

YSS独自のOCR技術

特殊なフォントの認識や図形認識も可能な独自技術を搭載。

これまでにない紙データの活用を可能にします。

高精度のOCR

独自技術により、文字の認識精度を高めました。

また、特殊なフォントであっても文字として認識し、データ抽出できます。

OCR 範囲の自動抽出

座標ではなく、見出し、段組、文字パターン、フォント種別、フォントサイズなどの条件を組み合わせることで、フォーマットの異なる画像から特定の項目のみ抽出できます。

OCR 結果の自動判定

OCR 元画像と、抽出結果テキストを重ね合わせて比較。

OCR の誤りを自動で検出します。

OCR 結果判定の一覧 PDF 出力

OCR 結果の自動判定内容を一覧として PDF に出力します。

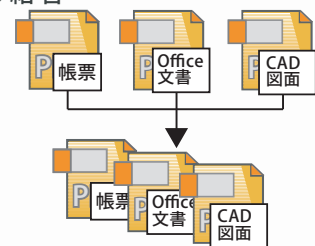
結果の手直しや、修正前の証跡としてご利用頂けます。

PDFメイクアップ

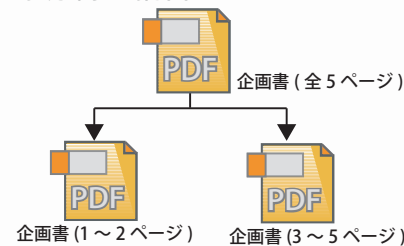
PDFへ様々な加工を可能にするPDF加エライブラリです。

システムから PDF へ手軽に文字の追記や PDF 同士の結合が行えます。

① 結合



② 分割・抽出



【効果】

- ・ PDF 同士の結合
- ・ PDF から特定ページのみ抽出など、様々な場面でご利用頂けます。

■ ドキュメント関連 (PDF / 画像解析 / 電子化) 受託開発

PDF 高度技術や画像解析技術について製品だけではカバーできない、かゆいところに手の届くカスタマイズをお受けいたします。

PDF ソリューションお問い合わせ先

YSS 株式会社 ワイ・エス・エス

〒210-0007  
 神奈川県川崎市川崎区駅前本町3-1 NMF川崎東口ビル8F  
 TEL : 044-211-9332  
 FAX : 044-211-9331  
 E-mail : eigyo@iothe.co.jp  
 URL : http://www.iothe.co.jp/

YOKOGAWA



横河電機株式会社様

Case Study

大量の画像データを社内システムへ自動的に取り込む技術として、YSS の独自技術をご採用頂きました。

使用製品・技術情報

IMG to Digi

YSS 画像解析技術

YSS OCR エンジン

PDF メイクアップ

ご利用ユーザ様情報

会社名 : 横河電機株式会社  
 所在地 : 〒180-8750 東京都武蔵野市中町 2-9-32  
 従業員 : 連結 19,685 名  
           個別 4,298 名 (2013年3月31日現在)  
 URL : http://www.yokogawa.co.jp/

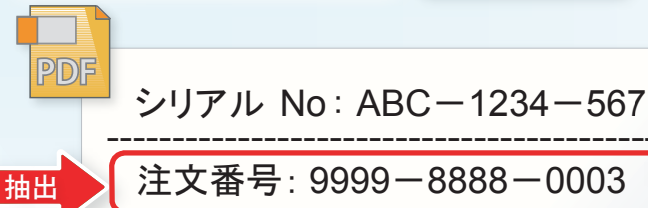


# 処理概要≫画像 PDF からテキストデータを自動抽出し、既存システムで利用可能な形に再構成

## システム概要

### 自動データ抽出 Point 1

Point 1



- ・見出し
- ・段組
- ・英数字など文字パターン
- ・フォント種別（複数指定可）
- ・フォントサイズ

の条件指定を組み合わせることで、**自動的に OCR 範囲を決定します。**  
フォーマットが統一されておらず、OCR 対象を座標で特定できない画像イメージにも対応します。

OCR 元のスキャン画像と、サイズを合わせて形状を再現した OCR 結果テキストのイメージを重ねて比較します。  
PDF 比較で培った YSS 独自技術にて **OCR 結果の誤りが自動検出できるようになりました。**

### OCR 結果自動判定 Point 2

Point 2



確認

### OCR 結果一覧表 Point 2

Point 2

PDF	9999-8888-7777	
PDF	9999-8888-7777	OK
PDF	9999-8888-7777	
画像	9999-8888-0003	
比較結果	9999-8888-0003	NG
OCR 結果	9999-8888-0008	
	⋮	
		判定結果

自動判定した OCR 結果を元画像と並べ、一覧表として PDF 出力。  
目視でのテキスト修正を容易にします。

### OCR 結果手直し

## データ抽出技術

- 特殊なフォントでも認識できる独自 OCR を搭載。画像からの確実にテキストデータを抽出します。
- OCR で抽出する範囲を、複数の指定条件から自動判定。フォーマットの異なる用紙からでも、座標を指定せずにデータ化したい箇所だけ OCR 実行できます。
- OCR 結果を、元画像と重ね合わせて自動誤り判定。結果を一覧表として出力し、目視での修正作業コストを大幅に削減します。

### Point 1

#### 自動データ抽出



複数の指定条件から抽出位置を自動判定

OCR 対象とするテキスト位置を座標ではなく、見出し・段組・文字パターン・フォントなどの組み合わせから自動判定します。

### Point 2

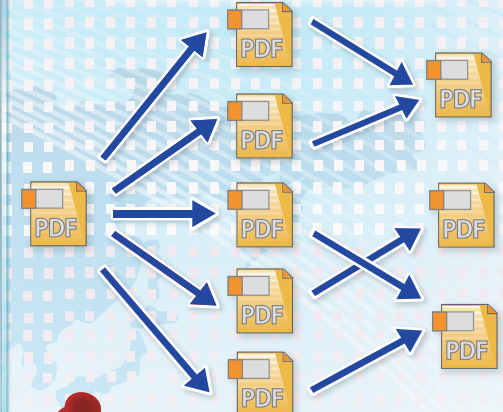
#### OCR 結果自動判定 & OCR 結果一覧表



元画像と OCR 結果を自動比較し、誤りが無いかを「OK」・「NG」・「疑い有り」の 3 種類で一覧表として出力します。

### Point 3

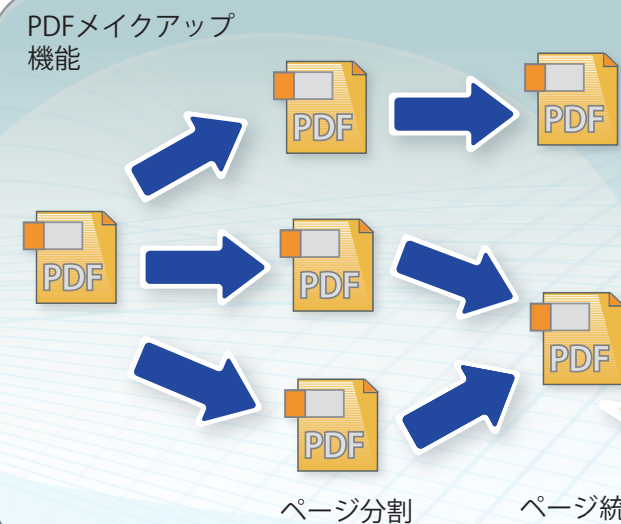
#### PDF 分割 / 統合



OCR 結果を元に、一つのドキュメント (PDF) から関連するページをまとめ、複数のドキュメントとして再構成します。

### PDF 分割 / 統合 Point 3

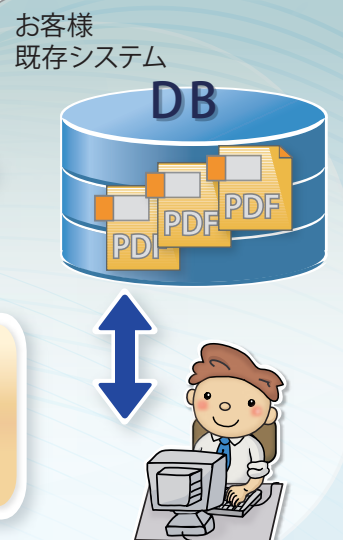
Point 3



### データ登録

OCR で抽出したデータを元に、既存システムで運用できる形へ PDF のページを再構成します。

### システム運用



## 導入のメリット

- 紙媒体でしか取り扱えなかったドキュメントを OCR で電子データ化し、社内システム間の **スムーズな連携を実現。**
- 自動データ抽出、自動結果判定による **OCR 作業の大幅な省力化と、手作業によるミス発生を低減。**
- 全て YSS 自社開発の製品でシステム構成する事で、お客様からの **お問い合わせへ迅速に対応。**