

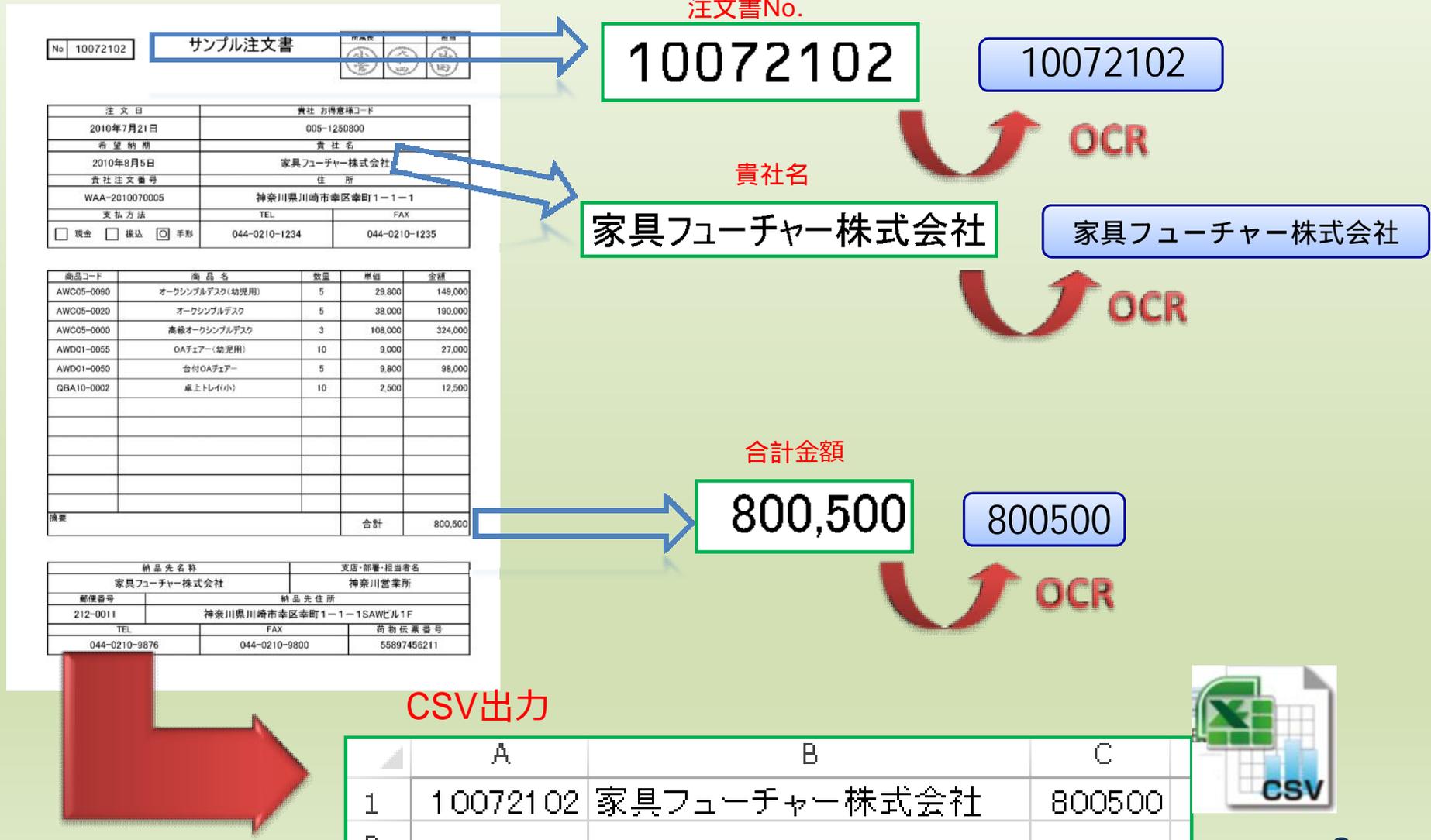
# 帳票OCR「WinReader p.Form」の ご紹介とOCR精度向上のコツ

2014.5

メディアドライブ株式会社  
営業部 第2営業グループ  
山本 晃弘

# 1. WinReader p.Formについて

「WinReader p.Form」は業務用帳票OCRソフトウェアです。  
 帳票に記入された活字文字を認識して、CSVファイル等にデータ出力できます。



## 2. OCRの手順(共通)

### OCRソフトウェアの起動

スキャナで帳票を画像ファイルとして取り込む

画像取り込みが終わったら、OCR処理が開始(手動・自動選択可)

OCR処理後、認識結果の確認・修正(承認)を行う

確認・修正が終わったら、テキストとPDF(TIF、JPEGも可)を出力



### 3 . WinReader p.Form システム構成図

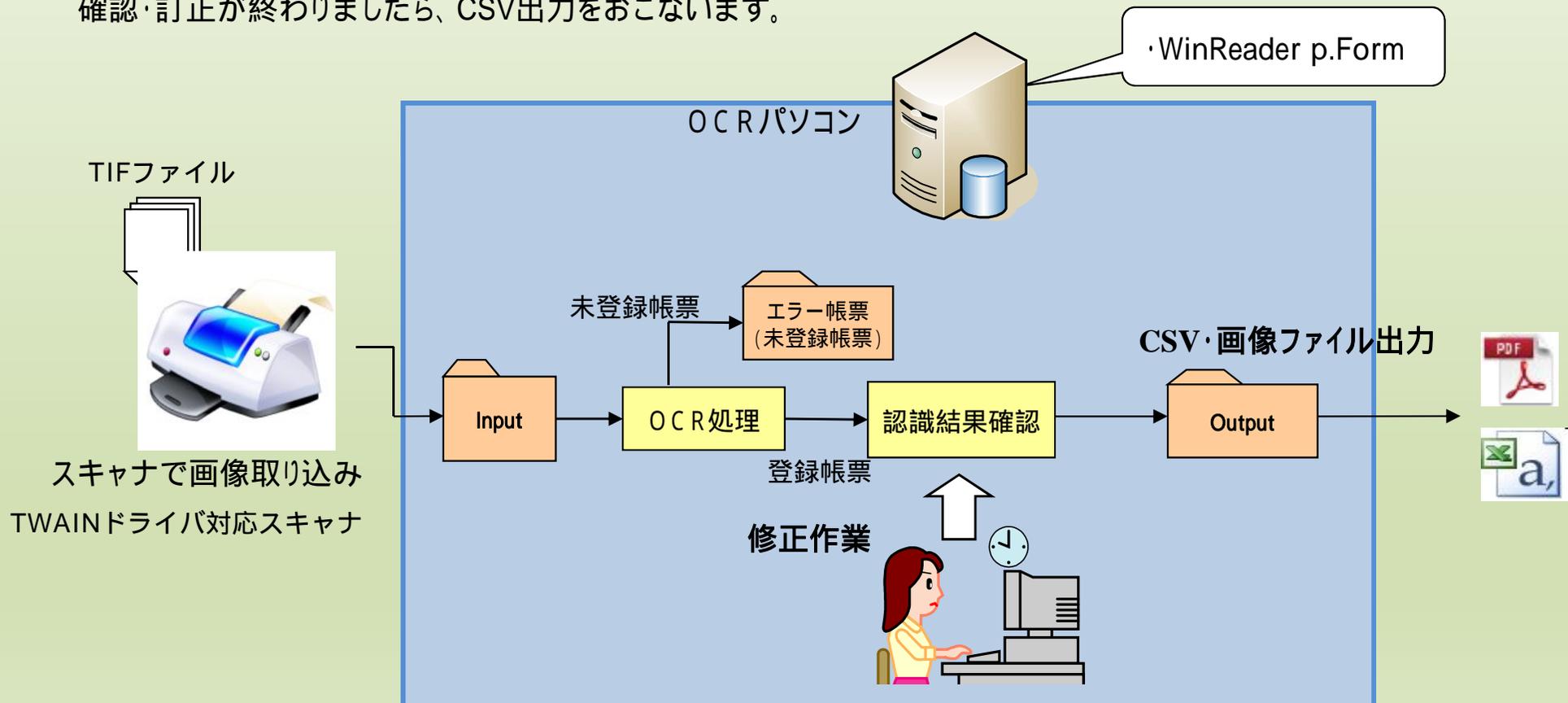
スキャナで取り込んだ画像をWinReader p.Formに取り込みます。

取り込んだ画像はあらかじめ登録された帳票と照合し、一致した画像はOCR認識が行われます。

OCR認識の結果が正しいかどうかを確認する画面をご用意しているため、表示された画像を確認しながら訂正します。

画面で確認できるため、紙を見ながら入力する必要はありません。

確認・訂正が終わりましたら、CSV出力をおこないます。



## 4. OCRの設定はユーザで可能

専門的な知識を必要とせず、ユーザでOCRの設定を行うことができます。

認識箇所を指定（マウスでクリック）

記載されている文字の種類を選択します。

The image shows a document with several fields labeled f1 through f10. A red box highlights field f1, which contains the text "No". A red arrow points from this box to the "フィールド設定" dialog box. The dialog box has several sections: "フィールド種別" with radio buttons for "固定", "手書", "マークチェック", "イメージ", "フリー", and "QR/バーコード"; "文字情報の設定" with input fields for "文字数" (1) and "修正候補数" (2), and a dropdown for "単語照合" (ういる); "文字属性の設定" with radio buttons for "活字" and "手書", and checkboxes for "記数", "力", "英", "漢", "ひ", "小数部桁数" (0), "詳細...", and "小文字"; "フィールド情報の設定" with a "名称" field and checkboxes for "データ検索時のキー", "混在帳票のIDフィールド", "データチェック式", and "空白不可"; "出力処理の設定" with radio buttons for "文字コード" (半角, 全角), checkboxes for "空白削除" (前, 後), "0詰め", "文頭", "行末", "文末", "固定長" (0), "バド", "イメージ", and "ここでは出力OFF"; and buttons for "番号変更...", "辞書処理...", "詳細設定...", "OK", "削除", and "キャンセル".

出力に付与する文字を入力  
します。  
例) 「,(カンマ)」  
カンマ区切りのテキストを  
出力する際に入力

# 5. オーバーレイ表示で、認識結果を確認・修正

画面上でOCR結果に誤りがないか確認し、修正後にPDF、CSV出力を行うことが可能です。

注文日	貴社 お得意様コード	
2010年7月21日	005-1250800	005-1250800
希望納期	貴社名	
2010年8月5日	家具フューチャー株式会社	
貴社注文番号	住所	
WAA-2010070005	神奈川県川崎市幸区幸町1-1-1	
支払方法	TEL	FAX
<input type="checkbox"/> 現金 <input type="checkbox"/> 振込 <input checked="" type="checkbox"/> 手形	044-0210-1234	044-0210-1235

認識結果は青文字で  
認識した文字の上部に  
表示

画像出力  
(PDF、TIFF、JPEG)



CSV出力



Adobe Reader window showing a scanned document titled "サンプル注文書". The document contains a table with the same data as the OCR table above. The text is clear and matches the original document.

txtout.csv window showing a CSV file with multiple rows of data. Each row contains the OCR results for a specific order, with the recognized text overlaid in blue on the original document image. The data matches the table above.

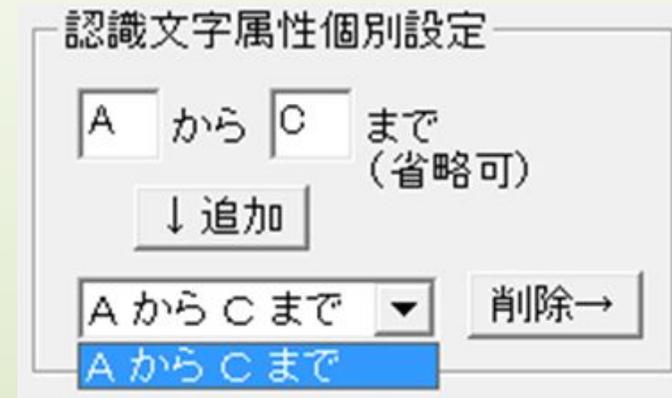
## 6. 認識設定の調整

### 文字種の限定

認識する文字種を限定することで、認識率を向上させる



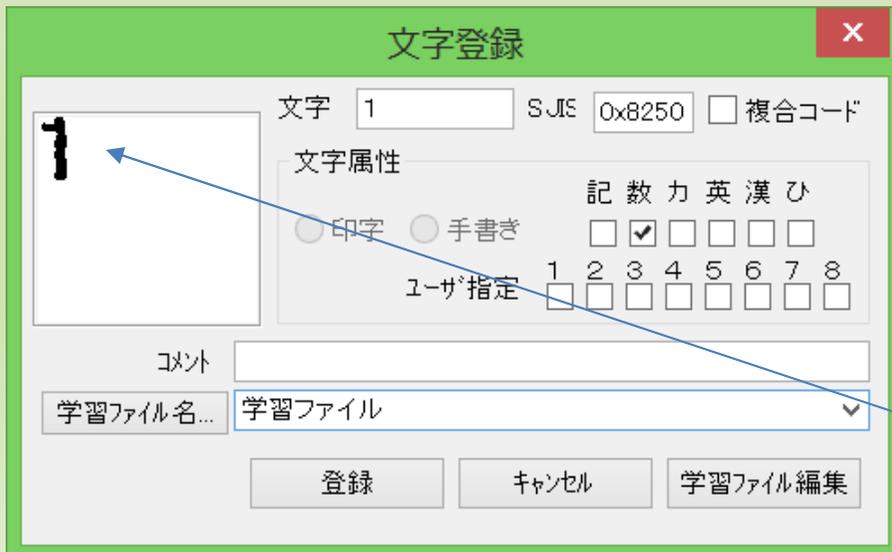
文字の種類を限定



さらに認識対象文字を限定することも可能。

### 文字学習機能の活用

文字を学習することにより、認識率を向上させる



文字の形状を記憶して、今後同じ形の文字を認識する場合に活用する

学習する文字を選択して、登録を行う

## 6. 認識設定の調整

### ノイズ除去サイズ設定

汚れなどのノイズを認識しないような設定を行うことが可能

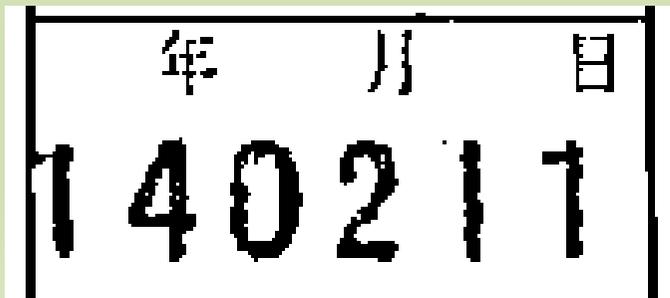


OCRでは、ノイズも文字として認識してしまう

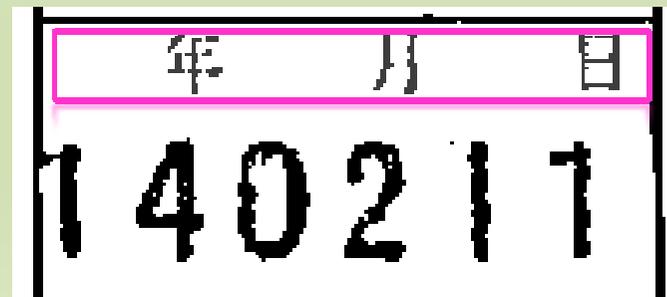
ノイズサイズ設定で記入された数値より、小さい独立した点や線は無視されて認識を行う

### 認識除外エリア

文字認識をしたくない箇所を選択して、認識対象から外すことが可能



上記の年・月・日の文字は不要である場合



認識除外エリアとして指定することによって、年・月・日を認識しないように設定が可能

## 7. 誤認識の判別

### データチェック (数値範囲)

数値の範囲を指定することで、範囲外の文字が認識された場合にエラーとする

1000以上、6000未満であればOK、範囲外はNGと設定した例



OKの時



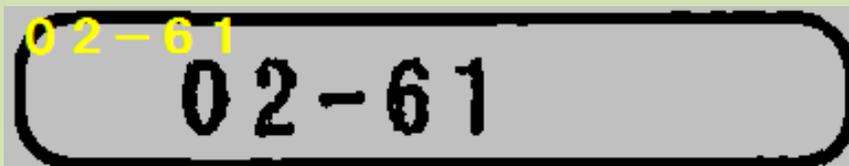
NGの時

誤認識が発生

### リジェクトレベルの調整

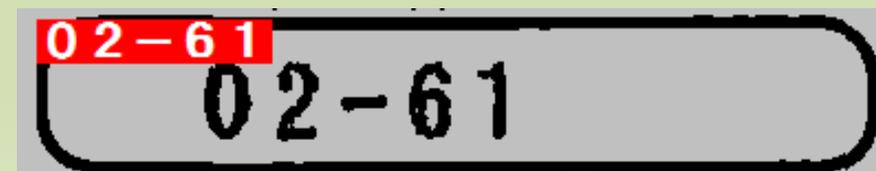
リジェクト (誤認識の可能性) の閾値の調整が可能

リジェクトレベル低



エラー無し

リジェクトレベル高



リジェクトエラー発生

## 8 . WinReader p.Form(パッケージ製品)で運用する利点

---

登録した帳票を素早くOCR処理が実行できる  
確認・訂正画面を持っているため、認識結果の確認が容易である  
帳票のレイアウトを判別し、レイアウト毎にCSVを出力できる  
パッケージ製品なので問題発生時の切り分けを行いやすい

## 9 . 導入前の注意点

---

帳票が変更・種類の増加が発生した時に、都度対応となる  
ユーザで設定して頂くか、メディアドライブで有償にて作業を実施  
混在した帳票を自動的に認識する場合は、最大30種類まで  
レイアウトによっては認識できない帳票がある  
枠線が黒く、はっきりしている必要がある  
薄い・途中で線が切れている・枠線と文字がくっついていると認識できない可能性がある

## 10. OCR機能 拡張製品

WinReader p.Formの上位製品として、「FormOCR」が御座います。御用途に合わせて製品をお選び頂きます。

OCRソフトウェア簡易比較	FormOCR v.6.0	WinReader p.Form
<b>認識</b>		
手書き認識		×
活字認識		
バーコード、QRコード認識		×
混在帳票対応 (IDフィールド使用) 混在した用紙の判別に使用		(手書き数字IDは不可)
<b>特定帳票対応</b>		
FAX画像補正 (帳票レイアウトの変更が必要) FAXで用紙を受け取る場合に必要	(オプション)	×
ドロップアウトカラー帳票認識 用紙に黒枠が無い場合に必要	(オプション)	×
<b>その他</b>		
複数訂正システム 複数台の端末で訂正を行う場合に必要	(オプション)	×

WinReader p.Form定価：¥220,000～

FormOCR定価：¥660,000～

WinReader p.FormからFormOCRへのアップグレードはできません。