

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2010-510563

(P2010-510563A)

(43) 公表日 平成22年4月2日(2010.4.2)

(51) Int.Cl.			F I			テーマコード (参考)		
<b>G06F</b>	<b>17/21</b>	<b>(2006.01)</b>	G06F	17/21	530A	5B109		
<b>H04N</b>	<b>1/00</b>	<b>(2006.01)</b>	H04N	1/00	C	5C062		
<b>G06Q</b>	<b>10/00</b>	<b>(2006.01)</b>	G06F	19/00	300B			
<b>G06F</b>	<b>3/12</b>	<b>(2006.01)</b>	G06F	3/12	N			

審査請求 未請求 予備審査請求 未請求 (全 17 頁)

(21) 出願番号 特願2009-536703 (P2009-536703)  
 (86) (22) 出願日 平成19年11月7日 (2007.11.7)  
 (85) 翻訳文提出日 平成21年5月13日 (2009.5.13)  
 (86) 国際出願番号 PCT/EP2007/061965  
 (87) 国際公開番号 W02008/058871  
 (87) 国際公開日 平成20年5月22日 (2008.5.22)  
 (31) 優先権主張番号 0622863.9  
 (32) 優先日 平成18年11月16日 (2006.11.16)  
 (33) 優先権主張国 英国 (GB)

(71) 出願人 390009531  
 インターナショナル・ビジネス・マシーンズ・コーポレーション  
 INTERNATIONAL BUSINESS MACHINES CORPORATION  
 アメリカ合衆国10504 ニューヨーク州 アーモンク ニュー オーチャードロード  
 (74) 代理人 100108501  
 弁理士 上野 剛史  
 (74) 代理人 100112690  
 弁理士 太佐 種一  
 (74) 代理人 100091568  
 弁理士 市位 嘉宏

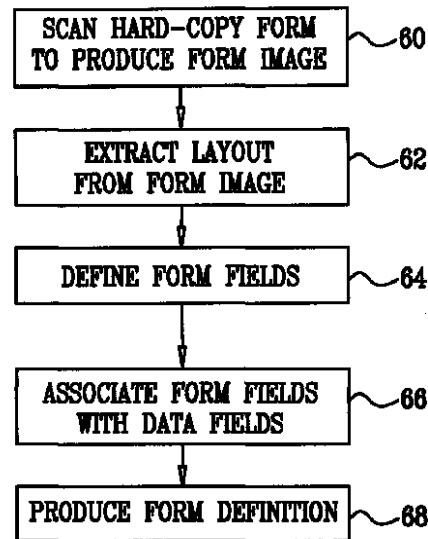
最終頁に続く

(54) 【発明の名称】 ハード・コピーの書式からの書式定義の自動発生

(57) 【要約】

書式発生のための、コンピュータによって実施される方法は、ハード・コピーの書式の画像を捕捉することと、この画像を自動的に処理して画像内の書式フィールドおよびこの書式フィールドの各々に関連付けられたテキスト文字を識別することと、を含む。フィールドに情報を入力するための各記入領域を定義する書式フィールドの幾何学的座標を求める。書式フィールド名を識別するために、テキスト文字に光学的文字認識(OCR)を適用する。書式フィールド名と対応するデータ・オブジェクトのオブジェクト名との関連付けを決定する。書式フィールドの記入領域の幾何学的座標を、書式フィールドに対応するデータ・オブジェクトのオブジェクト名と組み合わせて、書式定義を発生する。

【選択図】 図2



**【特許請求の範囲】****【請求項 1】**

書式発生のための、コンピュータによって実施される方法であって、

ハード・コピーの書式の捕捉画像を自動的に処理して、前記画像における書式フィールドおよび前記書式フィールドの各々に関連付けられたテキスト文字を識別するステップと、

前記フィールドに情報を入力するための各記入領域を定義する前記書式フィールドの幾何学的座標を求めるステップと、

書式フィールド名を識別するために、前記テキスト文字に光学的文字認識（OCR）を適用するステップと、

前記書式フィールド名と対応するデータ・オブジェクトのオブジェクト名との関連付けを決定するステップと、

前記書式フィールドの前記記入領域の前記幾何学的座標を、前記書式フィールドに対応する前記データ・オブジェクトの前記オブジェクト名と組み合わせて、書式定義を発生するステップと、

を含む、前記方法。

**【請求項 2】**

前記書式定義における前記決定した関連付けに応答してデータ・レポジトリから前記データ・オブジェクトの値を自動的に読み取るステップと、前記値が前記書式フィールドの前記記入領域に記入された前記書式のコピーを出力するステップと、を含む、請求項 1 に記載の方法。

**【請求項 3】**

前記書式の前記コピーを出力するステップが、前記コピーをプリントすること、前記コピーをファクシミリによって送信すること、および前記コピーを電子メールによって送信することの少なくとも 1 つを含む、請求項 2 に記載の方法。

**【請求項 4】**

前記データ・レポジトリがエンタープライズ・リソース・プランニング（ERP）・システムのデータベースを含む、請求項 2 に記載の方法。

**【請求項 5】**

前記ハード・コピーの書式が第 1 の ERP システムによって生成され、前記書式定義が、前記第 1 の ERP システムに取って代わる第 2 の ERP システムの一部である、請求項 1 に記載の方法。

**【請求項 6】**

前記画像を自動的に処理するステップが、前記ハード・コピーの書式のレイアウトに関連した特徴を抽出すること、および前記特徴を前記書式定義に含ませることを含む、請求項 1 に記載の方法。

**【請求項 7】**

前記レイアウトに関連した特徴が、ロゴ、タイトル、フレーム、コーナー、陰影パターン、および線の少なくとも 1 つを含む、請求項 6 に記載の方法。

**【請求項 8】**

前記記入領域がテキスト入力領域およびチェック欄の少なくとも 1 つを含む、請求項 1 に記載の方法。

**【請求項 9】**

前記関連付けを決定するステップが、前記書式フィールド名を前記データ・オブジェクトの前記オブジェクト名と自動的に照合することを含む、請求項 1 に記載の方法。

**【請求項 10】**

書式発生システムであって、

ハード・コピーの書式の捕捉画像を自動的に処理して、前記画像における書式フィールドおよび前記書式フィールドの各々に関連付けられたテキスト文字を識別し、前記書式フィールドに情報を入力するための各入力領域を定義する前記書式フィールドの幾何学的

10

20

30

40

50

座標を求め、書式フィールド名を識別するために前記テキスト文字に光学的文字認識（OCR）を適用し、前記書式フィールド名と対応するデータ・オブジェクトのオブジェクト名との関連付けを決定し、前記書式フィールドの前記入力領域の前記幾何学的座標を前記書式フィールドに対応する前記データ・オブジェクトの前記オブジェクト名と組み合わせるよう構成されたプロセッサを含む、前記システム。

【請求項 1 1】

前記プロセッサが、前記書式定義における前記決定した関連付けに回答してデータ・レポジトリから前記データ・オブジェクトの値を自動的に読み取るよう構成され、更に、前記値が前記書式フィールドの前記記入領域に記入された前記書式のコピーを出力するよう構成された出力デバイスを含む、請求項 1 0 に記載のシステム。

10

【請求項 1 2】

前記出力デバイスが、プリンタ、ファクシミリ、および電子メール接続の少なくとも一つを含む、請求項 1 1 に記載のシステム。

【請求項 1 3】

前記データ・レポジトリがエンタープライズ・リソース・プランニング（ERP）・システムのデータベースを含む、請求項 1 1 に記載のシステム。

【請求項 1 4】

前記ハード・コピーの書式が第 1 の ERP システムによって生成され、前記プロセッサが、前記第 1 の ERP システムに取って代わる第 2 の ERP システムに前記書式定義を供給するよう構成されている、請求項 1 0 に記載のシステム。

20

【請求項 1 5】

前記プロセッサが、前記ハード・コピーの書式のレイアウトに関連した特徴を書式画像から抽出し、前記特徴を前記書式定義に含ませるよう構成されている、請求項 1 0 に記載のシステム。

【請求項 1 6】

前記レイアウトに関連した特徴が、ロゴ、タイトル、フレーム、コーナー、陰影パターン、および線の少なくとも一つを含む、請求項 1 5 に記載のシステム。

【請求項 1 7】

前記記入領域がテキスト入力領域およびチェック欄の少なくとも一つを含む、請求項 1 0 に記載のシステム。

30

【請求項 1 8】

前記プロセッサが、前記書式フィールド名を前記データ・オブジェクトの前記オブジェクト名と比較することによって前記関連付けを決定するよう構成されている、請求項 1 0 に記載のシステム。

【請求項 1 9】

書式発生のためのコンピュータ・ソフトウェアであって、プログラム命令が記憶されたコンピュータ読み取り可能媒体を含み、この命令が、コンピュータによって実行されると、請求項 1 から 9 のいずれかに記載されたステップを前記コンピュータに実行させる、前記コンピュータ・ソフトウェア。

【発明の詳細な説明】

40

【技術分野】

【0001】

本発明は、一般に、書式処理システムに関し、更に具体的には、書式定義の自動発生のための方法およびシステムに関する。

【背景技術】

【0002】

多くの団体および企業では、業務プロセスの一部として書式が利用される。当技術分野において、書式を自動的に処理するための様々な方法およびシステムが知られている。例えば、米国特許第 5, 228, 100 号は、書式表示に対する入力データを受け入れるためのシステムを記載している。このシステムは、スキャナを用いて書式文書の画像をスキ

50

ヤンし、書式画像を生成する。プロセッサが、この書式画像における文字および線のパターンを認識し、文書の物理的および論理的な構造データを生成する。プロセッサは、論理的な構造データに基づいて、書式表示にデータを入力するためのプログラムを自動的に生成する。システムは、物理的な構造データによって定義された書式表示を提示するための画面と、データ入力装置と、データが記入された書式表示に基づいてプリントアウト文書を生成するためのプリンタと、を含む。

【発明の概要】

【発明が解決しようとする課題】

【0003】

従って、本発明の1つの態様によって、書式発生のための、コンピュータによって実施される方法が提供される。

10

【課題を解決するための手段】

【0004】

この方法は、ハード・コピーの書式の捕捉画像を自動的に処理して、画像内の書式フィールドおよびこの書式フィールドの各々に関連付けられたテキスト文字を識別することを含む。フィールドに情報を入力するための各記入領域を定義する書式フィールドの幾何学的座標を求める。書式フィールド名を識別するために、テキスト文字に光学的文字認識(OCR)を適用する。書式フィールド名と対応するデータ・オブジェクトのオブジェクト名との関連付けを決定する。書式フィールドの記入領域の幾何学的座標を、書式フィールドに対応するデータ・オブジェクトのオブジェクト名と組み合わせて、書式定義を発生する。

20

【0005】

本発明の一実施形態によれば、この方法は、書式定義における決定した関連付けにตอบสนองしてデータ・レポジトリからデータ・オブジェクトの値を自動的に読み取ることと、この値が書式フィールドの記入領域に記入された書式のコピーを出力することと、を更に含む。

【0006】

本発明の別の実施形態によれば、書式のコピーを出力するステップは、コピーをプリントすること、コピーをファクシミリによって送信すること、およびコピーを電子メールによって送信することの少なくとも1つを含む。

30

【0007】

本発明の更に別の実施形態によれば、データ・レポジトリは、エンタープライズ・リソース・プランニング(ERP)・システムのデータベースを含む。

【0008】

本発明の一実施形態によれば、ハード・コピーの書式は第1のERPシステムによって生成され、書式定義は、第1のERPシステムに取って代わる第2のERPシステムの一部である。

【0009】

本発明の別の実施形態によれば、画像を自動的に処理するステップは、ハード・コピーの書式のレイアウトに関連した特徴を抽出することと、特徴を書式定義に含ませることと、を含む。

40

【0010】

本発明のこの実施形態によれば、レイアウトに関連した特徴は、ロゴ、タイトル、フレーム、コーナー、陰影パターン、および線の少なくとも1つを含む。

【0011】

本発明の一実施形態によれば、記入領域はテキスト入力領域およびチェック欄の少なくとも1つを含む。

【0012】

本発明の別の実施形態によれば、関連付けを決定するステップは、書式フィールド名をデータ・オブジェクトの前記オブジェクト名と自動的に照合することを含む。

50

## 【 0 0 1 3 】

本発明の第2の態様によって、書式発生システムが提供される。このシステムは、ハード・コピーの書式から捕捉した画像を自動的に処理して、画像における書式フィールドおよび書式フィールドの各々に関連付けられたテキスト文字を識別し、書式フィールドに情報を入力するための各入力領域を定義する書式フィールドの幾何学的座標を求め、書式フィールド名を識別するためにテキスト文字に光学的文字認識（OCR）を適用し、書式フィールド名と対応するデータ・オブジェクトのオブジェクト名との関連付けを決定し、書式フィールドの入力領域の幾何学的座標を書式フィールドに対応するデータ・オブジェクトのオブジェクト名と組み合わせる書式定義を発生するように構成されたプロセッサを含む。また、書式発生システムは、ハード・コピーの書式の画像を捕捉するように構成された画像捕捉デバイスも含むことができる。

10

## 【 0 0 1 4 】

本発明の第3の態様によって、書式発生のためのコンピュータ・ソフトウェアが提供される。このソフトウェアは、プログラム命令が記憶されたコンピュータ読み取り可能媒体を含む。この命令は、コンピュータによって実行されると、このコンピュータに、ハード・コピーの書式の捕捉画像を受け入れさせ、この画像を自動的に処理して画像内の書式フィールドおよび書式フィールドの各々に関連付けられたテキスト文字を識別させ、書式フィールドに情報を入力するための各入力領域を定義する書式フィールドの幾何学的座標を求めさせ、書式フィールド名を識別するためにテキスト文字に光学的文字認識（OCR）を適用させ、書式フィールド名と対応するデータ・オブジェクトのオブジェクト名との関連付けを決定させ、書式フィールドの入力領域の幾何学的座標を書式フィールドに対応するデータ・オブジェクトのオブジェクト名と組み合わせる書式定義を発生させる。

20

## 【 0 0 1 5 】

以下に示す本発明の実施形態の詳細な説明を図面と共に考慮することによって、本発明は十分に理解されるであろう。

## 【 図面の簡単な説明 】

## 【 0 0 1 6 】

【 図 1 】 本発明の一実施形態に従った、書式処理システムを概略的に示すブロック図である。

【 図 2 】 本発明の一実施形態に従った、ハード・コピーの書式から書式定義を発生させるための方法を概略的に示すフロー・チャートである。

30

【 図 3 】 本発明の一実施形態に従った、記入済み書式をプリントするための方法を概略的に示すフロー・チャートである。

【 図 4 】 本発明の一実施形態に従った、ハード・コピーの書式の抜粋を示す図である。

【 図 5 】 本発明の一実施形態に従った、プリントした記入済み書式の抜粋を示す図である。

【 図 6 】 本発明の一実施形態に従った、ハード・コピーの書式の別の抜粋を示す図である。

【 図 7 】 本発明の一実施形態に従った、プリントした記入済み書式の別の抜粋を示す図である。

40

## 【 発明を実施するための形態 】

## 【 0 0 1 7 】

本発明の実施形態は、既存のハード・コピーの書式に基づいて書式定義を自動発生させるための方法およびシステムを提供する。これらの方法およびシステムは、例えば、ある団体が、あるエンタープライズ・リソース・プランニング（ERP）・システムから別のものに切り換える場合に、または他のデータ処理アプリケーションにおいて、利用可能である。これについては以下で更に説明する。

## 【 0 0 1 8 】

ERPシステムを変更する場合、新しいERPシステムにおいて、この団体が用いる書式を定義する必要がある。書式の内容およびレイアウトの手作業による定義は、グラフィ

50

ック・エディタの利用によって助けられることもあるが、通常は退屈で、時間がかかり、誤りを犯しやすいプロセスである。これは、この団体が多数の複雑な書式を用いている場合は特に当てはまる。従って、人の関与を最小限に抑え、定義にかかる時間を短縮するために、書式定義プロセスを自動化することが望ましい。

【 0 0 1 9 】

ほとんどの場合、以前のERPシステムのハード・コピーの書式が利用可能である。本発明の実施形態は、利用可能なハード・コピーの書式を用いて、新しいERPシステムにおいて用いることができる同等の書式を定義する。

【 0 0 2 0 】

いくつかの実施形態においては、まず、既存のハード・コピーの書式をスキャンして、コンピュータ読み取り可能画像を生成する。書式発生装置によって、この画像のレイアウトを分析し、書式のフィールドを識別する。各書式フィールドごとに、通常は光学式文字認識(OCR)プロセスを用いて、フィールド名を抽出する。更に、フィールドの記入領域、すなわち情報が入力されることになる空白領域の幾何学的座標を識別する。いくつかの実施形態においては、ロゴ、フレーム、および他のオブジェクト等、書式のレイアウトに関連する特徴も画像から抽出する。

10

【 0 0 2 1 】

次いで、書式発生装置は、識別した書式フィールドを、ERPデータベース内のフィールド等の適切なデータ・オブジェクトとリンクさせる。例えば、書式発生装置は、画像から抽出した書式フィールド名と合致するオブジェクト名をデータベース内で見つけようとする。これに加えて、またはこれの代わりに、ユーザが手作業で、書式フィールドとデータ・オブジェクトとの間の関連付けを実行することができる。ユーザは、自動的な関連付けの結果を検証し、または、書式発生装置が合致するオブジェクト名を探し出すことができない場合は関連付けを手作業で実行するように要求され、あるいはその両方とすることができる。関連付けプロセスの最後には、各書式フィールドはERPデータベース内の各データ・オブジェクトに関連付けられる。

20

【 0 0 2 2 】

書式発生装置は、識別された書式フィールド名および記入領域位置を含む書式定義を生成し、これは更に、各書式フィールドと対応するデータ・オブジェクトとの間に見出された関連付けを含む。また、書式定義は、書式のレイアウトに関連した特徴も含む。ERPシステムは、この書式定義に従って、データベースの適切なデータ・フィールドからデータを検索することによって、これ以降、書式に記入し書式をプリントアウトすることができる。

30

【 0 0 2 3 】

図1は、本発明の一実施形態に従った書式処理システム20を概略的に示すブロック図である。システム20は、エンタープライズ・リソース・プランニング(ERP)・システム24を含む。これは、とりわけ、情報を処理し、記入済み書式すなわち情報が記入された書式をプリントアウトする。ERPシステム24がプリントした書式は、インボイス、受領書、注文書、見積書、および報告書等、いずれかの適切な記入済み書式を含むことができる。以下の説明では、単一種類の書式の処理に言及するが、システム20は通常、多数の種類の手書をサポートする。

40

【 0 0 2 4 】

書式に記入することができる情報は、ERPデータベース28等のデータ・レポジトリに記憶される。記入されてプリントされる各書式は、例えば、データベース28において多数のデータ・オブジェクトを保持するレコード、または異なるレコード内に保持されるデータ・オブジェクトの組み合わせに相当するものとする。データ・オブジェクトが含む情報は、書式の異なるフィールド内に現れる。データベース28内の各データ・オブジェクトはオブジェクト名を有し、更にこれは、データ・フィールドを記述する短いテキスト記述または1つ以上のキーワードあるいはその両方を含む場合がある。

【 0 0 2 5 】

50

ERPシステムは、書式定義またはテンプレートに従って書式を生成しプリントする。いくつかの実施形態においては、書式定義は、書式の物理的レイアウトおよびその内容の双方を規定する。書式定義は、様々な書式フィールドを規定する。定義は通常、各書式フィールドのレイアウトに関連した特性を規定する。この特性は、例えば、書式上のフィールド表題および記入領域の位置座標、または、フィールド表題および内容をプリントするために用いられるフォント、あるいはその両方等である。また、定義は、フレーム、コーナー、陰影パターン、および線等のフィールドに関連したグラフィック・オブジェクト、ならびにロゴおよび書式識別子等の書式全体の特性も規定することができる。特に、書式定義は、書式フィールドに記入される情報を保持するERPデータベース28内の各データ・オブジェクトに書式フィールドをリンクする関連付けまたはリンクを規定する。

10

**【0026】**

従って、記入済み書式をプリントアウトするために、ERPシステム24は、書式定義における関連付けによって規定された、データベース28内の適切なデータ・オブジェクトについて照会することによって、記入する情報を検索する。ERPシステムは、書式定義に定義されたレイアウトに従って記入済み書式を構築する。具体的には、検索したデータベース情報を、書式フィールドの適切な記入領域に挿入する。

**【0027】**

システムは、プリンタ等の出力デバイス32を用いて、記入済み書式36をプリントアウトする。あるいは、システムは、画像ファイルとして記入済み書式を生成し、これを電子メールを用いてその宛先に送信することができる。これ以外に、出力デバイス32がファクシミリを含むことも可能である。

20

**【0028】**

書式定義は、既存のハード・コピー書式40から導出される。書式40は、捕捉デバイス44によって、書式画像と称するコンピュータ読み取り可能画像に変換される。捕捉デバイス44は、スキャナ、ファクシミリ、または、ハード・コピーの書式をコンピュータ読み取り可能画像に変換するための他のいずれかの適切なデバイスを含むことができる。

**【0029】**

書式画像は、書式発生装置48によって処理される。書式発生装置は、書式画像を処理し、データベース28と相互作用し、更にユーザ・ワークステーション56を介してユーザ52と相互作用する場合もあり、これによって書式定義を生成する。書式定義は、ERPシステム24に供給され、システム24はこれを用いて記入済み書式をプリントアウトする。

30

**【0030】**

通例、書式発生装置48は汎用コンピュータを含む。これは、本発明に記載する機能を実行するようにソフトウェアにおいてプログラムされる。ソフトウェアは、例えばネットワークを介して電子形態でコンピュータにダウンロードすることができ、あるいは、CD-ROM等の有形媒体上でコンピュータに供給することができる。いくつかの実施形態においては、書式発生装置は、ERPシステム24と同じコンピューティング・プラットフォームを共有するソフトウェア・プロセスとして実施される。あるいは、書式発生装置48およびERPシステム24は、別個のコンピューティング・プラットフォームを用いることも可能である。

40

**【0031】**

図2は、本発明の一実施形態に従った、書式定義の自動発生のための方法を概略的に示すフロー・チャートである。この方法は、捕捉ステップ60において、捕捉デバイス44がハード・コピーの書式40をコンピュータ読み取り可能書式画像に変換することから開始する。

**【0032】**

レイアウト分析ステップ62において、書式発生装置48は書式画像のレイアウトを分析する。書式発生装置は、ロゴ、タイトル、フレーム、コーナー、および線等、書式の異なるレイアウト関連の特徴を抽出する。更に、書式発生装置は、画像内のテキスト領域す

50

なわちテキスト文字を含む領域を識別し、光学式文字認識（OCR）を用いてテキストを復号する。（以下の説明では、記入されていない書式としてハード・コピーの書式40に言及するが、本発明において記載する方法およびシステムは、ハード・コピーの記入済み書式にも適用可能である。例えば、書式フィールド名と記入された内容とを区別するようにOCRソフトウェアを構成することができる。）

【0033】

フィールド定義ステップ64において、書式発生装置48は、書式画像から抽出した情報を用いて、異なる書式フィールドを識別し、それらの特性を書式定義に定義する。具体的には、書式発生装置は、識別した各書式フィールドの記入領域の書式フィールド名および座標を識別する。記入領域は、例えば、テキスト情報を入力するための空白領域およびチェック欄を含むことができる。

10

【0034】

関連付けステップ66において、書式発生装置は、各書式フィールドをデータベース28内の各データ・オブジェクトに関連付ける。いくつかの実施形態においては、書式発生装置は、上述のステップ62においてOCRを用いて抽出した書式フィールド名を用いる。通常、オブジェクト名は、書式フィールドにある程度似ている。それらは同様のエンティティを記述するからである。従って、書式発生装置は、書式フィールドおよびデータ・オブジェクトの名前を照合することによって、それらに関連付けることができる。先に述べたように、オブジェクト名は追加のテキストまたはキーワードを含む場合があり、これによってうまく合致が見つかる可能性を高めることができる。

20

【0035】

あるいは、ユーザ52は、ワークステーション56を用いて手作業で書式フィールドをデータ・オブジェクトに関連付けることができる。これに加えて、またはこれの代わりに、ユーザは、書式発生装置が決定した自動関連付けの結果を検証することができる。ユーザが手作業の関連付けを実行した場合であっても、完全に手作業の書式定義プロセスと比較すると、書式定義にかかる時間が著しく短縮されることに留意すべきである。

【0036】

定義出力ステップ68において、書式発生装置48は書式定義を生成する。書式定義は、各書式フィールドの定義を含む。具体的には、書式定義は、書式フィールドとデータベース28内のデータ・オブジェクトとの間の関連付けを含む。書式発生装置は、書式定義をERPシステム24に供給する。

30

【0037】

図3は、本発明の一実施形態に従った、記入済み書式36をプリントするための方法を概略的に示すフロー・チャートである。この方法は、データベース照会ステップ80において、書式定義に規定された関連付けに従ってERPシステム24が適切なデータ・オブジェクトを検索することから開始する。

【0038】

書式構築ステップ82において、ERPシステムは、検索した情報を用いて記入済み書式を構築する。ERPシステムは、書式定義に規定されたレイアウト関連フィールド特性に従って、書式をレイアウトする。システムは、適切なデータ・オブジェクトから検索した情報を、各書式フィールドの記入領域に入力する。次いで、出力ステップ84において、ERPシステムは、出力デバイス32を用いて記入済み書式をプリントする。

40

【0039】

図4は、本発明の一実施形態に従った、ハード・コピーの書式の抜粋88を示す図である。本例において、抜粋は、健康保険請求書式の一部である。この抜粋は、患者の住所の番地、市、州、郵便番号、および電話番号を入力するための5個の書式フィールドを含む。

【0040】

抜粋88を分析する場合、書式発生装置48は、この5個のフィールドを識別する。各フィールドにおいて、書式発生装置は、書式フィールド名を含むテキスト領域92および

50



記入領域 9 6 を識別する。書式発生装置は、OCR を用いてテキスト領域 9 2 内のテキストを抽出し、記入領域 9 6 の座標を求める。書式発生装置は、以下の書式フィールド名および座標を生成する。

【表 1】

書式フィールド名	記入領域の座標
患者の住所 (番地)	87, 429, 615, 465
患者の住所 (市)	87, 516, 544, 548
患者の住所 (州)	555, 516, 615, 548
患者の住所 (郵便番号)	87, 516, 320, 612
患者の住所 (電話番号)	331, 516, 615, 612

10

## 【0041】

本例において、ERP データベース内の特定の患者のレコードは、以下のデータ・オブジェクトおよびオブジェクト名を含む。

【表 2】

ID	番地	市	州	郵便番号	電話番号
29875	123 メイン通り	ニューヨーク	ニューヨーク	67676	(212) 555-1234

## 【0042】

書式フィールド名およびオブジェクト名は同一でないことに留意すべきである。それにもかかわらず、書式発生装置は、書式フィールド名「患者の住所 (番地)」を、オブジェクト名「番地」に関連付けることができる。なぜなら、双方の名に文字列「番地」が現れるからである。他のフィールドについても、同様の関連付けを求めることができる。

20

## 【0043】

図 5 は、本発明の一実施形態に従った、プリントした記入済み書式の抜粋 100 を示す図である。記入済み書式は、上述した方法を用いて、図 4 のハード・コピーの書式に基づいて生成される。抜粋 100 を生成するために、ERP システム 24 は、データベース 28 から所望の患者のレコードを検索し、書式定義に従って書式をレイアウトする。具体的には、システムは、規定された関連付けに従って、レコードのデータ・アイテムを適切な記入領域に挿入する。

30

## 【0044】

図 6 は、本発明の一実施形態に従った、ハード・コピーの書式の抜粋 104 を示す図である。患者の個人的な現況を入力するためのフィールドの本例においては、書式発生装置によって、単一の書式フィールドが、テキスト領域 108 および記入領域 112 を有する 6 個のサブ・フィールドと見なされる。この場合の記入領域はチェック欄を含む。書式画像を分析した後、書式発生装置 48 は、以下の書式フィールド名および座標を識別する。

【表 3】

書式フィールド名	記入領域の座標
患者の現況 (独身)	200, 100, 220, 120
患者の現況 (既婚)	250, 100, 270, 120
患者の現況 (その他)	300, 100, 320, 120
患者の現況 (有業)	200, 170, 220, 190
患者の現況 (全日制の学生)	250, 170, 270, 190
患者の現況 (定時制の学生)	300, 170, 320, 190

40

## 【0045】

図 7 は、本発明の一実施形態による、上述の図 6 のハード・コピーの書式に基づいた、プリントされた記入済み書式の抜粋 112 を示す図である。抜粋 112 においては、データベース 28 から検索した情報に基づいて、適切なチェック欄が ERP システム 24 によ

50

ってチェックされている。

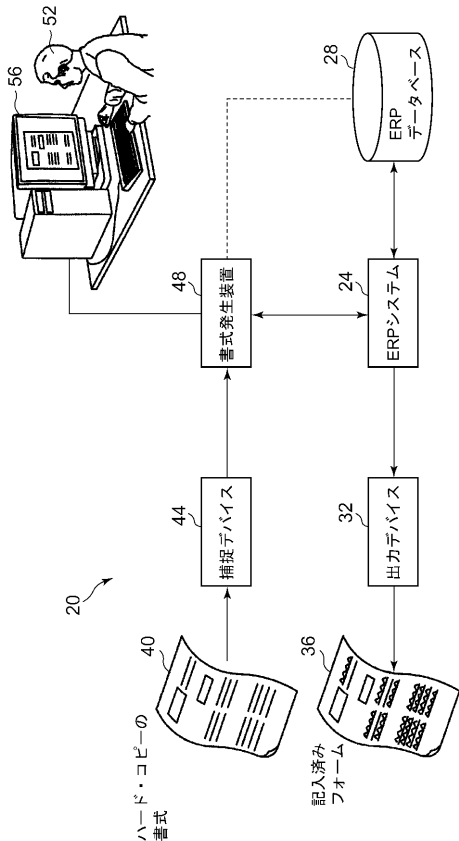
【0046】

本発明において記載した実施形態は、主に書式処理の用途に対応するが、本発明の原理は、喪失または破損した書式定義の回復および新しいレイアウトへの書式のカスタマイズまたは再定義等の用途にも使用可能である。また、本発明において記載した方法およびシステムは、新しい書式の設計または変更あるいはその両方のために使用可能である。例えば、紙の書式をスキャンして、既存のユーザ・プロファイルから自動的に抽出された情報を入力することができる。

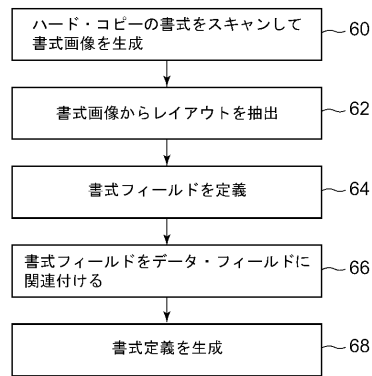
【0047】

従って、上述した実施形態は例示のために引用したに過ぎず、本発明はこれまで具体的に図示し記載したものに限定されないことは認められよう。本発明の範囲は、上述した様々な特徴の組み合わせおよび細分化した組み合わせの双方を含み、更に、これまでの記載を読むことで当業者に想起されるが従来技術には開示されていない変形および変更を含む。

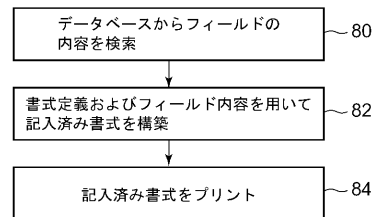
【図1】



【図2】



【図3】



【図4】

88

5. 患者の住所 (番地) ~92 96	
市 ~92 96	州 ~92
郵便番号 ~92 96	電話番号 (市外局番を含む) ~92 96

【図6】

104

8. 患者の現況			
108 独身 <input type="checkbox"/>	108 既婚 <input type="checkbox"/>	108 112 その他 <input type="checkbox"/>	
108 有業 <input type="checkbox"/>	108 全日制の学生 <input type="checkbox"/>	108 112 定時制の学生 <input type="checkbox"/>	

【図5】

100

5. 患者の住所 (番地)	
123 メイン通り	
市 ニューヨーク	州 ニューヨーク
郵便番号 67676	電話番号 (市外局番を含む) (212)555-1234

【図7】

112

8. 患者の現況		
独身 <input type="checkbox"/>	既婚 <input checked="" type="checkbox"/>	その他 <input type="checkbox"/>
有業 <input checked="" type="checkbox"/>	全日制の学生 <input type="checkbox"/>	定時制の学生 <input type="checkbox"/>

【手続補正書】

【提出日】平成21年6月15日(2009.6.15)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

書式発生のための、コンピュータによって実施される方法であって、

ハード・コピーの書式の捕捉画像を自動的に処理して、前記画像における書式フィールドおよび前記書式フィールドの各々に関連付けられたテキスト文字を識別するステップと、

前記フィールドに情報を入力するための各記入領域を定義する前記書式フィールドの幾何学的座標を求めるステップと、

書式フィールド名を識別するために、前記テキスト文字に光学的文字認識(OCR)を適用するステップと、

前記書式フィールド名と対応するデータ・オブジェクトのオブジェクト名との関連付けを決定するステップと、

前記書式フィールドの前記記入領域の前記幾何学的座標を、前記書式フィールドに対応する前記データ・オブジェクトの前記オブジェクト名と組み合わせて、書式定義を発生するステップと、

前記書式定義における前記決定した関連付けに回答してデータ・レポジトリから前記データ・オブジェクトの値を自動的に読み取るステップと、前記値が前記書式フィールドの前記記入領域に記入された前記書式のコピーを出力するステップと、

を含む、前記方法。

【請求項 2】

前記データ・レポジトリがエンタープライズ・リソース・プランニング（ERP）・システムのデータベースを含む、請求項1に記載の方法。

【請求項 3】

前記ハード・コピーの書式が第 1 の ERP システムによって生成され、前記書式定義が、前記第 1 の ERP システムに取って代わる第 2 の ERP システムの一部である、請求項1に記載の方法。

【請求項 4】

前記画像を自動的に処理するステップが、前記ハード・コピーの書式のレイアウトに関連した特徴を抽出すること、および前記特徴を前記書式定義に含ませることを含む、請求項1に記載の方法。

【請求項 5】

前記レイアウトに関連した特徴が、ロゴ、タイトル、フレーム、コーナー、陰影パターン、および線の少なくとも 1 つを含む、請求項4に記載の方法。

【請求項 6】

前記記入領域がテキスト入力領域およびチェック欄の少なくとも 1 つを含む、請求項 1 に記載の方法。

【請求項 7】

前記関連付けを決定するステップが、前記書式フィールド名を前記データ・オブジェクトの前記オブジェクト名と自動的に照合することを含む、請求項 1 に記載の方法。

【請求項 8】

書式発生システムであって、

ハード・コピーの書式の捕捉画像を自動的に処理して、前記画像における書式フィールドおよび前記書式フィールドの各々に関連付けられたテキスト文字を識別し、前記書式フィールドに情報を入力するための各記入領域を定義する前記書式フィールドの幾何学的座標を求め、書式フィールド名を識別するために前記テキスト文字に光学的文字認識（OCR）を適用し、前記書式フィールド名と対応するデータ・オブジェクトのオブジェクト名との関連付けを決定し、前記書式フィールドの前記入力領域の前記幾何学的座標を前記書式フィールドに対応する前記データ・オブジェクトの前記オブジェクト名と組み合わせる書式定義を発生し、前記書式定義における前記決定した関連付けに応答してデータ・レポジトリから前記データ・オブジェクトの値を自動的に読み取るように構成されたプロセッサを含み、更に、前記値が前記書式フィールドの前記記入領域に記入された前記書式のコピーを出力するように構成された出力デバイスを含む、前記システム。

【請求項 9】

前記データ・レポジトリがエンタープライズ・リソース・プランニング（ERP）・システムのデータベースを含む、請求項8に記載のシステム。

【請求項 10】

前記ハード・コピーの書式が第 1 の ERP システムによって生成され、前記プロセッサが、前記第 1 の ERP システムに取って代わる第 2 の ERP システムに前記書式定義を供給するように構成されている、請求項8に記載のシステム。

【請求項 11】

前記プロセッサが、前記ハード・コピーの書式のレイアウトに関連した特徴を書式画像から抽出し、前記特徴を前記書式定義に含ませるように構成されている、請求項8に記載のシステム。

【請求項 12】

前記レイアウトに関連した特徴が、ロゴ、タイトル、フレーム、コーナー、陰影パターン、および線の少なくとも 1 つを含む、請求項11に記載のシステム。

【請求項 13】

前記記入領域がテキスト入力領域およびチェック欄の少なくとも 1 つを含む、請求項8

に記載のシステム。

【請求項 14】

前記プロセッサが、前記書式フィールド名を前記データ・オブジェクトの前記オブジェクト名と比較することによって前記関連付けを決定するように構成されている、請求項8に記載のシステム。

【請求項 15】

書式発生のためのコンピュータ・ソフトウェアであって、プログラム命令が、請求項 1 から7のいずれかに記載された方法の各ステップを前記コンピュータに実行させる、前記コンピュータ・ソフトウェア。

## 【 国際調査報告 】

## INTERNATIONAL SEARCH REPORT

 International application No  
 PCT/EP2007/061965

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> INV. G06K9/20		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) G06K		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 407 935 A (HITACHI LTD [JP]; HITACHI SOFTWARE ENG [JP]) 16 January 1991 (1991-01-16) abstract; figures 1-106 columns 1-4 column 25, line 43 - column 29, line 50 column 36, line 4 - column 40, line 57 column 41, line 13 - column 42, line 27	1-19
X	WO 96/13009 A (HO JANET CHUNG KONG [US]) 2 May 1996 (1996-05-02) abstract; figures 1-11 pages 2,3 page 8, line 17 - page 10, line 30 -/--	1-19
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C.		
<input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents:		
*A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. *Z* document member of the same patent family		
Date of the actual completion of the international search 11 February 2008		Date of mailing of the international search report 18/02/2008
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer Alecu, Teodor Iulian

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2007/061965

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 625 465 A (LECH ROBERT [US] ET AL) 29 April 1997 (1997-04-29) abstract; figures 1-10 the whole document -----	1-19
A	WO 98/47098 A (READSOFT AB [SE]; ANDERSSON JAN [SE]) 22 October 1998 (1998-10-22) the whole document -----	1-19
A	US 2002/106128 A1 (ZLOTNICK AVIAD [IL]) 8 August 2002 (2002-08-08) the whole document -----	1-19

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2007/061965

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
EP 0407935	A	16-01-1991	DE 69033312 D1	11-11-1999
			JP 3179570 A	05-08-1991
			JP 2000155795 A	06-06-2000
			US 5228100 A	13-07-1993
WO 9613009	A	02-05-1996	AU 685337 B2	15-01-1998
			AU 4133996 A	15-05-1996
			CA 2201716 A1	02-05-1996
			EP 0799454 A1	08-10-1997
			JP 10507857 T	28-07-1998
			US 5619708 A	08-04-1997
US 5625465	A	29-04-1997	NONE	
WO 9847098	A	22-10-1998	AT 247306 T	15-08-2003
			AU 6861798 A	11-11-1998
			DE 69817171 D1	18-09-2003
			DE 69817171 T2	17-06-2004
			DK 976092 T3	08-12-2003
			EP 0976092 A1	02-02-2000
			ES 2207824 T3	01-06-2004
			PT 976092 T	31-12-2003
			SE 511242 C2	30-08-1999
			SE 9701183 A	02-10-1998
			US 2002106128	A1
EP 1358622 A1	05-11-2003			
WO 02063546 A1	15-08-2002			



## フロントページの続き

(81)指定国 AP(BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW

(72)発明者 ナヴォン、ヤーコヴ

イスラエル国 4 0 6 9 6 アイン・ヴェレド ピーオービー 1 9 0

(72)発明者 ショットランド、アリ

イスラエル国 3 4 9 8 0 ハイファ アントヴァーパン・ストリート 9

Fターム(参考) 5B109 NA01 NA04 ND02

5C062 AA02 AA05 AA30 AB02 AB07 AB17 AB22 AB38 AC04 AC34  
AC58 BD09