



# OCRガイドブック

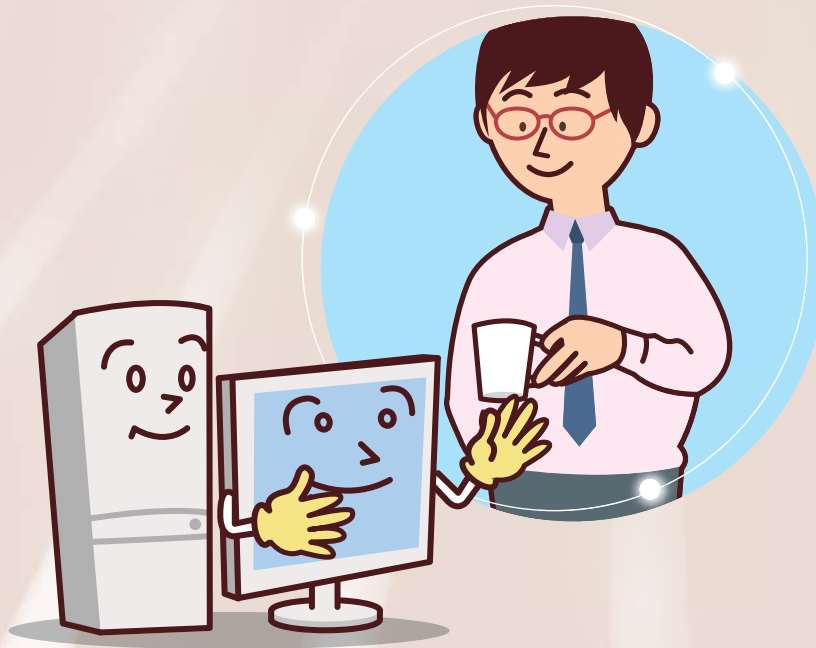
——— OCRを上手にご活用いただくために ———

**JEITA**

社団法人 電子情報技術産業協会

何故 OCR なのか	1
OCR の現在	2
将来の OCR	3
OCR の種類と特徴	4
<b>OCR 対応業務の紹介</b>	
[1] 一般企業	6
(1.1)異種混在読取	
(1.2)手作り帳票	
[2] 自治体	7
(2.1)シーリングはがき	
(2.2)デジタルペン	
[3] 流通業	8
(3.1)生協	
[4] 金融・保険	8
(4.1)スタンドスキャナー	
(4.2)総合振込依頼書	
(4.3)本人確認用OCR	
[5] 医療機関	10
(5.1)チェックマーク読み取り	
[6] 製造	10
(6.1)FAX	
[7] 学校・教育	11
(7.1)マークシート	
[8] 文書 OCR	11
(8.1)名刺OCR	
(8.2)デジカメOCR	
(8.3)ファイリング連携	
(8.4)PDF OCR	
(8.5)家計簿データ	
用語の解説	14

## ● 何故OCRなのか ●



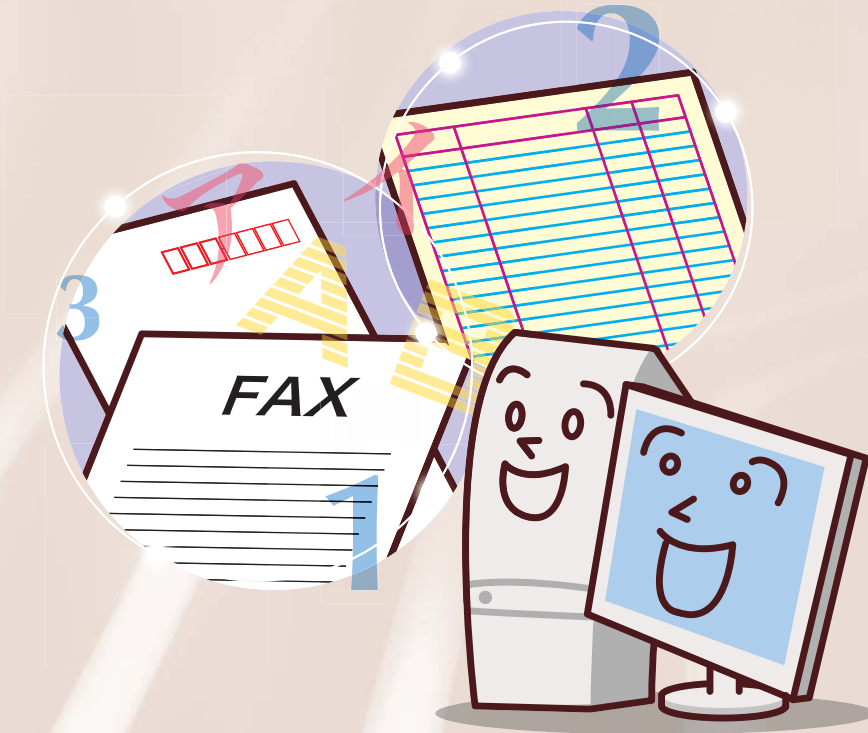
世の中ではペーパレス化、電子伝票といった動きがありますが、私達の机の上は書類や伝票類であふれんばかりになってしまいました。何故でしょう？それは、紙が人間にとって記録性や保管性、一覧性などに優れているからにほかなりません。また、人と人(子供と大人、文化の異なる人同士…)とのコミュニケーションには紙が最も優れたメディアであるからなのです。

その紙を媒体とするのがOCR(光学的文字読取装置)です。パソコンは入力手段なくしてはただの箱。その入力手段には利用形態に適した様々な機器が開発され、活用されており、その代表的なものがキーボードでありマウスです。しかし、これらを使っての入力は人手による作業となり、一般企業等では、労働環境の改善や社員教育の他、キーパンチャーの確保やその人件費等の入力コストの増大などへの対応が課題となってきています。

また、金融危機を発端とする景気の低迷や物価の下落、石油製品の高騰などが重なり、どのように業務の効率化を図り、経費節減を実現するかが重要となってきています。そこでどう入力作業を効率化し、入力コストを削減していくかという課題解決に大きな関心が寄せられています。一般的にキーボードからの入力では、紙に手書きされた情報やプリンタで紙に印字された情報を対象に人の手を介して行われますが、誰が考えても、「その紙から直接にデータを取り込めれば、こんな効率化は他にない。」と思われるでしょう。それができるのが正にOCRなのです。

現在の社会では、データの最終形態は電子データとして様々に加工され、社会的に役立つことを求められております。紙という最も優れた媒体から電子データを効率的に作り出す部分で、そこにOCRが存在し、入力省力化の旗手として多くの分野で今、活躍しています。

## ● OCRの現在 ●

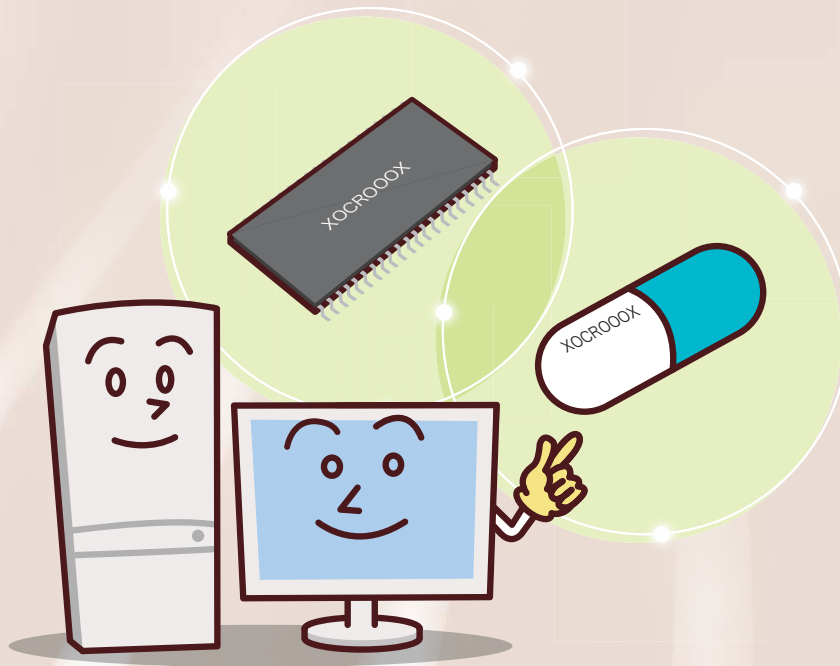


OCRは、「郵便番号読取装置」や「納税通知書の読取処理」などの身近なところでも使われ続けており、さらにコンピュータ技術や文字認識技術の進歩により、ますます使いやすくなり、利用範囲の拡大が可能になってきています。以前は利用者側の負担となる制限が多く、例えば専用用紙を必要とし、限られた印刷メーカーだけが可能なドロップカラーによる印刷が必要でした。しかし、いまや利用者のニーズに合わせて様々な機能が利用可能になり、また様々な形態のOCRが開発され、利用内容に合った機種を選べる様にもなりました。発展してきた内容を説明します。

1. ワードプロソフト、表計算ソフト、ファイリングソフト、翻訳ソフト等の市販のアプリケーション・ソフトはもちろんのこと、利用者が使用している業務システムとも容易にデータの交換が可能になっています。
2. OCR専用で作られた用紙の他に、現在ユーザが使用している帳票(既存の帳票)もそのままOCR対象帳票とすることが可能なものもあります。また、帳票をワードプロソフトなどの作画・罫線機能などで簡単に作って利用することも可能になりました。
3. 新聞や雑誌などの段組(レイアウト)だけでなく、読取対象文字種も漢字のみならず欧文も混在するなど複雑な段組の文書も直ちに電子データ化し、ワードプロの文書ファイルやインターネット用PDFやHTMLの電子化文書に変換を行えるものも多数あります。
4. 特殊な用途として、名刺やレシート等を読み取るOCRも普及しています。名刺に関しては、携帯電話で読み取ることもできる商品も出ています。
5. FAXで送信されてきたデータも自動受信し、電子データ化して資料請求や通信販売などのデータとして直ちに利用できる形にしてくれるものもあります。
6. スキャナー部では、対応の紙質の範囲も広がり、薄い紙からシーリングハガキなど分厚いものまで対応可能なものも出現しました。上質紙やOCR紙のほかPPC用紙や再生紙等も利用可能になりました。
7. OCRシステムの価格も、スタンドアロンタイプ中心の昔と比べ、認識処理部が専用ハードウェアからソフトウェア化が進んだこともあり、価格帯が広がり、目的に応じた選択が可能になりました。ハードウェア(デバイス)タイプでも価格低減がなされています。
8. 文字の他にマーク読取やバーコード、2次元コードの読取機能を併せ持つものもあります。
9. スキャナーを利用して画像をスキャンするものの他に、ペンに組み込んだ電子回路で筆跡を読み取り、文字認識を行う装置もあります。

※このように、利用者のニーズにあったOCRシステムを選べる時代になりました。

## ● 将来のOCR ●



次章に現在のOCR対応業務を紹介しています。このように様々な形態のものが出現し、機能面や性能面で、また価格などの面でも選択肢の広がってきたOCRですが、今後はどのような方向に発展していくのでしょうか。

これまでは紙と切り離せない存在でしたが、FAX-OCRでは受信した電子データのまま処理がなされ、元の用紙は手元にないという形態のものが主流です。また用紙以外の媒体を対象とするものでは、製薬会社などで小さなカプセルの曲面に印字された商品コードやロットナンバーを読み取り、印字品質チェックと実際の物量カウントなどに利用されています。IC製造ライン等でも非接触で同様の捺印検査などが文字認識技術の応用として利用されています。また、デジタルカメラやビデオカメラなどを通して得られた画像の中の文字や記号・マーク等をコード化するのに認識技術が利用されています。

情報化社会が発展し、紙でなくても情報伝達が出来るようになってきました。それに伴い従来型のOCRは次第に形や機能が変わっていくことが予想されます。そのような未来に活躍するOCRとは完全に人間の目にとって替わるような知能型ロボットになっているはずですが、まさに人間の代理人になって情報を求め、どのようなイメージでも瞬時にコンピュータで検索したり人間が理解できるデータに変換したりOCRです。それはOCRという名前ではなくなっているかもしれません。このような時代がくるまでOCRは紙と情報化社会の橋渡しを続けていくのです。

# OCRの 種類と 特徴

現在のOCRは、ユーザそれぞれのニーズに合わせて  
選べるように進歩してきました。  
業務内容に合わせて最適なOCRを選択することにより、  
効率的な入力を実現することが可能になります。



## ① OCR システム構成要素からの分類

### デバイスタイプ

- 文字を読み取るために「専用イメージスキャナー」あるいは「専用の認識ハードウェア」のいずれかまたは両方を持つものを『デバイスタイプ』と称します。
- 一般的には従来型OCRの延長線上にあり、大量一括処理や高速処理の実現に向き、次に説明するソフトタイプに比べ高価格となっています。

### ソフトタイプ

- 文字認識をコンピュータ（パソコン等を含む）上のソフトウェアで行うものを『ソフトタイプ』と称します。
- ソフトタイプはイメージ入力を汎用イメージスキャナー・デジタルカメラ・複合機により行います。製品としては文字認識ソフトウェアだけのものもあり、低価格のものが多いのが特長と言えます。

## ② 利用面からの分類

### 伝票(帳票)処理用OCR

#### (1) 特長

- 定型業務に即して作られた、予め決まった形式の帳票や伝票を読み取ります。
- 読み取ったデータは業務ソフトウェアと連携して、データ処理を行うことが可能です。
- 専用イメージスキャナーを持つタイプが主流で、大量高速な処理が可能です。

#### (2) 技術進歩

- 様々な業務に対応できる様に、多様な入力媒体(紙、イメージデータ等)を扱うことが可能です。
- 市販ソフトウェア(表計算、データベースソフト等)やユーザアプリケーションとの連携が簡単にできるプログラムインターフェースやファイル出力形式(CSV形式など)を備えています。
- 読み取り対象となる帳票の紙質や印刷インク等の制限が大幅に緩和されたため、ワープロソフトでの罫線機能などで作成した帳票や、市販の帳票もそのまま利用可能なものもあります。
- 利用者がその用途に合わせた最適なシステムを組めるよう、機能や価格、性能などで選択肢が広がっています。処理の対象やその量および処理結果に求められる精度など、最適なものを適正な価格で選択できるようになっています。
- 様々な種類の帳票を混在させて読み取ることができます。帳票毎に分類して、排出するソーター機構を備えたものもあります。

### 文書OCR

#### (1) 特長

- 書籍、新聞、名刺などのあらかじめ形式の決まっていない印刷文書を読み取ります。
- 読み取ったデータは、市販ソフトウェアと連携した利用が容易です。例えば、読み取り結果をワープロソフト、表計算ソフト、Webブラウザなどで扱える形式で直接出力できます。

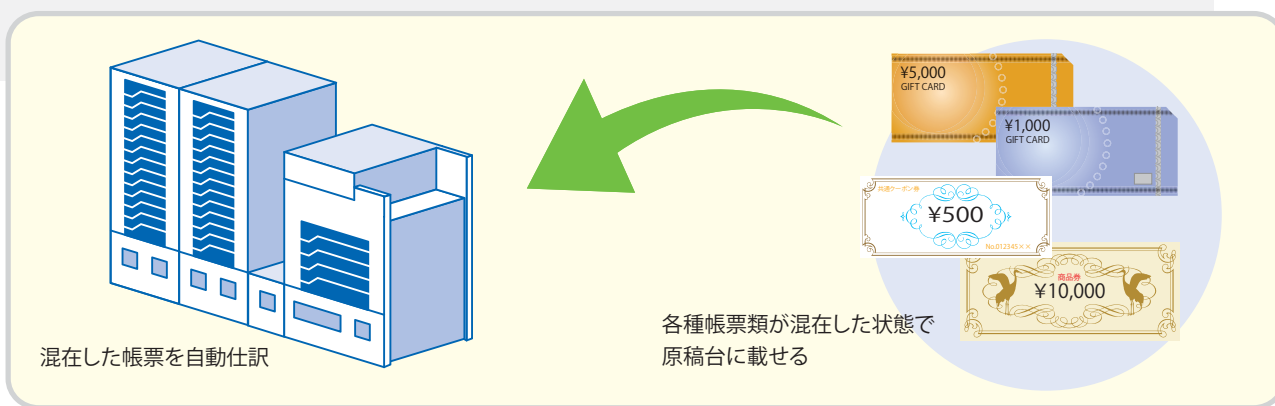
#### (2) 技術進歩

- 文書のレイアウト(段組)を自動的に解析、理解することで、表や文章と図や写真を自動的に分離して読むことができます。そのため、入り組んだ複雑な文書や表の読み取りが可能です。
- 検索用キーワードの自動入力が可能となります。イメージファイリングシステムの入力時、タイトルや要約などを読み取り、検索用キーワードとしてイメージと同時にデータベースへの登録ができます。
- 汎用のイメージスキャナーを使用することが可能なため、低コストになっています。
- 個人ユーザから企業ユーザまでの幅広いユーザに対応できるような、商品のバリエーションが増えています。

1. 一般企業

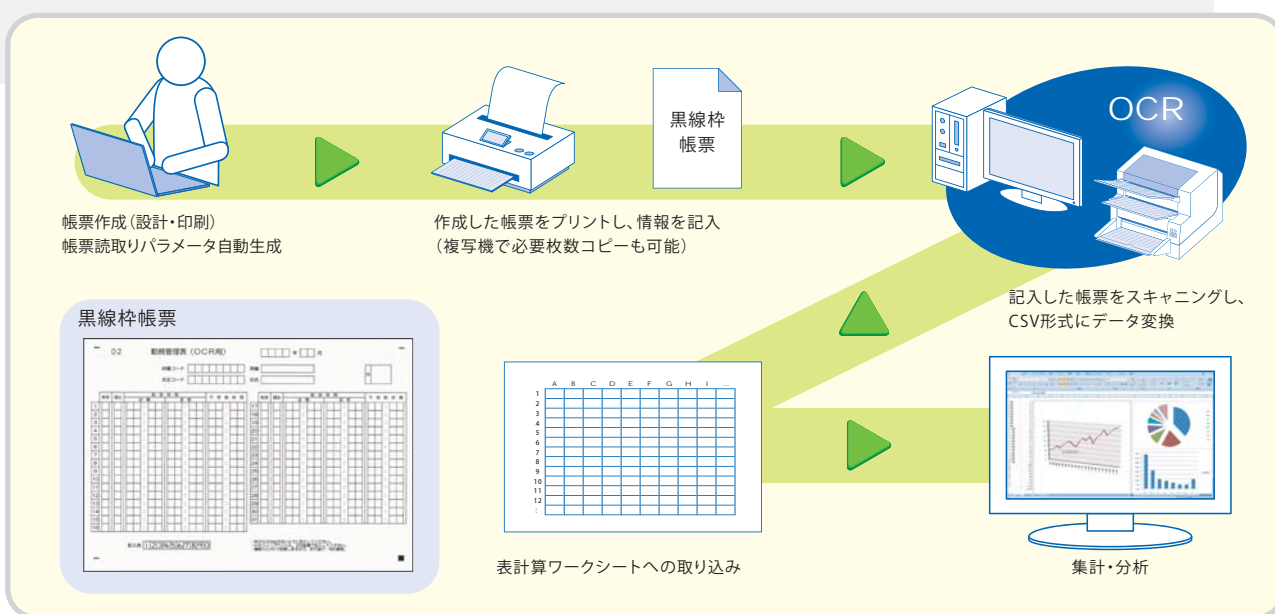
■ 帳票の仕訳(異種混在読取)

商品券・タクシー券など、多種類の帳票を混在したまま原稿台に載せ、種別ごとに集計して別のポケットへ分けて出力することができます。また種別毎に限らず、宛先や、収納機関名など、仕訳方法を自由に設定することができます。



■ 手作り帳票(黒線枠帳票の作成と読取り、集計と分析)

アンケート調査など臨時の非定型業務では、手作りの黒線枠帳票をワープロ感覚で作成・利用することができます。帳票は複写機などでコピーし、帳票読取りのための定義パラメータは専用ソフトで自動生成させるなど、その読取り処理もスムーズに行えます。また、読取り結果は表計算ソフトなどに取り込み、集計・分析も素早く行えます。

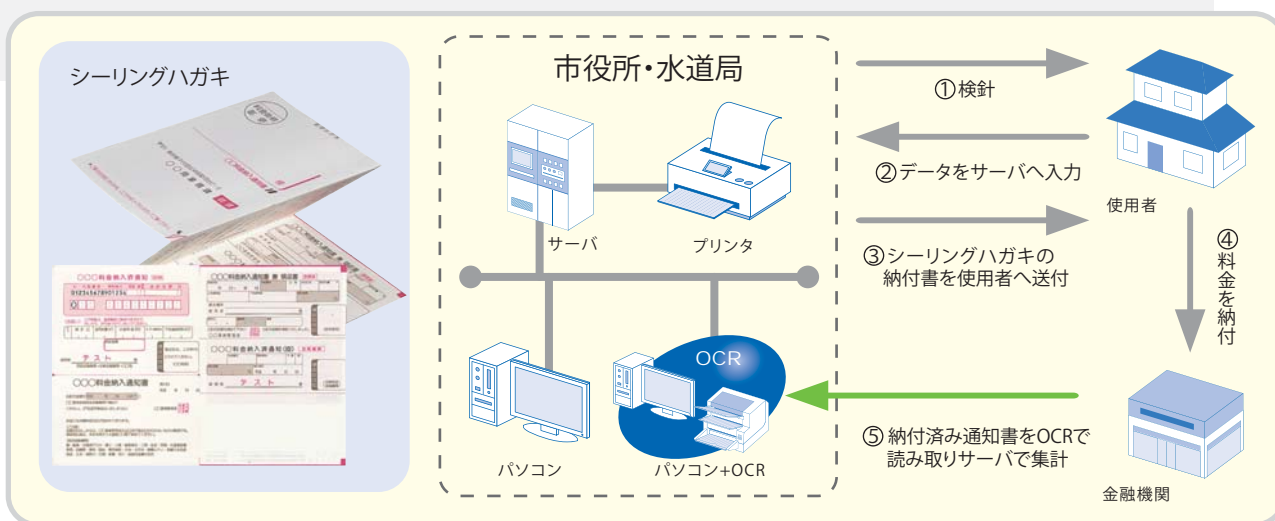




2.自治体

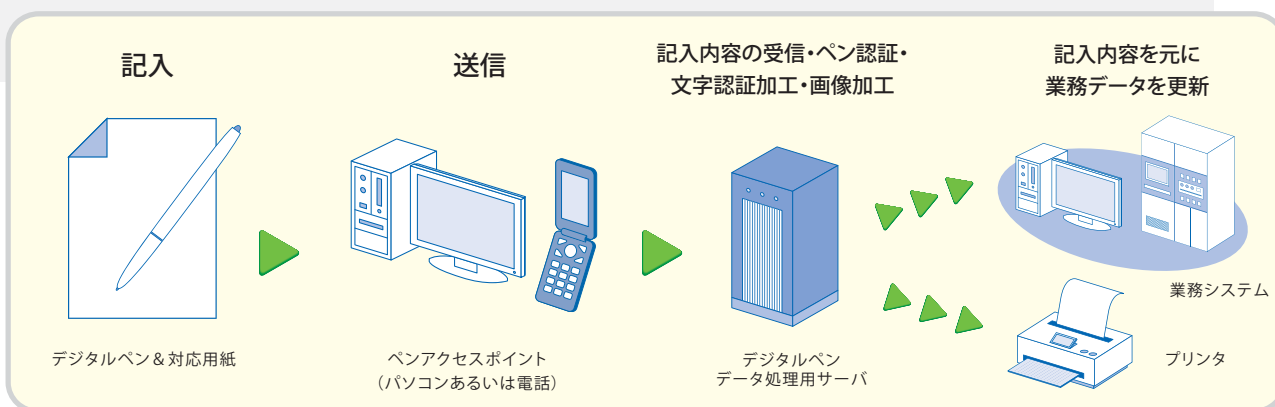
■納付済み通知書を入力(シーリングハガキ読み取り)

自治体から発行されるシーリングハガキの納付書を使用者に郵送。金融機関で料金納付行い、その納付済み通知書を自治体に送付。そこで、OCRにより消し込み処理。ハガキ化による郵送代金の節減、封入作業の省力化を実現しています。



■デジタルペン(手書き文字の即時入力)

デジタルペンとその対應用紙を使用し、手書きで様々な場所からの文字入力を実現できます。また、記入された文字情報はPCやBluetooth対応の携帯電話からシステムへ素早く送信できます。窓口業務での各種登録手続きや製造現場での製品検査業務、配送現場での請求書発行、屋外調査の報告書入力などでスムーズな電子化を行えます。

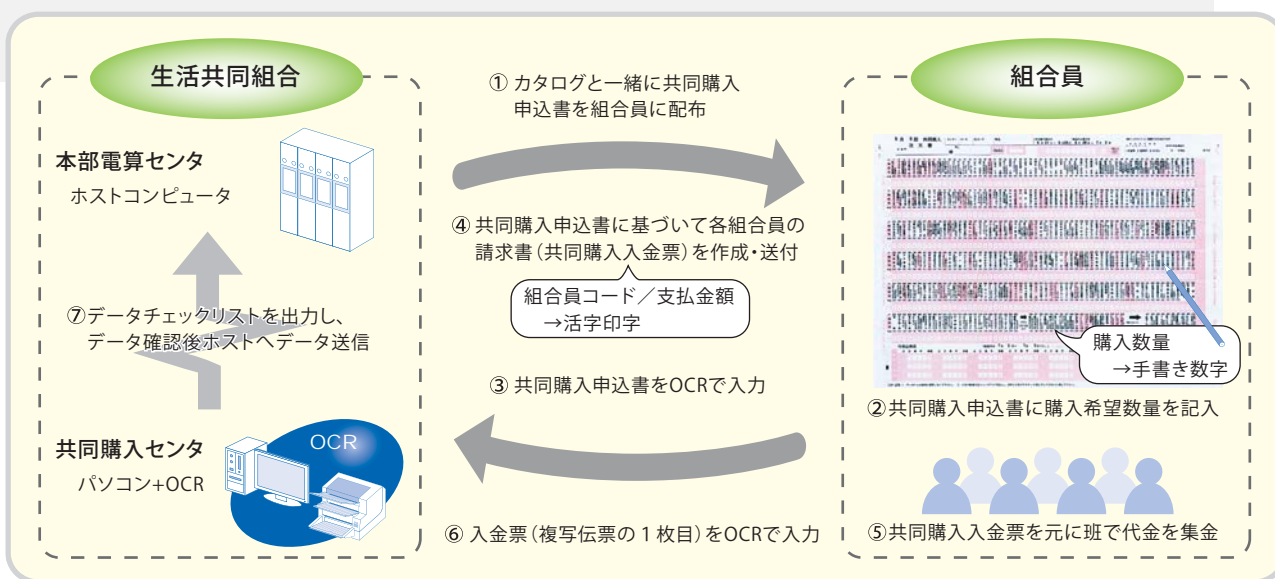


## OCR対応業務の紹介

### 3. 流通業

#### 共同購入申込書を入力(大型帳票(A3サイズ)読み取り)

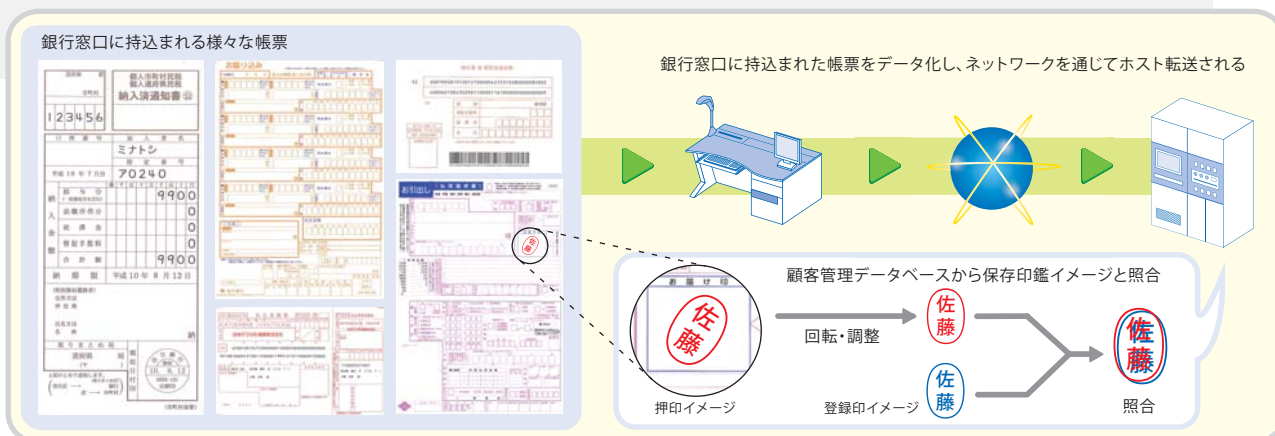
生協などで使用される注文書をOCRにより一括入力。そのデータをもとに商品の発注・配送を行い、また入金金の消し込みもOCRで行います。381mm幅のスプロケット穴付きストックフォーム用紙の入力ニーズに対応し、さらにA3サイズの大型帳票の入力も可能なので、メニューが多く記載された注文書もスプロケット穴を付けたまま効率よく処理できます。



### 4. 金融・保険

#### 銀行窓口業務

銀行の窓口では各種振込(税金・料金)、新規口座開設、預金引出等の業務をお客様と対面しながら行いますが、OCR技術が円滑な処理を実現させています。窓口で持ち込まれた帳票をイメージ入力させて自動的に種類を識別し業務フローを指示、各種確認までの作業を滞ることなく素早く完結できます。



# OCR対応業務の紹介

## ■総合振込依頼書(非定型・一般帳票の読み取り)

多くの企業等から依頼される総合振込の処理は、個々の企業毎にその様式も異なり、帳票定義パラメータによる読み取りを困難にしています。この場合、読み取らせたい項目名などを事前登録し、帳票構造を認識させる機能を利用する事で、項目名とデータ部の対応付けを自動処理しながら、必要なデータ欄のみを読み取らせる事ができます。

**様式の異なる振込依頼書でも必要なデータを読み取ることが可能**

## ■本人確認用OCR

運転免許証や外国人登録証などの記載項目を自動的に読み取り、顧客情報の入力や本人確認手続きを円滑に行います。氏名のふりがなや郵便番号など、免許証に直接記載されていない項目も、自動的に辞書引きして表示させることができます。本籍欄を自動的に塗りつぶして画像を保存することもできます。

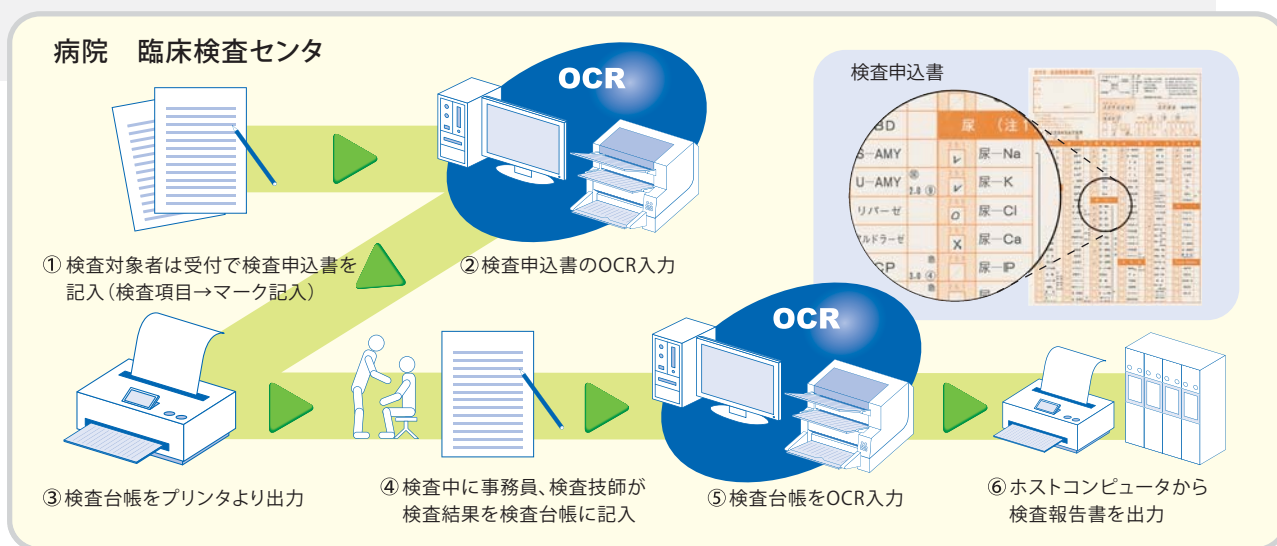
**OCR**

## ((( OCR対応業務の紹介 )))

### 5. 医療機関

#### ■臨床検査依頼書を入力

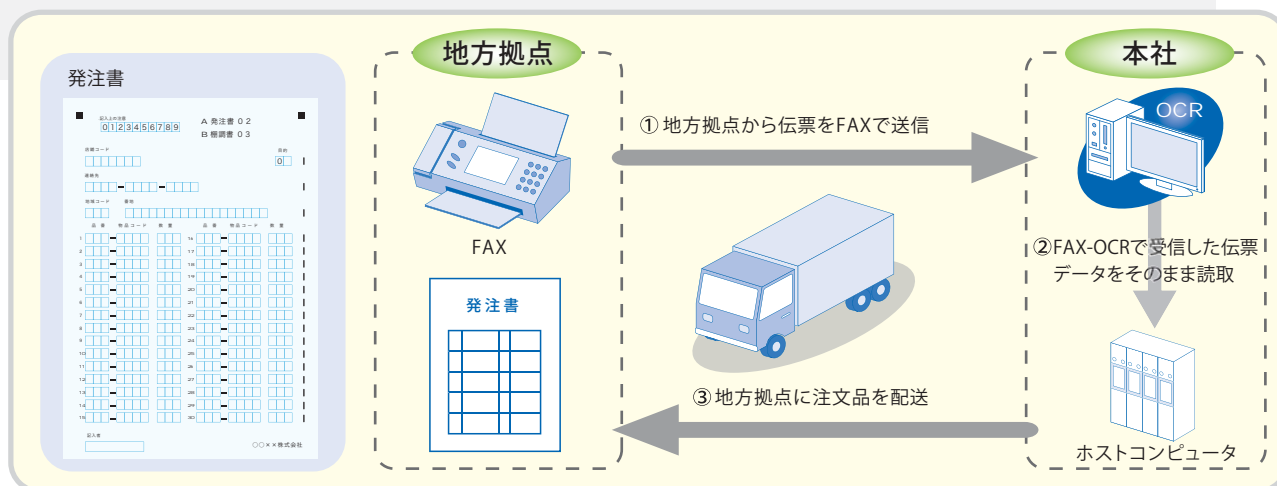
病院の検査センターで使用する検査申込書／検査台帳をOCRにより一括入力。入力する項目は手書き数字と手書きマークの記入が主となるため、高精度な読み取りが可能です。事務員や検査技師の事務処理を軽減し検査終了から結果報告までの時間を大幅に短縮します。



### 6. 製造

#### ■発注書を入力 (FAX-OCR読み取り)

全国販売特約店のFAXから送られてくる発注書を紙を介さずダイレクトにOCR入力。遠隔地のFAXが入力端末となるので夜間でも受付可能であり、発注／納品業務における時間と距離を超えたリアルタイム処理を実現します。



## 7. 学校・教育

### ■試験答案用紙を入力(マークシート読み取り)

学校・教育関係の試験等で使用されるマークシート方式による試験答案用紙をOCRにより一括入力。正確且つスピーディーに処理可能です。また、漢字読取機能により、入学願書の受付業務にも利用可能です。



## 8. 文書OCR

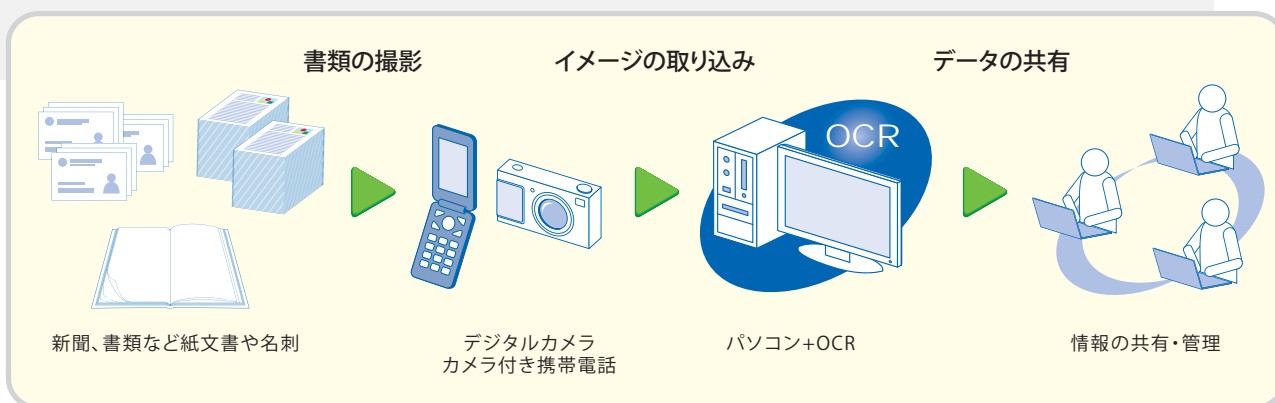
### ■名刺データの入力・管理に活用

小型スキャナーやデジタルカメラで名刺を取り込み、名刺に記載されている内容を文字認識(OCR)技術によりテキストデータに変換し、データベースに登録・管理できる、名刺OCRソフトが普及しています。登録したデータは検索や宛名印刷、メールソフト・地図ソフト・路線検索ソフトなどとも連携して活用できます。



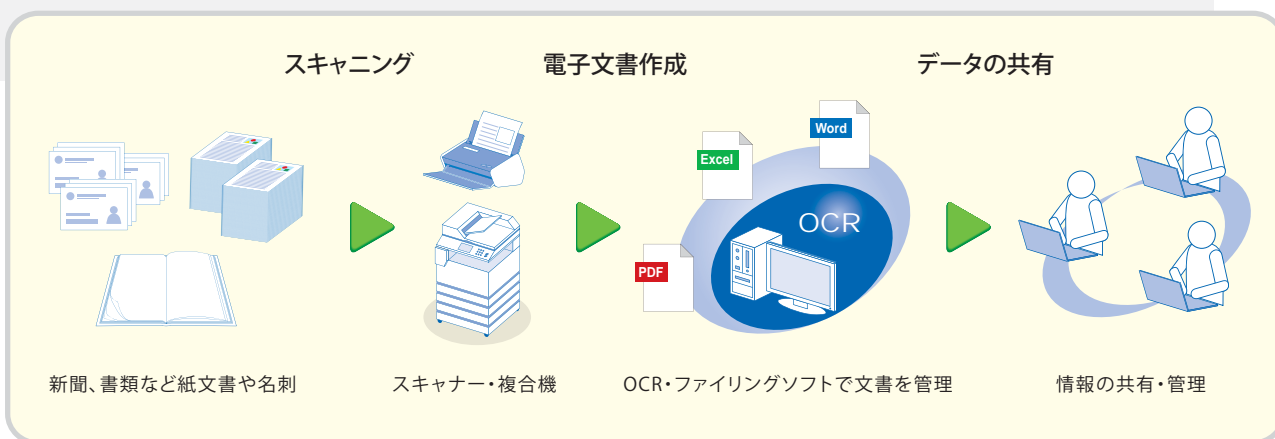
### ■デジタルカメラやカメラ付き携帯電話でOCRを活用

「デジタルカメラ」や「カメラ付き携帯電話」の高解像度化が進み、スキャナーや複合機等の画像入力装置がなくても、簡単に高画質な画像の撮影ができるようになりました。「デジタルカメラ」や「カメラ付き携帯電話」を利用することで、手軽に紙文書の電子化や名刺情報のデータ化を行うことが可能です。



### ■情報共有、ファイリングに活用

様々な紙文書をスキャナーや複合機から入力し、OCRソフトと連携することで、直接、テキスト付きPDF、Word、Excelなどの電子文書を作成します。また、ファイリングしたイメージデータに、文書OCRを利用して、検索するためのキー(タイトルや内容)を自動的に付与しています。企業の事務所をはじめ、役所、病院、学校などで情報共有に活用され、セキュリティ管理も充実します。

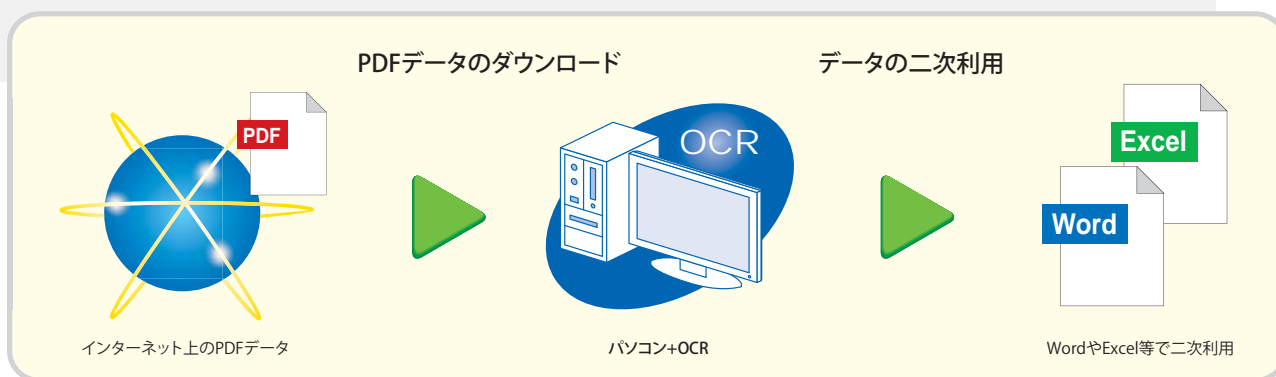


## ((( OCR対応業務の紹介 )))

### ■PDFデータをOCR処理することで二次活用が可能に

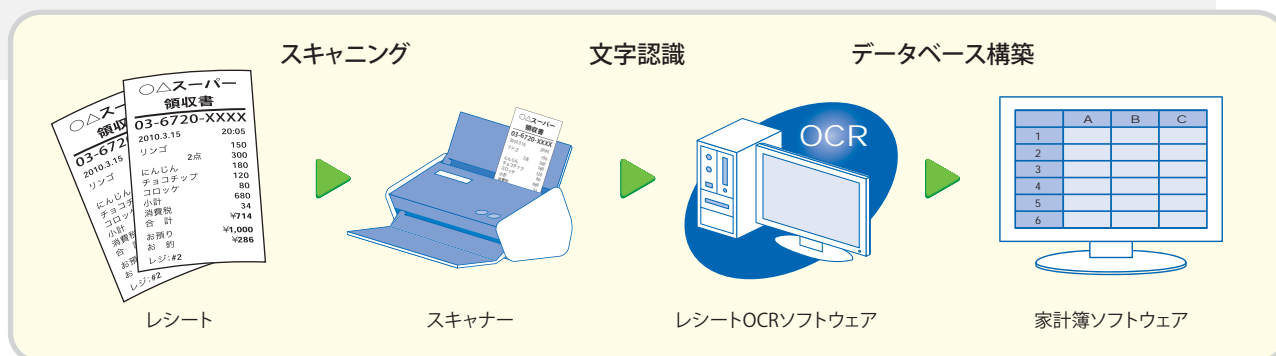
PDFデータは文書データの配布手段として世界中に普及しています。

近年、身近になったPDFデータを編集／加工したいという「二次活用」ニーズが高まり、PDFデータからWordやExcelデータを生成する用途でOCR技術が利用されています。



### ■家計簿データの入力に活用

小型スキャナーでレシートを取り込み、印字内容を文字認識 (OCR) してテキストデータに自動変換し、家計簿ソフトに入力できるレシートOCRがあります。小さい文字で印字したレシート内容をキーボード入力する必要がなく便利です。



ここでは、OCRシステムの基本的な用語の説明をしています。他の用語についても知りたい時は、[www.jeita.or.jp/](http://www.jeita.or.jp/)にある「OCR・イメージスキャナー用語集」にアクセスしてください。

## イメージ

- **イメージ** OCR関連では、読み取り対象の文字やマークなどを含む画像のこと。
- **解像度** 画像の分解可能な細かさの程度の表し方。
- **d.p.i** 解像度の単位。1インチ(約25.4mm)当たり何点(何画素)存在するかを示す。プリンタ等の印字密度やスキャナー等の解像度を表現する。  
(dot per inch:ディ・ピー・アイ)

## OCRの種類

- **帳票OCR** 各種の帳票を読み取るOCRのこと。伝票処理用OCRとも単にOCRともいう。使用する帳票は、従来は記入枠や印字ガイド枠がドロップアウトカラーで印刷されたOCR入力専用の帳票であったが、最近では既存の一般帳票を扱うケースも増えてきている。
- **文書OCR** 日本語などの印刷文書(新聞、雑誌、名刺、印刷原稿等)を読み取るOCRのこと。レイアウト解析技術によりフォーマット定義パラメータを使用せず文書を読み取ることが可能。
- **ソフトOCR** パソコンなどで動作するソフトウェアで認識処理を行うOCRのこと。
- **FAX-OCR** 帳票をFAXで送信し、電話回線等を通して帳票イメージを受信して、文字認識を行うOCRシステムである。

## OCRの機能

- **レイアウト解析** 文書を図、表、文章の領域に分類し、文章領域については、見出しや段組等の解析処理を行うこと。文書構造解析ともいう。
- **マルチフォント読取** 特定のフォント(字形)だけでなく、複数のフォントを混在で読み取れること。ターンアラウンド帳票や統一伝票など複数の顧客で印字された帳票を一括して読み取る等、フォントの統一が困難な場合にマルチフォント読取機能を採用する。
- **オムニフォント読取** 複数の異なるフォントの活字を読み取ること。マルチフォントより多い100種類以上のフォントの活字を混在で読み取る場合に使われることが多い。特に文書OCRでは、様々なフォントの活字で印刷された文書を読み取る必要がある。
- **学習機能** 文字認識や知識処理の辞書を自動的、あるいはユーザの操作で変更/追加する機能のこと。一般的に、辞書の変更/追加を行うことにより性能向上を期待するが、変更/追加するデータを十分に吟味しなければ、既存の辞書に悪影響をおよぼし、性能低下となる場合もあり、注意を要する。
- **後処理** 一文字単位の認識処理の精度には同形文字などの限界があり、認識精度を高めるため、認識処理後に種々のチェックや辞書との単語照合等を実施すること。チェックディジット・チェック、範囲チェック、単語照合、知識処理などがある。



## ● 知識処理

認識結果が、単語または意味のある文字列となるように、知識辞書等を使って認識結果を変更する、または候補を絞り込む処理のこと。後処理ともいう。

## OCRの運用

## ● エラー(誤読)

読み取り結果が、記入時点や印字時点で意図した文字と異なる場合や、論理的なチェックに引っ掛かった場合、「エラー」(Error)と称する。

## ● リジェクト

読取不能のこと(Reject)。リジェクトには、それぞれの文字が読取不能の場合と、論理チェック等で抽出されたものがある。例えば各項目の金額と、その合計金額が記載された帳票の場合、各項目の金額の読み取り結果を合計して合計金額の読み取り結果と一致しなければ帳票リジェクトとすることができる。年月日のデータのような範囲チェックに掛かったものはフィールドリジェクトとなり、確認修正時にクリーンデータに修正することもできる。

## ● フォント

一群の文字に対して統一的にデザインされた字形の一組のこと。字体ともいう。OCRで特別に高精度で読み取るための字体が英数字としてJIS X-9001(JIS OCR-A)、JIS X-9002(JIS OCR-B)、カタカナとしてJIS X-9003(JIS OCR-K)が規格化されている。

## ● 処理速度

単位時間当たりの処理量。一般に1分間の帳票読み取り枚数。(枚/分)で示す。読取速度ともいう。

**帳票OCRでの処理速度**

下記2つの条件での数値がよく使用される。

(1)

- ・帳票サイズ: A4
- ・読み取り文字数: 300字(30字/行、10行)
- ・読み取り文字種: 手書き数字
- ・データチェック: なし
- ・測定期間: 一括読取モードで帳票を読み込み始めてから認識結果を上位のコンピュータに出力し終わるまでの時間。またはファイルに書き込むまでの時間。

(2)

- ・帳票サイズ: 最小サイズ帳票(A8程度)
- ・読み取り文字数: 10字(10字/行、1行)
- ・読み取り文字種: 活字数字
- ・データチェック: なし
- ・測定期間: 一括読取モードで帳票を読み込み始めてから認識結果を上位のコンピュータに出力し終わるまでの時間。またはファイルに書き込むまでの時間。

## ● 確認修正

OCRの認識結果を確認し、必要があれば訂正する作業。OCRシステムでは、ハード、ソフトの処理を問わず、最も重要な処理要素であり避けては通れないものである。

## ● 即時修正

OCR処理において、リジェクトや論理チェックエラーが発生した時点で、読み取り結果の修正操作を行う処理方法のこと。読み取り動作を中断してオペレータに修正を求めため、読み取り完了までの処理時間は長くなるが、修正入力ミスが低減できる。

## ● 一括修正

一括読取(認識結果を無修正で一括してファイル等に出力する)によりファイル等に蓄積した未修正の読み取りデータを、認識処理とは別フェーズで読み取り結果や対象文字イメージ等をディスプレイに表示し、内容の確認や修正、データ追加等を行う処理方法のこと。

- **ターンアラウンド修正** コンピュータで必要事項を印刷した帳票を顧客に送り、記入されたり、払込み等で戻ってきたりした帳票をOCRで読み取り、消し込み業務等をする処理。
- **マルチ修正** OCR認識結果の確認修正作業を複数のクライアントで行うシステムのこと。LANで接続された構内のクライアントで修正する形態が多いが、最近はインターネットを介して遠隔地からマルチ修正を行うケースもある。
- **トランザクション処理** 帳票を1枚読み取る毎に上位装置に伝送し、処理する方式。上位装置でデータチェックを行ったり、帳票ID認識により異種フォーマットの帳票も一組の帳票として入力したりすることができる。

## OCR帳票

- **帳票設計** OCRの規約/制限事項を盛り込んで、帳票原稿を作成すること。OCRで読み取る帳票には以下のような制約事項がある場合がある。
  - ・帳票のサイズ、色、紙質、厚さ
  - ・文字の種類/文字サイズ/文字枠のサイズ、色、形状
  - ・ドロップアウトカラー/その他の色の使用制限
  - ・ラインマーク(黒・非ドロップアウトカラー)
  - ・文字間隔/行間隔/行配列、フィールド間隔/フィールド数とクリアエリア
  - ・その他(最大文字数など)
- **帳票仕様** 正常な読み取り動作を保証するために必要とされる、帳票に対する要求仕様。OCRの導入時に、最初に手掛けなければならないのが、OCR帳票作成であり、事前に入手した帳票設計仕様書に記載されている詳細規程内容に合わせた帳票設計をすることが、導入成功のポイントとなる。OCRの帳票設計仕様から外れた帳票は、搬送性能の悪化や認識性能の悪化要因となり、注意を要する。
- **ドロップアウトカラー** 帳票上に予め印刷もしくは印字されている画像の色のうち、人間の目には見えるが、イメージスキャナー等で読み取った時に、画像として現れない色のこと。ドロップアウトカラーとして見える色は機種(イメージセンサ、光源、フィルタの組み合わせ)により異なるが、一般ボールペンの黒もドロップアウトしてしまい、使えないものもある。
- **PCS**  
(Print Contrast Signal:ピーシー・エス) 用紙と印刷色、記入文字等との相対的な濃度の比(コントラスト)を表す数値のこと。用紙そのもの(印刷や記入の無い紙面から)の反射の明るさを0として、用紙からの反射の全く無い「真っ黒」を1として用紙上の印刷部分や記入部分の「濃度」を相対的に表す。
- **クリアエリア** OCR読み取り対象の範囲に設ける余白。文字切出し等で必要。黒色禁止領域またはマージンエリアともいう。
- **黒色禁止領域** OCR用紙で文字切出しや行マーク等の検出を行うのに必要な文字枠や印字枠の周辺に設けるべき余白のこと。この領域内には、ドロップアウトしない色での印刷や書込みは厳禁で、違反すると文字の切出しや黒枠の消し込みに支障をきたし、桁ズレや誤読の原因となる。
- **帳票サイズ** OCRが読み取り可能な帳票の大きさの仕様のこと。用紙/原稿サイズともいう。カタログに記載される仕様としては、「縦/横の最小および最大寸法」、または「A4、B4等の用紙寸法の一般呼称」が用いられる。
- **紙質** 当該OCRが使用できる帳票用紙の光学的および機械的特性。カタログに記載される用紙仕様としては、「OCR用紙」、「上質紙」、「普通紙」、「再生紙」、「PPC用紙」等の様に、用紙の呼称で表現される。「OCR用紙」は、JIS規格で光学的/機械的特性が規定されているが、「上質紙」、「普通紙」、「再生紙」、「PPC用紙」については、光学的/機械的特性や品質が規格化されていないため、使用用紙の銘柄指定や事前評価を推奨するOCRメーカーが多い。

- **OCR用紙** 文字、記号を光学的に読み取るデータ処理に用いる用紙で、認識率の確保のために製紙過程でのゴミの混入がなく、平滑性、帯電防止性、耐摩耗性、耐伸縮性等に優れ、腰の強い(こわさ/粘り気のある)、裁断時にも誤差が少ない用紙。最近のOCRでは、上質紙や普通紙も許容されており、ターンアラウンド処理などの特定の用途にのみOCR用紙が採用されている。
- **シーリングはがき** 税収納伝票などで利用されており、宛先欄・通知欄・領収欄などの3面を圧着しているはがきのこと。シーリングはがきのメリットは、印刷時に3面を同時に打ち出せることと圧着によってセキュリティ(プライバシー)を保護することも容易に実現できる点(一旦剥がされると二度とくっつかない)。
- **上質紙** 化学パルプだけで製造した紙。印刷、筆記などの用途に用いる。
- **普通紙** 一度使用した紙を原料として、再利用して製造した紙。再生過程の印刷インクの残滓等による、OCRでの「ゴミ(ノイズ)」発生等に注意が必要。
- **帳票厚さ** 用紙の厚さ。実際には用紙1枚の物理的な厚さではなく、一定面積あたりの重量で示す。
- **坪量** 帳票厚さの単位。1㎡あたりの重量(g/㎡)で表示される。一般にOCRでは50~150g/㎡の紙を許容するものが多い。
- **連量** 従来使われていた帳票厚さの単位。四六版サイズ(788mm×1091mm)での1000枚分の重量(kg単位)を連量または四六版連量という。
- **フォーマット定義パラメータ** OCRにおいて読み取るべき文字、帳票等の情報を記述したもの。フォーマット定義、フォーマット定義体、フォーマットテーブル、フォーマットプログラム、OCRフォーマット、FCなどともいう。特に帳票OCRでは、読み取る帳票、認識対象文字、認識方法等の情報を事前に登録しておかねばならず、登録された定義に基づき帳票の読み取りを行う。具体的には以下のような内容を定義する。
  - ・帳票の外形寸法
  - ・帳票ID位置/字種/字体/桁数
  - ・読取フィールドの位置、フィールド内の字種/字体/桁数
  - ・その他OCR本体で処理される論理チェックの内容等

