

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5661663号
(P5661663)

(45) 発行日 平成27年1月28日 (2015. 1. 28)

(24) 登録日 平成26年12月12日 (2014. 12. 12)

(51) Int. Cl.

F I

G06T 11/60 (2006.01)
G09B 7/02 (2006.01)
G06K 9/20 (2006.01)
H04N 1/387 (2006.01)

G06T 11/60 100C
G09B 7/02
G06K 9/20 340L
H04N 1/387

請求項の数 3 (全 11 頁)

(21) 出願番号 特願2012-25664 (P2012-25664)
(22) 出願日 平成24年2月9日 (2012. 2. 9)
(65) 公開番号 特開2013-161463 (P2013-161463A)
(43) 公開日 平成25年8月19日 (2013. 8. 19)
審査請求日 平成26年1月17日 (2014. 1. 17)

(73) 特許権者 000006150
京セラドキュメントソリューションズ株式
会社
大阪府大阪市中央区玉造 1 丁目 2 番 2 8 号
(74) 代理人 100097113
弁理士 堀 城之
(74) 代理人 100162363
弁理士 前島 幸彦
(72) 発明者 南 貴博
大阪府大阪市中央区玉造 1 丁目 2 番 2 8 号
京セラミタ株式会社内

審査官 真木 健彦

最終頁に続く

(54) 【発明の名称】 情報抽出装置

(57) 【特許請求の範囲】

【請求項 1】

学習教材の画像データから情報を抽出し、抽出した情報に基づいて問題を生成する情報抽出装置であって、

前記画像データを解析することで、図面領域と文字領域とを特定すると共に、前記図面領域の周辺に配置された前記文字領域を図面周辺領域として特定する画像データ解析手段と、

前記図面領域、前記図面周辺領域及び前記文字領域のそれぞれについて、文字認識処理を行い、前記図面領域、前記図面周辺領域及び前記文字領域のそれぞれに含まれている文字列をテキストデータに変換する文字認識手段と、

前記図面領域及び前記図面周辺領域のテキストデータから図番及びキーワードを抽出するキーワード抽出手段と、

前記図番と関連づけて前記図面領域を前記画像データから図面イメージとして抽出する図面イメージ抽出手段と、

前記図面イメージ内の前記キーワードに対応する領域を消去する図面イメージ内キーワード消去手段と、

前記文字領域のテキストデータを検索し、前記図番が含まれる文を図面説明文として特定する文字領域検索手段と、

前記図番と関連づけて前記図面説明文に対応する文字列イメージを前記画像データの
前記文字領域から抽出する文字列イメージ抽出手段と、

10

20

前記図面説明文に前記キーワードが存在する場合、前記文字列イメージから前記キーワードに対応する領域を消去する文字列イメージ内キーワード消去手段と、

前記図面領域内のテキストデータと前記図面説明文とのいずれかに前記キーワードが存在する前記図番の前記図面イメージ及び前記文字列イメージを用いて、前記図面イメージ及び前記文字列イメージを前記図番毎にレイアウトした穴埋め問題を生成するレイアウト手段とを具備することを特徴とする情報抽出装置。

【請求項 2】

前記レイアウト手段は、前記図面領域内のテキストデータと前記図面説明文とのいずれにも前記キーワードが存在しない前記図番の前記図面イメージ及び前記文字列イメージを用い、複数の前記図面イメージ及び複数の前記文字列イメージをランダムな順序でレイ

10

【請求項 3】

前記文字領域検索手段は、前記図番が含まれる文と、該図番が含まれた文の前後の一文で且つ前記キーワードが含まれている文とを前記図面説明文として特定することを特徴とする請求項 1 又は 2 記載の情報抽出装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、教科書等の学習教材の画像データから情報を抽出し、抽出した情報に基づいて問題を生成する情報抽出装置に関する。

20

【背景技術】

【0002】

学校等の学習現場において、補助教材として教材出版社等から出版されたドリルや問題集が用いられているが、使用する問題集によっては、教科書の内容に即していないことがある。そこで、教科書等の学習教材から簡単に問題を生成する技術が提案されている。例えば特許文献 1 には、問題の解答となる文字列をマーキングで指定した問題原稿の画像を読み取り、抽出した文字列を用いて作成した解答集の電子データと、マーキングの位置の文字列を消去した問題文の電子データとを生成する技術が提案されている。

【先行技術文献】

【特許文献】

30

【0003】

【特許文献 1】特開 2007 - 4523 号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

近年の学習教材には、グラフや写真等の図面が多く採用されており、各種試験においても図面を用いたものが多くなっている。しかしながら、特許文献 1 の従来技術においては、文字列のみを対象に問題文を生成しているため、学習教材内にレイアウトされている図面に関する問題を生成することができないという問題点があった。また、従来技術においては、作成者が事前に文字列をマーキングしておかなければならず、繁雑な作業が必要となってしまう。

40

【0005】

本発明の目的は、上記問題点に鑑み、従来技術の問題を解決し、学習教材にマーキング等の準備することなく、図面がレイアウトされている学習教材の画像データを用いて、図面に関する問題を生成することができる情報抽出装置を提供することにある。

【課題を解決するための手段】

【0006】

本発明の情報抽出装置は、学習教材の画像データから情報を抽出し、抽出した情報にも基づいて問題を生成する情報抽出装置であって、前記画像データを解析することで、図面領域と文字領域とを特定すると共に、前記図面領域の周辺に配置された前記文字領域

50

を図面周辺領域として特定する画像データ解析手段と、前記図面領域、前記図面周辺領域及び前記文字領域のそれぞれについて、文字認識処理を行い、前記図面領域、前記図面周辺領域及び前記文字領域のそれぞれに含まれている文字列をテキストデータに変換する文字認識手段と、前記図面領域及び前記図面周辺領域のテキストデータから図番及びキーワードを抽出するキーワード抽出手段と、前記図番と関連づけて前記図面領域を前記画像データから図面イメージとして抽出する図面イメージ抽出手段と、前記図面イメージ内の前記キーワードに対応する領域を消去する図面イメージ内キーワード消去手段と、前記文字領域のテキストデータを検索し、前記図番が含まれる文を図面説明文として特定する文字領域検索手段と、前記図番と関連づけて前記図面説明文に対応する文字列イメージを前記画像データの文字領域から抽出する文字列イメージ抽出手段と、前記図面説明文に前記キーワードが存在する場合、前記文字列イメージから前記キーワードに対応する領域を消去する文字列イメージ内キーワード消去手段と、前記図面領域内のテキストデータと前記図面説明文とのいずれかに前記キーワードが存在する前記図番の前記図面イメージ及び前記文字列イメージを用いて、前記図面イメージ及び前記文字列イメージを前記図番毎にレイアウトした穴埋め問題を生成するレイアウト手段とを具備することを特徴とする。

10

さらに、本発明の情報抽出装置において、前記レイアウト手段は、前記図面領域内のテキストデータと前記図面説明文とのいずれにも前記キーワードが存在しない前記図番の前記図面イメージ及び前記文字列イメージを用い、複数の前記図面イメージ及び複数の前記文字列イメージをランダムな順序でレイアウトした選択問題を生成しても良い。

20

さらに、本発明の情報抽出装置において、前記文字領域検索手段は、前記図番が含まれる文と、該図番が含まれた文の前後の一文で且つ前記キーワードが含まれている文とを前記図面説明文として特定しても良い。

【発明の効果】

【0007】

本発明によれば、画像データを解析することで、図面領域、図面周辺領域及び文字領域をそれぞれ特定し、図面領域、図面周辺領域及び文字領域のそれぞれに含まれている文字列をテキストデータに変換し、図面領域及び図面周辺領域のテキストデータから図番及びキーワードを抽出し、図面領域を画像データから図面イメージとして抽出し、図面イメージ内のキーワードに対応する領域を消去し、文字領域のテキストデータから図番が含まれる文を図面説明文として特定し、図面説明文に対応する文字列イメージを画像データの文字領域から抽出し、文字列イメージからキーワードに対応する領域を消去し、図面イメージ及び文字列イメージをレイアウトした穴埋め問題を生成するように構成することにより、学習教材にマーキング等の準備をすることなく、図面がレイアウトされている学習教材の画像データを用いて、図面に関する問題を生成することができるという効果を奏する。

30

【図面の簡単な説明】

【0008】

【図1】本発明に係る情報抽出装置の実施の形態の構成を示すブロック図である。

【図2】本発明に係る情報抽出装置の実施の形態の問題頁生成動作を説明するためのフローチャートである。

40

【図3】本発明に係る情報抽出装置の実施の形態の問題頁生成動作を説明するための説明図である。

【図4】本発明に係る情報抽出装置の実施の形態の問題頁生成動作を説明するための説明図である。

【図5】本発明に係る情報抽出装置の実施の形態の問題頁生成動作を説明するための説明図である。

【図6】本発明に係る情報抽出装置の実施の形態の問題頁生成動作を説明するための説明図である。

【発明を実施するための形態】

50

【 0 0 0 9 】

次に、本発明の実施の形態を、図面を参照して具体的に説明する。

本実施の形態の情報抽出装置 1 0 0 は、パーソナルコンピュータ等の情報処理装置であり、図 1 を参照すると、問題頁生成制御部 1 と、操作部 2 と、画像データ読み取り部 3 と、記憶部 4 と、印字部 5 とがシステムバス 6 によって接続されている。

【 0 0 1 0 】

操作部 2 は、キーボード等の入力手段であり、画像データ読み取り部 3 による原稿の読み取り動作に係る各種指示入力、問題頁生成制御部 1 による問題頁生成動作に係る各種指示入力、印字部 5 による印字動作に係る各種指示入力等が行われる。

【 0 0 1 1 】

画像データ読み取り部 3 は、原稿をスキャンして画像データを取得するスキャナー装置である。画像データ読み取り部 3 によって取得された画像データは、記憶部 4 に記憶される。なお、画像データを取得する手段は、上記に限られず、インターネット等のネットワークと接続可能なインターフェース部経由で画像データを取得する手段であっても良く、フラッシュメモリや D V D 等の各種記録媒体経由で画像データを取得する手段であっても良い。

【 0 0 1 2 】

記憶部 4 は、半導体メモリや H D D (Hard Disk Drive) 等の記憶手段であり、画像データ読み取り部 3 によって取得された画像データが記憶されると共に、各種の管理情報が記憶されている。

【 0 0 1 3 】

印字部 5 は、問題頁生成制御部 1 によって生成された問題頁を記録紙に印字して出力するプリンタ等の出力手段である。なお、本実施の形態では、問題頁の出力手段として印字部 5 を採用したが、出力手段としてディスプレイ等の表示画面を採用してもよい。

【 0 0 1 4 】

問題頁生成制御部 1 は、R O M (Read Only Memory)、R A M (Random Access Memory) 等を備えたマイクロコンピュータ等の情報処理部である。R O M には情報抽出装置 1 0 0 の動作制御を行うための制御プログラムが記憶されている。そうして、問題頁生成制御部 1 は、R O M に記憶されている制御プログラムを読み出し、制御プログラムを R A M に展開させることで、操作部 2 から指示入力に応じて問題頁生成動作を実行する。

【 0 0 1 5 】

問題頁生成制御部 1 の機能ブロックは、画像データ解析部 1 1 と、文字認識部 1 2 と、キーワード抽出部 1 3 と、図面イメージ抽出部 1 4 と、図面イメージ内キーワード消去部 1 5 と、文字領域検索部 1 6 と、文字列イメージ抽出部 1 7 と、文字列イメージ内キーワード消去部 1 8 と、レイアウト部 1 9 とからなる。

【 0 0 1 6 】

画像データ解析部 1 1 は、画像データを解析することで、図面領域、図面周辺領域及び文字領域をそれぞれ特定する。画像データ解析部 1 1 では、まず、画像データを図面領域と文字領域とに分割する。当該分割方法としては、各種の方法が提案されているが、例えば、図面と文字の大きさの違いや輪郭の形状によって図面領域と文字領域とを分割することができる。次に、画像データ解析部 1 1 は、図面領域に隣接する比較的面積が狭い文字領域を特定し、当該文字領域を各図面領域に対応する図面周辺領域とする。

【 0 0 1 7 】

文字認識部 1 2 は、画像データ解析部 1 1 によって特定された図面領域、図面周辺領域及び文字領域のそれぞれについて、文字認識 (O C R) 処理を行い、図面領域、図面周辺領域及び文字領域のそれぞれに含まれている文字列をテキストデータに変換する。

【 0 0 1 8 】

キーワード抽出部 1 3 は、文字認識部 1 2 によって認識された図面領域及び図面周辺領域内のテキストデータに対し、形態素解析を行うことで、図番と、キーワードとを抽出する。図番は、図面を示す「図」、「グラフ」、「写真」、「f i g」等の単語に検索こ

10

20

30

40

50

とで抽出される。また、文字列内の「固有名詞」や「普通名詞」がキーワードとして抽出される。

【0019】

図面イメージ抽出部14は、キーワード抽出部13によって抽出された図番と関連づけして、該当する図面イメージ（図面領域）を画像データから図番毎にそれぞれ抽出する。

【0020】

図面イメージ内キーワード消去部15は、図面領域内のテキストデータにキーワード抽出部13によって抽出されたキーワードが存在する場合、図面イメージ抽出部14によって抽出された図面イメージからキーワードに対応する領域を消去する。

10

【0021】

文字領域検索部16は、文字認識部12によって認識された文字領域内のテキストデータを検索することで、キーワード抽出部13によって抽出された図番及びキーワードが含まれた文を図面説明文として特定する。図面説明文は、図番毎に特定される。まず、文字領域検索部16は、文字領域内の文字列（テキストデータ）を検索することで、キーワード抽出部13によって抽出された図番が含まれた文を図面説明文として特定する。次に、文字領域検索部16は、図番が含まれた文の前後の文にキーワード抽出部13によって抽出されたキーワードが含まれているか否かを検索し、キーワードが含まれている場合には、当該文も図面説明文として特定する。なお、キーワードが含まれているか否かを検索する範囲は、例えば図番が含まれた文の前後一文というように予め設定しておくようにしても良く、ユーザーが操作部2から設定できるようにしても良い。

20

【0022】

文字列イメージ抽出部17は、キーワード抽出部13によって抽出された図番と関連づけして、文字領域検索部16によって特定された図面説明文に対応する文字列イメージを画像データから図番毎にそれぞれ抽出する。

【0023】

文字列イメージ内キーワード消去部18は、文字領域検索部16によって特定された図面説明文にキーワード抽出部13によって抽出されたキーワードが存在する場合、文字列イメージ抽出部17によって抽出された文字列イメージからキーワードに対応する領域を消去する。

30

【0024】

レイアウト部19は、図面領域内のテキストデータと、文字領域検索部16によって特定された図面説明文とのいずれかにキーワード抽出部13によって抽出されたキーワードが存在する図番の図面イメージ及び文字列イメージを用い、穴埋め問題頁を生成する。この場合には、図面イメージと文字列イメージとのいずれか又は両方において、キーワードに対応する領域が消去されている。また、レイアウト部19は、図面領域内のテキストデータと、文字領域検索部16によって特定された図面説明文との両方にキーワード抽出部13によって抽出されたキーワードが存在しない図番の図面イメージ及び文字列イメージを用い、選択問題頁を生成する。

【0025】

次に、問題頁生成制御部1による問題頁生成動作について図2乃至図6を参照して詳細に説明する。なお、図3乃至図6において、イメージデータ内の文字は、ゴシック体で示し、テキストデータに変換されている文字は、行書体で示している。

40

【0026】

図2を参照すると、画像データ解析部11は操作部2によって指示された画像データ20を記憶部4から取得し（ステップA1）、取得した、画像データ20を解析することで、図面領域21、図面周辺領域22及び文字領域23をそれぞれ特定する（ステップA2）。図3（a）に示す画像データ20の場合には、図3（b）に点線で示す領域が、それぞれ図面領域21、図面周辺領域22及び文字領域23として特定される。

【0027】

50

次に、文字認識部 12 は、図面領域 21、図面周辺領域 22 及び文字領域 23 のそれぞれについて、文字認識 (OCR) 処理を行い、図 4 (a) に示すように、図面領域 21、図面周辺領域 22 及び文字領域 23 のそれぞれに含まれている文字列 (イメージデータ) をテキストデータ 24 に変換する (ステップ A3)。なお、文字認識部 12 によって変換されたテキストデータ 24 には、図面領域 21 及び文字領域 23 内の位置情報が含まれている。すなわち、テキストデータ 24 内の特定の文字列から、当該文字列に対応する図面領域 21 及び文字領域 23 内の領域を特定することができるようになっている。

【0028】

次に、キーワード抽出部 13 は、図面領域 21 及び図面周辺領域 22 内のテキストデータ 24 に対し、形態素解析を行うことで、図番と、キーワードとを抽出する (ステップ A4)。図 4 (a) に示すテキストデータの場合には、図 4 (b) に示すように、図番として「グラフ 1」、「図 2」、「写真 3」が抽出される。また、図番「グラフ 1」に対応するキーワードとして「大阪」、「東京」、「広島」が、図番「図 2」に対応するキーワードとして「日経平均株価」がそれぞれ抽出される。なお、図番「写真 3」には、キーワードとなる文字列が図面領域 21 及び図面周辺領域 22 内のテキストデータ 24 に含まれないため、キーワードが抽出されない。

【0029】

次に、図面イメージ抽出部 14 は、ステップ A4 で抽出された図番と関連づけして、図 3 (b) に示す画像データ 20 の図面領域 21 を図面イメージ 25 としてそれぞれ抽出する (ステップ A5)。図 4 (c) には、図番「グラフ 1」、「図 2」、「写真 3」にそれぞれ対応する図面イメージ 25 がそれぞれ抽出された例が示されている。

【0030】

次に、図面イメージ内キーワード消去部 15 は、図 4 (a) に示す図面領域 21 内のテキストデータ 24 にステップ A4 で抽出されたキーワードが存在する場合、図 4 (d) に示すように、ステップ A5 で抽出された図面イメージ 25 からキーワードに対応する領域を消去する (ステップ A6)。図 4 (d) には、図番「グラフ 1」に対応する図面イメージ 25 において、キーワード「大阪」、「東京」、「広島」に対応する領域が消去された例が示されている。なお、図 4 (d) では、消去された領域に枠が生成されている。このように、消去された領域に目印 (枠、アンダーライン、色分け等) を施すことで、ユーザーはキーワードが消去されていることを容易に把握することができる。

【0031】

次に、文字領域検索部 16 は、図 4 (a) に示す文字領域 23 内のテキストデータ 24 を検索することで、ステップ A4 で抽出された図番及びキーワードが含まれた文を図面説明文として特定する (ステップ A7)。次に、文字列イメージ抽出部 17 は、図 5 (a) に示すように、キーワード抽出部 13 によって抽出された図番と関連づけして、ステップ A7 で特定された図面説明文に対応する文字列イメージ 26 を、図 3 (b) に示す画像データ 20 の文字領域 23 から図番毎にそれぞれ抽出する (ステップ A8)。図 5 (a) には、図番「グラフ 1」、「図 2」、「写真 3」にそれぞれ対応する文字列イメージ 26 がそれぞれ抽出された例が示されている。

【0032】

次に、文字列イメージ内キーワード消去部 18 は、ステップ A7 で抽出された特定された図面説明文に、ステップ A4 で抽出されたキーワードが存在する場合、図 5 (b) に示すように、ステップ A8 で抽出された文字列イメージ 26 からキーワードに対応する領域を消去する (ステップ A9)。図 5 (b) には、図番「グラフ 1」に対応する文字列イメージ 26 において、キーワード「大阪」、「東京」に対応する領域が、図番「図 2」に対応する文字列イメージ 26 において、キーワード「日経平均株価」に対応する領域がそれぞれ消去された例が示されている。

【0033】

次に、レイアウト部 19 は、図面領域 21 内のテキストデータ 24 と、ステップ A7 で特定された図面説明文とのいずれかに若しくは両方に、ステップ A4 で抽出されたキー

10

20

30

40

50

ワードが存在する図番の場合には、当該図番の図面イメージ 2 5 及び文字列イメージ 2 6 を用い、図 6 (a) に示すような、穴埋め問題頁 2 7 をイメージデータとして生成する (ステップ A 1 0) 。穴埋め問題頁 2 7 には、図面イメージ 2 5 及び文字列イメージ 2 6 が図番毎にレイアウトされている。図番「グラフ 1 」においては、図面領域 2 1 内のテキストデータ 2 4 と、ステップ A 7 で特定された図面説明文との両方にステップ A 4 で抽出されたキーワードが存在する。従って、ステップ A 6 でキーワードに対応する領域が消去されている図面イメージ 2 5 と、ステップ A 9 でキーワードに対応する領域が消去されている文字列イメージ 2 6 とが穴埋め問題頁 2 7 にレイアウトされる。図番「図 2 」においては、ステップ A 7 で特定された図面説明文にのみステップ A 4 で抽出されたキーワードが存在する。従って、ステップ A 5 で抽出された図面イメージ 2 5 と、ステップ A 9 で

10

【 0 0 3 4 】

また、レイアウト部 1 9 は、図面領域 2 1 内のテキストデータ 2 4 と、ステップ A 7 で特定された図面説明文との両方に、ステップ A 4 で抽出されたキーワードが存在しない図番の場合、当該図番の図面イメージ 2 5 及び文字列イメージ 2 6 を用い、図 6 (b) に示すような、選択問題頁 2 8 をイメージデータとして生成する (ステップ A 1 1) 。選択問題頁 2 8 は、図面イメージ 2 5 と文字列イメージ 2 6 との整合を問う問題であり、複数の図面イメージ 2 5 がレイアウトされる図面レイアウト領域 2 9 と、複数の文字列イメージ 2 6 がレイアウトされる説明文レイアウト領域 3 0 とが区別されており、それぞれのレイアウト領域において、複数の図面イメージ 2 5 と複数の文字列イメージ 2 6 とがランダムな順序でレイアウトされる。図番「写真 3 」においては、図面領域 2 1 内のテキストデータ 2 4 と、ステップ A 7 で特定された図面説明文との両方にステップ A 4 で抽出されたキーワードが存在しない。従って、選択問題頁 2 8 をレイアウトされる。図 6 (b) では、図番「写真 3 」の図面イメージ 2 5 が図面レイアウト領域 2 9 の 4 番目「 (D) 」に、文字列イメージ 2 6 が説明文レイアウト領域 3 0 の 1 番目「 (A) 」にそれぞれレイアウトされている。

20

【 0 0 3 5 】

レイアウト部 1 9 によって生成された穴埋め問題頁 2 7 及び選択問題頁 2 8 とは、印字部 5 によって記録紙に印字されて出力される。なお、穴埋め問題頁 2 7 の解答として、キーワードに対応する領域が消去されていない図面イメージ 2 5 及び文字列イメージ 2 6 をレイアウトした穴埋め解答頁を生成するようにしても良い。また、選択問題頁 2 8 の解答として、同じ順序で図面イメージ 2 5 及び文字列イメージ 2 6 をレイアウトした選択解答頁を生成するようにしても良い。これらの穴埋め解答頁及び選択解答頁は、纏めノートとしても活用することができる。

30

【 0 0 3 6 】

また、本実施の形態では、図面領域 2 1 及び図面周辺領域 2 2 内のテキストデータ 2 4 に基づいて、キーワードを特定するように構成したが、文字認識部 1 2 において、同時に文字の属性も認識させ、文字領域 2 3 内のテキストデータ 2 4 から強調箇所に関する検索を行い、強調箇所をキーワードとするようにしても良い。なお、強調箇所とは、色文字や、太字等の文中の他の記述と比べて属性が異なっている箇所である。この場合には、文字列イメージ内キーワード消去部 1 8 によってキーワードに対応する領域が消去された文字列イメージ 2 6 のみが穴埋め問題頁 2 7 にレイアウトされることになる。

40

【 0 0 3 7 】

以上説明したように本実施の形態においては、画像データ 2 0 を解析することで、図面領域 2 1 と文字領域 2 3 とを特定すると共に、図面領域 2 1 の周辺に配置された文字領域 2 3 を図面周辺領域 2 2 として特定する画像データ解析部 1 1 と、図面領域 2 1 、図面周辺領域 2 2 及び文字領域 2 3 のそれぞれについて、文字認識処理を行い、図面領域 2 1 、図面周辺領域 2 2 及び文字領域 2 3 のそれぞれに含まれている文字列をテキストデータ 2 4 に変換する文字認識部 1 2 と、図面領域 2 1 及び図面周辺領域 2 2 のテキストデ

50

ーター２４から図番及びキーワードを抽出するキーワード抽出部１３と、図番と関連づけて図面領域２１を画像データ２０から図面イメージ２５として抽出する図面イメージ抽出部１４と、図面イメージ２５内のキーワードに対応する領域を消去する図面イメージ内キーワード消去部１５と、文字領域２３のテキストデータ２４を検索し、図番が含まれる文を図面説明文として特定する文字領域検索部１６と、図番と関連づけて図面説明文に対応する文字列イメージ２６を画像データ２０の文字領域２３から抽出する文字列イメージ抽出部１７と、図面説明文にキーワードが存在する場合、文字列イメージ２６からキーワードに対応する領域を消去する文字列イメージ内キーワード消去部１８と、図面領域２１内のテキストデータ２４と図面説明文とのいずれかにキーワードが存在する図番の図面イメージ２５及び文字列イメージ２６を用いて、図面イメージ２５及び文字列イメージ２６を図番毎にレイアウトした穴埋め問題頁２７を生成するレイアウト部１９とを備えている。これにより、学習教材にマーキング等の準備をしなくても、図番及びキーワードを抽出し、抽出した図番及びキーワードに基づいて、図面がレイアウトされている学習教材の画像データから図面に関する穴埋め問題を簡単に生成することができるという効果を奏する。

10

【００３８】

さらに、本実施の形態では、レイアウト部１９において、図面領域２１内のテキストデータ２４と図面説明文とのいずれにもキーワードが存在しない図番の図面イメージ２５及び文字列イメージ２６を用い、複数の図面イメージ２５及び複数の文字列イメージ２６をランダムな順序でレイアウトした選択問題頁２８を生成するように構成されている。これにより、図面がレイアウトされている学習教材の画像データから図面に関する選択問題を簡単に生成することができるという効果を奏する。

20

【００３９】

なお、本発明が上記各実施の形態に限定されず、本発明の技術思想の範囲内において、各実施の形態は適宜変更され得ることは明らかである。また、上記構成部材の数、位置、形状等は上記実施の形態に限定されず、本発明を実施する上で好適な数、位置、形状等にすることができる。なお、各図において、同一構成要素には同一符号を付している。

【符号の説明】

【００４０】

- １ 問題頁生成制御部
- ２ 操作部
- ３ 画像データ読み取り部
- ４ 記憶部
- ５ 印字部
- ６ システムバス
- １１ 画像データ解析部
- １２ 文字認識部
- １３ キーワード抽出部
- １４ 図面イメージ抽出部
- １５ 図面イメージ内キーワード消去部
- １６ 文字領域検索部
- １７ 文字列イメージ抽出部
- １８ 文字列イメージ内キーワード消去部
- １９ レイアウト部
- ２０ 画像データ
- ２１ 図面領域
- ２２ 図面周辺領域
- ２３ 文字領域
- ２４ テキストデータ
- ２５ 図面イメージ

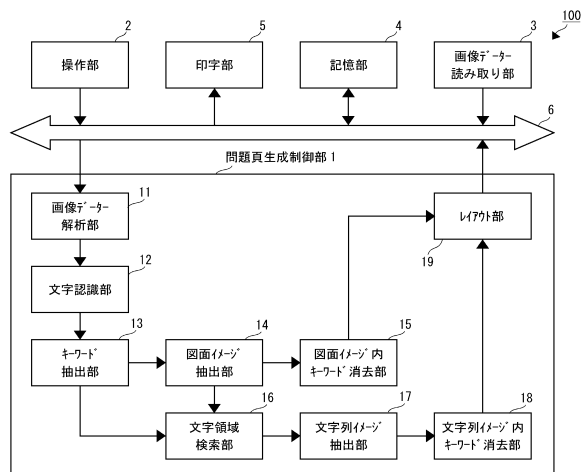
30

40

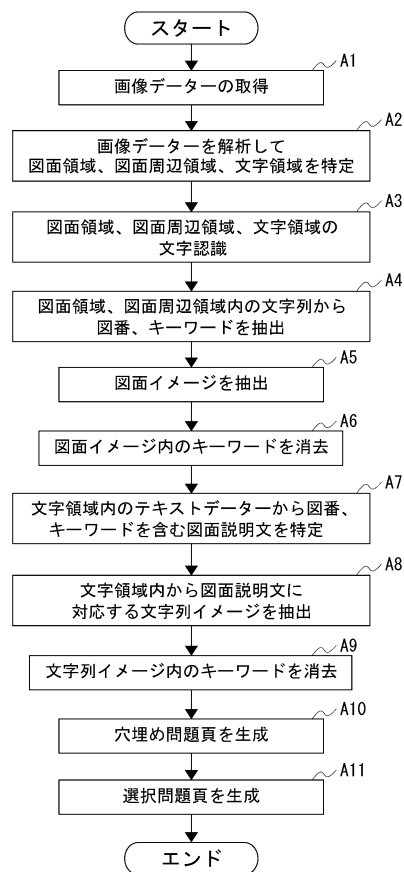
50

- 2 6 文字列イメージ
- 2 7 穴埋め問題頁
- 2 8 選択問題頁
- 2 9 図面レイアウト領域
- 3 0 説明文レイアウト領域
- 1 0 0 情報抽出装置

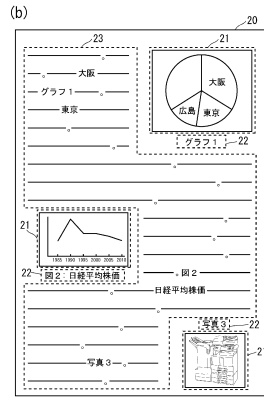
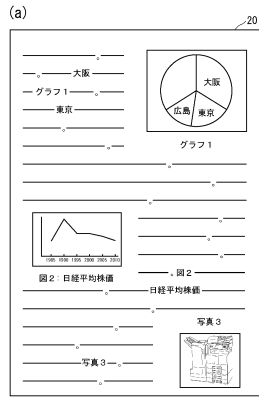
【図 1】



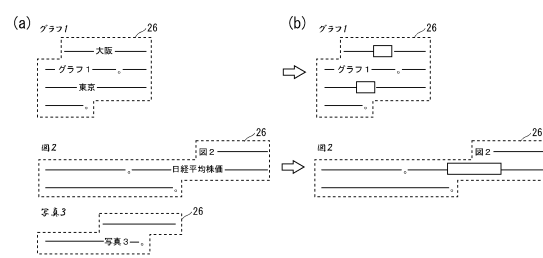
【図 2】



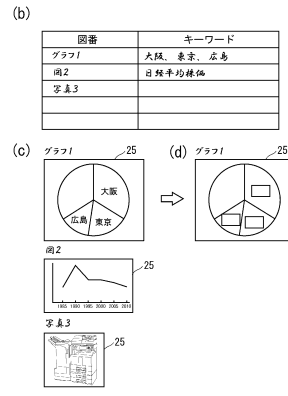
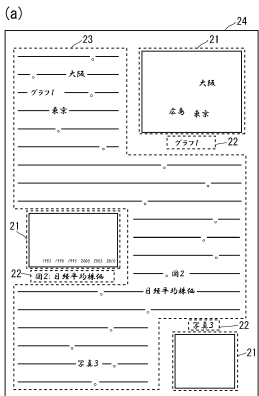
【図 3】



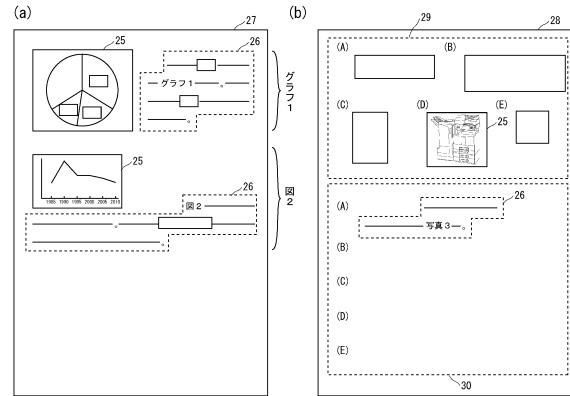
【図 5】



【図 4】



【図 6】



フロントページの続き

(56)参考文献 特開平 07 - 134540 (JP, A)

特開 2011 - 141597 (JP, A)

磯本 征雄 長谷川 信, 教科書通読支援のための小テストの自動生成, 電子情報通信学会技術研究報告 Vol. 109 No. 335, 日本, 社団法人電子情報通信学会, 2009年12月4日, ET2009-56, P.31-36

(58)調査した分野(Int.Cl., DB名)

G06T 11/60

G06F 17/20 - 17/26

G09B 7/00 - 7/12