

業務効率化ソリューション「Prexifort-OCR」のご紹介

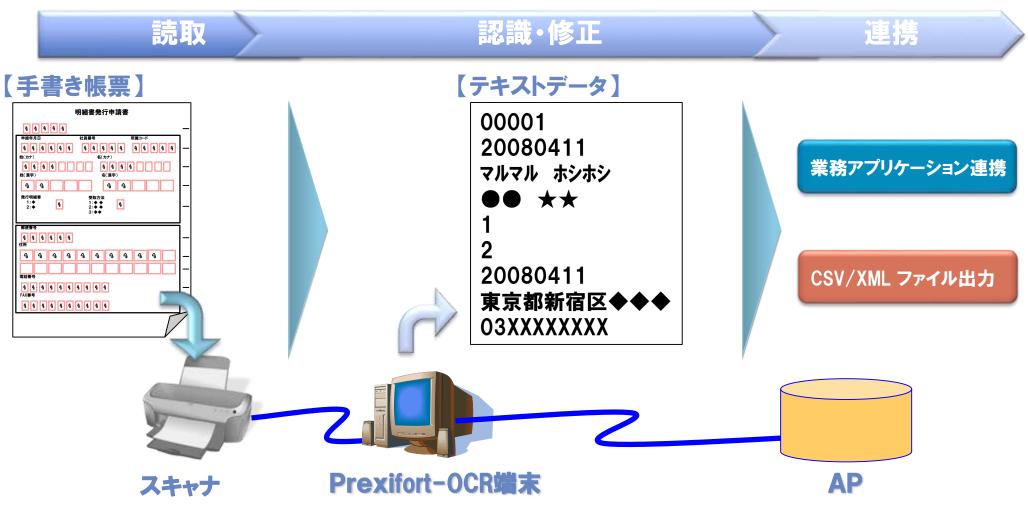
株式会社NTTデータ 第二公共事業本部 第四公共事業部 第二統括部 RPAソリューション担当

NTT Data



Copyright © 2017NTT DATA Corporation

Prexifort-OCR(プレキシフォート)は、スキャナ等で画像化した帳票の手書き文字や印字文字から、帳票定義された箇所の文字情報を抽出・認識し、テキストデータに変換するOCR認識ソフトウェアです。





Prexifort-OCRは、実力・実績とも、日本国内のソフト型OCR市場NO.1の製品です。

ソフト型OCR市場シェア NO.1

中央府省等での40年以上の実績

文字認識力·速度 NO.1

旧郵政研究所主催「文字認識技術コンテスト」最優秀賞

安さ NO.1

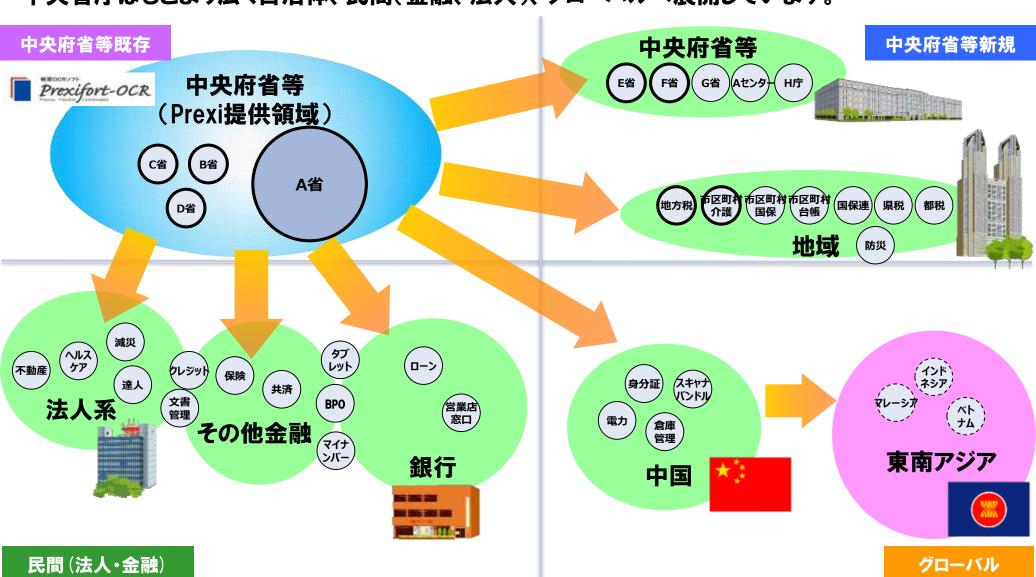
他社よりも安くしますので、ご相談ください

Prexifort-OCR 注力領域

Copyright © 2017 NTT DATA Corporation



Prexifort-OCRは、中央省庁基幹システムを通じた長年にわたるOCRの実績を元に戦略商品化し、中央省庁はもとより広く自治体、民間(金融、法人)、グローバルへ展開しています。



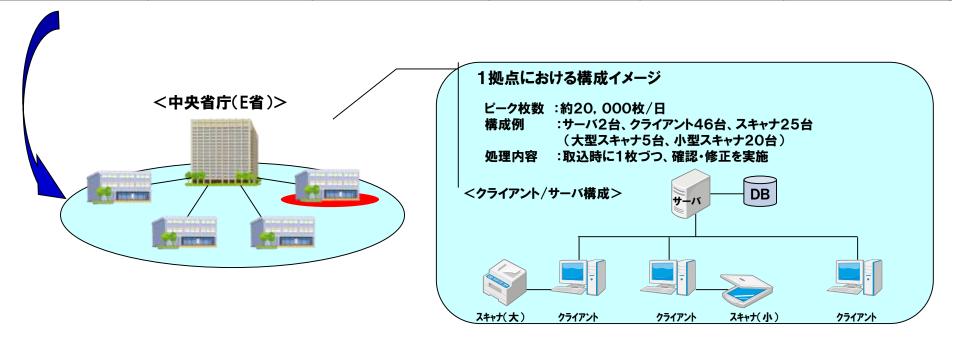
4

Prexifort - OCRの実績



Prexifort-OCRは現在、官公庁を中心に約30,000ライセンス以上が稼働中です。 また昭和46年から40年にわたり商用で提供してきた点でも、安心してご利用いただけます。

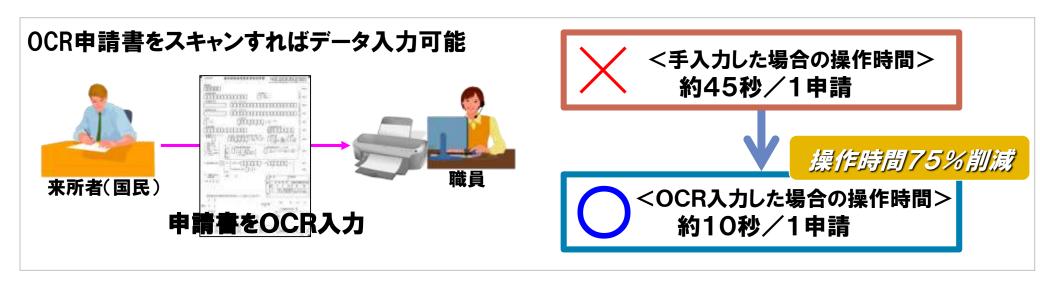
導入先	稼働時期	サーバ	クライアント	スキャナ	構成
中央省庁(A省)	平成24年	_	約420ライセンス	約420台	スタンドアロン
中央省庁(B省)	平成25年	_	約930ライセンス	約930台	スタンドアロン
中央省庁(C省)	平成25年	_	約480ライセンス	約480台	スタンドアロン
中央省庁(D省)	平成25年	_	約420ライセンス	約420台	スタンドアロン
中央省庁(E省)	平成26年	約7,800ライセンス	約23,300ライセンス	約23,200台	クライアントサーバ



2.導入効果



ピーク時に1事務所で1日約2万件発生する申請受付・入力業務の処理には、 Prexifort-OCRによる効率化が欠かせません。



1日に2万件の申請書を受け付ける事務所では、Prexifort-OCRは25人分の作業に相当



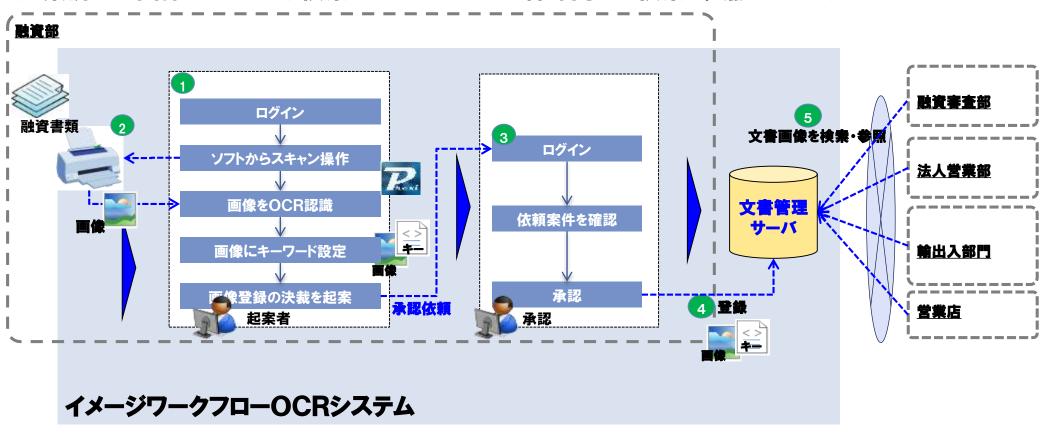


OCRが、画像ファイルの検索キーワードを自動設定します。

OCRは、画像ファイルに変換処理を加え、ファイルサイズを最小化します。

ワークフローに画像ファイルが取り込まれ、簡易に起案・承認できます。

起案済みの画像ファイルは、検索キーワードにより、各部門から検索・参照できます。



導入効果 タブレットによる本人確認書類の読取



スマートデバイス(タブレット)で、カメラで撮影した本人確認書類画像のOCR認識が可能です。 お客様窓口における本人確認処理の作業効率化に貢献します。

顧客

- ①タブレットで本人 確認書類をカメラ撮影
- ②画像をOCR処理⇒確認・修正



窓口職員

③ ②にて顧客が入力した情報が 業務画面に表示される。⇒登録



顧客/窓口職員の 作業時間の削減

顧客の待ち時間の 有効活用

活用シーン

- ・銀行口座の開設
- ・カード発行
- •診療所、病院受付
- ・レンタル・リース店舗受付 等

0

3.商品の概要

Copyright © 2017NTT DATA Corporation



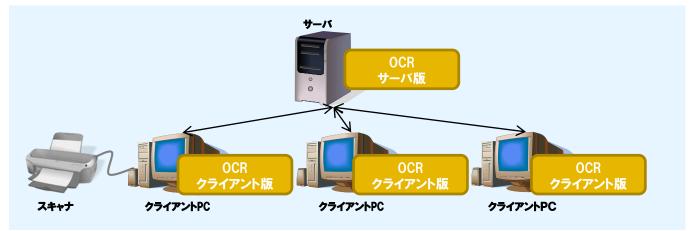
Prexifort-OCRはお客様システムに合わせてご利用いただけます。

<u>スタンドアロン構成</u>

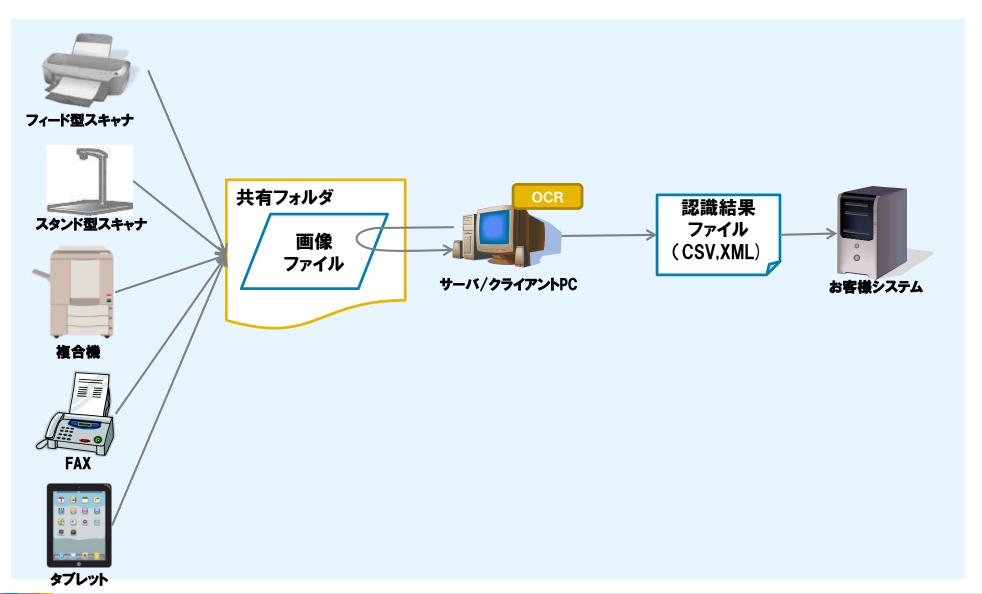
- ・クライアントPCでOCR認識を行います。
- ・認識結果ファイル(CSV,XML)を取り込むことでお客様システムとの連携も可能です。
- ・画像ファイルの読み取りも可能です。



- ■クライアント・サーバ構成 ・サーバでOCR認識を行います。
 - ・複数クライアントPCから認識結果の確認・修正が可能です。



各種スキャナ/複合機/FAX/タブレット等で取得した画像ファイルのOCR認識も可能です。



Prexifort - OCR ~ 認識力の特長 ~



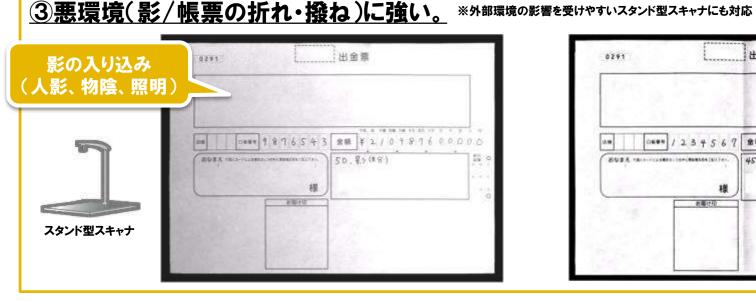
Prexifort-OCRは文字認識力No.1のOCRソフトです。

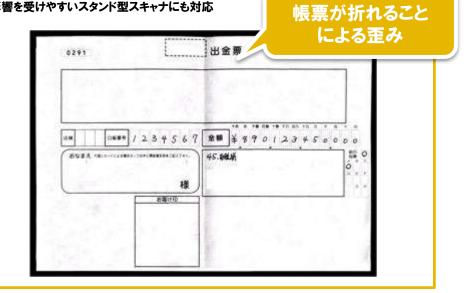
■認識力を支える3つの強み

手書きANKS:99%以上 手書き漢字:95%以上 ※OCR帳票を利用した場合









13

Prexifort - OCR ~ ハイブリッドOCR ~



ハイブリッドOCR採用により、運用上最も重要な「誤読すり抜け率」を0%近くまで低減いたします。



■ハイブリッドOCR

文字認識方式の異なる2つのOCRエンジンの認識結果を突合



エンジンAのみでは「誤読」だった結果を、エンジンBの結果と突合することで「不読」とすることができます。
⇒不読の場合は「?」と表示し、オペレータによる再確認を

⇒ 不読の場合は「?」と表示し、オペレータによる再確認を促すため、「誤読すりぬけ率」を0%近くまで低減します。

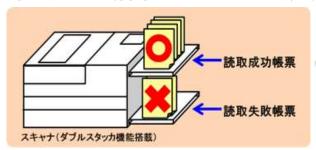
記入文字	エンジンA	エンジンB	結果
	7	7	7
7	?	7	?
画像	7	1	?
	1	?	?

※認識結果

正読:記入された文字を正しく認識した場合 誤読:記入された文字を誤って認識した場合 不読:記入された文字を認識できなかった場合 スキャナの標準規格TWAINに対応しており、どのようなスキャナでも接続可能です。 特にPrexifort-OCR対応スキャナをご利用いただくことで、スキャナの特殊機能まで ご利用いただけます。

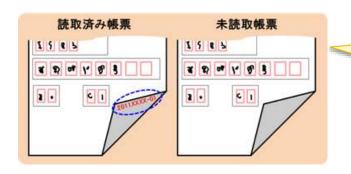
対応スキャナの特殊機能

●ダブルスタッカ機能 読取結果に応じて排出するスタッカを振り分けることが可能



読取りの成功/失敗を自動で仕分けすることで、 失敗帳票の抽出作業を削減。

●帳票裏面への通番印字が可能



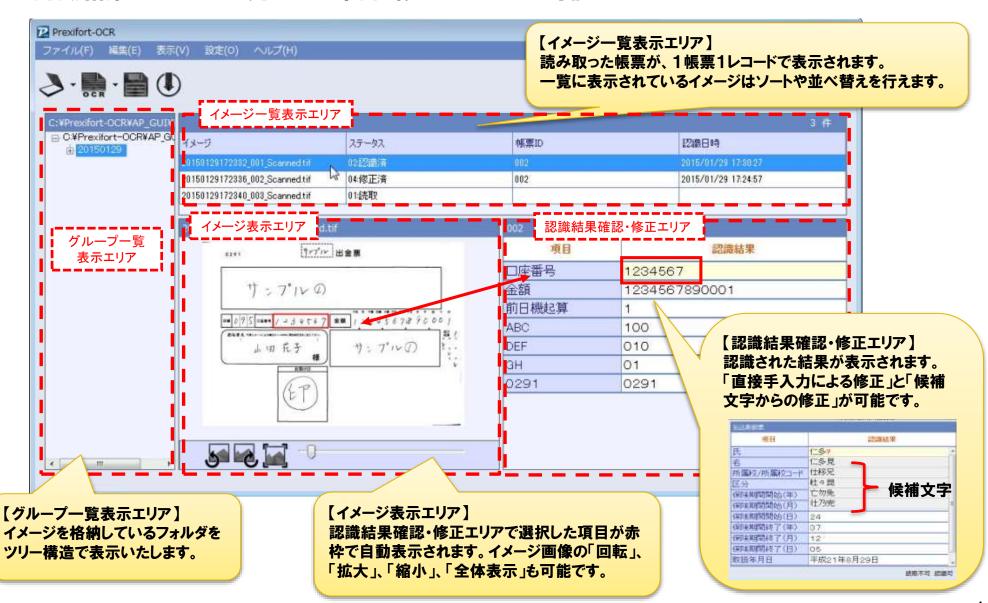
処理枚数/処理日等の通番を印字することで、 帳票管理作業の効率性を向上。

15

Prexifort - OCR ~「認識・修正」の特長 ~



OCR認識結果をイメージを見ながら確認・修正することが可能です。





XMLやCSV等、様々な認識結果出力形式に対応しており、業務アプリケーションとの連携が容易です。

OCR認識結果連携

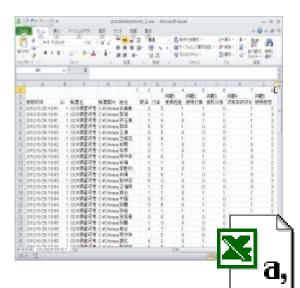
XMLファイル出力

認識結果をXML形式等で出力することが可能です



CSVファイル出力

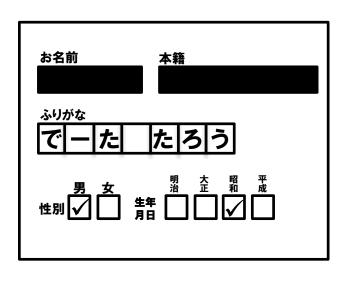
認識結果をCSVファイルで出力する ことが可能です



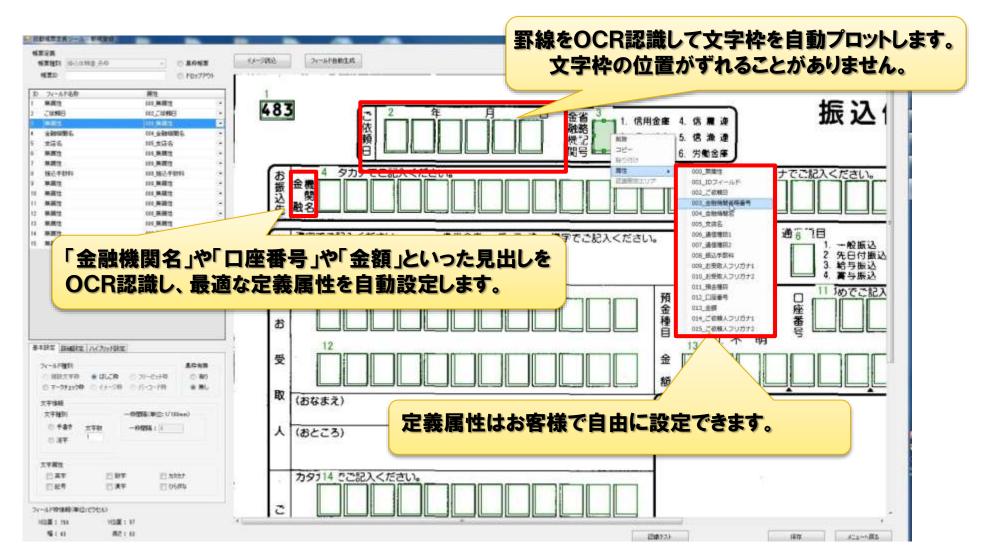
帳票画像連携

マスキング出力

事前に設定された特定のフィールドを黒色で塗りつぶして出力することが可能です



帳票定義を自動化することで、お客様自身で簡単に素早く帳票定義できるようになります。 私製帳票等の、新たな帳票を定義する際の作業負荷を大きく軽減いたします。





豊富なOCRシステム構築支援ツールを備えており、アプリケーション開発も容易です。

SDKを用意

SDKによりきめ細やかな連携が可能

OCR認識結果確認画面を簡単に作成できる画面コンポーネント や、スキャナと連動してイメージを取得する読取API、イメージデー タから認識結果を取得する認識APIなど、各種APIを提供します。 お客様ニーズに合わせたシステム開発が容易できます。



各種辞書メンテナンス

辞書の活用により、容易に認識率を向上

知識処理機能で照合する際に必要な各種辞書ファイルをメンテ ナンスできる支援ツールを提供します。姓名・郵便番号・ユーザ 辞書などの各種辞書ファイルに対して、追加、変更、削除などを 行うことができます。



19

動作環境



■クライアント環境

動作環境OS	・Windows 7 Professional (32bit版/64bit版) ・Windows 8 Professional (32bit版/64bit版) ・Windows 8.1 Professional (32bit版/64bit版) ・NET Framework4.5 以上	%1 %1 %1
CPU	・動作環境OSが正常に動作する環境	
メモリ容量	・1GB 以上(32bit版OSの場合)・2GB 以上(64bit版OSの場合)	
ハードディスク容量	・空きディスク容量600MB以上	

^{※1 64}bit版OSではWOW64で動作します。

■認識対象文字

	・数字、ひらがな、カタカナ、英字(大文字/小文字)、 記号
手書き	・漢字(JIS第一水準2965字、JIS第二水準字454字、 記号(仝々)2 種)
	·簡体字(第一水準3755字、第二水準3008字)
	・数字、ひらがな、カタカナ、英字(大文字/小文字)、 記号
活字(印字)	•漢字(JIS第一水準2965字、JIS第二水準3390字、 記号(仝々)2 種)
	·簡体字(第一水準3755字、第二水準3008字) ·繋体字(13053字)
チェックマーク	・レ点、 塗りつぶし、 オーバーライトマークチェック
バーコード	・JAN8、JAN13、ITF、CODE39、CODE128、Codabar (NW-7)、カスタマバーコード、QRコード
MICR	・MICR E-13Bフォント及び、CMX-7フォントに対応

■サーバ環境

動作環境OS	・Windows Server 2008 R2 (64bit版) ・Windows Server 2012 R2 (64bit版) ・NET Framework 4.5 以上
CPU	・動作環境OSが正常に動作する環境
メモリ容量	·2GB以上
ハードディスク容量	・空きディスク容量600MB以上

■入力/出力

入力画像タイプ	白黒二値 ・グレースケール 256階調 ・8bit/24bitフルカラー
入力画像形式	BMP(DIB) ・JPEG(オプション)・TIF/TIFF(マルチページ形式はオプション)・PDF(オプション)
入力画像解像度	・200~600dpi (「手書き: 200dpi」、「活字300dpi以上」を推奨)

出力形式	•CSV形式,XML形式
イメージ出力形式	•BMP (DIB) ,TIF/TIFF

Copyright © 2017 NTT DATA Corporation

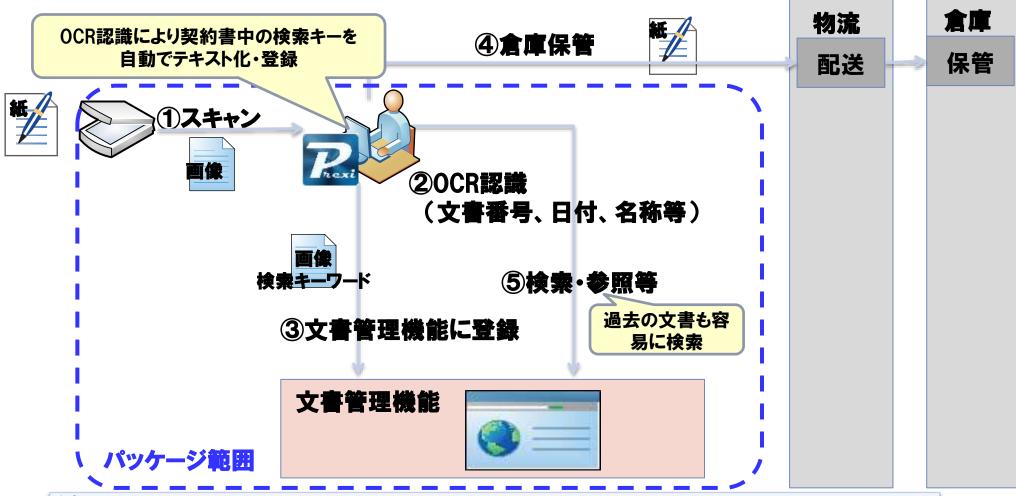
4.OCR文書管理オプション

Copyright © 2017NTT DATA Corporation

OCR文書管理オプションのイメージ



書類の画像データをOCRすることで属性や検索キーワードが自動登録できます。 文書管理における書類の検索・参照が容易になります。



活用シーン

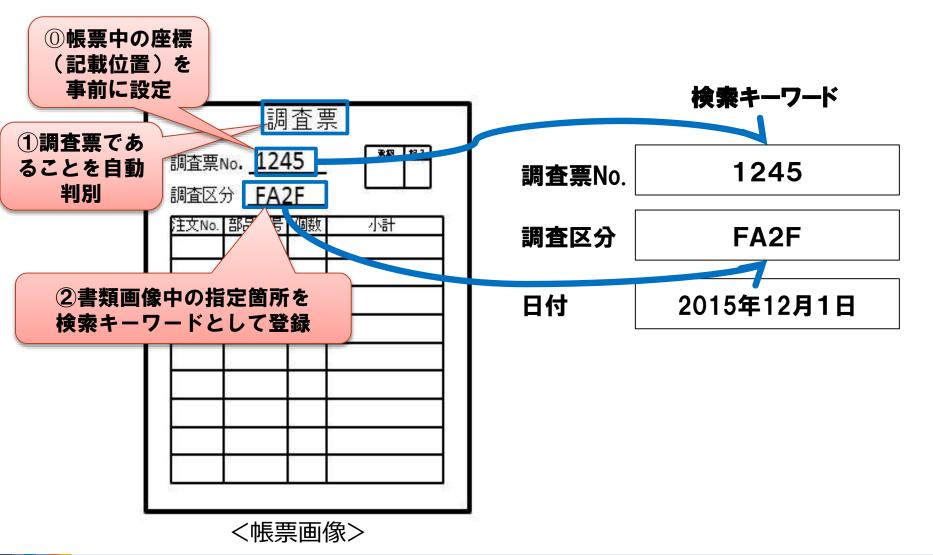
- ・領収書や請求書、契約書を紙のまま管理しており、保管や捜索に苦労されているお客様
- ・画像登録時の検索キーワードの設定作業を効率化したいお客様

22

OCRによる検索キーワードの自動検出・登録方法①帳票手義



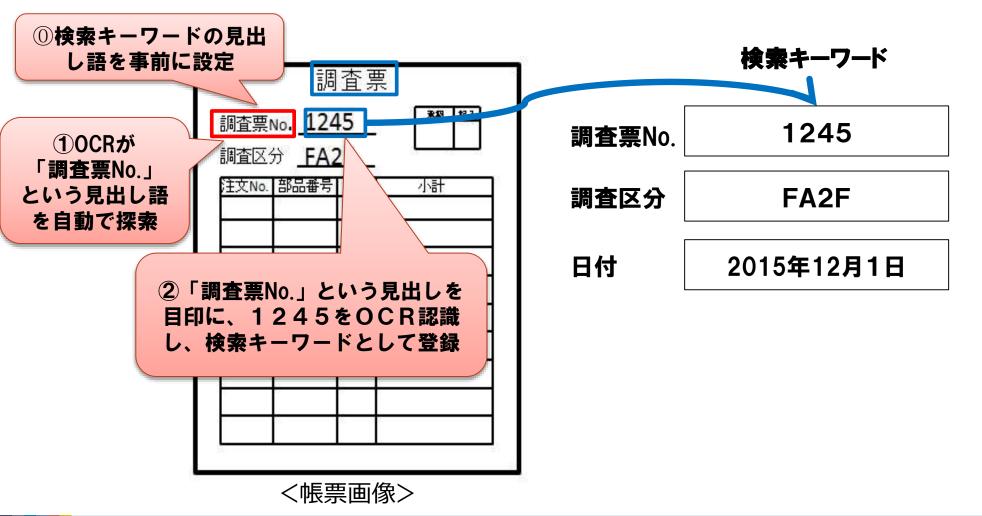
書類が定型で、OCR対象箇所の座標(記載位置)を予め特定できる場合、帳票定義を行う方法により、検索キーワードをOCR認識・登録します。認識成功率が最も高くなります。



OCRによる検索キーワードの自動検出・登録方法②見出し語

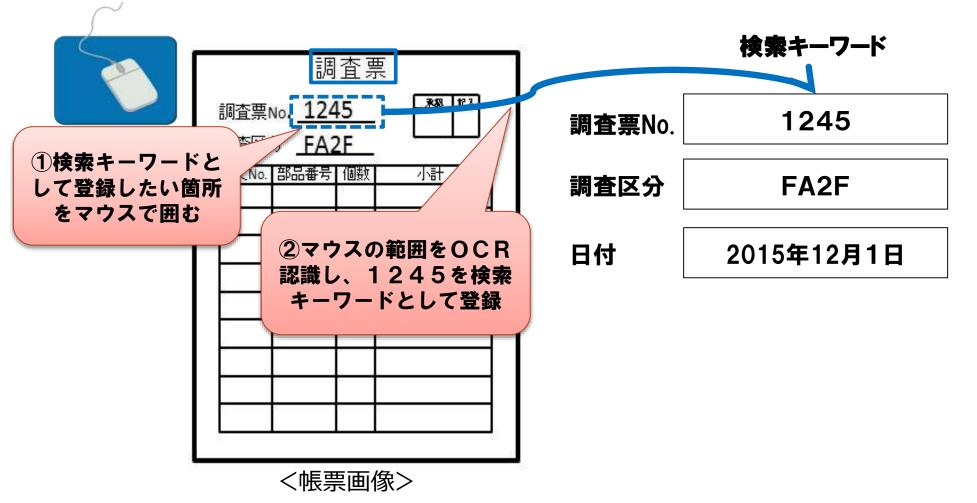


書類が非定型でOCR対象箇所の座標(記載位置)を特定できない場合も、 予め検索キーワードの見出し語を登録しておき、見出し語を自動探索する方法により、 検索キーワードをOCR認識・登録することができます。



OCRによる検索キーワードの自動検出・登録方法③マウスで囲う NTT Data

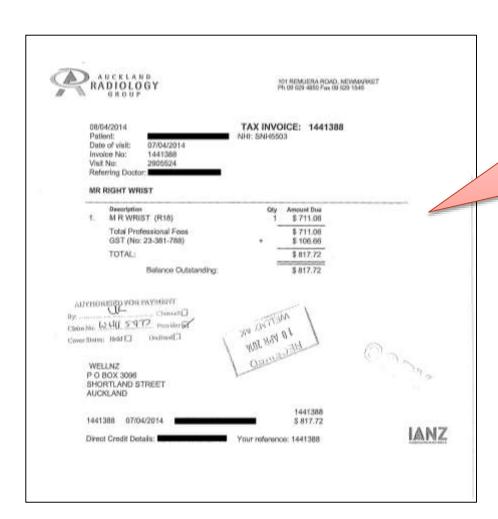
検索キーワードの見出し語さえ特定できない場合や、見出し語の探索に失敗した場合も、 帳票画像上でマウスで囲んだ箇所をOCR認識する方法により、 簡単にOCR認識して検索キーワードとして登録できます。



OCR認識対象の帳票類



検索キーワードのOCR認識に適応する書類の例をご紹介します。

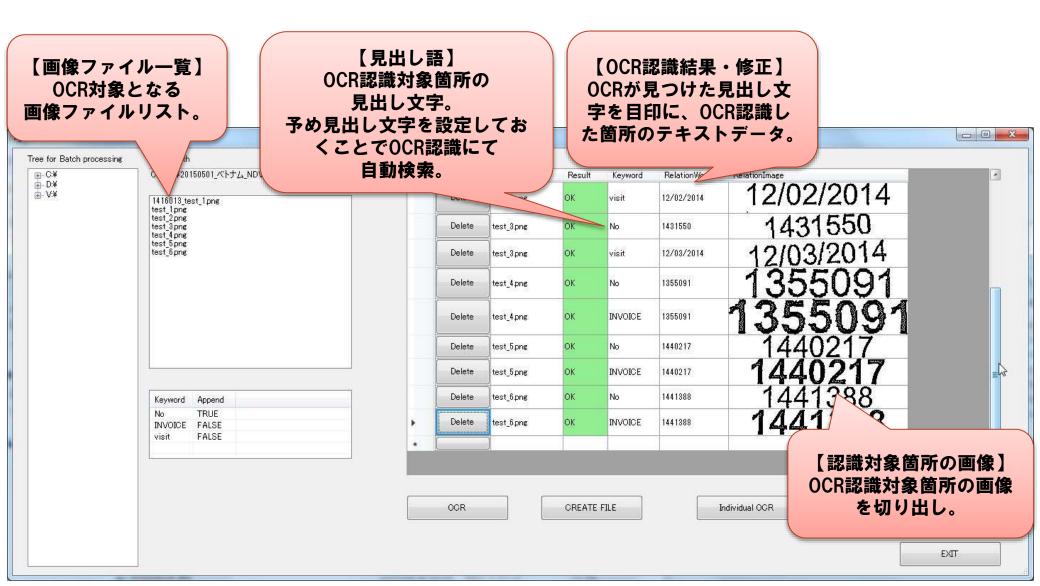


認識率がほぼ100%となり、 目視確認を省略できる、活字英数字の 「契約書番号・契約日付や、請求書番号・ 日付等」が、OCRによる検索キーワード の登録対象として最適と想定

キーワード認識OCR画面イメージ ※デモ用サンプル



OCR認識結果と認識箇所の画像を一覧に表示することで、目視確認及び修正を容易にします。



電子帳簿保存法対応ソリューション Prexifort-OCR「e文書Plus」

各拠点

または、本部





入力要件

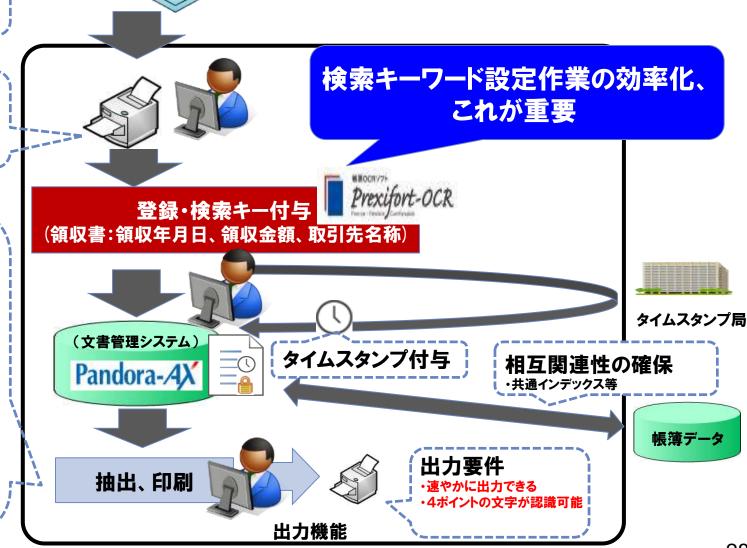
・4つの方式より選択

スキャナ要件

- ·解像度:200dpi相当以上
- ・24ビットカラー以上

記録管理要件

- ・帳簿書類間の関連性の確保
- ·検索機能
- ・取引年月日その他の日付、取引金額、 その他記録項目の検索
- ・日付、金額の範囲指定検索
- ・二つ以上の任意の項目を組合せ検索
- ・タイムスタンプ一括検証機能
- ・解像度、階調、大きさに関する情報を保存
- ・アカウント情報保持
- ・スキャンユーザ情報取得/保管
- ・ヴァージョン管理
- ・アカウント情報保持
- ・ユーザIDの操作ログを生成/保管
- ・削除履歴を生成/保管



Copyright © 2017 NTT DATA Corporation

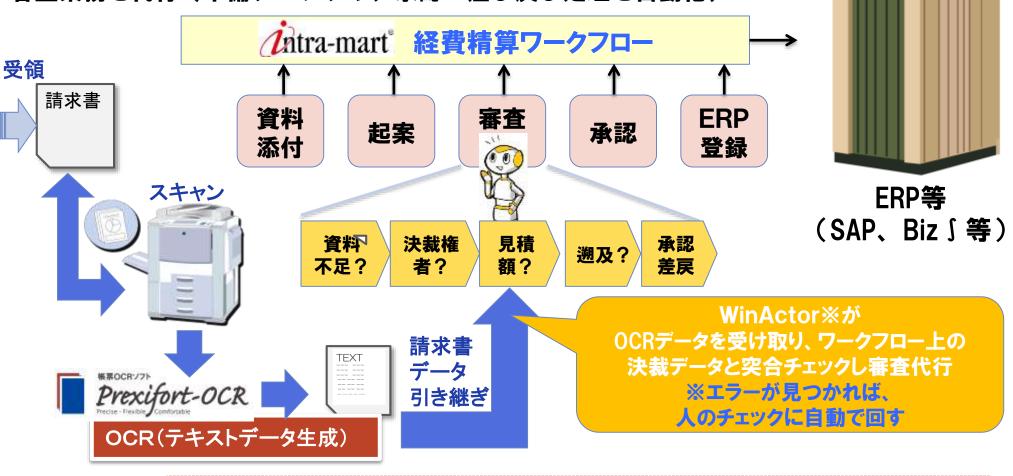
5.イメージワークフローOCR

Copyright © 2017NTT DATA Corporation

請求処理業務の自動化イメージ

NTTDATA

OCRにより請求書情報(会社名・金額・銀行口座等)をデータ化、 WinActorが請求書データと電子決裁データをマッチングし、 審査業務を代行(不備チェックや、承認・差し戻し処理を自動化)

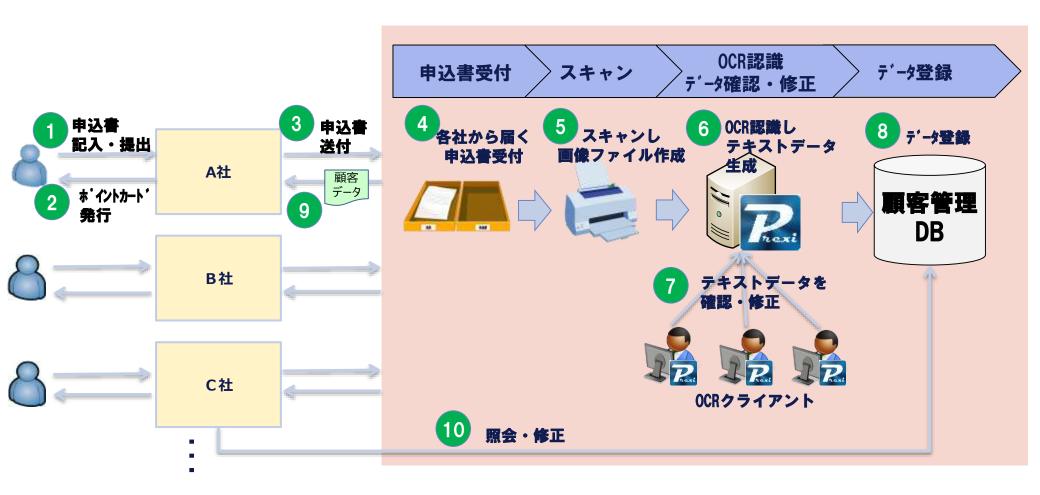


Windows操作ロボット(すり)

※RPAソリューション「WinActor」は、 各種Windows操作を代行するソフトウエア型のロボットです。



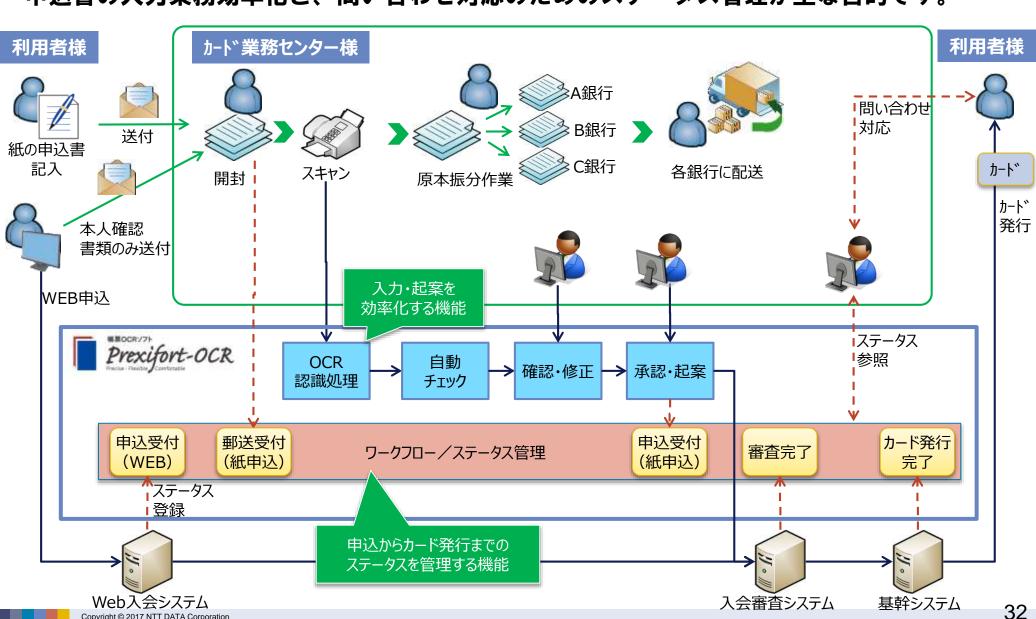
各社がそれぞれ実施している顧客情報の入力・管理業務の集約に取り組まれてます。 申込書に記載の顧客情報を効率的に入力するため、OCRシステムの導入を決められました。



事例:某クレジットカード会社様 入会申込イメージワークフローシステム



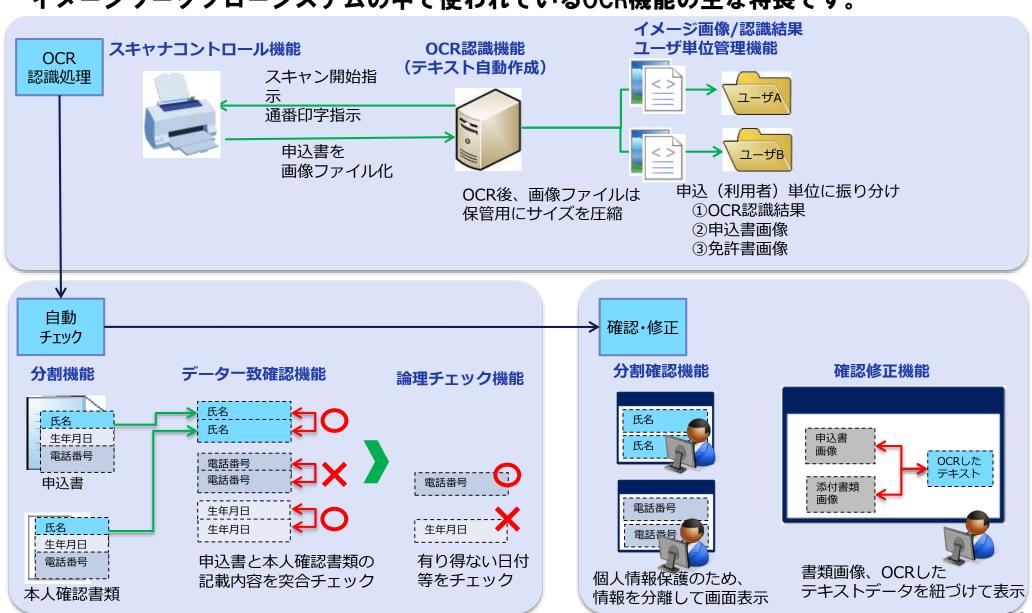
申込書の入力業務効率化と、問い合わせ対応のためのステータス管理が主な目的です。



OCR機能の特長



イメージワークフローシステムの中で使われているOCR機能の主な特長です。

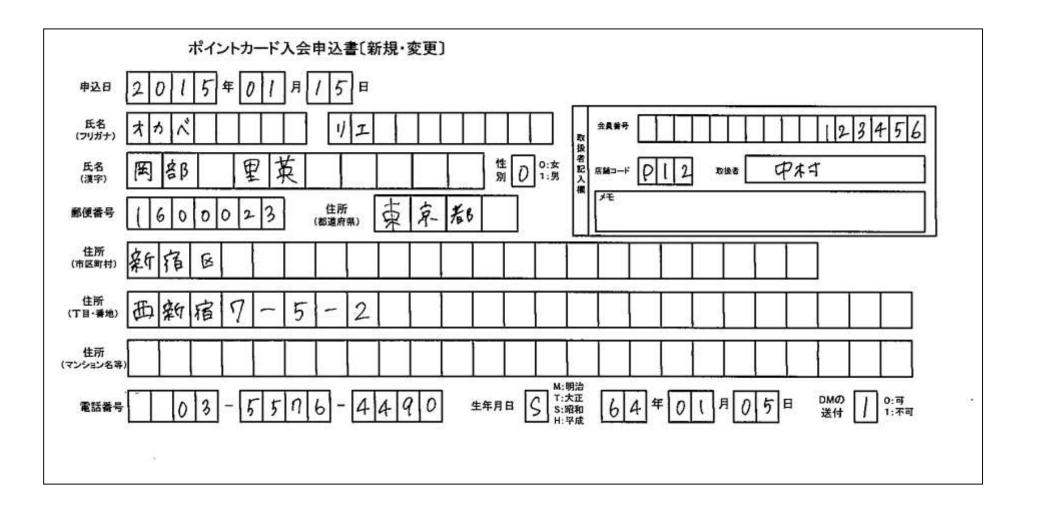


6.ご参考、帳票やスキャナ等

Copyright © 2017NTT DATA Corporation

帳票サンプル(入会申込書【推奨帳票形式】)





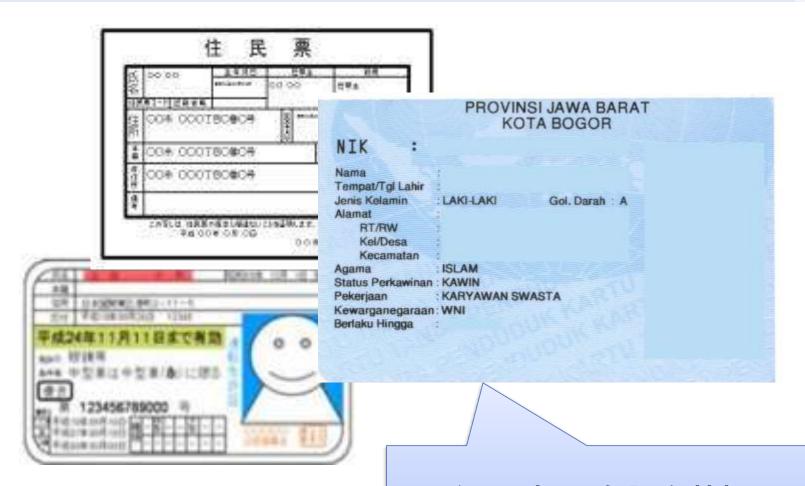
35

帳票サンプル(健康診断書)



出力例	19 12	DE IN	12 H	受診結果						B文章区本額込2-28-	1 6
	la nometro la strata		生年月日	昭和34年12月24日	健診年月日	2008年4月1日	項目		今回受診	- 03-1231-5578 前回受除	前々回受診
カナ氏名	ナンコウジョウジ		性別/年齢	男 48 歳	特定健康診查 受診券番号	1234567890)	le ser	2008年 4月 1日 軽度貨幣 軽度なた輸催性		
	脳血管について提	住願あり			ADDET		心電図検査	所見	軽度なQT延長 確定できない不整要		
既往歷	駆曲告について設住懸あり の表達について設住務あり 等不全へ入工適称について設住務あり 対由について設住療あり 不聖職 行相戦和 得年・一指陽油店		様定できない予報報 キースワグナー: 0 シェイエS: 0 シェイエH: 0 輸送機費 所見 輸送内性網的定: 異常なし								
提業祭	出任 似糖 斯敦		模堆歷	369	飲酒	ほとんど飲まない					
自覚症状	遊稿・耳鳴り 下痢と便秘 手足のしびれ感						メタボリック シンドローム		予備群族出		
物質症状	肝臓肥大(+) 不整断あり 心音不純								メタボリックシンドローム判	定による基準該当と予備を	YEN 15
	48		基準値	今回受診 2008年 4月 1日	前回受診	前々回受診		英	P 26 79	7-4	E 20 28 15
	杂長	tons		170.2			裏側 VS性≥85cm 女性≥90cm				+
	体重	Oogi		62,6			# # #				
身体計選	数据	(cm)		67.1			以下のうち2項目以上に該当 以下のうち1項目が該当				5 ち1項目が第3
	BAI		18.5~24.9	H 26,5			動坊 中性動防≥150mg/d1 かつ/または 細Lコレステロール<40mg/d1 かつ/または 服薬中の方				
血圧	収縮期血圧	bwetgt	~130	122			旗圧 蒋	被開仇日	F≥130mHg かつ/または 核液	資酬真圧主85mmRg かつ/	または 服薬中の方
mit.	拡張期血圧	(wedig)	~85	65			糖代謝 交	施巧血素	徳≥110mg/dl(へモグロビンAlco	0場合≥5.質) かつ/変	たは、服薬中の方
	中性脂肪	(mg/cfl.)	~150	11 165					由圧は正常です。		
血中脂質核查	HDL-31/370-6	ing/dl)	90~	45			0.0000000000000000000000000000000000000	5281	胎質検査に異常が認められます		
	LDL-3/290-A	(mg/dil)	70~139	110			検査分野別判	定	肝機能は正常です。 由験検査は止塞です。		
	GOT	(0/1)	\$0-40	16					原検査は正常です。		
肝機能検査	GPT	(tl/1)	5~45	17		-	a guecousses		今回健認の医師の判断は「要募 資金検査: 質問意の回答結果に		
	y -GTP	(0/1)	~60	15			医師の判断		心能因榜查:前年度結果、及び	f質問章の回答結果により:	ALMs
血糖經濟	空腹時血糖 (*以青年10時間光度)	mg/dil	70~110	105			V-14402-201	2000	继收粮查: 前年疫結果, 及5份	(陳政府により実施	
	ARS DE SATE	(1)	4.3~5.8	4.6			判断した医師		〇〇 健太		
层検査	糖			(~)			今回非特定提	18:			
	蛋白		_	(-)			非特定健設に 保滑血	て異常的	前となった検査があります。医師 (一) 由抗	にご相談ください。 クレアチニン (mg/dil)	0.63
	赤血球数	(75/m)	450~530	489			血清原	77	(mg/42) # 8.5 \$277	レステロール (rg/dl)	210
貧血検査	血色素量	(g/dL)	14.0~18.0	15.8			総蛋白			P (U/1) フェリチン (ng/nl)	217 195
	ヘイトクラット 一直	(9)	39.8~51.8	40.6			血清ア	レブミン	(q/dl) H 12.6	1000 - SV-51	

Copyright © 2017 NTT DATA Corporation



インドネシア・中国の身分証に対応、 マイナンバーカードも対応予定

精度向上策(帳票設計)



OCRに適した帳票の設計例をご紹介いたします。

項番	精度低減要因	精度向上策
	運用上発生しうる雑文字 シタクタ 文字のかすれ	ガイド文字の設定や、記入例を記載することで、雑な文字が記入されることを抑制できます。
1	0658	記入文字が固定される場合は、認識対象外とすることで認識精度が向上します。 1256012743541
	文字のはみだし	文字枠を適切な大きさにすることで、文字枠への接触やはみ出しを防ぎます。 2 1 0 0 0 0 000 円
	文字枠無し(フリーピッチ)	文字枠を設定することで、文字が切り出しやすくなります。
2	やまもとたろう	で ま も と た 3 3
3	フィールド内への印字 2 5 0 0 0 0	フィールド外に印字することで、認識を阻害しなくなります。
4	印字を囲むマーク形式 生年月日 (昭和)(平成)	マーク枠を設けていただくことで、認識精度が向上します。 ② ™ ② ™ ■ ○ ▼ ■ ○ ▼ ■ ○ ▼ ■ ○ ▼ ■ ○ ▼ ■ ○ ▼ ■ ○ ▼ ■ ○ ▼ ■ ○ ▼ ■ ○ ▼ ■ ○ ▼ ■ ○ ■ ○ ▼ ■ ○ ■ ○ ■ ○ ■ ○ ■ ○ ■ ○ ■ ○ ■ ○ ■ ○ ■

推奨スキャナラインナップ(コダック社製品)~1



機種	外観	<mark>認識速度</mark> (2~300dpi)	1日あたりの 読取限度目安	読取範囲	原稿搭載 容量
ScanMate i1150 (銀行窓口向け)	(330×162×246mm, 3.2Kg)	30枚/分 ※最初の10枚は約 60%UP!	3千枚	最大:A4 ※A3フラットベッド の利用可	50枚
i1190E	(330×162×246mm, 3.2Kg)	40枚/分	5千枚	最大:A4 ※A3フラットベッド の利用可	75枚
i2420	(330×162×246mm, 5.5Kg)	40枚/分	5千枚	最大:A4 ※A3フラットベッド の利用可	75枚
i2620	(330×162×246mm, 5.5Kg)	60枚/分 (200dpi)	7千枚	最大:A4 ※A3フラットベッド の利用可	100枚
i2820	(330×162×246mm, 5,5Kg)	60枚/分 (200dpi)	8千枚	最大:A4 ※A3フラットベッド の利用可	100枚

Copyright © 2017 NTT DATA Corporation

推奨スキャナラインナップ(コダック社製品)~②



機種	外観	認識速度 (2~300dpi)	1日あたりの 読取限度目安	読取範囲	原稿搭載 容量
i2900	(431.8×371×237.5mm, 15.8Kg)	60枚/分	1万枚	最大:A4	250枚
i3200	(431.8×371×237.5mm, 15.8Kg)	50枚/分	1万5千枚	最大:A3	250枚
i3400	(431.8×371×237.5mm, 15.8Kg)	90枚/分	2万枚	最大:A3	200枚
i4250	(489×467.3×347.9mm, 30.4Kg)	110枚/分	4万枚	最大:A3	500枚
i4650	(489×467.3×347.9mm, 30.4Kg)	130枚/分	5万枚	最大:A3	500枚

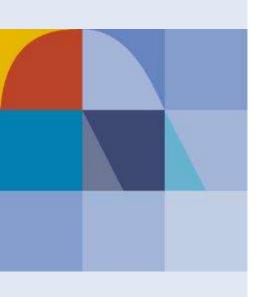
DATA Corporation

推奨スキャナラインナップ(コダック社製品)~3



機種	外観	認識速度 (2~300dpi)	1日あたりの 読取限度目安	読取範囲	原稿搭載 容量
i4850	(489×467.3×347.9mm, 30.4Kg)	150枚/分	5万枚	最大:A3	500枚
i5200	(693×510×434mm, 57.6Kg)	140枚/分	無制限	最大:A3	500枚
i5600	(693×510×434mm, 57.6Kg)	170枚/分	無制限	最大:A3	750枚
i5850	(693×823×1067~1308mm, 186Kg)	210枚/分	無制限	最大:A3	750枚

Copyright © 2017 NTT DATA Corporation



NTT Data Global IT Innovator