

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5324669号
(P5324669)

(45) 発行日 平成25年10月23日(2013.10.23)

(24) 登録日 平成25年7月26日(2013.7.26)

(51) Int.Cl. F I
G O 6 K 9/03 (2006.01) G O 6 K 9/03 Z

請求項の数 22 (全 17 頁)

(21) 出願番号	特願2011-548186 (P2011-548186)	(73) 特許権者	505281067
(86) (22) 出願日	平成22年1月25日 (2010.1.25)		グーグル インコーポレイテッド
(65) 公表番号	特表2012-516508 (P2012-516508A)		GOOGLE INC.
(43) 公表日	平成24年7月19日 (2012.7.19)		アメリカ合衆国 カリフォルニア州 94
(86) 国際出願番号	PCT/US2010/021965		043 マウンテン ビュー アンフィシ
(87) 国際公開番号	W02010/088182		アター パークウェイ 1600
(87) 国際公開日	平成22年8月5日 (2010.8.5)	(74) 代理人	100077539
審査請求日	平成25年1月23日 (2013.1.23)		弁理士 飯塚 義仁
(31) 優先権主張番号	61/147,901	(74) 代理人	100114742
(32) 優先日	平成21年1月28日 (2009.1.28)		弁理士 林 秀男
(33) 優先権主張国	米国 (US)	(74) 代理人	100125265
(31) 優先権主張番号	12/366,547		弁理士 貝塚 亮平
(32) 優先日	平成21年2月5日 (2009.2.5)		
(33) 優先権主張国	米国 (US)		
早期審査対象出願		最終頁に続く	

(54) 【発明の名称】 出版物からOCR認識されたテキストとそれに対応するイメージをクライアント装置において選択に表示すること

(57) 【特許請求の範囲】

【請求項 1】

出版物を表示するためにコンピュータによって実行される方法であって、
 出版物のソースイメージに対して施された光学式文字認識 (OCR) 処理に応じて生成された複数のテキストセグメントを含む文書を受信することと、

ユーザによって使用されるクライアント装置のディスプレイ上に1以上の前記テキストセグメントを表示することと、

表示されたテキストセグメントに対する前記ユーザによる選択に応じて、該選択されたテキストセグメントを前記OCR処理によって生成させた元である前記ソースイメージ内のイメージセグメントを自動的に取り出すことと、

前記クライアント装置の前記ディスプレイ上で、前記選択されたテキストセグメントを前記取り出されたイメージセグメントで置き換えること、
 を具備する方法。

【請求項 2】

更に、前記ユーザによる前記イメージセグメントの選択に応じて、前記クライアント装置の前記ディスプレイ上で、該選択されたイメージセグメントを前記テキストセグメントで置き換えること、
 を具備する請求項1の方法。

【請求項 3】

前記文書は、更に、それに関連するテキストセグメントの品質を示すがらくた度スコア

を含み、

前記 1 以上の前記テキストセグメントを表示することは、

前記関連するテキストセグメントの前記がらくた度スコアを或る閾値と比較することと、

前記閾値以下の前記がらくた度スコアに応じて、前記関連するテキストセグメントを表示することと、

前記閾値を超える前記がらくた度スコアに応じて、前記クライアント装置の前記ディスプレイ上で、前記関連するテキストセグメントを、該関連するテキストセグメントを前記 O C R 処理によって生成させた元であるイメージセグメントで、自動的に置き換えること、

10

を含む請求項 1 の方法。

【請求項 4】

前記文書は、前記複数のテキストセグメントを前記ソースイメージ内の複数のイメージセグメントに関連付ける位置情報を含み、

前記イメージセグメントを取り出すことは、

前記表示されたテキストセグメントについての前記文書内での位置情報を特定することと、

前記イメージセグメントを要求するリクエストであって、該特定された位置情報を含む該リクエストを送信すること、

を更に具備する請求項 1 の方法。

20

【請求項 5】

前記イメージセグメントを要求する前記リクエストは遠隔のサーバーに送信され、該イメージセグメントが該遠隔のサーバーから取り出される請求項 4 の方法。

【請求項 6】

前記位置情報は、前記表示されたテキストセグメントに含まれる文章を見せるソースイメージ内の領域を記述する情報である請求項 4 の方法。

【請求項 7】

更に、前記表示されているテキストセグメントの前記ユーザによる選択に応じて、該表示されているテキストセグメントを前記文書に関連付けて記憶すること、

を具備する請求項 1 の方法。

30

【請求項 8】

出版物を表示するために実行可能なコンピュータプログラムコードを記憶したコンピュータ読取可能な記憶媒体であって、該プログラムは、コンピュータに、

出版物のソースイメージに対して施された光学式文字認識 (O C R) 処理に応じて生成された複数のテキストセグメントを含む文書を受信する手順と、

ユーザによって使用されるクライアント装置のディスプレイ上に 1 以上の前記テキストセグメントを表示する手順と、

表示されたテキストセグメントに対する前記ユーザによる選択に応じて、該選択されたテキストセグメントを前記 O C R 処理によって生成させた元である前記ソースイメージ内のイメージセグメントを自動的に取り出す手順と、

40

前記クライアント装置の前記ディスプレイ上で、前記選択されたテキストセグメントを前記取り出されたイメージセグメントで置き換える手順、

を実行させることを特徴とするコンピュータ読取可能な記憶媒体。

【請求項 9】

前記プログラムは、更に、前記コンピュータに、前記ユーザによる前記イメージセグメントの選択に応じて、前記クライアント装置の前記ディスプレイ上で、該選択されたイメージセグメントを前記テキストセグメントで置き換える手順、

を実行させることを特徴とする請求項 8 のコンピュータ読取可能な記憶媒体。

【請求項 10】

前記文書は、更に、それに関連するテキストセグメントの品質を示すがらくた度スコア

50

を含み、

前記 1 以上の前記テキストセグメントを表示する前記手順は、

前記関連するテキストセグメントのための前記がらくた度スコアを或る閾値と比較する手順と、

前記閾値以下の前記がらくた度スコアに応じて、前記関連するテキストセグメントを表示する手順と、

前記閾値を超える前記がらくた度スコアに応じて、前記クライアント装置の前記ディスプレイ上で、前記関連するテキストセグメントを、該関連するテキストセグメントを前記 O C R 処理によって生成させた元であるイメージセグメントで、自動的に置き換える手順、

を前記コンピュータに実行させることからなる請求項 8 のコンピュータ読取可能な記憶媒体。

【請求項 1 1】

前記文書は、前記複数のテキストセグメントを前記ソースイメージ内の複数のイメージセグメントに関連付ける位置情報を含み、

前記イメージセグメントを取り出す前記手順は、

前記表示されたテキストセグメントについての前記文書内での位置情報を特定する手順と、

前記イメージセグメントを要求するリクエストであって、該特定された位置情報を含む該リクエストを送信する手順と、

を前記コンピュータに更に実行させることからなる請求項 8 のコンピュータ読取可能な記憶媒体。

【請求項 1 2】

出版物をユーザに提供するためにコンピュータによって実行される方法であって、

出版物のソースイメージに対して施された光学式文字認識 (O C R) 処理に応じて生成された複数のテキストセグメントを含む文書を受信することと、ここで、前記文書は、前記複数のテキストセグメントに対応する前記ソースイメージの複数のセグメントを特定する位置情報も含んでおり、

クライアント装置に前記文書を提供することと、

前記文書内に含まれる 1 テキストセグメントを前記 O C R 処理によって生成させた元である 1 イメージセグメントの前記位置情報を含むリクエストを前記クライアント装置から受信することに応じて、該 1 イメージセグメントを該クライアント装置に自動的に提供すること、

を具備し、前記クライアント装置に提供された前記 1 イメージセグメントは、該クライアント装置のディスプレイ上で前記 1 テキストセグメントと置き換えられるようになっている、方法。

【請求項 1 3】

前記文書を受信することは、更に、

それに関連するテキストセグメントの品質を示すがらくた度スコアを生成することと

、
前記関連するテキストセグメントと前記がらくた度スコアを含む前記文書を生成すること、

を含む請求項 1 2 の方法。

【請求項 1 4】

前記がらくた度スコアを生成することは、

1 つの言語モデルに基づいて前記関連するテキストセグメント中の各文字毎の言語条件付き文字確率の 1 組を計算することと、

前記関連するテキストセグメント中の各文字毎の前記言語条件付き文字確率の 1 組に基づいて、該関連するテキストセグメントのための前記がらくた度スコアを生成すること

、

10

20

30

40

50

を含む請求項 1 3 の方法。

【請求項 1 5】

更に、前記送信する文書内に制御モジュールを含めることを具備し、該制御モジュールは、前記クライアント装置において前記文書の表示を制御し、かつ、該クライアント装置のユーザとの相互作用に応じて、前記 1 テキストセグメントを前記 OCR 処理によって生成させた元である 1 イメージセグメントを要求するリクエストを自動的に生成するように、前記クライアント装置において実行されるものである、請求項 1 3 の方法。

【請求項 1 6】

前記位置情報は、前記テキストセグメント内に含まれる文章を提示する前記ソースイメージ内の領域を記述するものである、請求項 1 2 の方法。

【請求項 1 7】

出版物をユーザに提供するために実行可能なコンピュータプログラムコードを記憶したコンピュータ読取可能な記憶媒体であって、該プログラムは、コンピュータに、

出版物のソースイメージに対して施された光学式文字認識 (OCR) 処理に応じて生成された複数のテキストセグメントを含む文書を受信する手順と、ここで、前記文書は、前記複数のテキストセグメントに対応する前記ソースイメージの複数のセグメントを特定する位置情報も含んでおり、

クライアント装置に前記文書を提供する手順と、

前記文書内に含まれる 1 テキストセグメントを前記 OCR 処理によって生成させた元である 1 イメージセグメントの前記位置情報を含むリクエストを前記クライアント装置から受信することに応じて、該 1 イメージセグメントを該クライアント装置に自動的に提供する手順、

を実行させることを特徴とし、前記クライアント装置に提供された前記 1 イメージセグメントは、該クライアント装置のディスプレイ上で前記 1 テキストセグメントと置き換えられるようになっている、コンピュータ読取可能な記憶媒体。

【請求項 1 8】

前記文書を受信する前記手順は、更に、

それに関連するテキストセグメントの品質を示すならくた度スコアを生成する手順と、

前記関連するテキストセグメントと前記ならくた度スコアを含む前記文書を生成する手順、

を前記コンピュータに実行させることを特徴とする請求項 1 7 のコンピュータ読取可能な記憶媒体。

【請求項 1 9】

前記ならくた度スコアを生成する前記手順は、

1 つの言語モデルに基づいて前記関連するテキストセグメント中の各文字毎の言語条件付き文字確率の 1 組を計算する手順と、

前記関連するテキストセグメント中の各文字毎の前記言語条件付き文字確率の 1 組に基づいて、該関連するテキストセグメントのための前記ならくた度スコアを生成する手順、

を前記コンピュータに実行させることを特徴とする請求項 1 8 のコンピュータ読取可能な記憶媒体。

【請求項 2 0】

更に、前記送信する文書内に制御モジュールを含める手順を前記コンピュータに実行させ、該制御モジュールは、前記クライアント装置において前記文書の表示を制御し、かつ、該クライアント装置のユーザとの相互作用に応じて、前記 1 テキストセグメントを前記 OCR 処理によって生成させた元である 1 イメージセグメントを要求するリクエストを自動的に生成するように、前記クライアント装置において実行されるものである、請求項 1 8 のコンピュータ読取可能な記憶媒体。

【請求項 2 1】

前記取り出されたイメージセグメントが、前記クライアント装置の前記ディスプレイに基づいて調整される、請求項 1 の方法。

【請求項 2 2】

前記取り出されたイメージセグメントのサイズ及び解像度の少なくとも一方が、前記クライアント装置の前記ディスプレイの解像度に基づいて調整される、請求項 2 1 の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、光学式文字認識（OCR）、特に、OCRを使用して抽出されたテキスト、および、前記テキストの抽出元であるオリジナルのイメージ（画像）を表示する分野に関する。

10

【背景技術】

【0002】

印刷された文書が光学式文字認識（OCR）を使用してスキャンされ、編集可能なテキストに変換される。コンピュータ画面で文書を読む場合、ユーザは、典型的には、画像バージョンよりOCR認識されたバージョンを好む。OCR認識されたテキストは、文書画像に比べて、サイズが小さく、従って、より効率的にコンピュータネットワークを介して送信可能である。さらに、前記OCR認識されたテキストは、編集可能であり（例えば、コピーおよびペーストをサポートしており）また検索可能であり、（例えば、ローカルに利用可能なフォントを使用して）明瞭に、且つ、（コンピュータスクリーンに合うよう調整されたレイアウトを使用して）弾力的に表示可能である。上記利点は、携帯電話およびミュージックプレーヤー等の携帯装置を使って読書することを好むユーザにとって特に有益である。

20

【0003】

しかしながら、多くの場合、OCR認識されたテキストにはエラーが存在する。このようなエラーは、文書における欠陥、スキャン処理時に発生したアーチファクト（artifact：人口物）、および、OCRエンジンの欠点による場合がある。これらのエラーは、OCR認識されたテキストの使用および利点享受を妨害し、前記テキストの利点を損なうことがある。従って、OCR処理によって発生したエラーの影響を最小化しながら、OCR認識されたテキストを使用する利益を実現するための方法が要求されている。

30

【発明の概要】

【0004】

ここに開示する実施例は、OCRを使用して画像から抽出されたテキストを表示するための方法（及びそれに対応するシステムとコンピュータプログラム製品）を含んでいる。

【0005】

1つの観点においては、OCR認識された文書（document：ドキュメント）はOCR認識されたテキストの複数セグメントの集合として生成される。該文書内の各テキストセグメントには、イメージの連続からなる元の文書における元のテキストイメージにおける該テキストセグメントを内容とする矩形のイメージ区間を一意的に識別（同定）する、情報のタグが付けられる。該文書は、1テキストセグメントの表示を、ユーザの選択に応じて、OCR認識されたテキストとそれに対応するイメージ区間（イメージセグメント）との間で読み手がトグル式に切り換えることができるようにするプログラムコードを含んでいる。

40

【0006】

別の観点において、各テキストセグメント毎にがらくた度スコア（点数）が計算される。OCR認識された文書内の各テキストセグメントは、そのがらくた度スコアのタグが付けられる。OCR認識された文書がロードされるとき、埋め込まれたプログラムコードが各テキストセグメントのがらくた度スコアを或る閾値と比較する。或るテキストセグメントのがらくた度スコアが閾値よりも低いならば、該プログラムコードは該テキストセグメントを表示する。そうでない場合は、該プログラムコードは該テキストセグメントに代え

50

てイメージセグメントを表示する。ユーザはテキストセグメントを選択することにより、この表示をトグル式に切り換えることができる。

【 0 0 0 7 】

本明細書で説明する特徴および利点は、全てを包含しているとは限らず、特に、追加される多くの特徴および利点は、図面、明細書、および特許請求の範囲から当該技術に精通する者には明らかであろう。更に、本明細書で使用する言葉は、原則として可読性と説明目的のために選定されており、本発明の主題を描写し、または制限するよう選定したものではないことに留意されたい。

【図面の簡単な説明】

【 0 0 0 8 】

10

【図 1】本開示に係る一実施例に従うコンピューティング環境のハイレベルなブロック図。

【 0 0 0 9 】

【図 2】本開示に係る一実施例に従う図 1 に示すコンピューティング環境において使用される 1 台のコンピュータの一例を示すハイレベルなブロック図。

【 0 0 1 0 】

【図 3】本開示に係る一実施例に従う文書供給システム内の複数モジュール例を示すハイレベルなブロック図。

【 0 0 1 1 】

【図 4】本開示に係る一実施例に従う文書供給システムの動作例を示す流れ図。

20

【 0 0 1 2 】

【図 5】本開示に係る一実施例に従う文書供給システムによって生み出される制御モジュールの動作例を示す流れ図。

【 0 0 1 3 】

【図 6 A】本開示に係る一実施例に従う文書供給システムによって生成されるウェブページを読むユーザ体験を例示する一画面の図。

【図 6 B】同ウェブページの別画面の図。

【図 6 C】同ウェブページの別画面の図。

【発明を実施するための形態】

【 0 0 1 4 】

30

ここに説明するコンピューティング環境は、OCR 認識されたテキストを読む読み手が OCR 認識されたテキストのセグメントとテキストセグメントを含むソースイメージの 1 セグメントとの間でトグル式に切り換えることを可能にする。

【 0 0 1 5 】

図面及び以下の記述は単に説明の目的で或る実施例を記述している。ここに説明された構成及び方法についての変形がここに記述された原理から逸脱することなく実現され得ることを、以下の説明から当業者は容易に認識するであろう。いくつかの実施例に対する言及が詳細になされ、それらの一例が添付図面において示されている。どこで使用される場合であっても、類似のまたは似たような参照番号が図面で使用され、類似のまたは似たような機能性を示す。

40

システム環境

【 0 0 1 6 】

図 1 は、本開示に係る一実施例に従うコンピューティング環境 100 のハイレベルなブロック図であり、該コンピューティング環境 100 は、印刷された出版物 (publication: 公表物) を OCR 認識されたテキストに変換し、読み手が該 OCR 認識されたテキスト及び所望に応じてそれに対応するソース (元の) イメージを見れるようにする。図示するように、コンピューティング環境 100 は、スキャナー 110、OCR エンジン 120、文書供給システム (文書サーバーシステム) 130、クライアント装置 140 を含んでいる。本記述を単純化し明確化するために、各エンティティ (実体、装置) の 1 つのみが図示される。コンピューティング環境 100 においては、その他のエンティティ (実体、装

50

置)も同様に存在する。或る実施例において、OCRエンジン120及び文書供給システム130は1つのエンティティ(実体、装置)内に組み込まれる。

【0017】

スキャナー110は、印刷された出版物(例えば書籍、新聞)を光学的にスキャンし該印刷された出版物をデジタルのテキストイメージに変換する。スキャナー110の出力はOCRエンジン120に送られる。

【0018】

OCRエンジン120は、ソースイメージを編集可能なテキスト(以下、OCR認識されたテキストという)に変換(翻訳)するように構成されたハードウェア装置及び/又はソフトウェアプログラムである。OCRエンジン120は、コンピュータアルゴリズムを使用してソースイメージを処理し、それに対応するOCR認識されたテキストを生成する。

10

【0019】

加えて、OCRエンジン120は、ソースイメージ中の該OCR認識されたテキストに対応するイメージセグメントを記述する位置情報を生成し出力する。例えば、テキストの各セグメント(例えばパラグラフ、コラム、タイトル)毎に、OCRエンジン120は、該テキストセグメントに対応するソースイメージのセグメントをユニークに特定する境界ボックス(囲み)を記述する1組の値を提供する。この境界ボックスを記述する1組の値は、 $x-y$ 軸上の四角形の右上角の2次元座標値と、該四角形の幅及び高さを含む。従って、境界ボックスは、該テキストセグメントに対応するイメージセグメントとして、ソースイメージ中の領域をユニークに特定する。

20

【0020】

OCRエンジン120は、また、OCR認識されたテキストの品質を評価する信頼性レベルを生成してよい。加えて、OCRエンジン120は、フォーマット情報(例えばフォント種類、フォントサイズ、スタイルなど)のようなその他の情報を生成してよい。OCRエンジン120の一例は、ABBYYファインリーダーOCR(登録商標)、ADOBEアクロバットキャプチャ(登録商標)、MICROSOFTオフィス文書イメージング(登録商標)などである。OCRエンジン120の出力は文書供給システム130に送られる。

【0021】

30

文書供給システム130は、印刷された出版物の電子的表現物をユーザに提供するように構成されたコンピュータシステムである。文書供給システム130はOCRエンジン120から受信した情報を記憶する。この情報は、OCR認識されたテキスト、ソースイメージ、該ソースイメージ内のセグメントにOCR認識されたテキストのセグメントを関連付ける位置情報、及び信頼性レベルを含む。一実施例において、文書供給システム130は受信した情報を使用して、OCR認識されたテキストの各テキストセグメントの「がらくた度スコア」を計算する。この「がらくた度スコア」は全体の品質を評価するものである。加えて、文書供給システム130は、クライアント装置140によって実行されることが出来る制御モジュール132を含む。この制御モジュール132は、クライアント装置140のユーザがテキストセグメントとそれに対応するイメージセグメントとを選択的にトグル式に切り換えることができるようにするものであり、これにより、ユーザがOCR認識されたテキスト又は印刷された出版物のソースイメージ中の該テキストに対応する部分のいずれかを見ることができるようになる。

40

【0022】

一実施例において、文書供給システム130は、ユーザが、クライアント装置140を使用して、OCR認識された印刷出版物をウェブページとして読むためのウェブサイトを提供する。印刷された出版物の特定の部分についてのリクエストをクライアント装置から受信すると、文書供給システム130は、リクエストされた出版物の部分の内容とする文書(例えばウェブページ)を生成する。一実施例において、該文書はリクエストされた出版物の部分中に複数テキストセグメントを含む(例えば本の一章分のテキスト)。加えて

50

、該文書は、該テキストセグメントをそれに対応するイメージセグメントに関連付ける位置情報を含む。また、該文書は前記制御モジュール 1 3 2 を含む。文書供給システム 1 3 0 は、該生成した文書を、それをリクエストしているクライアント装置 1 4 0 に提供する。

【 0 0 2 3 】

クライアント装置 1 4 0 は、文書供給システム 1 3 0 に文書をリクエストし、その応答として受信した文書を表示するように構成されている。その機能は、クライアント装置 1 4 0 上で実行するウェブブラウザ（例えば、Microsoft Internet Explorer（登録商標）、Mozilla FireFox（登録商標）、Apple Safari（登録商標）など）のような閲覧アプリケーション 1 4 2 によって提供される。閲覧アプリケーション 1 4 2 は文書供給システム 1 3 0 から受信した文書に含まれる制御モジュール 1 3 2 を実行し、それはユーザがテキストセグメントの表示とそれに対応するイメージセグメントの表示とを交替でトグル式に切り換えることができるようにする。

【 0 0 2 4 】

スキャナー 1 1 0 は OCR エンジン 1 2 0 と通信可能に接続され、OCR エンジン 1 2 0 は文書供給システム 1 3 0 と通信可能に接続され、文書供給システム 1 3 0 はクライアント装置 1 4 0 と通信可能に接続される。これらの接続のいずれかが配線結線式又は無線式ネットワークであってよい。このネットワークの一例は、インターネット、イントラネット、WiFi ネットワーク、WiMAX ネットワーク、携帯電話ネットワーク、若しくはこれらの組み合わせを含む。

コンピュータアーキテクチャ

【 0 0 2 5 】

図 1 に示されたエンティティは 1 又は複数のコンピュータを使用して実装される。図 2 はコンピュータ 2 0 0 の一例を示すハイレベルなブロック図である。コンピュータ 2 0 0 はチップセット 2 0 4 に結合された少なくとも 1 つのプロセッサ 2 0 2 を含む。チップセット 2 0 4 は、メモリコントローラハブ 2 2 0 と入出力（I/O）コントローラハブ 2 2 2 を含む。メモリ 2 0 6 とグラフィックアダプタ 2 1 2 はメモリコントローラハブ 2 2 0 に結合され、ディスプレイ 2 1 8 はグラフィックアダプタ 2 1 2 に結合される。記憶装置 2 0 8、キーボード 2 1 0、ポインティングデバイス 2 1 4、ネットワークアダプタ 2 1 6 は I/O コントローラハブ 2 2 2 に結合される。コンピュータ 2 0 0 の別の実施例は、異なるアーキテクチャを持つ。

【 0 0 2 6 】

記憶装置 2 0 8 は、ハードディスクドライブ、コンパクトディスク読出し専用メモリ（CD-ROM）、DVD、ソリッドステートメモリデバイスなどのコンピュータ読取可能な記憶媒体である。メモリ 2 0 6 はプロセッサ 2 0 2 によって使用される命令及びデータを保持する。ポインティングデバイス 2 1 4 は、マウス、トラックボール、その他のタイプのポインティングデバイスであり、コンピュータシステム 2 0 0 にデータを入力するためにキーボード 2 1 0 と組み合わせて使用される。グラフィックアダプタ 2 1 2 はディスプレイ 2 1 8 上にイメージその他の情報を表示する。ネットワークアダプタ 2 1 6 は 1 又は複数のコンピュータネットワークにコンピュータシステム 2 0 0 を結合する。

【 0 0 2 7 】

コンピュータ 2 0 0 は、ここで述べる機能を提供するためにコンピュータプログラムモジュールを実行するようになっている。本書において、「モジュール」とは、特定の機能を提供するために使用されるコンピュータプログラム論理のことをいう。従って、モジュールは、ハードウェア内、ファームウェア内、及び/又はソフトウェア内に実装されることができる。一実施例において、プログラムモジュールは記憶装置 2 0 8 内に格納され、メモリ 2 0 6 上にロードされ、プロセッサ 2 0 2 によって実行される。

【 0 0 2 8 】

図 1 のエンティティによって使用されるコンピュータ 2 0 0 の種類は、実施形態及び該エンティティに要求される処理能力に依存して変り得る。例えば、文書供給システム 1 3

10

20

30

40

50

0 は、本書で述べる機能を提供するために協働するマルチ・ブレード・サーバーで構成されるかもしれない。別の例として、クライアント装置 1 4 0 は限られた処理能力を持つ携帯電話で構成されるかもしれない。コンピュータ 2 0 0 は、上述した構成要素のあるもの、例えばキーボード 2 1 0、グラフィックアダプタ 2 1 2 やディスプレイ 2 1 8 など、を欠いていてもよい。

文書供給システムのアーキテクチャ例の概観

【 0 0 2 9 】

図 3 は、一実施例に従う文書供給システム内の複数モジュール例を示すハイレベルなブロック図である。文書供給システムのいくつかの実施形態は、本書で述べるものとは異なる及び / 又は別のモジュールを有する。同様に、その機能は、本書で述べるものとは異なるやり方で、別の実施形態に従い複数モジュール間で分散化され得る。図示したように、文書供給システム 1 3 0 は、テキスト評価エンジン 3 1 0、コード発生モジュール 3 2 0、文書生成モジュール 3 3 0、入出力管理モジュール (以下 I / O モジュールという) 3 4 0、データ記憶部 3 5 0 を含む。

【 0 0 3 0 】

テキスト評価エンジン 3 1 0 は、OCR エンジン 1 2 0 によって提供された情報に基づいてテキストセグメントのがらくた度スコアを生成する。がらくた度スコアは該テキストセグメントの全体的な品質を評価する数値である。一実施例において、がらくた度スコアは 0 乃至 1 0 0 の範囲であり、0 は高いテキスト品質を示しており、1 0 0 は低いテキスト品質を示す。

【 0 0 3 1 】

がらくた度スコアを生成するために、テキスト評価エンジン 3 1 0 の一実施例は、テキストセグメント内の各文字 (キャラクタ) 毎に言語条件付き文字確率の 1 組を生成する。各言語条件付き文字確率は、該文字とテキストセグメントにおいて該文字に先行している 1 セットの文字群とが或る言語モデルに如何によく調和しているかを示す。該文字に先行している 1 セットの文字群は、典型的には、少数 (例えば 4 乃至 8 文字程度) に限定され、例えば複合語あるいはその他の結合語中の文字群が該モデルに基づき強い確率値が付与される。言語条件付き文字確率はテキスト品質のその他の指針 (例えば OCR エンジン 1 2 0 によって提供される前記信頼性レベル) と組み合わせられて、テキストセグメント内の各文字毎のテキスト品質スコア (点数) を生成するようにしてよい。そのような値の計算は、テキスト品質の位置特定分析を考慮に入れる。

【 0 0 3 2 】

テキスト評価エンジン 3 1 0 は、1 テキストセグメント内の複数文字に対応付けられた 1 組のテキスト品質スコアを組み合わせ、該テキストセグメントの品質を特徴付ける「がらくた度スコア」を生成する。テキスト評価エンジン 3 1 0 は、1 テキストセグメント内の複数文字に対応付けられた 1 組のテキスト品質スコアを平均化して「がらくた度スコア」を生成するようにしてもよい。

【 0 0 3 3 】

コード発生モジュール 3 2 0 は、クライアント装置 1 4 0 における文書の表示を制御する制御モジュール 1 3 2 を生成するか又は提供する。一実施例において、制御モジュール 1 3 2 は、JAVASCRIPT (登録商標)、JAVA (登録商標) 又は Perl (登録商標) のようなプログラム言語を使用して書かれたブラウザで実行可能なコードを使用して実装される。コード発生モジュール 3 2 0 は、Google Web Toolkit (登録商標) のようなアプリケーションを含むか又はそれと通信することができ、及び / 又は開発者が該制御モジュール 1 3 2 を開発できるようにする統合化開発環境 (IDE) を提供できる。実施例によっては、コード発生モジュール 3 2 0 は、制御モジュール 1 3 2 の予め作成されたインスタンスを格納する。該制御モジュール 1 3 2 はクライアント装置 1 4 0 に提供される文書内に含ませることができるか、若しくはクライアント装置 1 4 0 が文書供給システム 1 3 0 に文書をリクエストするとき該コード発生モジュール 3 2 0 がリアルタイムに制御モジュール 1 3 2 を形成することができる。

【 0 0 3 4 】

文書生成モジュール 3 3 0 は、リクエストしているクライアント装置 1 4 0 に出版物の部分を提供する文書を生成する。一実施例において、生成した文書はハイパーテキストランゲージ (H T M L) を使用して形成されたウェブページである。その他の実施例では、ポータブル文書フォーマット (P D F) のようなウェブページではない文書及び / 又は H T M L 以外の言語を使用して形成されたウェブページを生成する。

【 0 0 3 5 】

文書を生成するために、文書生成モジュール 3 3 0 は、クライアント装置 1 4 0 によってリクエストされた出版物と部分を特定し、該部分を構成するテキストセグメントをデータ記憶部 3 5 0 から取り出す。文書生成モジュール 3 3 0 は、複数のテキストセグメントを有する文書を作成し、また、該文書内の各テキストセグメント毎に当該テキストセグメントをソースイメージからのイメージセグメントに関連付ける位置情報を持つタグを付ける。文書生成モジュール 3 3 0 は、また、各テキストセグメント毎にそれに対応するがらくた度スコアを持つタグを付ける。加えて、文書生成モジュール 3 3 0 は、コード発生モジュール 3 2 0 によって提供された制御モジュール 1 3 2 を該文書内に埋め込む。文書生成モジュール 3 3 0 は、O C R 認識されたテキストが利用可能になるとき該文書を生成してよい。別の例として、文書生成モジュール 3 3 0 は、オンデマンドで (例えばクライアント装置 1 4 0 からのリクエストに応じて) 該文書を動的に生成してよい。

【 0 0 3 6 】

I / O モジュール 3 4 0 は、文書供給システム 1 3 0 の入力及び出力を管理する。例えば、I / O モジュール 3 4 0 は、O C R エンジン 1 2 0 から受信したデータをデータ記憶部 3 5 0 に記憶し、テキスト評価エンジン 3 1 0 を作動させてそれに対応する「がらくた度スコア」を生成させる。別の例として、I / O モジュール 3 4 0 はクライアント装置 1 4 0 からのリクエストを受信し、文書生成モジュール 3 3 0 を作動させてリクエストされた文書をそれに応じて提供させる。イメージセグメントを要求するリクエストを文書供給システムが受信すると、I / O モジュール 3 4 0 はデータ記憶部 3 5 0 から該イメージセグメントを取り出し、それをクライアント装置 1 4 0 に提供する。一実施例において、I / O モジュール 3 4 0 はイメージセグメントをクライアント装置 1 4 0 に返す前に該イメージセグメントを処理する。例えば、I / O モジュール 3 4 0 は、該文書を表示するクライアント装置 1 4 0 の画面の解像度に基づいてイメージセグメントのサイズ及び / 又は解像度を調整してよい。

【 0 0 3 7 】

データ記憶部 3 5 0 は文書供給システム 1 3 0 によって使用されるデータを記憶する。そのようなデータの一例は、O C R 認識されたテキスト及びそれに関連する情報 (例えば、がらくた度スコア、位置情報など)、ソースイメージ、生成された文書などを含む。データ記憶部 3 5 0 は関連データベース又は任意のその他のタイプのデータベースであってよい。

文書及び制御モジュール

【 0 0 3 8 】

一実施例によれば、文書供給システム 1 3 0 は制御モジュール 1 3 2 を埋め込んだ文書を生成する。文書は、それに対応するイメージセグメントを識別する情報のタグが付けられた複数テキストセグメントを含んでいる。これらのテキストセグメントは、また、ソースイメージにおける元の文章を模擬するようにデザインされたフォーマット情報のタグが付けられる。そのようなフォーマット情報はフォント種類、フォントサイズ、スタイル (例えばイタリック、ボールド、アンダーラインなど) を含む。

【 0 0 3 9 】

制御モジュール 1 3 2 の一実施例は、該文書に関連するイベントを取り扱うイベントハンドラーを含む。例えば、該文書がクライアント装置 1 4 0 でウェブブラウザにロードされことに応じて (オンロード・イベント)、制御モジュール 1 3 2 は H T M L テキストタグを使用してそこに含まれたテキストセグメントの表示を生成する。別の例では、ユーザ

による 1 テキストセグメントの選択に応じて、制御モジュール 132 は該テキストセグメントとそれに対応するイメージセグメントとの間で該表示をトグル式に切り換える。

【0040】

一実施例において、ウェブブラウザによってウェブページがロードされるとき、埋め込まれた制御モジュールが各テキストセグメントのがらくた度スコアを閾値と比較し、当該テキストセグメントが表示するのに十分な品質のものであるかを判定する。がらくた度スコアが該閾値以下であれば、該制御モジュールは下記のような HTML コードを使用して該テキストセグメントを表示する。

```
<p id='pageID.40.paralD.1.box.103.454.696.70.garbage.40'> <i>The courtyard of the Sheriff's house. A chapel. A shed in which is a blacksmith's forge with fire. A prison near which is an anvil, before which Will Scarlet is at work making a sword.</i></p>
```

10

上記 HTML コードは、次に示すイタリック体のテキストを含む。「The courtyard of the Sheriff's house. A chapel. A shed in which is a blacksmith's forge with fire.

A prison near which is an anvil, before which Will Scarlet is at work making a sword.」（その保安官の家の中庭。教会。その中の小屋は鍛冶屋の炎の鍛冶場。その近くの牢屋は鉄床で、その前でウィル・スカーレットが剣を作る仕事をしている。）このパラグラフは次に示す情報 “id=

pageID.40.paralD.1.box.103.454.696.70.garbage.40' ” のタグが付けられ、これは、対応するイメージセグメントが 40 ページ (pageID.40) のパラグラフ 1 (paralD.1) に位置し、該イメージセグメントの左上角が (103, 454)、該イメージセグメントが高さ 696 ピクセルで幅 70 ピクセルの長さ、それに関連する「がらくた度スコア」が 40 (garbage.40)、ということを示している。

20

【0041】

もし「がらくた度スコア」が閾値を超えていたら、制御モジュール 132 は、イメージセグメントを自動的に取り出し、次に示すような HTML コードを使用して、テキストセグメントの代わりに該イメージセグメントを表示する。

```
<p id='pageID.40.paralD.1.box.103.454.696.70.garbage.40'></p>
```

上記 HTML コードは、上記パラグラフと同じ文章を内容とするイメージセグメントを取り出し、テキストセグメントの代わりに該イメージセグメントを表示する。上記 bookID は、文書生成モジュール 330 によって該文書内にハード的にコード化 (hardcode) され得る。この閾値はユーザによって設定し得るか、又は該文書内にプリセットされ得る。

30

【0042】

ユーザは、また、該文書がテキストセグメント又はイメージセグメントを表示するかを特定することができる。例えば、ユーザはテキストセグメントをアクティブ化するためにキーボード又はポインティングデバイスを使用することができ、若しくはタッチ感知スクリーン上のテキストセグメントをタップすることができる。ユーザの選択に応じて、制御モジュール 132 は、テキストセグメントとそれに対応するイメージセグメントとの間で表示を動的にトグル式に切り換える。表示がテキストセグメントからイメージセグメントに切り換えられるとき、制御モジュール 132 は、該イメージセグメントをユニークに識別する情報 (例えばページ番号、パラグラフ番号、パインディングボックスなど) を用いて文書供給システム 130 に該イメージセグメントをリクエストし、ウェブページに該イメージセグメントのイメージタグを挿入し、OCR 認識されたテキストに代えて該イメージセグメントをユーザに表示する。表示されないとしても、ユーザがトグルバックしたときにそれに対応するテキストが容易に表示され得るように、テキストセグメントはローカル・バリアブル (変数) 内に記憶される。

40

【0043】

典型的には、イメージセグメントが表示されるとき、制御モジュール 132 はディスプレイが 100% であるように構成し、画面スクリーンの全幅を満たすようにイメージセグ

50

メントがサイズ修正されるべきであることを示す。しかし、1テキストセグメント（例えば短い言葉又は「第1章」のようなタイトルライン）が非常に短いとき（例えば1行の50%未満）、制御モジュールは画面スクリーン幅に対して同様の比率で該イメージを表示するように構成され得る。

文書供給システムのための方法の概観

【0044】

図4は、一実施例に従い、ユーザによって見るためにクライアント装置140に文書を相互作用的に提供するための文書供給システム130用の方法400を図示する流れ図である。その他の実施例として、該方法400の各ステップを図示とは異なる順序で実行することもできる。更に、その他の実施例として、本書で述べたものとは異なる及び/又は追加のステップを含んでいてもよい。文書供給システム130は、方法400の各ステップの複数のインスタンスを同時に及び/又は並行して実行することができる。

10

【0045】

最初に、文書供給システム130は、OCR認識されたテキスト、イメージ、関連情報（例えば位置情報、信頼性レベル）をOCRエンジン120から受信する（410）。文書供給システム130は、各OCR認識されたテキストセグメント毎のがらくた度スコアを（例えばテキスト評価エンジン310を通して）計算し（420）、文書内に含まれるべき制御モジュール132を（例えばコード発生モジュール320を通して）生成する（430）。

【0046】

文書供給システム130は、出版物の一部分（例えば本の一章）のリクエストをクライアント装置140から受信し（440）、該リクエストされた部分を構成する複数テキストセグメントをデータ記憶部350から取り出し、該テキストセグメントを含むウェブページのような文書を生成する（450）。これらのテキストセグメントには位置情報及びがらくた度スコアを含む関連属性のタグが付けられる。生成した文書はまた制御モジュール132を含む。文書供給システム130は、生成した文書を、それをリクエストしたクライアント装置140に送信する（460）。

20

【0047】

上述のように、ユーザは対応するテキストセグメントに代えてイメージセグメントを見るために、該文書と相互作用することができる。クライアント装置140で実行される制御モジュール132は或るイメージセグメントを表示するようリクエストを受信すると、該イメージセグメントをユニークに識別（特定）するパラメータを伴ってイメージリクエストを文書供給システム130に送信する。文書供給システム130は、該イメージリクエストを受信し（470）、該リクエストされたイメージセグメントを取り出し（480）、表示のためにクライアント装置140に送信する（490）。該イメージリクエストは、該文書を表示する画面スクリーンの解像度のような付加的情報を提供してもよい。文書供給システム130は、そのような情報に基づいてイメージセグメントを処理し（例えばサイズ修正、解像度調整など）、その後、処理済のイメージセグメントを表示のためにクライアント装置140に送信する（490）ようにしてもよい。

30

制御モジュールのための方法の概観

40

【0048】

図5は、一実施例に従い、文書内に含まれる制御モジュール132の動作500を図示する流れ図である。制御モジュール132は、クライアント装置140での閲覧アプリケーション142（例えばウェブブラウザ）によって文書が表示されるときに、該アプリケーションによって実行される。別の実施例において、制御モジュール132の機能が、閲覧アプリケーション142それ自体によって提供される（例えばプラグイン・アプレットによって）。従って、制御モジュール132は、文書供給システム130によってクライアント装置140に送られる文書内に必ずしも含まれている必要はない。

【0049】

図示のように、文書がロードされるとき、制御モジュール132は文書の表示を生成す

50

る(510)。上述のように、制御モジュール132は各テキストセグメントのがらくた度スコアを閾値と比較し、該テキストセグメント又はそれに対応するイメージセグメントを表示するか同化を決定する。

【0050】

制御モジュール132は、表示されたセグメントへのユーザによる選択を監視し検出する(520)。制御モジュール132は、選択されたセグメントがテキストセグメントとして又はイメージセグメントとして現在表示されているかどうかを判定する(530)。表示されているセグメントがテキストセグメントであれば、制御モジュール132はそれに対応するイメージセグメントをリクエストし(540)、該リクエストしたイメージセグメントを受信し(550)、受信したイメージセグメントをテキストセグメントに代えて表示する(560)。そうでなければ、制御モジュール132はイメージセグメントのイメージタグをテキストセグメントで置き換える(570)。一実施例において、制御モジュール132は、表示されていないテキストセグメントを文書内にローカルに記憶しておき(例えば、ローカルのJavaScript(登録商標)バリアブル)、ユーザが次に表示をテキストに戻すよう切り換えるときに、文書供給システム130からテキストセグメントをリクエストして取り出す必要がないようにする。表示を切り換えた後、制御モジュール132は、ユーザ選択の監視に戻る。

実例

【0051】

図6A～6Cは、本開示に係る一実施例に従い、文書と相互作用するユーザ体験を例示する画面ショットを示す図である。この例においては、該文書はウェブページである。図6Aに示すように、アップル社(登録商標)のiPhone(登録商標)クライアント装置を使用して、ユーザが「クリスマス・キャロル：過去のクリスマスの幽霊であること」と題するOCR認識された本に関して生成されたウェブページを取り出している。このウェブページは本の120～130頁を内容とする。

【0052】

ユーザはパラグラフ610のイメージセグメントを見ることを望み、該パラグラフの表示をタップする(叩く)。これに応じて、制御モジュール132はパラグラフ610のテキストセグメントを図6Bに示すように割れ目イメージ620に置き換える。この割れ目イメージ620は、「元の本の画像を読み込み中... (直前の画像に戻るためにはこの画像をタップして下さい。)」という文章を見せる。この割れ目イメージ620は、どう戻るかについての簡潔なガイドを提供すると同時にユーザが動作を理解するのを手助けするようにデザインされている。例えば、クライアント装置140のネットワーク接続が貧弱な場合、パラグラフ610のオリジナルイメージを読み込むのに時間がかかるかもしれない。ユーザは割れ目イメージ620を叩いて動作をキャンセルしテキストセグメントの閲覧に戻ることができる。この割れ目イメージ620はまた感覚的な読み込み時間を減少させる手助けとなる。

【0053】

イメージセグメント630が取り出されるとき、制御モジュール132は図6Cに示すようにテキストセグメントに代えてイメージセグメント630に交換する。その後ユーザが再びタップすると図6Aに示すようにテキストセグメントに戻る。

【0054】

上記説明のいくつかの部分は、アルゴリズム的处理又は動作に関する実施例を述べている。これらのアルゴリズムの説明および表現は、当該分野に精通する者により使用されて、自らの業務内容を当該分野に精通する他の者に最も効率的に伝えるための手段である。これらの動作を、機能的、計算的または論理的に説明したが、これらは、プロセッサによって実行される命令群からなるコンピュータプログラム又はそれと等価の電子回路を介して実装されることが理解される。更に、モジュールとして、これら動作の構成に言及することが、一般性を失わせることなく、時には便利であると証明されてもいる。説明された動作及びこれらに関連するモジュールはソフトウェア。ファームウェアあるいはそれら

10

20

30

40

50

の組み合わせで実施され得る。

【 0 0 5 5 】

本書で使用する「一実施例」または「1つの実施例」とは、該実施例に関連した特定の要素、特徴、構造又は特性が少なくとも1つの実施例内に含まれることを意味する。本明細書における各所での「一実施例において」という表現の出現は、必ずしもそのすべてが同一の実施例を言及するものではない。

【 0 0 5 6 】

いくつかの実施例は、「結合された」又は「接続された」という表現を使用して述べられているかもしれない。これらの用語は、互いに類義語として意図されないことを理解されたい。例えば、いくつかの実施例は、2以上の要素が直接的に物理的に又は電氣的に互いに接触していることを示すために、「接続された」の語を使用して述べているかもしれない。別の実施例は、2以上の要素が直接的に物理的に又は電氣的に接触していることを示すために、「結合された」の語を使用して述べているかもしれない。しかし、「結合された」の語は、また、2以上の要素が直接的に互いに接触していないが、互いに協働又は相互作用していることを意味するものであってもよい。実施例はこれらの文脈に限定されない。

10

【 0 0 5 7 】

本書で使用されている「備える」「具備している」「含む」「含んでいる」「持つ」「持っている」の語又はその他の如何なるそれらの変形語は、非排他的な包摂をカバーしようとするものである。例えば、1組の要素を掲げてなるプロセス、方法、物又は装置は、必ずしもそれらの要素のみに限定されるものではなく、その他の明示的に掲げられていない要素又はそのようなプロセス、方法、物又は装置につきものその他の要素を含んでいてよい。更に、特に反対の意を表さない限り、「又は」とは、包括的な「又は」及び非排他的な「又は」のことである。例えば、A又はBという条件は、Aが真（又は存在する）かつBが偽（又は存在しない）、Aが偽（又は存在しない）かつBが真（又は存在する）、及びAもBも真（又は存在する）、のいずれか1つによって満足させられるものである。

20

【 0 0 5 8 】

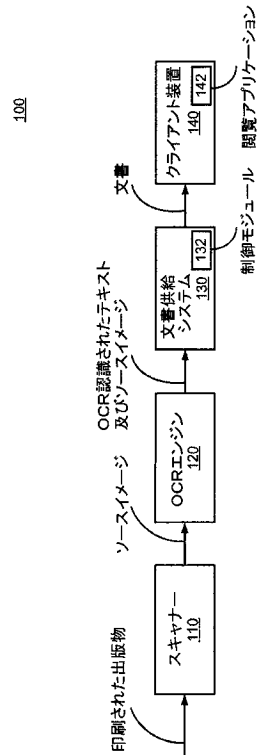
加えて、「1」又は「或る」の使用は、本書中の実施例の構成要素及び成分を記述するために使用される。これは、単に便宜と、開示の一般性を与えるためになされる。この記述は1又は少なくとも1つを含むと読まれるべきであり、また、単一のとは、それが明らかにその反対を意味していない限り、複数も含む。

30

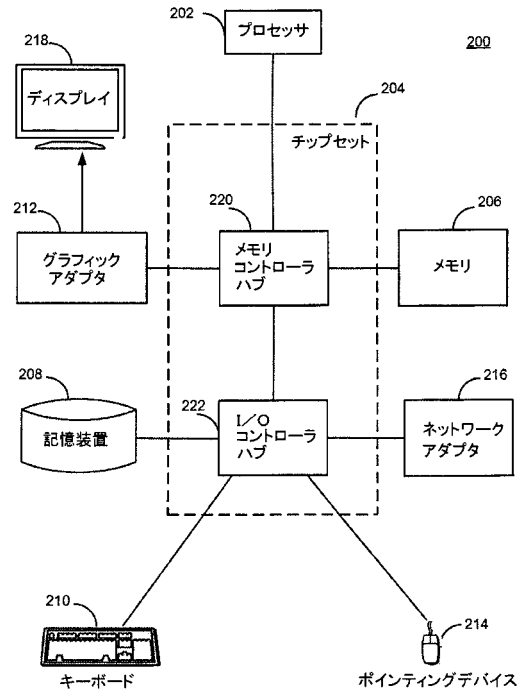
【 0 0 5 9 】

本書での開示を読解すると、この技術分野の熟練者は、OCR認識されたテキストを表示するためのシステム及び処理についての更に追加の変形構成及び機能的設計が認識できるであろう。従って、特定の実施例及び応用が図示され説明されているが、本発明は、ここに説明した構成及び要素に正確に限定されるものではなく、添付の請求の範囲で定義される精神と範囲から逸脱することなく、当業者にとって明白な様々な変形、変更、変化が、ここに開示した方法及び装置の構成、動作、詳細においてなされ得る。

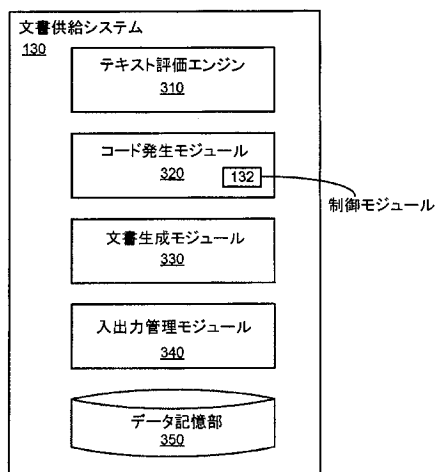
【図 1】



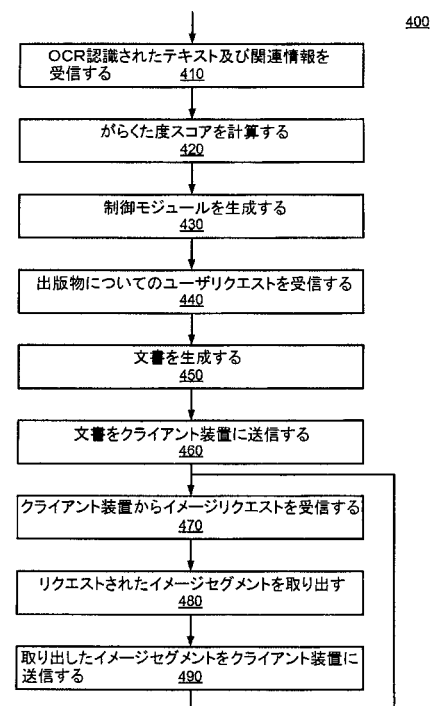
【図 2】



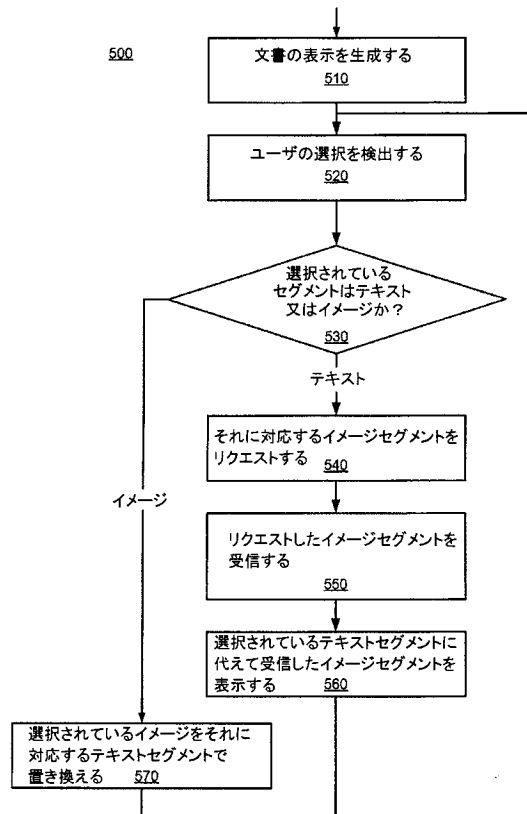
【図 3】



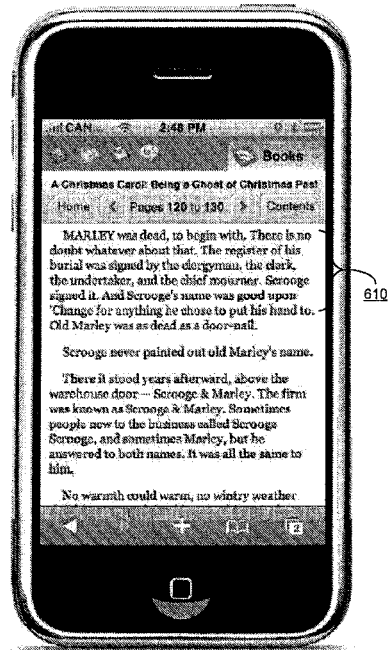
【図 4】



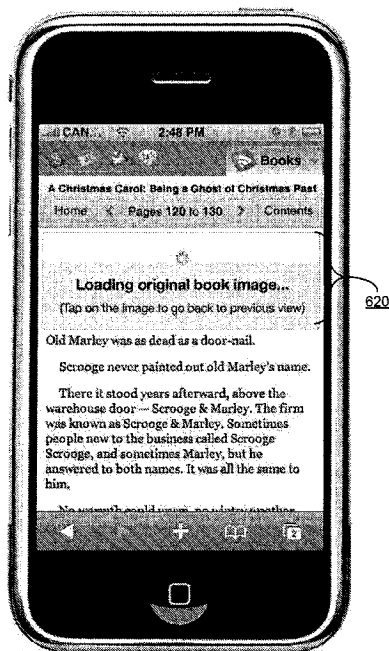
【図 5】



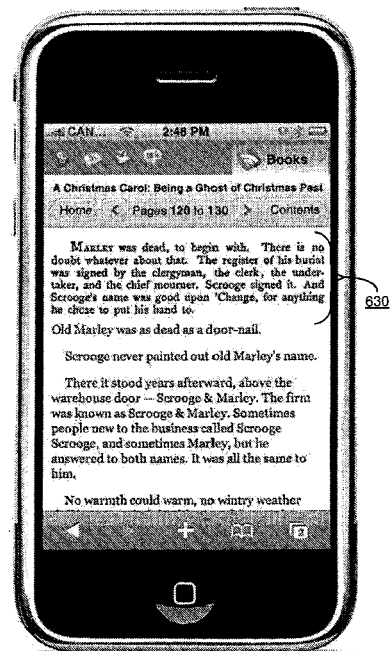
【図 6 A】



【図 6 B】



【図 6 C】



フロントページの続き

- (72)発明者 ラトナカール, ビレシュ
アメリカ合衆国 カリフォルニア州 94043 マウンテン ビュー アンフィシアター パー
クウェイ 1600, グーグル インコーポレイテッド内
- (72)発明者 ポパト, アショク
アメリカ合衆国 カリフォルニア州 94043 マウンテン ビュー アンフィシアター パー
クウェイ 1600, グーグル インコーポレイテッド内
- (72)発明者 ハウゲン, フランシス
アメリカ合衆国 カリフォルニア州 94043 マウンテン ビュー アンフィシアター パー
クウェイ 1600, グーグル インコーポレイテッド内

審査官 板垣 有紀

- (56)参考文献 特開2002-049890(JP, A)
特開2005-352735(JP, A)
特開平05-081467(JP, A)
特開平07-249098(JP, A)
特開2006-031299(JP, A)

- (58)調査した分野(Int.Cl., DB名)
G06K 9/03