Name	black	A2	
			10		
Scale	1:25	Part Name	PUMP STATION	Drw. No.	DRW160AZ022-835101
				Ident Nr	974114701

例：図面番号_図面名称を自動付与！

img-123123001.tif
img-123123002.tif
img-123123003.tif
img-123123004.tif
img-123123005.tif

リネーム前

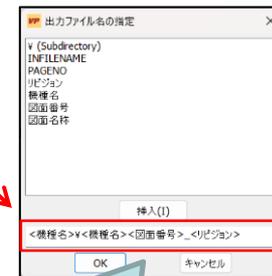
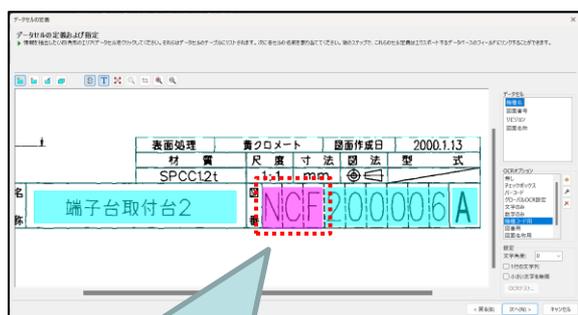
DRW160AZ022-835.101_PUMP STATION.tif
DRW160AZ022835.101-101_CASE.tif
DRW140AZ005-710.105_B-17 FLYING FORTRESS.tif
DrW260AZ028-220.209_Peilstab.tif
DrW260AZ020-620.208_Halter.tif

リネーム後

3-3、属性値によるフォルダ仕分け

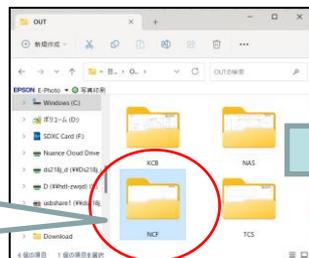
■大変なフォルダ仕分けも楽々！

抽出した属性値をそのままフォルダ名に割り当てることができます。例えば図面のタイトル欄から抽出した図面名や図面作成日などフォルダ名に割り当て、これらの属性値を流用したフォルダ仕分けが行えます。

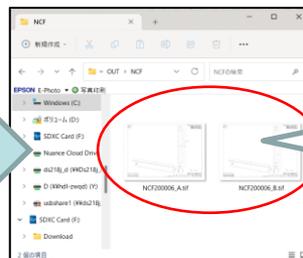


処理後

出力フォルダに機種名のフォルダが自動で作成されます。
例：NCFフォルダが作成されます。



該当するファイルのみが、このフォルダに自動的に仕分けされます。
例：機種名[NCF]の図面がNCFフォルダに仕分けされます。

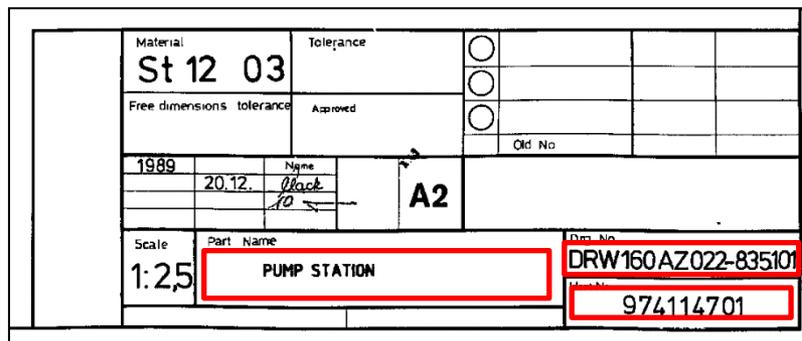
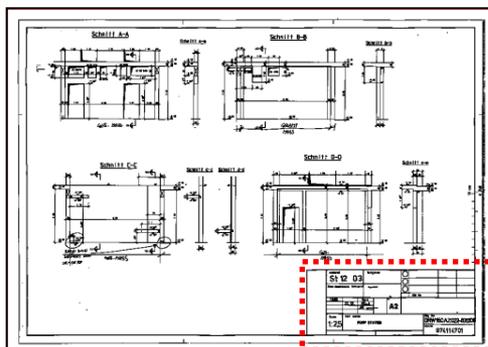


3-4、各種データベース形式による属性出力

■ 図面ごとに属性情報を手入力するのは大変！

紙図面の表題欄をOCR処理し、図面番号や図面名称等の情報をCSVファイルもしくは属性付きXDWファイルに出力します。
 ※インデックス処理（表題欄抽出&OCR処理）が行えるフォーマット形式はXDW,PDF,TIF,JPEG等です。

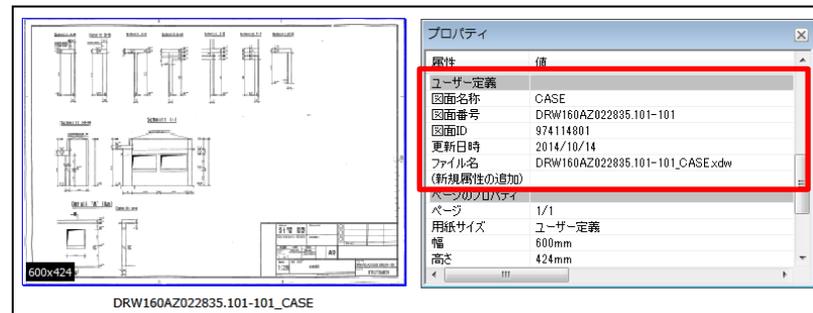
表題欄をOCR処理



CSVファイル出力

ファイル名	図面名称	図面番号	図面ID
DRW160AZ022835.101-101_CASE.tif	CASE	DRW160AZ022835.101-101	974114801
DRW160AZ022-835.101_PUMP STATION.tif	PUMP STATION	DRW160AZ022-835.101	974114701
DrW260AZ020-620.208_Halter.tif	Halter	DrW260AZ020-620.208	274224
DRW150AZ033-895.101-101_Scheibe.tif	Scheibe	DRW150AZ033-895.101-101	570133701

ユーザー定義属性付きXDWファイル出力



3-5、OCRの認識率アップ

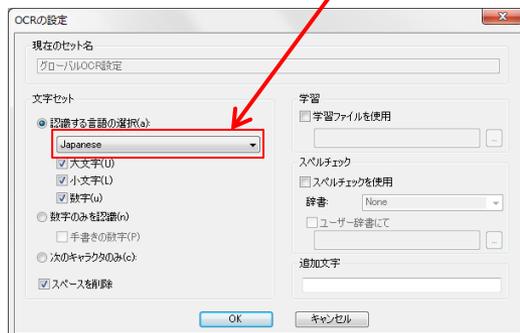
■OCRの文字種を各セルごとにセット可能！

表題欄内の各セルごとにOCRの文字種を指定することができます。例えば日本語、英数字、英字のみ、数字のみ、特定文字を指定することで文字認識の精度を飛躍的に高めることができます。

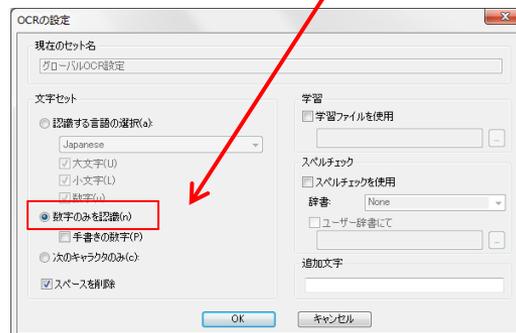
※インデックス処理（表題欄抽出&OCR処理）が行えるフォーマット形式はXDW,PDF,TIF,JPEG等です。

Material St 12 03		Tolerance		<input type="radio"/>			
Free dimensions tolerance		Approved		<input type="radio"/>			
				<input type="radio"/>			
		Old No.					
1989	DATE	Name					
Drawn	20.12.	Black					
Gertrud		Am					
Name							
Scale	Part Name		Drw. No.				
1:2.5	ケース		DRW 160 AZ 022-835.101-101				
			Ident. Nr.	974114801			
11/24/21	11/24/21		Jurch				

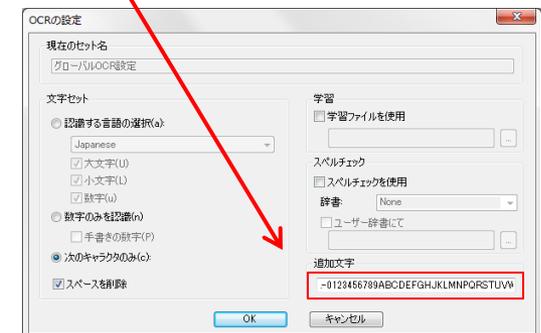
日本語認識



数字認識



特定文字を使用

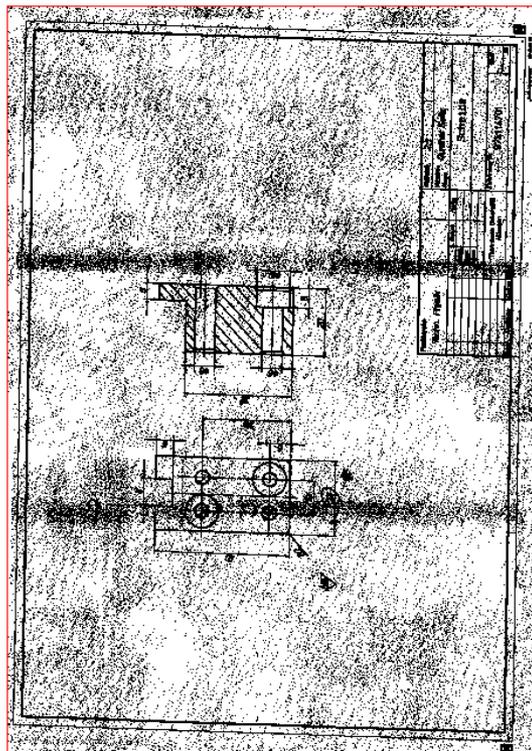


3-6、ラスタークリーンアップ

■品質の劣化した図面を自動クリーンアップ！

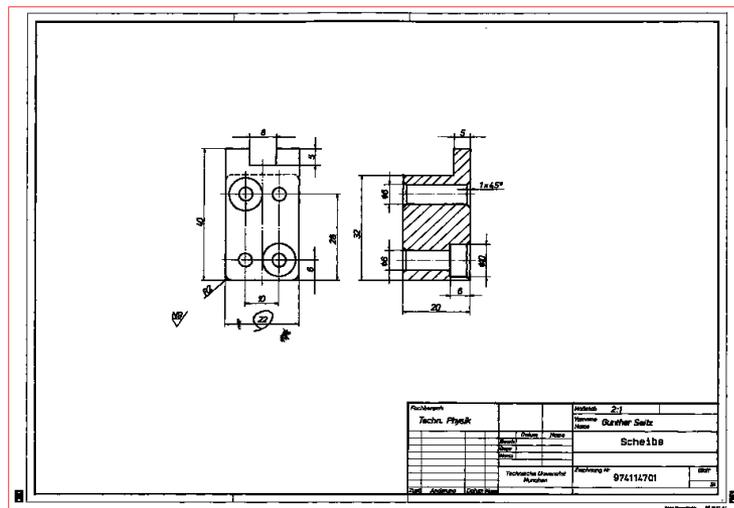
インデックス処理中に図面のクリーンアップ（自動ゴミ除去、自動傾き補正、自動回転等）をバッチ処理で行うことができます。これにより品質の向上した状態の最終図面を保管および管理することが可能となります。

※ラスタークリーンアップが行えるフォーマット形式はXDW,PDF,TIF,JPEG等です。



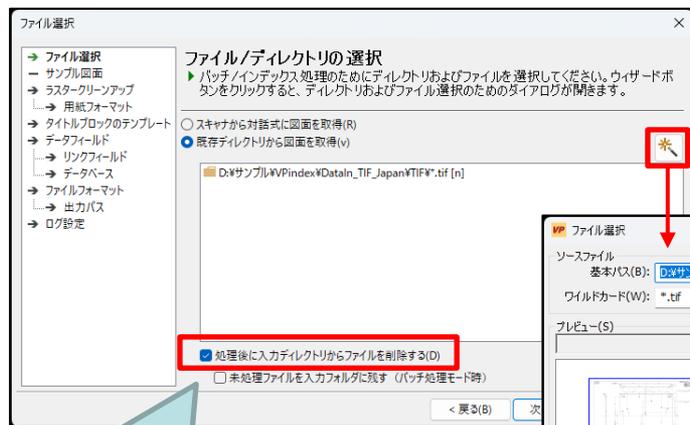
オリジナル

自動処理でゴミ除去、傾き補正および表題欄の位置を検知した上での自動回転等が行えます。



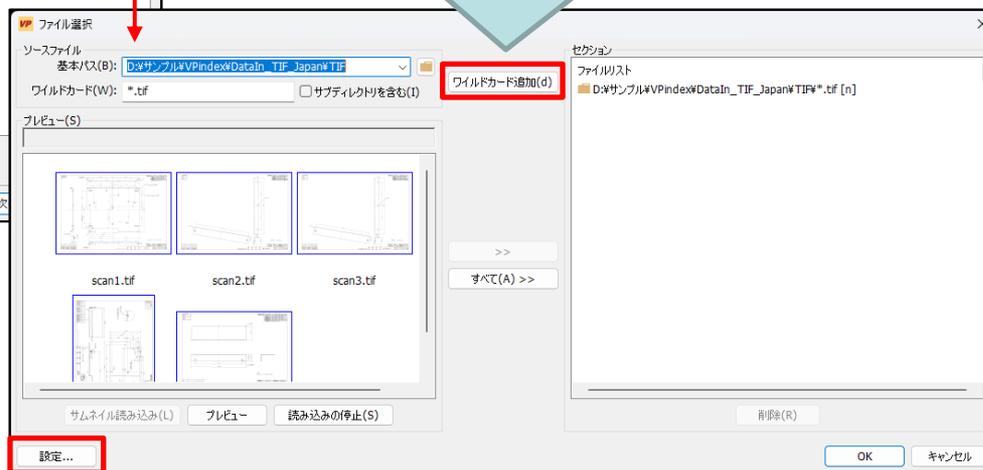
クリーンアップ後

4-1、処理（入力）ファイルの指定



インデックス処理後に元ファイルを自動的に削除することができます。

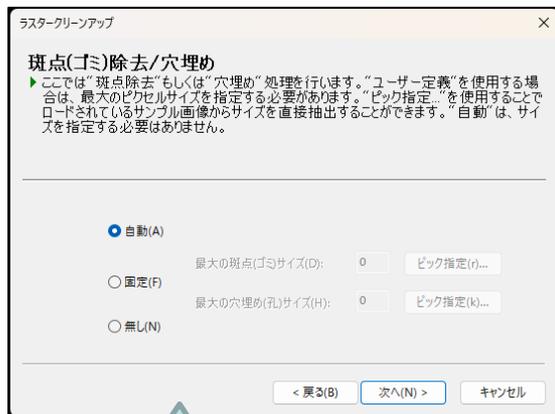
処理対象のファイルは、選択による指定以外に、指定した入力フォルダ内の全ファイルを処理対象とすることができます。
※拡張子によるフィルター指定有り。



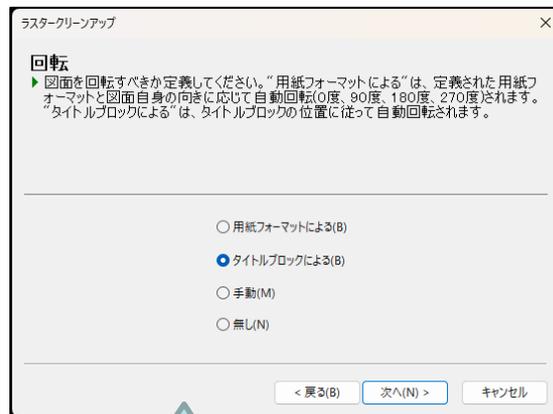
マルチページファイルの場合、マルチページを分解してシングルページにするか、マルチページを保持して1ページ目（もしくは指定ページ）をインデックス処理するか指定できます。

4-2、ラスタークリーンアップの定義

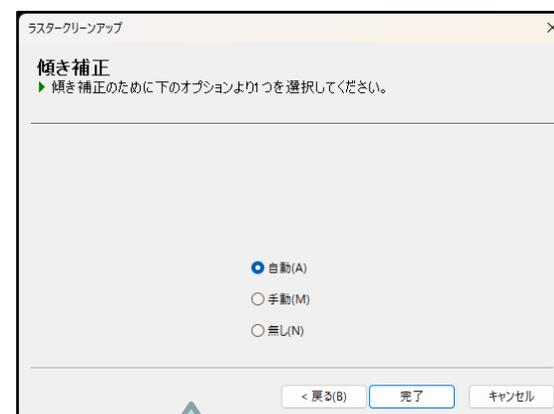
以下は代表的なラスタークリーンアップ機能です。
※その他、様々なラスタークリーンアップ機能をセット可能



スキャン時の汚れ（ゴミ）を自動もしくはサイズ指定にて除去します。

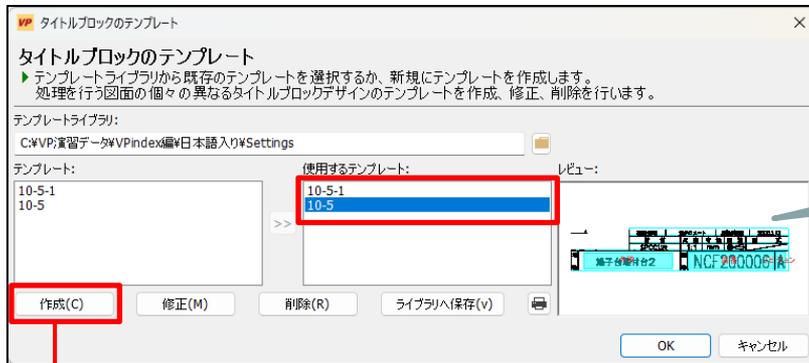


「タイトルブロックによる回転」を使用すると、図面全体の向きをタイトル欄の位置を判別して自動的に所定の向きに回転します。



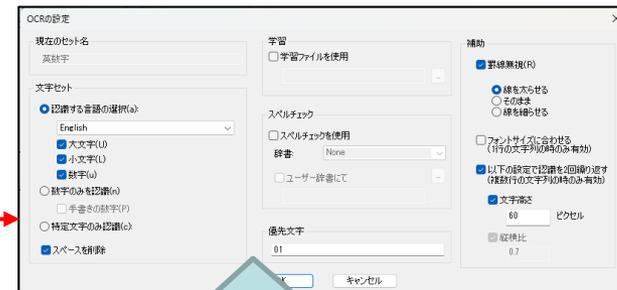
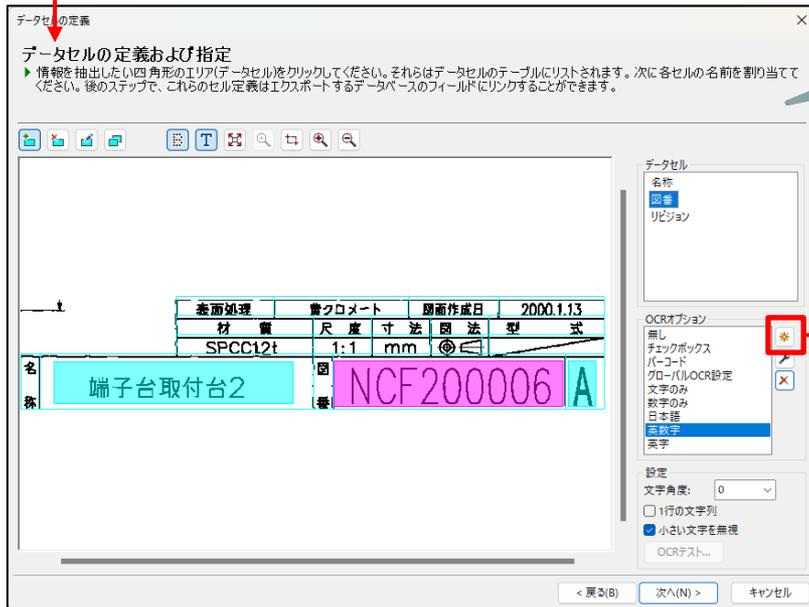
スキャン時の傾きを自動もしくは手動にて補正します。

4-3、テンプレート登録



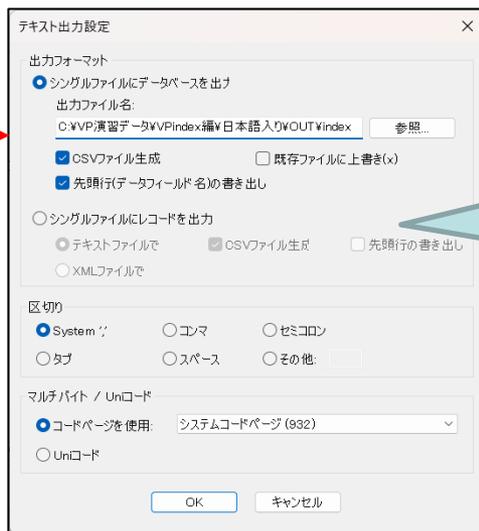
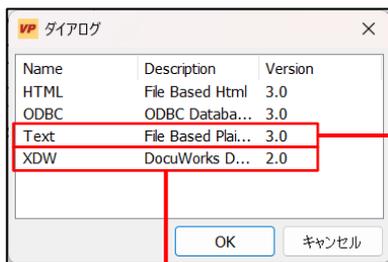
社内には存在する様々な図面表題欄の罫線パターン（テンプレート）を登録します。※複数パターンを登録可能

OCR処理させる場所およびセル名、各々の場所に対して使用する文字種（OCR詳細設定）をセットします。

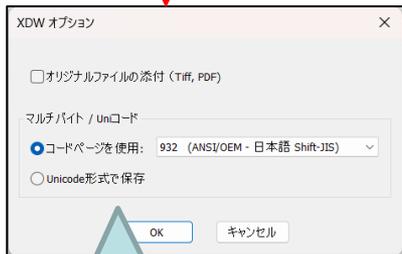


各セルに対して、認識する言語（日本語、英数字、数字）や優先文字、罫線無視などをセットします。これにより認識率の飛躍的な向上が期待できます。

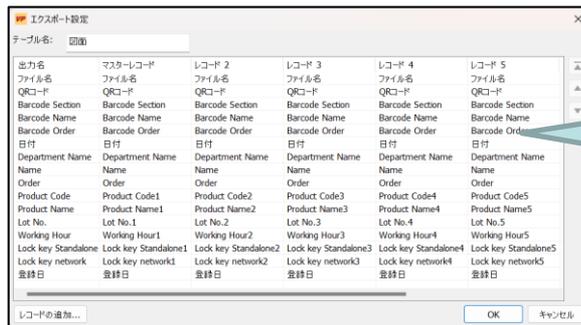
4-4、属性出力 (CSV) の定義



CSV出力時の各種設定を行います。
CSV出力は、1つのCSVファイルに全レコードを出力するか、もしくは入力ファイルと同名のCSVファイルを個々に出力するかを選択することができます。

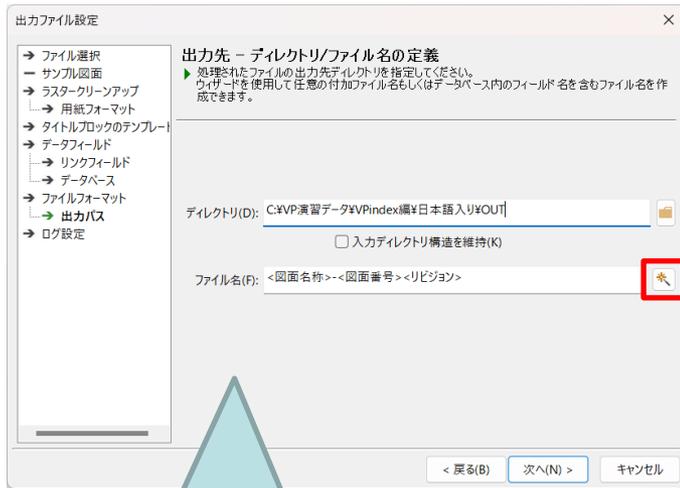


ユーザー定義属性付きのXDWファイルに出力します。

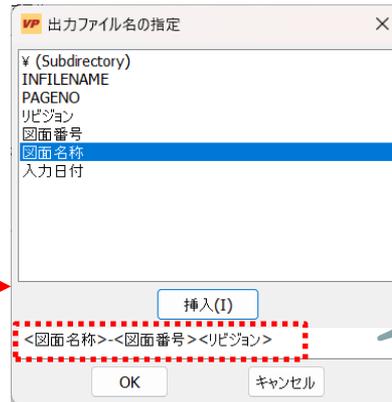


CSVの構造はカスタマイズ可能です。
例えばオリジナルと同様の帳票形式でCSV出力したい場合に有効です。
※本機能はVPindexのみの機能です。

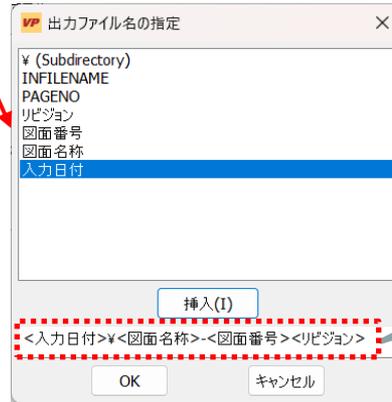
4-5、出力先およびファイル名付けの定義



リネーム後の出力先フォルダ、およびファイル名のルール付けを指定します。



属性項目を組み合わせ、ファイル名付けのルールをセットすることができます。



属性項目の後に¥マークをいれることで、属性値のフォルダが作成されます。これにより属性値によるフォルダ仕分けが可能となります。

5、ラスター編集機能 VPindexのみ

VPindexには、VPraster同等（CADアドオン版は除く）のラスター編集機能、ベクター編集機能が付いています。

ラスターデータを背景に、.dwg互換のCADデータを作図することができます。

また、ラスターベクター混在のデータを、AutoCAD準拠のフォーマット、ラスターDWG形式に保存することが可能です。

ラスター編集機能の詳細は、VPHybridCADの資料、もしくはカタログをご参照ください。

6、VPindexシリーズ ラインナップ

製品ラインナップ：

製品名	対話型OCR処理	バッチ形式OCR処理	帳票形式CSV出力	バーコードQRコード認識	各種ラスター編集	差分比較
VPindex	○	○	○	○	○	○
VPindex lite	○	×	×	×	×	×

VPindexシリーズはWindowsスタンドアロン環境でのみご使用いただけます。
その他、機能詳細は「製品カタログ」を御参照ください。

動作環境：

対応OS : Windows 10(32bit,64bit) / 11

お問合せ



Digital e-Tech Corporation

<http://www.d-etech.co.jp>

Always Ahead

株式会社デジタル・イーテック

〒418-0114

静岡県富士宮市下条627番地の1

TEL: 0544-58-0900

FAX: 0544-58-0904

Mail: VPinfo@d-etech.co.jp

URL: <https://www.d-etech.co.jp>