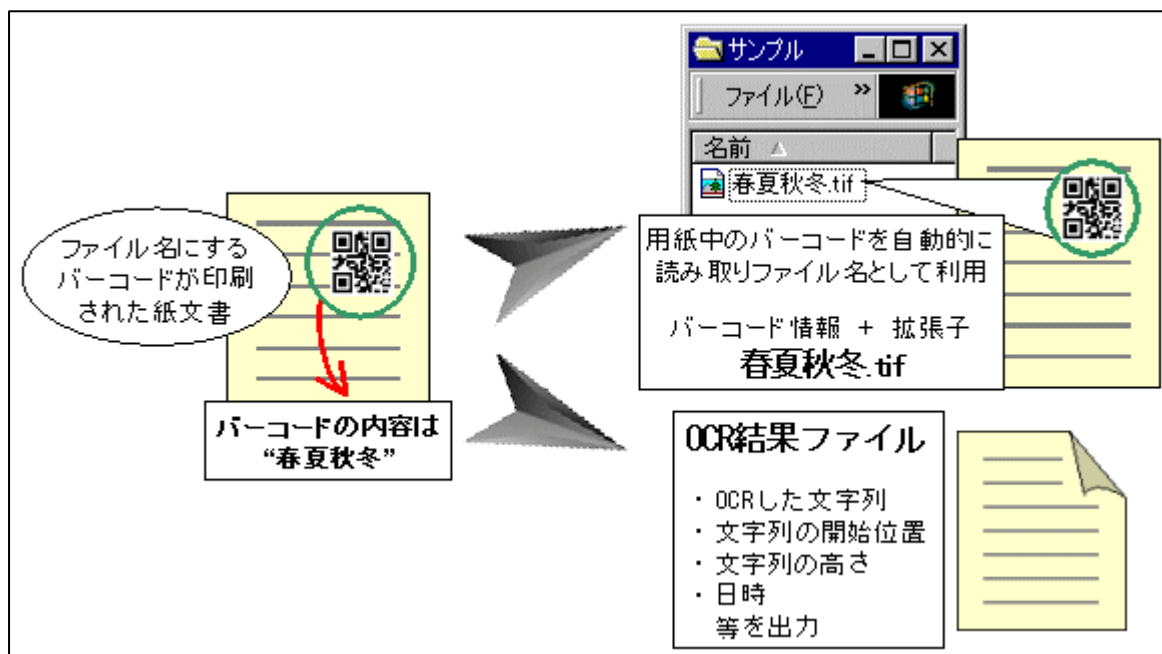


TWAIN対応スキャナからの KDbarscan(OCR) バーコード情報活用プログラム KDbarscan(OCR)

[概要]

TWAINに対応したスキャナからデータ(用紙)を取り込む処理を対象にしたアプリケーションです。バーコード画像を含む用紙をスキャナから取り込む時に**バーコードの情報を読み取り出力ファイル名に利用**する製品「KDbarscan/KDbarscan」に**OCR処理機能**を追加したキャプチャリングソフトです。TWAIN対応しているスキャナであればどのメーカーのスキャナからでも処理が可能です。



[バーコード情報の読み取りとは]

バーコード情報の読み取りとは、用紙の中からバーコード画像の部分を自動的に探しだし、バーコードの情報を読み取る処理です。スキャナからバーコードを可読できるイメージを取り込むことができればどのような用紙からでも精度の高い情報取得を行い利用することができます。

[認識対応バーコード一覧]

名称	内容
QR	QRコード。型番は任意。(2次元)
CODE39/128	Code39/128。桁数は任意。
GS1-128 (EAN128)	Code128のうち、UGC/EAN128規格に適合するもの。
PDF417	PDF417バーコード。(2次元)
JAN8/13	JANおよびEANコードのうち、8/13桁のもの。
ITF	ITF。桁数は任意。
NW-7	NW-7。桁数は任意。
CUSTOMER	郵便カスタマバーコード。
GS1 DataBar (RSS)	RSSバーコード。(7種類のタイプをサポート)
BLANK	空白ページの認識が可能。(空白をバーコードに置換)

I. 商品導入の背景、狙い

【特徴1】

スキャナより取り込んだデータ(モノクロ・カラー)からバーコード情報を読み取り、出力ファイル名にできるためバーコードから出力ファイル名の名前付けを自動的に行うことができます。

【特徴2】

スキャナより取り込んだデータ上の文字をテキスト文書へ変換を行うことができます。これにより、イメージファイル内の文字の検索やコピーなどの処理を行えるようになります。

【特徴3】

スキャナより取り込んだデータからバーコード情報を読み取り出力ファイル名やCSVファイルに出力することによって文字検索などテキスト情報として活用することができます。

【特徴4】

マルチページの取り込みと出力において読み取ったバーコード情報ごとに区別して出力する、同じ情報のファイルをまとめて出力することができるため出力ファイルの仕分けを自動的に行うことができます。

【特徴5】

バーコードの読み取りは、誤認識率が非常に低いことが特徴で、可読できるバーコードのイメージを貼り付けることができればどのようなイメージデータからでも精度の高い情報取得が行えます。

QRコード、PDF417バーコードについては誤り訂正レベルにより一部欠損がある場合でも認識が可能になります。

【特徴6】

キーワード別配信機能(オプション)を利用すると、スキャナから読み取ったバーコード情報とユーザから指定されたキーワードで検証を行い条件に応じた指定フォルダへの出力を行うことができます。

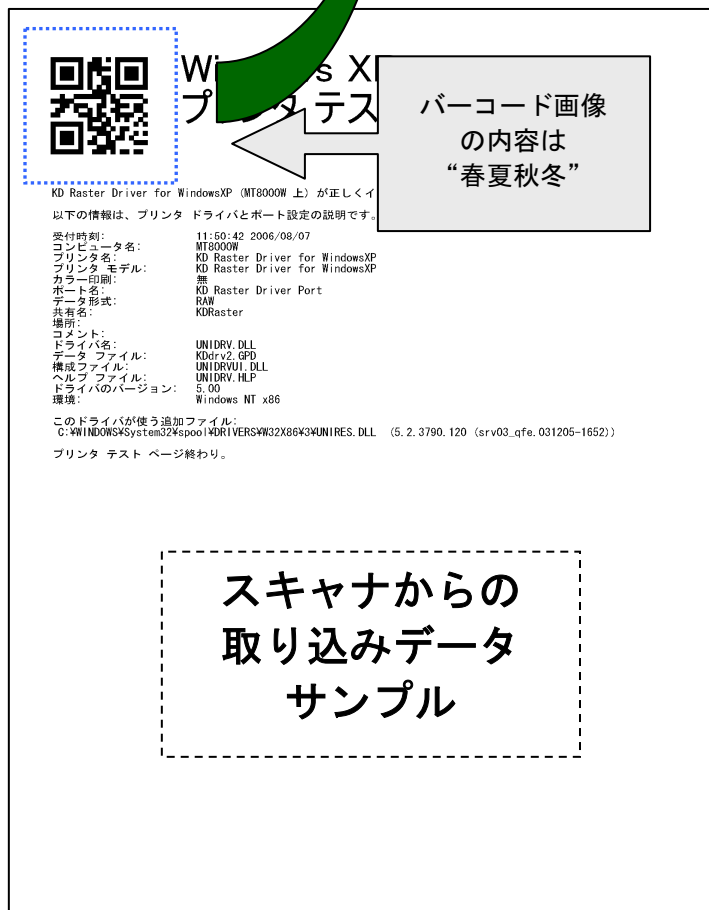
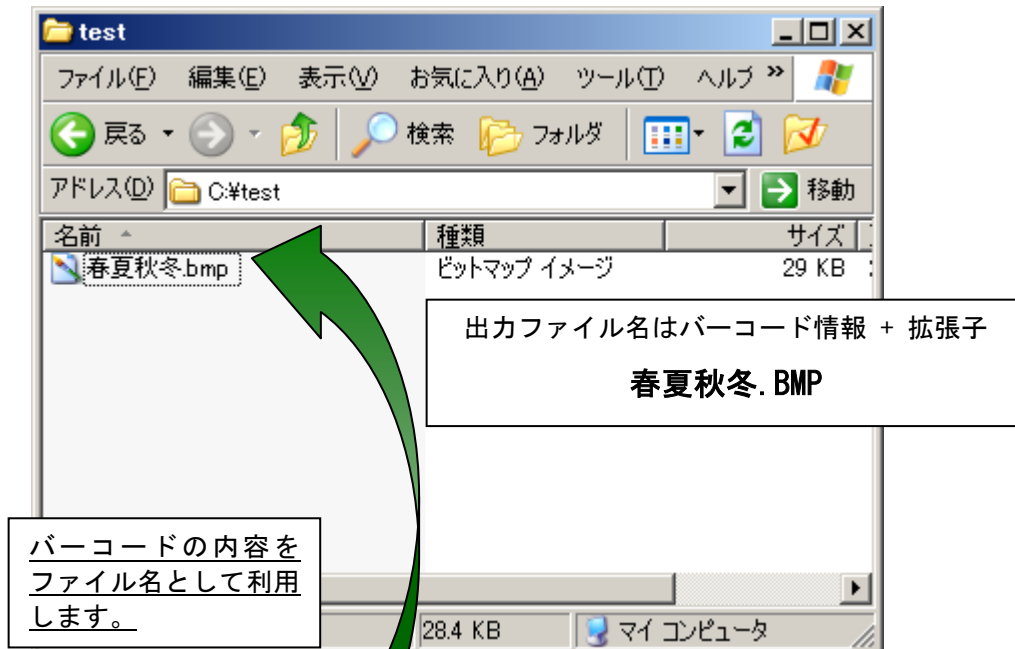
【特徴7】

スキャナより取り込んだデータをいろいろなフォーマット(TIFF(G4, PackBits, LZW, JPEG, 非圧縮)、JPEG、BMP、PDF、PDF_FLATE、DOCUWORKS(XDW)、CSV)から選択して出力することができます。カラーモード(カラー/モノクロ/グレースケール)はお使いのスキャナのTWIN32スキャナドライバの設定に依存します。

II. 搭載機能について

1. バーコード情報からのファイル名出力機能
スキャナから取り込むデータ(用紙)に存在するバーコード画像から情報を読み取り
出力ファイル名として利用します。

(例)



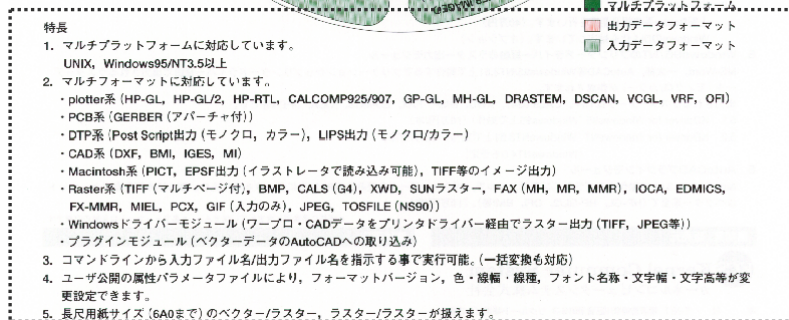
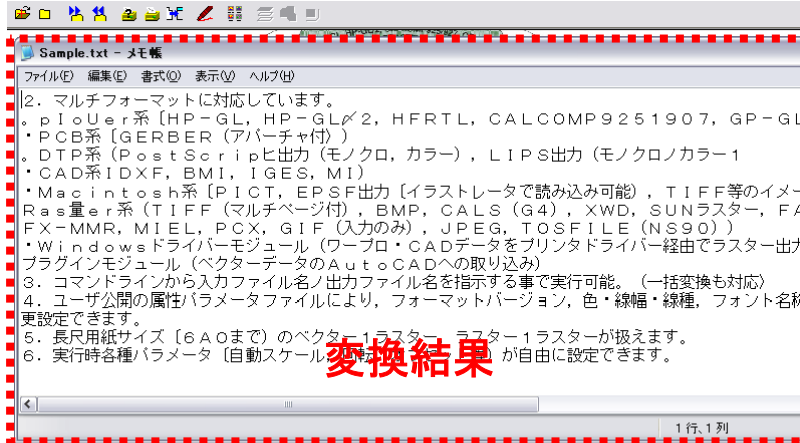
スキャナからの
取り込みデータ
サンプル

2. OCR処理機能

スキャナから取り込むデータ(用紙)の文字をOCR機能を使用してテキスト文書への変換を行います。

(例)

原図の下の枠をOCR処理し、テキストに出力します。



原 図

3. バーコード情報のCSVファイル出力機能

スキャナから取り込むデータ(用紙)に存在するバーコード画像から情報を読み取り CSVファイルに情報を出力します。

出力ファイル名の拡張子をCSVに変更したファイル名で出力します。

(例)

	出力ファイル	バーコード 種別	角度	バーコード 情報(内容)
①	test-1.tif	CODE39	0	1234A
②	test-1.tif	CODE128	90	12345
③	test-1.tif	QR	0	KernelComputerSystem
4				

スキャナから取り込んだデータ(用紙)よりバーコード情報をCSVファイルに出力

Windows XP
プリンタ テスト ページ *1234*

①

②

③

スキャナからの
取り込みデータ
サンプル

- ・ 用紙の中に存在する全ての(複数)のバーコードを認識します。
- ・ 異なる種別のバーコードを同時に認識します。
- ・ バーコードを0, 90, 180, 270度の4方向から認識します。
- ・ バーコードを認識する順番を上下左右から選択します。

出力する情報を以下の項目から選択することができます。

- ・ 処理対象となったファイル名、出力ファイル名
- ・ バーコードの内容(テキスト情報)、種別(CODE39, QR_CODE等)、角度(0, 90, 180, 270度)
- ・ バーコードイメージが描画されている位置(座標(x, y))
- ・ マルチページ出力においてバーコードが検出されたページ番号
- ・ 処理を行った日付時間

4. バーコード情報による仕分け機能

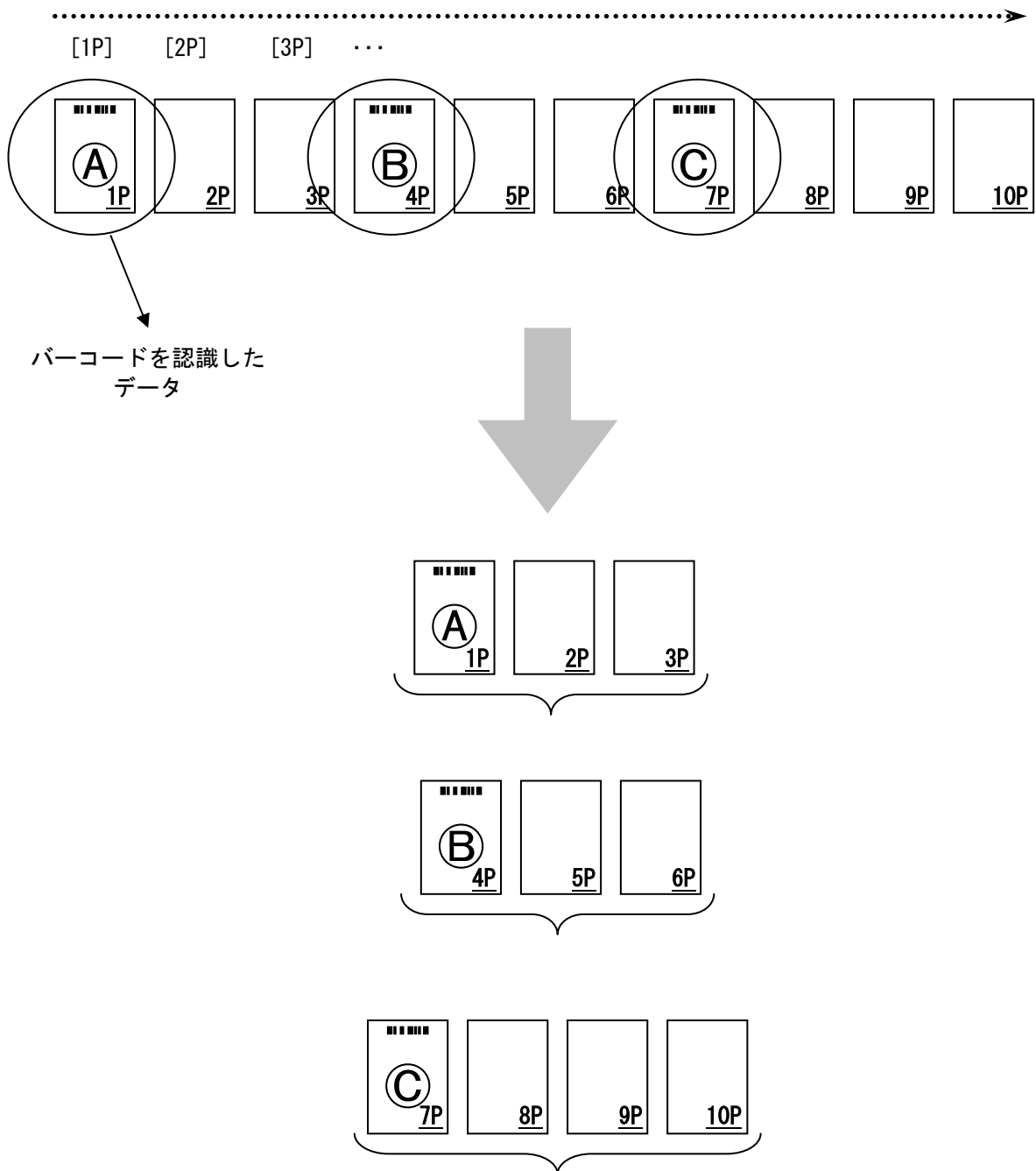
バーコード情報を含むマルチページデータに対してバーコード情報を検出したページを
基に仕分けして出力します。

バーコードページ毎に仕分けして出力する、バーコードの内容単位に仕分けすることが
できます。

(例①)バーコード情報を検出したページごとに区分けして出力を行います。

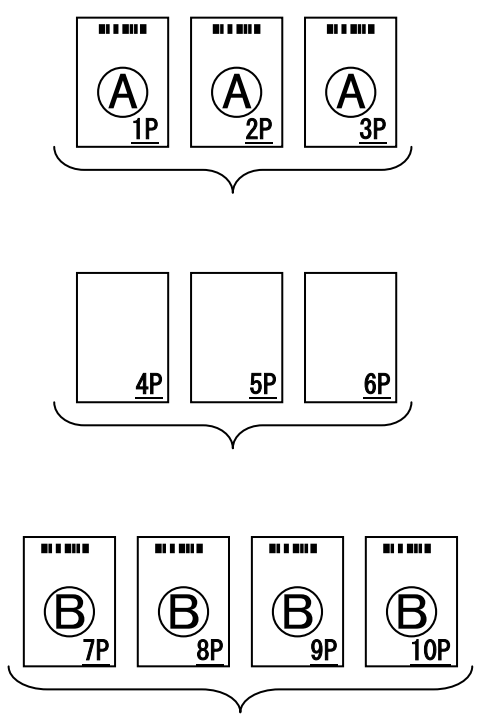
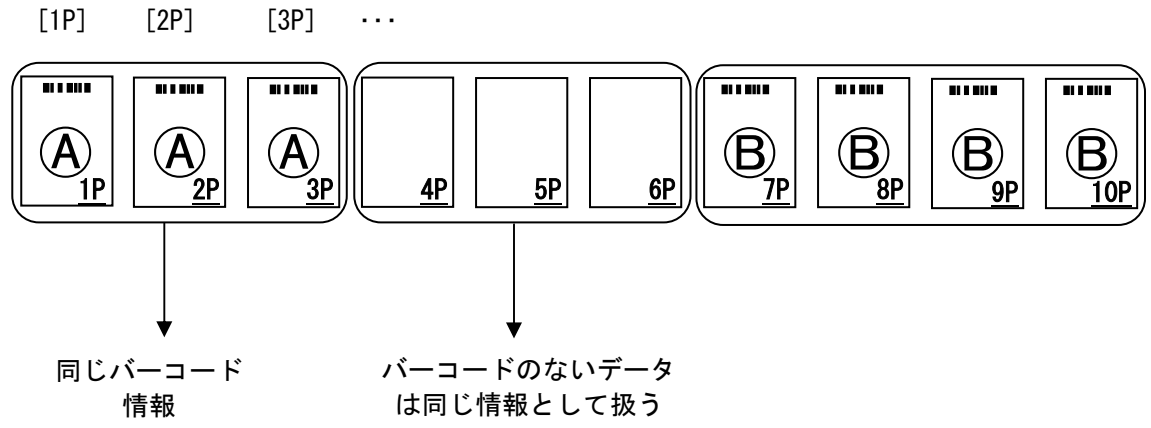
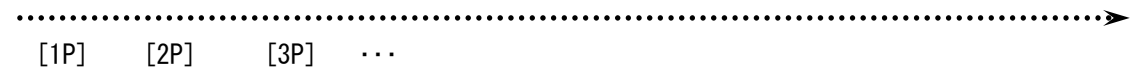
※バーコードページ毎

用紙の取り込み順番方向



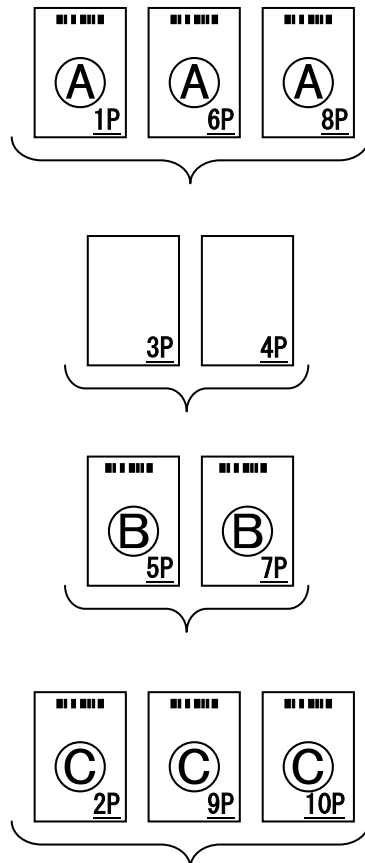
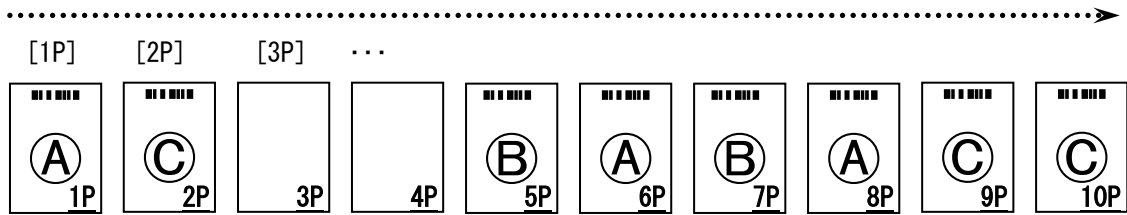
(例②) 先頭から順番に比較してバーコードの内容(同じ情報)単位に仕分けを行う。
※バーコードの内容単位

用紙の取り込み順番方向



(例③) 全てのページに対してバーコードの内容(同じ情報)単位に仕分けを行う。
※バーコードの内容単位(順不同)

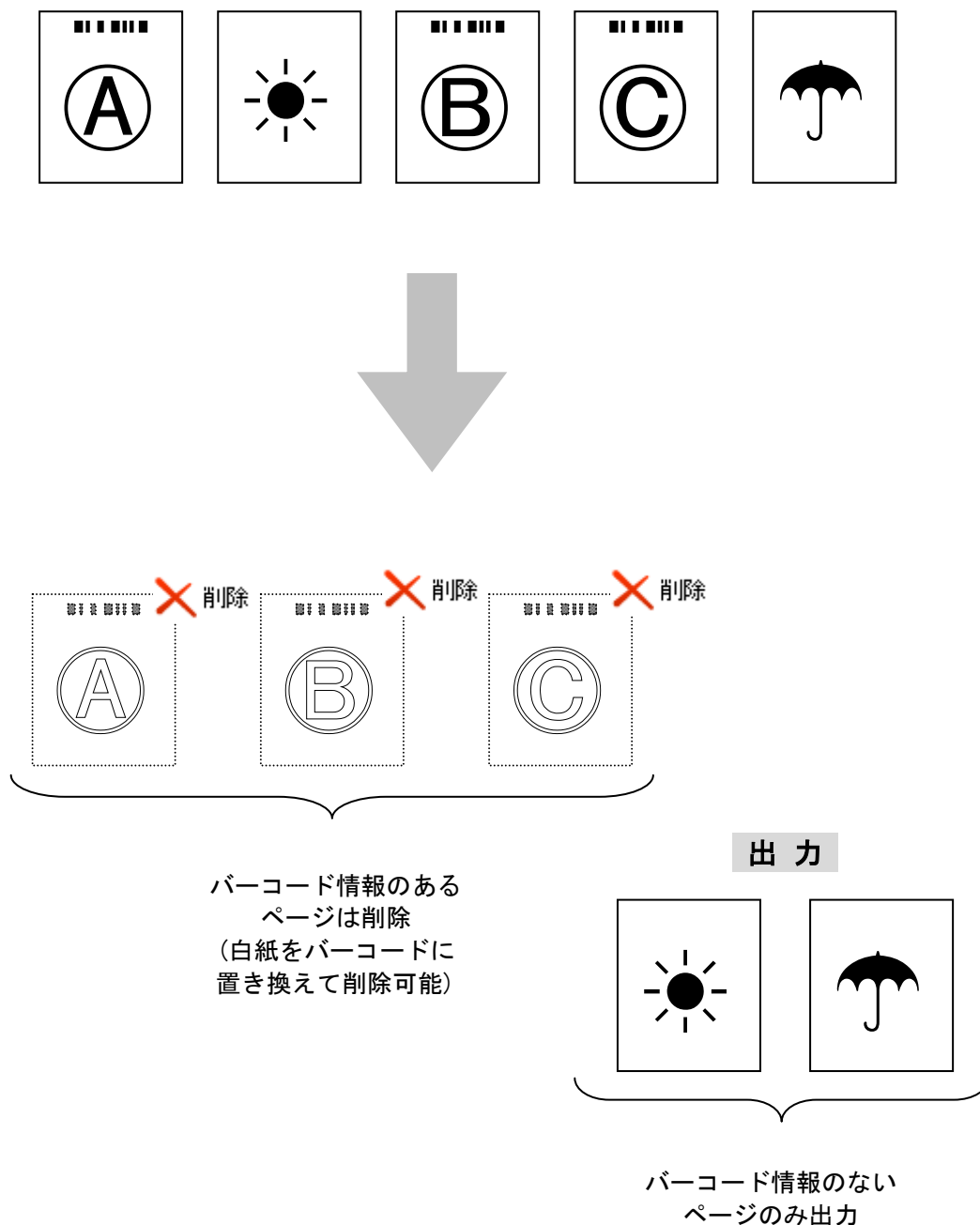
用紙の取り込み順番方向



5. バーコード情報による削除機能
バーコード情報(白紙の指定が可能)を含むデータに対してバーコード情報を検出したページを削除します。バーコードがあるページや白紙ページを削除する場合に利用します。(逆のパターンの「バーコード情報を検出できなかったページの削除」も設定可能です。)

※処理するデータ全てにバーコード情報が含まれる場合、データが削除されるため何も出力されません。

(例)



6. バーコード情報によるキーワード別配信機能 (※オプション機能になります)
スキャナから読み取ったバーコード情報とユーザから指定されたキーワードで
検証を行い条件に応じた指定フォルダへの出力、データの削除などを行います。

(例①)

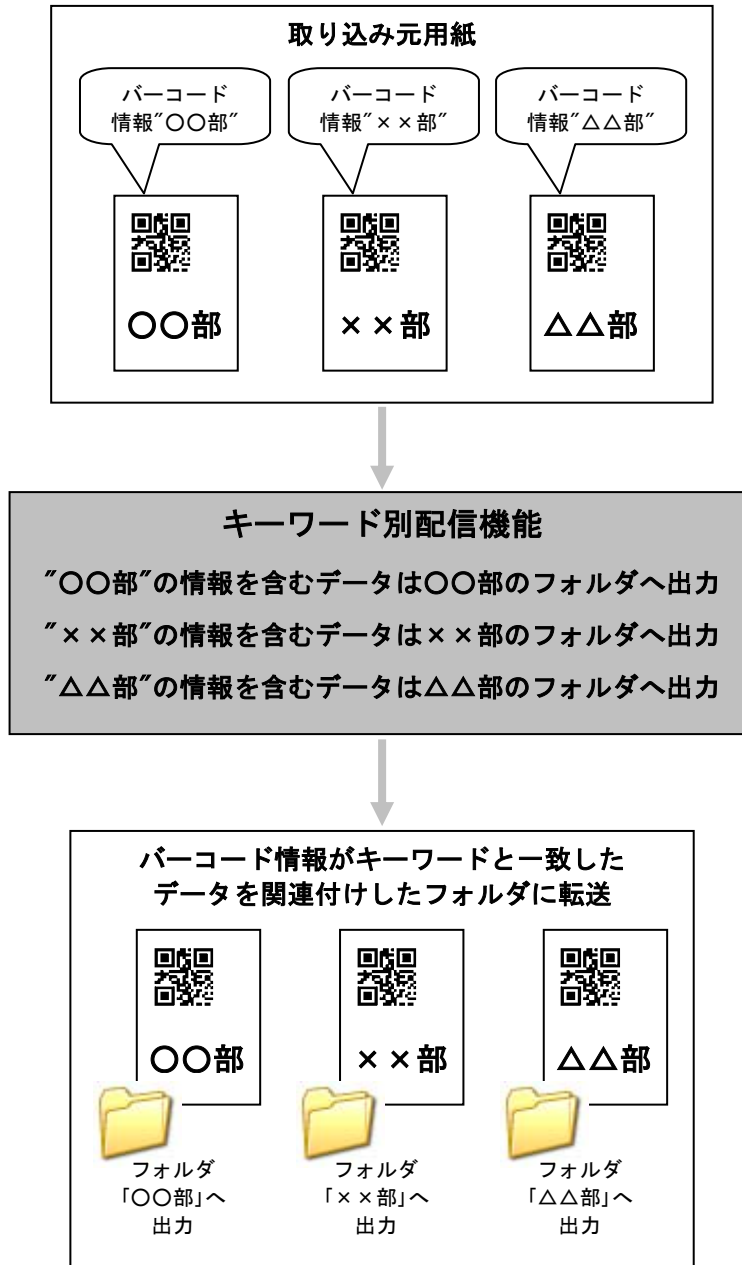
キーワードごとに出力先フォルダを指定し、バーコード情報に応じて配信を行う。

[設定条件]

“〇〇部”の情報を含むデータは「〇〇部」のフォルダへ出力する。

“××部”の情報を含むデータは「××部」のフォルダへ出力する。

“△△部”の情報を含むデータは「△△部」のフォルダへ出力する。



(例②)

キーワードに該当しないデータを指定のフォルダへ出力する。

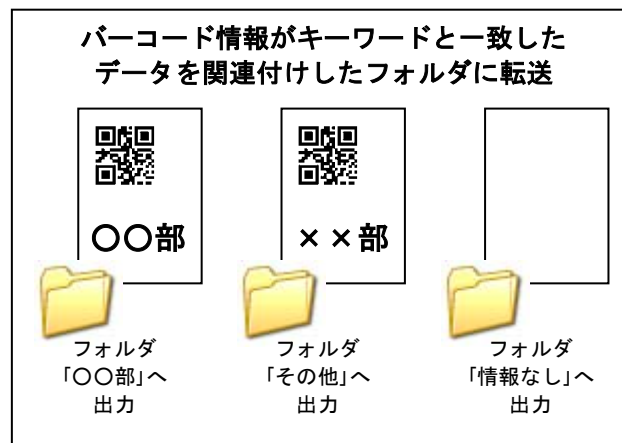
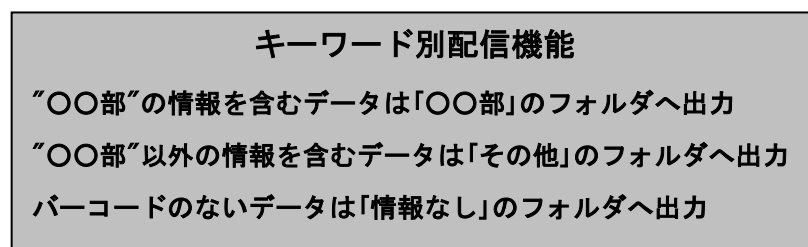
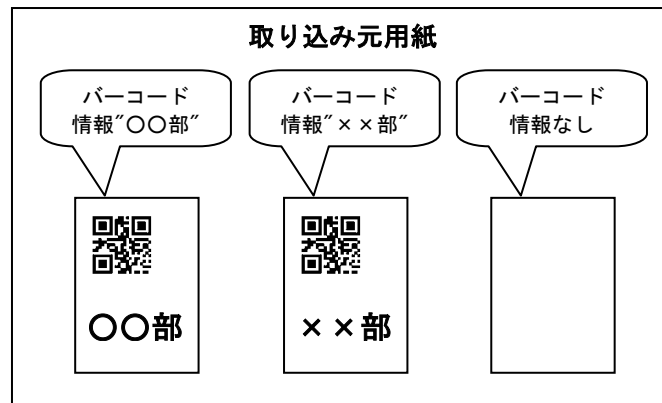
バーコードのないデータを指定のフォルダへ出力する。

[設定条件]

“〇〇部”の情報を含むデータは「〇〇部」のフォルダへ出力する。

“〇〇部”以外の情報を含むデータは「その他」のフォルダへ出力する。

バーコードのないデータは「情報なし」のフォルダへ出力する。



(例③)

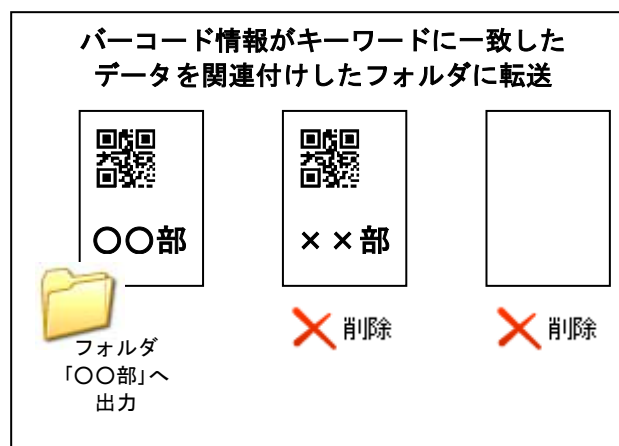
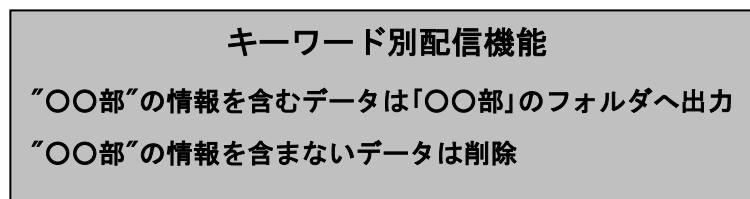
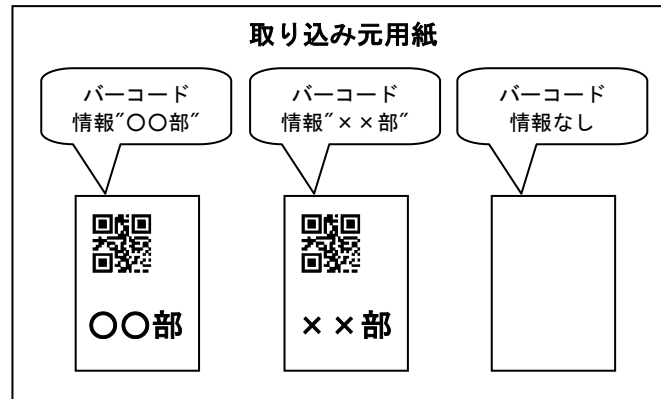
キーワードに該当しないデータを削除する。

バーコードのないデータを削除する。

[設定条件]

“〇〇部”の情報を含むデータは「〇〇部」のフォルダへ出力する。

“〇〇部”の情報を含まないデータは削除する。



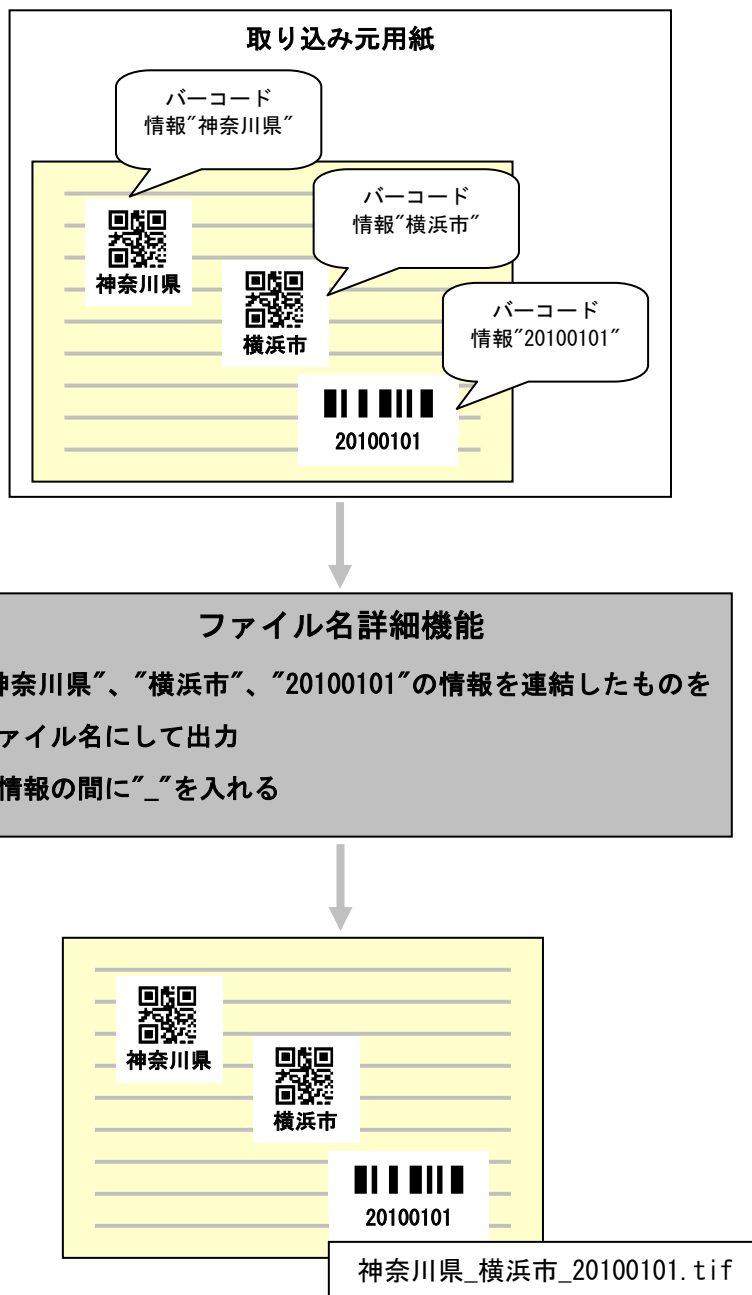
7. バーコード情報によるファイル名詳細機能(※オプション機能になります)
読み取ったバーコード情報から詳細なファイル名付けをしてデータを出力することができます。
複数個からなるバーコード情報や、1つのバーコード情報を分割しての部分的な情報など、特定の情報を自由に組み合わせてファイル名付けをすることができます。

(例)

用紙中のバーコードを連結したものをファイル名にして出力する。

[設定条件]

“神奈川県”、“横浜市”、“20100101”の情報を含むデータを読み取り、情報を連結したものをファイル名にする。各情報の間に“_”を入れる。



8. バーコード情報による階層出力機能 (※オプション機能になります)
読み取ったバーコード情報から階層構造のフォルダを作成し、そのフォルダにデータを出力することができます。
複数個からなるバーコード情報や、1つのバーコード情報を分割して部分的な情報など、特定の情報を自由に組み合わせて階層構造のフォルダ名を作成することができます。

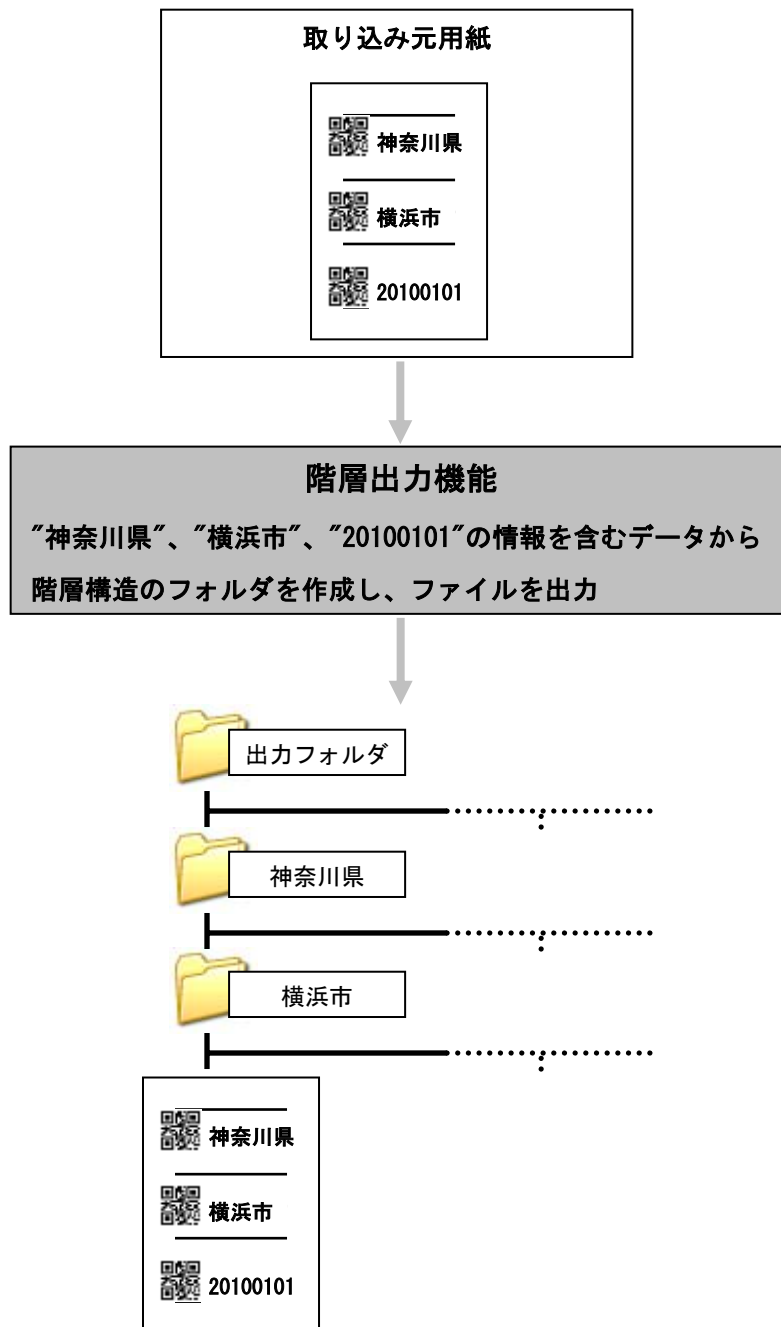
(例①)

それぞれのデータにある情報で作成した階層構造のフォルダにデータを出力する。

[設定条件]

出力フォルダに

“神奈川県”、“横浜市”、“20100101”の情報を含むデータから「神奈川県」、「横浜市」という階層構造のフォルダを作成し、ファイルを出力する。

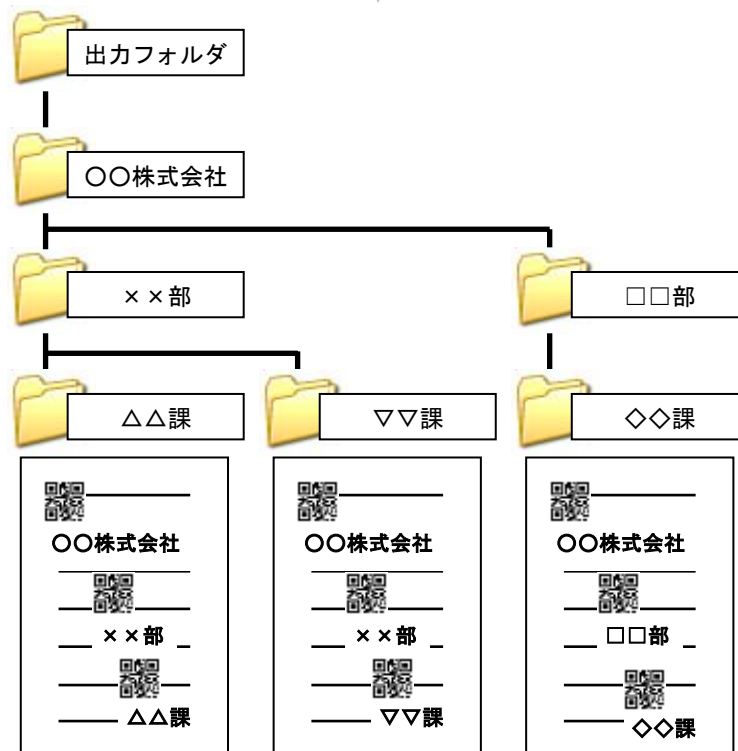
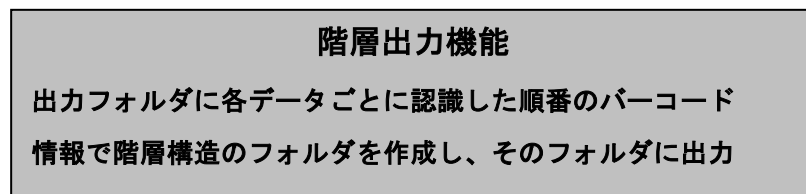
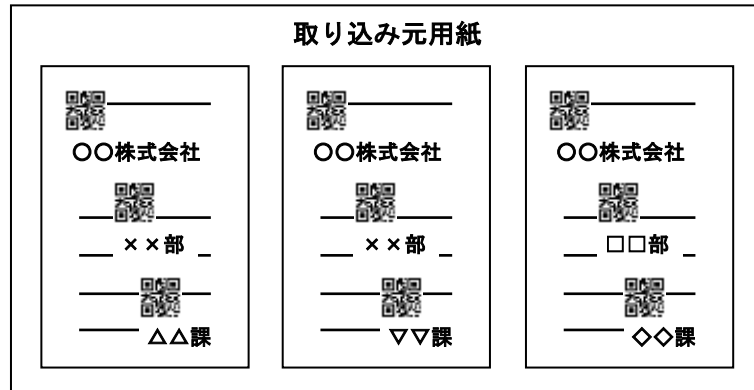


(例②)

それぞれのデータにある情報で作成した階層構造のフォルダにデータを出力する。

[設定条件]

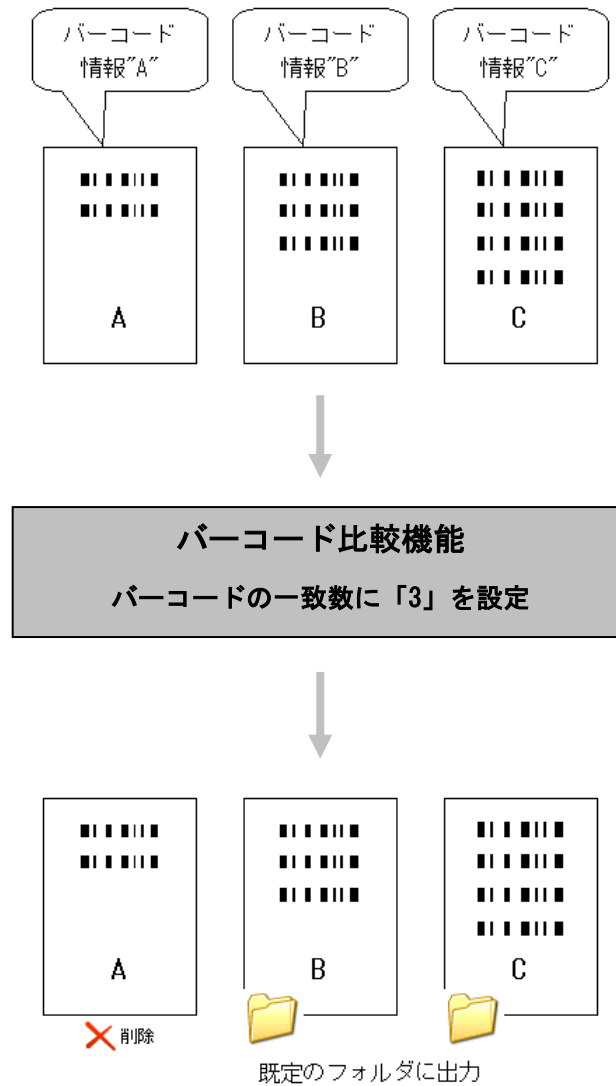
出力フォルダに各データごとに認識した順番のバーコード情報で階層構造のフォルダを作成し、そのフォルダに出力する。



9. バーコード情報による比較機能(※オプション機能になります)
スキャナから読み取った複数のバーコード情報が一致しているか比較を行い、指定個数以上一致している情報のみ出力する動作によって認識精度を高めた処理を行います。

(例)

3つ以上同じバーコード情報があるデータのみ出力する。



Ⅲ. バーコードの認識について

- ・バーコード情報の認識について全方向(90度単位の4方向(0度、90度、180度、270度))に対応しています。向きの異なるデータに対してのバーコード情報の認識が可能です。
- ・種類の異なるバーコードを同時に認識、処理することができます。
- ・バーコードを読み取る範囲を(X, Y)座標で指定することができます。
- ・同一ページに複数バーコードが存在する場合、全てのバーコードを認識し情報を取得します。最初に認識したバーコードの情報を出力ファイル名に利用するため認識する順位を範囲、上下左右方向から調整します。
- ・バーコードの認識については入力データの状態に依存します。認識率を高くするためには品質の高いデータを処理対象とするようにしてください。

[推奨]

スキャナからの取り込み解像度(取り込み用紙の品質) : 300DPI以上

- ・認識率が低下する要素として、バーコードを含むイメージデータにノイズがかかっている、バーコードのイメージ部分に欠損がある、対応していない方向(0、90、180、270度以外)に描画されている、データの品質(解像度)が低い場合(200DPI以下)などが挙げられます。
- ・FAXデータなどは、取り込み機器、出力機器の影響を受けるため認識率は一定ではありません。FAXデータを使用する場合、高品質以上で作成されたデータを推奨します。
- ・バーコードがあるにもかかわらず情報を読み取れていない(認識できていない)場合、以下の機能を利用してバーコードの認識率を上げる調整が行えます。
 - ①ノイズがある場合、ノイズ(ディザノイズ)除去機能を有効にする。
 - ②スキャナから取り込むときの出力解像度を上げる。
- ・バーコード情報を正しく読み取れない(誤認識している)場合、以下の機能を利用して誤認識を回避する調整が行えます。
 - ①バーコードの種別が予め分かっている場合、該当のバーコードの種別のみを認識対象にする。
 - ②バーコードの位置が予め分かっている場合、認識範囲を指定する。
 - ③検出するバーコードの桁数の範囲を指定する。(例 : 3桁以上20桁以下)

IV. OCR機能について


- ・ OCR処理エンジンは「メディアドライブ(株)活字文書OCRライブラリ」を使用しています。
- ・ A3サイズを超えるものをOCR処理するには、「A3より大きいデータをA3サイズに納める」のチェックを有効にする必要があります。また、この機能を利用する場合、認識率や処理速度が低下することがあります。
- ・ 文字列の開始地点、処理範囲について
 - ・ 左上を基準原点とします。
- ・ 処理範囲の設定で、短形の辺が20ピクセルより小さい場合はOCR処理は行われません。
- ・ OCR変換時の認識言語に「英語」が指定された場合、固定長フォント(MS ゴシックなど)で記述されている数値は、“(カンマ)”以降の“0(ゼロ)”が“0(オー)”になる場合があります。下記サンプルの場合、「MS P ゴシック」は10,000,000(千万)と認識され、「MS ゴシック」の場合は10(十)と000,000(オーが6つ)と認識される場合があります。

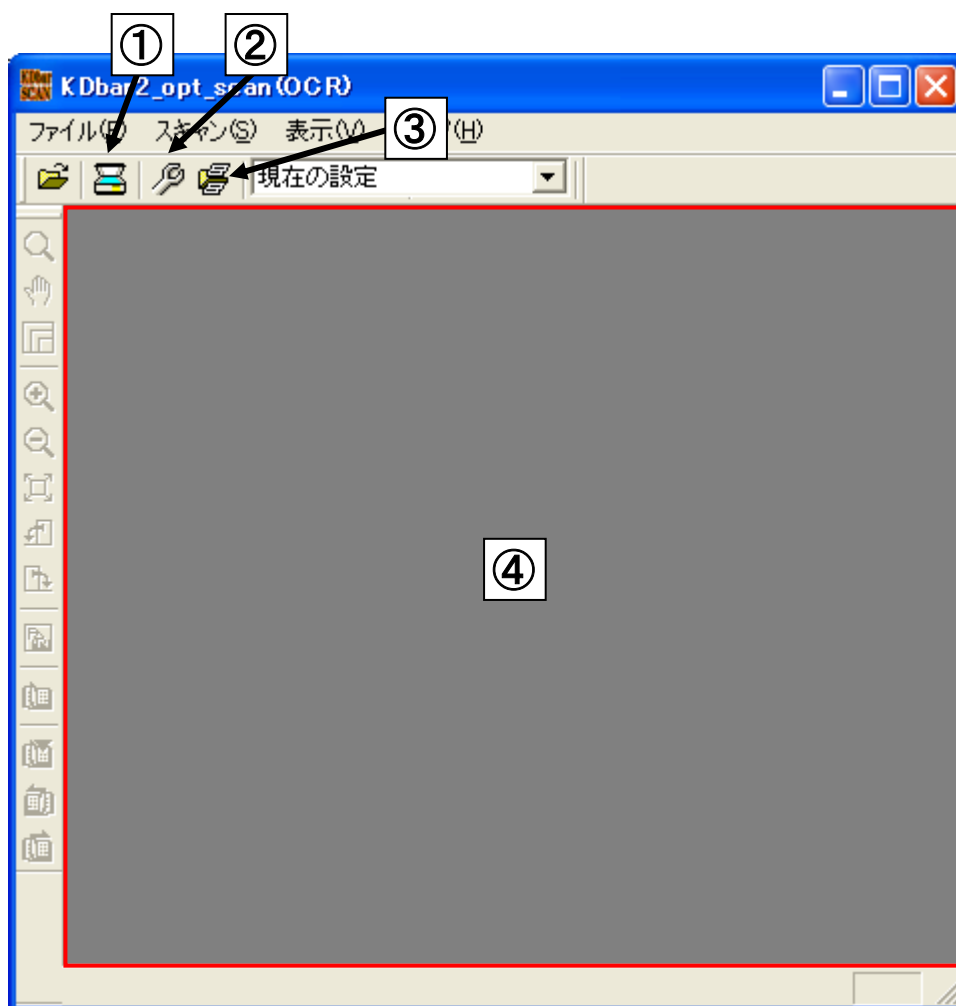
MS Pゴシック	MS ゴシック	MS Pゴシック	MS ゴシック
10,000,000	10,000,000	10,000,000	10,000,000
200,000	200,000	200,000	200,000





V. 操作画面

[メインダイアログ]


プログラムを起動すると下記ダイアログが表示されます。スキャナが利用可能な状態で

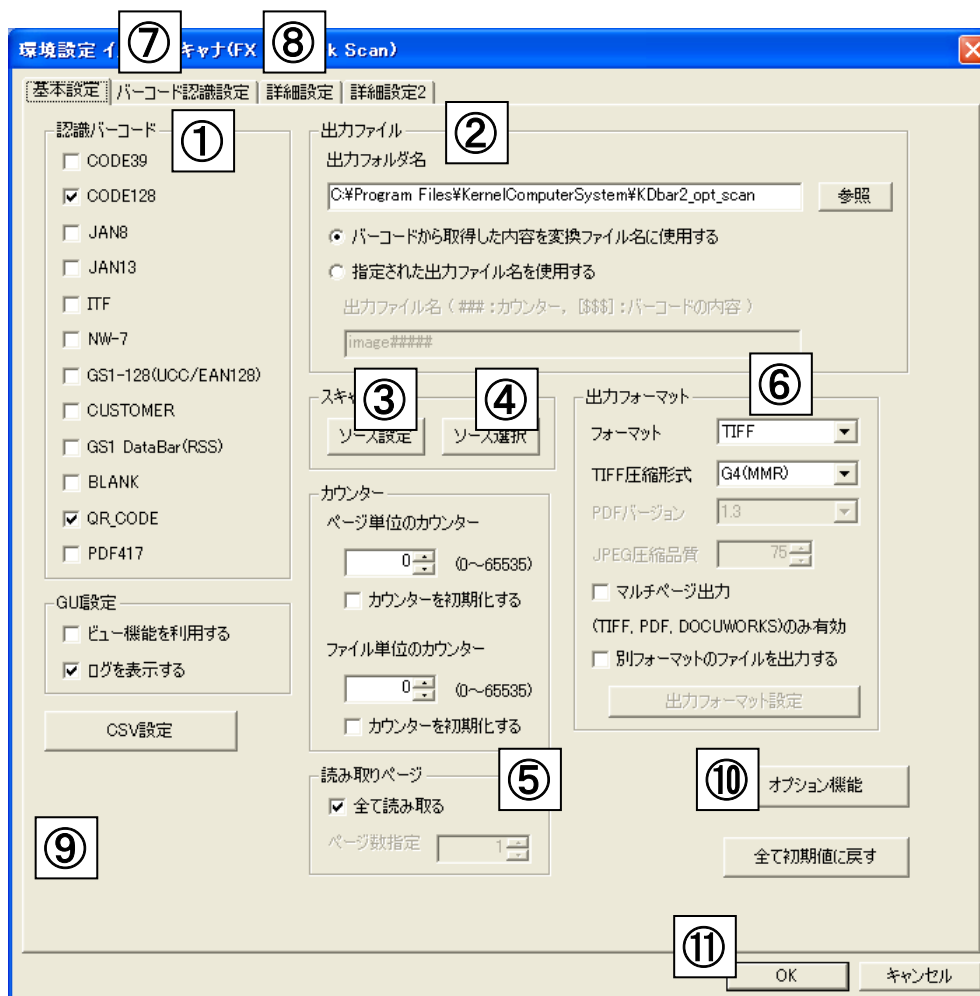
 ボタンをクリックするとスキャナにセットされた用紙の取り込み設定内容に沿った処理が行われ、指定フォルダに取り込んだデータのイメージファイルが作成されます。



- (1) スキャナが利用可能な状態で  ボタンをクリックするとスキャナにセットされた用紙の取り込みを開始し設定内容に従って処理を開始します。
- (2)  ボタンを押すとバーコードの認識に関する設定を行うダイアログが起動します。
※後述に詳細を説明
- (3)  ボタンを押すと設定情報のセーブ/ロードを行うダイアログが起動します。
既にセーブ情報がある場合、隣にある  コンボボックスからロードを行うことが可能です。
- (4) ビュー機能を利用する場合、ビューエリアに読み取った画像が表示されます。

[設定ダイアログ]

メインダイアログの  ボタンをクリックすると下記の設定ダイアログが表示されます。
バーコードの認識設定、出力するデータ等の設定をダイアログ上より行い処理に反映します。



- (1) 認識するバーコードの種類を選択します。複数同時選択が可能です。
- (2) バーコードから取得した内容をファイル名に使用するか選択し、出力先のフォルダを指定します。
- (3) ソースの設定ダイアログが表示されます。解像度、取り込み用紙サイズ、カラーモードなどスキャナドライバで対応している項目について設定を行います。
- (4) ソースの選択ダイアログが表示されます。装置のドライバ(データソース)一覧が表示されるので使用するスキャナドライバを選択してください。

- (5) スキャナにセットされた用紙を全て読み取る、または指定したページ数だけ読み取るかの設定を行います。

- (6) 出力ファイルのフォーマット指定を行います。

- (7) バーコードの認識に関する設定を行うダイアログを表示します。バーコードの読み取り方向、処理を行う領域等の指定を行います。

- (8) その他の設定を行うダイアログを表示します。ページ分割(仕分け)機能やページの削除等の設定を行います。

- (9) 読み取ったバーコード情報をCSVファイルに出力する設定を行います。

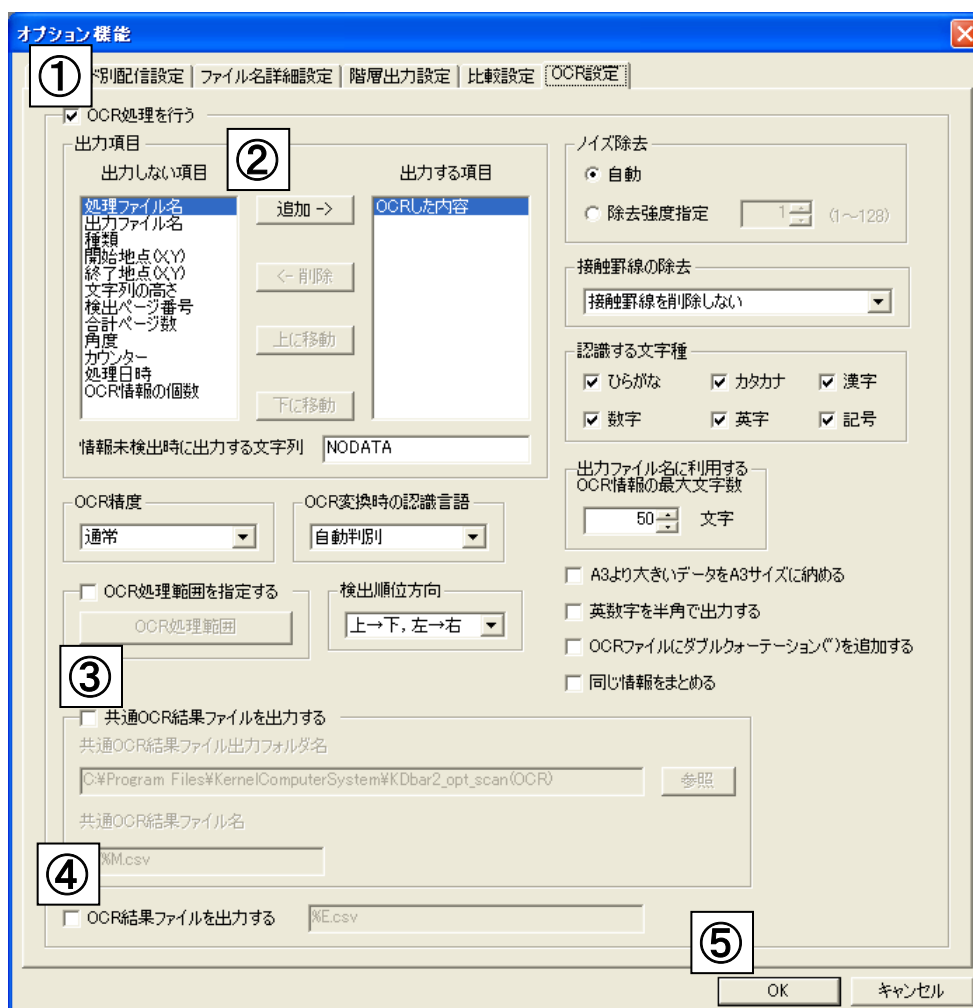
- (10) キーワード別配信・ファイル名詳細・階層出力・比較(オプション)機能やOCR処理機能に関する設定を行うダイアログを表示します。

- (11) 「OK」ボタンをクリックすると情報を保存してメインダイアログに戻ります。
「キャンセル」ボタンをクリックすると情報を保存せずメインダイアログに戻ります。

[OCR設定ダイアログ]

イメージデータにOCR(文字認識)処理を行い、テキストデータに変換して出力する際の設定を行います。

※OCR処理エンジンは「メディアドライブ(株)活字文書OCRライブラリ」を使用しています。



- (1) チェックを付けて有効にした場合、読み取ったOCR処理を行います。
- (2) 一覧の中からOCR結果ファイルに出力するものを選択してください。
- (3) チェックを付けて有効にした場合、OCR結果の情報を共通のテキストファイルに出力します。指定したテキストファイルに追記されていきます。
- (4) 各出力ファイルに対応したOCR結果をテキスト形式で出力します。同名のファイルが存在する場合、上書きされてしまいますので注意してください。
- (5) 「OK」ボタンをクリックすると情報を保存して設定ダイアログに戻ります。「キャンセル」ボタンをクリックすると情報を保存せず設定ダイアログに戻ります。

VI. 読み取り対応バーコード一覧

> GS1-128 (UCC/EAN-128)

GS1-128 (UCC/EAN-128) は共通商品コード (JAN、EAN、UPC) を補足するコードで、商品の製造年月日や製造ロット番号、出荷コンテナ番号などの商品関連情報や物流関連情報を追加、補足するためのコードです。各業界で利用され始めています。



> QR_CODE

日本で多く普及している2次元コードで、大容量データ、読み取りの高速化などに対応しています。一部が汚損して読み取れなくてもデータを復元することができます。



> CODE39

信頼性が高いシンボルであることから、主に産業分野の作業指示票や現品ラベルに使用されています。



> JAN8/JAN13

世界共通コードで生活用品のほぼ全てにマーキングされています。(書籍、雑誌業界、家電業界、アパレル業界などでも利用されています)



> NW-7

血液銀行、宅配便の伝票、DPE、図書館の貸し出し管理などで利用されています。



> ITF

標準物流コードとして、JIS化されています。



> CUSTOMER_BCODE

郵便事業の効率化を目的とした郵便物に印字するバーコードです。



> CODE128

アスキーコード128文字(数字、アルファベット大文字/小文字、記号、制御コード)全てをバーコード化することができます。([ESC], [STX], [ETX], [CR], [LF]などの制御コードもコード化できます)



> PDF417

PDF417 バーコードは高い情報量とエラー訂正機能を持つ複数行の可変長のコードです。各業界で利用され始めています。



> GS1 DataBar (RSS)

医薬品、医療材料などの小物商品の管理を目的とした省スペースのバーコードです。EAN UCC シンボルの派生バージョンで、14桁のグローバル・トレード・アイテム・ナンバー (GTIN) の値で構成されています。



Ⅶ. その他

【動作環境について】

- ＞ 「KDbat1_scan(OCR)」、「KDbat2_scan(OCR)」、「KDbat1_opt_scan(OCR)」、
「KDbat2_opt_scan(OCR)」を利用するにあたりTWAINに対応したスキャナが必要になります。
TWAINに対応しているスキャナであればどのメーカーのスキャナからでも処理が可能です。

対応OS

- ・ Microsoft Windows 2000(日本語版)が動作する環境
- ・ Microsoft Windows XP(日本語版)が動作する環境
- ・ Microsoft Windows Vista(日本語版)が動作する環境

【価格】

- ◇ 製品名：KDbat1_scan(OCR) (Light版)
1次元バーコードの読み取り + OCR処理
・ 販売価格：¥300,000(税別) / 1ライセンス

- ◇ 製品名：KDbat2_scan(OCR) (Standard版)
1次元バーコード, 2次元コード(QR, PDF417) + OCR処理
・ 販売価格：¥350,000(税別) / 1ライセンス

- ◇ 製品名：KDbat1_opt_scan(OCR) (Light版 + オプション)
1次元バーコードの読み取り + キーワード別配信・
ファイル名詳細・階層出力・比較 + OCR処理
・ 販売価格：¥350,000(税別) / 1ライセンス

- ◇ 製品名：KDbat2_opt_scan(OCR) (Standard版 + オプション)
1次元バーコード, 2次元コード(QR, PDF417) + キーワード別配信・
ファイル名詳細・階層出力・比較オプション + OCR処理
・ 販売価格：¥400,000(税別) / 1ライセンス