

VPindexシリーズによる 表題欄OCR処理（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。

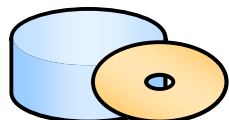
2、VPindexによるインデックス処理の流れ



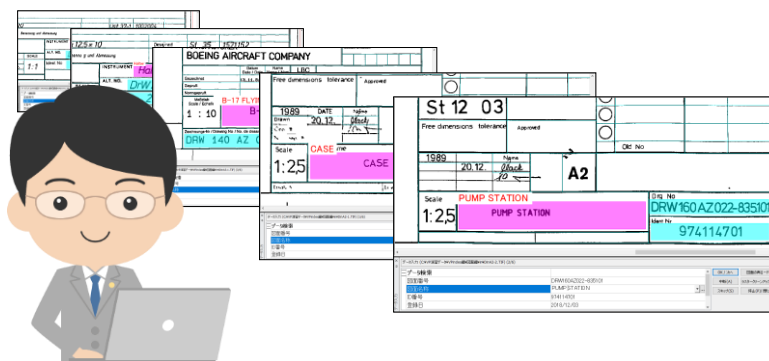
スキャンング → 自動クリーンアップ ⇒ 表題欄OCR処理 ⇒ データベース出力 → 各種管理システム



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)

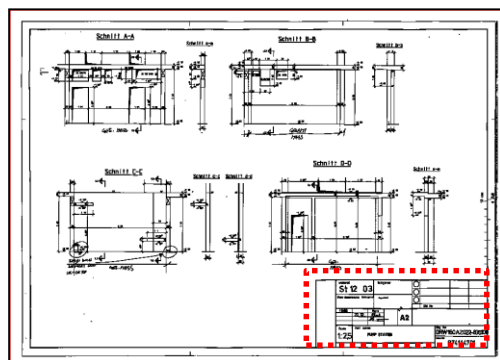
	A	B	C	D	E
1	ファイル名	図面番号	図面名称	ID番号	登録日
2	D:\SCAN\DRW160AZ022-835_PUMP STATION.tif	DRW160AZ022-835	PUMP STATION	974114801	2018/12/3
3	D:\SCAN\DRW160AZ022-836_CASE.tif	DRW160AZ022-836	CASE	974114701	2018/12/3
4	D:\SCAN\DRW140AZ005-710_B-17 FLYING.tif	DRW140AZ005-710	B-17 FLYING	974129001	2018/12/3
5	D:\SCAN\DRW260AZ028-220_Peilstab.tif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

3-1、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	10	
			A2		
Scale	1:25	Part Name	PUMP STATION	Drw. No.	DRW160AZ022-835101
				Ident Nr	974114701



図面名を自動付与！

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



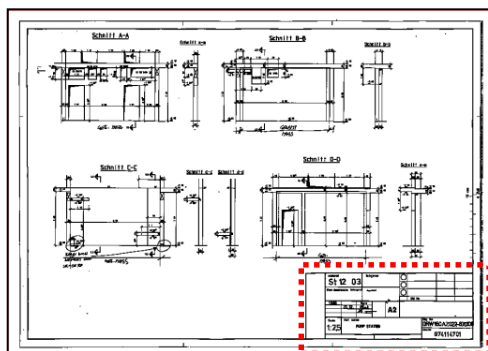
DRW160AZ022-835.101_PUMP STATION.tif
DRW160AZ022835.101-101_CASE.tif
DRW140AZ005-710.105_B-17 FLYING FORTRESS.tif
DrW260AZ028-220.209_Peilstab.tif
DrW260AZ020-620.208_Halter.tif

3-2、属性情報の自動抽出

■ 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



Material St 12 03		Tolerance		<input type="radio"/>	
Free dimensions tolerance		Approved		<input type="radio"/>	
1989		Name 20.12. Black 10		A2	
Scale 1:25		Part Name PUMP STATION		Drw. No. DRW160AZ022-835101	
				974114701	

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

属性付きXDWファイル出力

属性	値
ユーザー定義	
図面名称	CASE
図面番号	DRW160AZ022835.101-101
図面ID	974114801
更新日時	2014/10/14
ファイル名	DRW160AZ022835.101-101_CASE.xdw
(新規属性の追加)	
ページの幅	1/1
用紙サイズ	ユーザー定義
幅	600mm
高さ	424mm

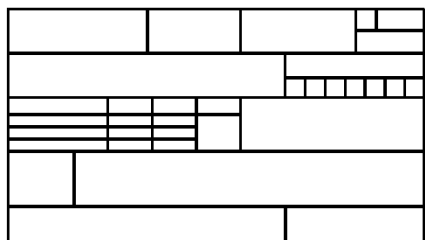
3-3、表題欄形状のパターン認識

■複数パターンを表題欄を自動識別!

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。

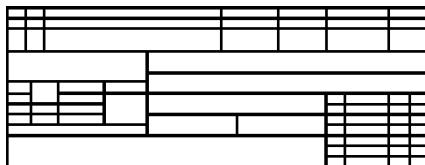
例：パターン1

Hersteller / Material / Matière		Gewicht / Weight / Poids		Teilenummer / Part No. / No. de pièce	
MANUFACTURER: BOEING AIRCRAFT COMPANY		35000 POUNDS LOADERS		Kategorie: Modell No. / No. mod.	
Gezeichnet / Drawn		Name / Name / Nom		Documentation / Description / Documentation	
20.05.1992		LBC		A3	
Verstärkung / Reinforcement / Renforcement		Bezeichnung / Description / Désignation			
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		



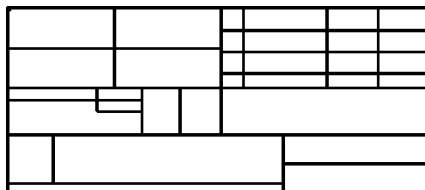
例：パターン2

Teil / Part		Abmessung / Dimension / Dimension		Gezeichnet / Drawn	
Rohr 4,3x12,5x10		ST 35		15/11/52	
INSTRUMENT					
Halter					
DATE		SCALE		DRW No.	
29.11		---		DrW260AZ020-620.208	
I. Approved		Ident. No.		274 224	

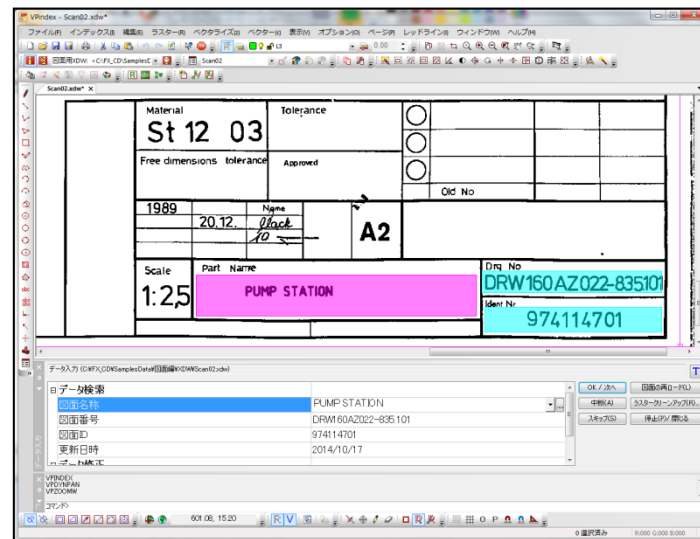


例：パターン3

Material		Tolerance			
St 12 03					
Free dimensions / Tolerance		Approved			
1989		Name		Old No.	
20.12		Black 10		A2	
Scale		Part Name		Drw No.	
1:25		PUMP STATION		DRW160AZ022-835101	
				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。



3-4、OCRの認識率アップ

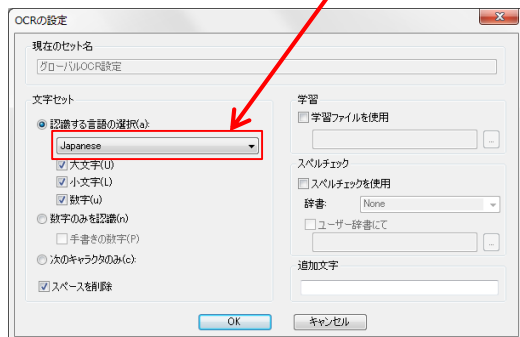
■OCRの文字種を各セルごとにセット可能！

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。

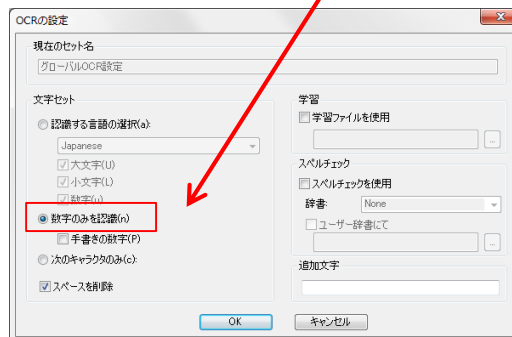
※インデックス処理（表題欄抽出&OCR処理）が行えるフォーマット形式はXDW,PDF,TIF,JPEG等です。

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		Black					
Name							
Scale	Part Name		Drw No				
1:25	ケース		DRW 160 AZ 022-835.101-101				
			Ident No	974114801			
11/21/2010	11/21/2010						

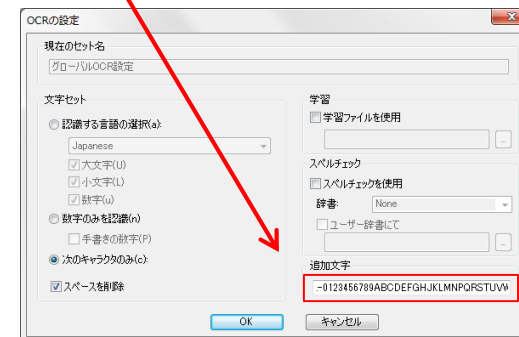
日本語認識



数字認識



特定文字を使用

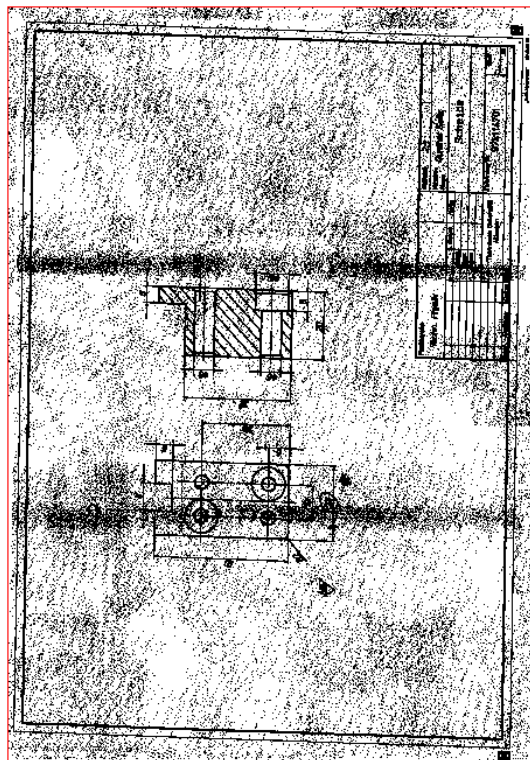


3-5、ラスタークリーンアップ

■品質の劣化した図面を自動クリーンアップ！

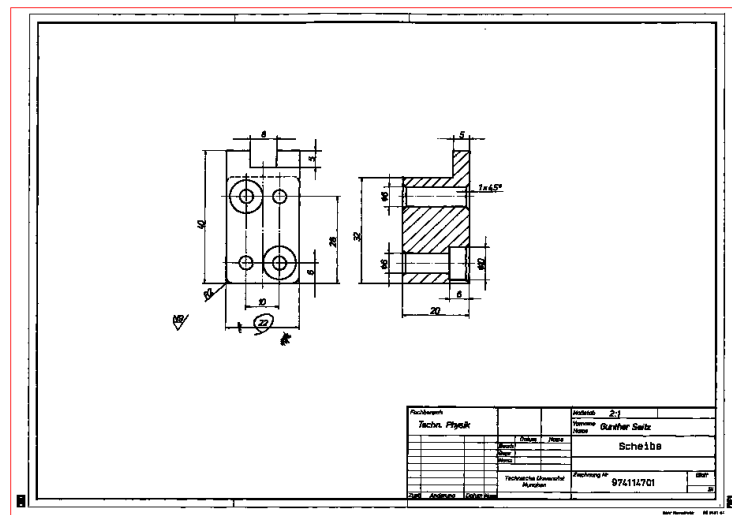
インデックス処理中に図面のクリーンアップ（自動ゴミ除去、自動傾き補正、自動回転等）をバッチ処理で行うことができます。これにより品質の向上した状態の最終図面を保管および管理することが可能となります。

※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、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。

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