

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6790712号  
(P6790712)

(45) 発行日 令和2年11月25日(2020.11.25)

(24) 登録日 令和2年11月9日(2020.11.9)

(51) Int. Cl.		F I			
<b>G06T</b>	<b>11/60</b>	<b>(2006.01)</b>	<b>G06T</b>	<b>11/60</b>	<b>100A</b>
<b>G06K</b>	<b>9/20</b>	<b>(2006.01)</b>	<b>G06K</b>	<b>9/20</b>	<b>340J</b>
<b>G06F</b>	<b>40/123</b>	<b>(2020.01)</b>	<b>G06F</b>	<b>40/123</b>	
<b>G06F</b>	<b>40/103</b>	<b>(2020.01)</b>	<b>G06F</b>	<b>40/103</b>	

請求項の数 9 (全 20 頁)

(21) 出願番号	特願2016-205442 (P2016-205442)	(73) 特許権者	000005223 富士通株式会社 神奈川県川崎市中原区上小田中4丁目1番1号
(22) 出願日	平成28年10月19日(2016.10.19)	(74) 代理人	100087480 弁理士 片山 修平
(65) 公開番号	特開2018-67153 (P2018-67153A)	(72) 発明者	福田 隆寛 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内
(43) 公開日	平成30年4月26日(2018.4.26)	審査官	楸 利孝
審査請求日	令和1年7月9日(2019.7.9)		

最終頁に続く

(54) 【発明の名称】 形状抽出プログラム、形状抽出方法及び形状抽出装置

(57) 【特許請求の範囲】

【請求項1】

文字を含む文書情報から、区切り線、前記文字のフォント、フォントサイズ、前記文字の色の1又は複数の組み合わせに基づいて部分領域を特定し、

特定した前記部分領域から文字を除いた形状を抽出し、

特定した前記部分領域に配置された文字を形態素解析して抽出した名詞を属性情報とし、該属性情報を抽出した前記形状に対応づけて記憶部に記憶させる、  
処理をコンピュータに実行させることを特徴とする形状抽出プログラム。

【請求項2】

前記部分領域が複数特定され、2つの部分領域のそれぞれに含まれる文字が所定の類似関係を満たす場合に、前記2つの部分領域を同じグループとして記憶部に記憶させる、ことを特徴とする請求項1に記載の形状抽出プログラム。

【請求項3】

前記形態素解析する文字は、フォントサイズが所定のサイズより大きい文字、又は、見出し文字、又は、特定の色に対応する文字である、ことを特徴とする請求項1に記載の形状抽出プログラム。

【請求項4】

前記文書情報は複数ページの文書情報を含み、特定した前記部分領域が配置されたページに対応するカテゴリ情報、前記部分領域が配置されたページに配置された広告の段数情報、特定した前記部分領域の文字の行数情報の少なくともいずれかが、特定した前記部分

10

20

領域に対応づけて前記記憶部に記憶される、  
ことを特徴とする請求項 1 ~ 3 のいずれか一項に記載の形状抽出プログラム。

【請求項 5】

前記文書情報は見開きのページ構成を有し、特定した前記部分領域が配置されたページが見開きの右側か、左側かの別を示す情報を前記部分領域に対応づけて前記記憶部に記憶させる、ことを特徴とする請求項 1 ~ 4 のいずれか一項に記載の形状抽出プログラム。

【請求項 6】

前記文書情報の曜日又は日付情報が、前記部分領域に対応づけて前記記憶部に記憶される、ことを特徴とする請求項 1 ~ 5 のいずれか一項に記載の形状抽出プログラム。

【請求項 7】

編集対象の文書ページの構成要素として前記記憶部に記憶された情報を選択し、表示することを特徴とする請求項 1 ~ 6 のいずれか一項に記載の形状抽出プログラム。

【請求項 8】

文字を含む文書情報から、区切り線、前記文字のフォント、フォントサイズ、前記文字の色の 1 又は複数の組み合わせに基づいて部分領域を特定し、

特定した前記部分領域から文字を除いた形状を抽出し、

特定した前記部分領域に配置された文字を形態素解析して抽出した名詞を属性情報とし、該属性情報を抽出した前記形状に対応づけて記憶部に記憶させる、

処理をコンピュータが実行することを特徴とする形状抽出方法。

【請求項 9】

文字を含む文書情報から、区切り線、前記文字のフォント、フォントサイズ、前記文字の色の 1 又は複数の組み合わせに基づいて部分領域を特定する特定部と、

特定した前記部分領域から文字を除いた形状を抽出する抽出部と、

特定した前記部分領域に配置された文字を形態素解析して抽出した名詞を属性情報とし、該属性情報を抽出した前記形状に対応づけて記憶部に記憶させる記憶制御部と、

を備える形状抽出装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、形状抽出プログラム、形状抽出方法及び形状抽出装置に関する。

【背景技術】

【0002】

新聞等の編集作業においては、予め編集する情報の内容等に応じて各ページの所望の位置に、文字や画像を配置するためのフレームやボックスと呼ばれる所望のサイズの枠を設定した後、枠内に文字や画像を流し込むことにより、作業が行われることが多い。また、最近では、フレームやボックスが予め所望の位置に配置されたひな形を用意しておき、選択して作業に利用することで、省力化を図ることも行われている。

【0003】

なお、特許文献 1 等には、文書画像処理を自動的に行う技術について開示されている。

【先行技術文献】

【特許文献】

【0004】

【特許文献 1】特開平 1 - 183784 号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、実際に製作した新聞等を人手でひな形化する場合、製作した新聞等から記事が配置されている領域を特定し、特定した領域の形状を抽出する必要があり、手間と労力がかかる。

【0006】

10

20

30

40

50

1つの側面では、本発明は、文書情報から文字が配置された領域の形状を簡易に抽出することが可能な形状抽出プログラム、形状抽出方法及び形状抽出装置を提供することを目的とする。

【課題を解決するための手段】

【0007】

一つの態様では、形状抽出プログラムは、文字を含む文書情報から、区切り線、前記文字のフォント、フォントサイズ、前記文字の色の1又は複数の組み合わせに基づいて部分領域を特定し、特定した前記部分領域から文字を除いた形状を抽出し、特定した前記部分領域に配置された文字を形態素解析して抽出した名詞を属性情報とし、該属性情報を抽出した前記形状に対応づけて記憶部に記憶させる、処理をコンピュータに実行させるプログラムである。

10

【発明の効果】

【0008】

文書情報から文字が配置された領域の形状を簡易に抽出することができる。

【図面の簡単な説明】

【0009】

【図1】一実施形態に係る新聞製作システムの構成を概略的に示す図である。

【図2】図2(a)は、情報処理装置のハードウェア構成を示す図であり、図2(b)は、ユーザ端末のハードウェア構成を示す図である。

【図3】情報処理装置及びユーザ端末の機能ブロック図である。

20

【図4】蓄積対象テーブルの一例を示す図である。

【図5】ひな形DBのデータ構造の一例を示す図である。

【図6】ひな形生成処理の一例を示すフローチャートである。

【図7】図6のステップS12の具体的な処理を示すフローチャートである。

【図8】図7のステップS122の処理を説明するための図である。

【図9】ユーザが製作した新聞紙面の一例を示す図である。

【図10】新聞紙面のレイアウトデータから得られる、配置領域の位置や大きさと種別を示す図である。

【図11】モジュール分割により新聞紙面の配置領域が分割された状態を太線枠にて示す図である。

30

【図12】書き割り化後の新聞レイアウトを示す図である。

【図13】図13(a)は、組み上がり状態(新聞発行時の状態)を示す表示画の一例であり、図13(b)は、書き割り状態(配置領域のみの状態)を示す表示画での一例である。

【図14】ひな形提示処理の一例を示すフローチャートである。

【図15】図15(a)、図15(b)は、図14のステップS50において紙面の情報を取得するために用いる画面の例を示す図である。

【図16】ひな形を提示する一覧画面の例を示す図(その1)である。

【図17】ひな形を提示する一覧画面の例を示す図(その2)である。

【図18】ひな形を表示部上に展開した状態を示す図である。

40

【発明を実施するための形態】

【0010】

以下、新聞製作システムの一実施形態について、図1～図18に基づいて詳細に説明する。

【0011】

図1には、一実施形態に係る新聞製作システム100の構成が概略的に示されている。図1に示すように、新聞製作システム100は、形状抽出装置及び編集支援装置としての情報処理装置10と、ユーザ端末60と、レイアウトデータ格納装置70と、を備える。新聞製作システム100に含まれる各装置は、インターネットなどのネットワーク80に接続されている。

50

## 【 0 0 1 2 】

情報処理装置 10 は、サーバ等を含み、ユーザが製作した文書情報（ここでは、新聞紙面のデータ（レイアウトデータ））から新聞製作に利用可能なひな形を自動的に生成するとともに、ユーザ端末 60 に対して適切なひな形を提供するための装置である。

## 【 0 0 1 3 】

図 2（a）には、情報処理装置 10 のハードウェア構成が示されている。図 2（a）に示すように、情報処理装置 10 は、CPU（Central Processing Unit）90、ROM（Read Only Memory）92、RAM（Random Access Memory）94、記憶部（ここでは HDD（Hard Disk Drive））96、ネットワークインタフェース 97、及び可搬型記憶媒体用ドライブ 99 等を備えている。これら情報処理装置 10 の構成各部は、バス 98 に接続されている。情報処理装置 10 では、ROM 92 あるいは HDD 96 に格納されているプログラム（形状抽出プログラムや編集支援プログラムを含む）、或いは可搬型記憶媒体用ドライブ 99 が可搬型記憶媒体 91 から読み取ったプログラム（形状抽出プログラムや編集支援プログラムを含む）を CPU 90 が実行することにより、図 3 に示す、各部の機能が実現される。なお、図 3 には、情報処理装置 10 の HDD 96 等に格納されているデータベースやテーブルについても図示されている。

10

## 【 0 0 1 4 】

図 3 には、情報処理装置 10 及びユーザ端末 60 の機能ブロック図が示されている。図 3 に示すように、情報処理装置 10 は、ひな形生成部 20 と、ひな形提示部 22 として機能する。

20

## 【 0 0 1 5 】

ひな形生成部 20 は、蓄積対象テーブル 30 を参照して、ユーザが製作した新聞紙面が蓄積対象か否かを判定し、蓄積対象である場合には、ユーザが製作した新聞紙面に基いて、次回以降の新聞紙面製作に利用可能なひな形を自動生成する。ひな形生成部 20 が自動生成したひな形（レイアウトデータ）は、ひな形 DB 32 に格納される。ここで、ひな形生成部 20 は、ひな形を自動生成する際に、ユーザが製作した新聞紙面を解析する。具体的には、ひな形生成部 20 は、ユーザが製作した新聞紙面の記事や見出し、画像が配置された部分領域（配置領域）の属性（配置領域が、記事、見出し、画像のいずれが配置された領域であるか）を判定する。また、各配置領域のカテゴリ（記事や見出し、画像の内容）を判定する。

30

## 【 0 0 1 6 】

また、ひな形生成部 20 は、ユーザが製作した新聞紙面の解析結果を用いて、新聞紙面をカテゴリごとに分割してモジュール化し、各モジュールにおいて配置領域を抽出する「モジュール分割」を実行する。なお、モジュールとは、同一のテーマ（カテゴリ）に属する記事（文字）や写真等のまとまりを意味する。更に、ひな形生成部 20 は、抽出した配置領域から文字や画像を削除して配置領域だけの状態にする「書き割り化」を実行し、書き割り化後の配置領域のレイアウトデータをひな形とする。

## 【 0 0 1 7 】

ひな形提示部 22 は、ユーザが新聞紙面の製作を開始する際に、ひな形 DB 32 を参照して、ユーザが製作しようとしている新聞紙面（編集対象の文書ページ）に適したひな形を選定し、選定したひな形の一覧画面をユーザ端末 60 に送信する。また、ひな形提示部 22 は、一覧画面でユーザにより選択されたひな形をユーザ端末 60 に送信する。

40

## 【 0 0 1 8 】

図 4 には、蓄積対象テーブル 30 のデータ構造の一例が示されている。図 4 に示すように、蓄積対象テーブル 30 は、「面名」と、「蓄積対象 / 対象外」のフィールドを有する。「面名」のフィールドには、面の名称（一面、二面、社会面、くらし面など）が格納される。「蓄積対象 / 対象外」のフィールドには、ひな形の情報を生成し、蓄積する対象であるか否か（対象 / 対象外）が格納される。したがって、ひな形生成部 20 は、蓄積対象テーブル 30 を参照することで、新聞紙面のうち、どの面のひな形を生成し、蓄積すればよいのかを判定することができる。

50

## 【 0 0 1 9 】

図5には、ひな形DB32の一例が示されている。ひな形DB32は、ひな形のデータをひな形を生成する際に利用した新聞紙面の属性と対応付けて記憶するデータベースである。具体的には、ひな形DB32は、図5に示すように「データ種別」、「作成日時」、「最終使用日時」、「使用回数」、「掲載年月日」、「掲載曜日」、「左右ページ」、「面名」、「版名」、「モジュール名」、「縦大きさ」、「横大きさ」、「記事部段数」、「レイアウト格納パス」、「表示画格納パス（書き割り）」、「表示画格納パス（組み上がり）」、「モジュール構成情報格納パス」、「記事数」、「記事行数」、「画像数」、「画像の縦横」の各フィールドを有する。

## 【 0 0 2 0 】

「データ種別」のフィールドには、「面」又は「モジュール」が格納される。「面」が格納されているデータ（行）は、新聞の面（1ページ）全体のレイアウト（ひな形）に関するデータを意味し、「モジュール」が格納されているデータ（行）は、面に含まれる各記事のレイアウト（ひな形）に関するデータを意味する。

## 【 0 0 2 1 】

「作成日時」のフィールドには、ひな形が作成された日時の情報が格納され、「最終使用日時」のフィールドには、ひな形が最後に使用された日時の情報が格納される。「使用回数」のフィールドには、ひな形が使用された回数の情報が格納され、「掲載年月日」、「掲載曜日」のフィールドには、ひな形の作成に利用した新聞が発行された日付及び曜日が格納される。なお、「作成日時」や「最終使用日時」、「使用回数」のフィールドに格納されている情報は、あまり利用されないひな形のデータを削除するか否かを判断する判断基準として利用することができる。

## 【 0 0 2 2 】

「左右ページ」のフィールドには、面及びフィールドが紙面の左側のページ及び右側のページのいずれに配置されたかの情報が格納され、「面名」のフィールドには、面の名称又はモジュールが配置された面の名称が格納される。「版名」のフィールドには、面及びモジュールが掲載された版の名称が格納され、「モジュール名」のフィールドには、モジュールの名称が格納される。なお、データ種別が「面」のデータ（行）においては、モジュール名は、面名と同一となる。

## 【 0 0 2 3 】

「縦大きさ」、「横大きさ」のフィールドには、面やモジュールの縦横の大きさが格納され、「記事部段数」のフィールドには、面において記事（広告を除く部分）が何段配置されているかの情報が格納される。記事部段数は、面が有する段数から広告の段数を差し引いた段数であるといえる。

## 【 0 0 2 4 】

「レイアウト格納パス」のフィールドには、レイアウトデータ（ひな形情報）が保存されている保存先のパスが格納される。なお、レイアウトデータのファイルは、情報処理装置10のストレージに格納される。なお、ストレージは、図2(a)のHDD96等であるため、以下においては、「ストレージ96」とも表記するものとする。例えば、レイアウトデータのファイルは、見出しや記事、画像、区切り線などが配置される配置領域のレイアウト（位置、大きさ、形状等）を管理するとともに、配置領域の属性（見出し、記事、画像、区切り線のいずれが配置されるか）を管理している。なお、本実施形態では、ストレージ96にレイアウトデータを保存し、ひな形DB32において保存先のパスを管理するため、ひな形DB32及びストレージ96においてレイアウトデータを読み出し可能に記憶しているといえる。

## 【 0 0 2 5 】

「表示画格納パス（書き割り）」、「表示画格納パス（組み上がり）」のフィールドには、書き割り後の表示画（例えば、図13(b)参照）と、書き割り前の表示画（例えば、図13(a)参照）を保存する保存先のパスが格納される。なお、表示画のファイルは、例えばjpeg形式のファイルであり、情報処理装置10のストレージ96に格納される。

10

20

30

40

50

「モジュール構成情報格納パス」のフィールドには、モジュールの構成情報が保存されている保存先のパスが格納される。モジュールの構成情報は、モジュールに含まれるオブジェクト（見出し、記事、画像など）の識別番号のリストである。オブジェクトの識別番号を用いて、レイアウトデータのファイルを検索することで、オブジェクトの位置や大きさ、形状を取り出すことができる。なお、モジュールの構成情報のファイルについても、情報処理装置 10 のストレージ 96 に格納される。なお、本実施形態では、表示画やモジュール構成情報をストレージ 96 に保存し、ひな形 DB 32 において保存先のパスを管理するため、ひな形 DB 32 及びストレージ 96 において表示画やモジュール構成情報を読み出し可能に記憶しているといえる。

**【 0026 】**

「記事数」のフィールドには、モジュールに含まれる記事の数が格納され、「記事行数」のフィールドには、モジュールに含まれる記事の行数が格納される。また、「画像数」のフィールドには、モジュールに含まれる画像の数が格納され、「画像の縦横」のフィールドには、モジュールに含まれる画像が縦長であるか横長であるかが格納される。

**【 0027 】**

図 1 に戻り、ユーザ端末 60 は、新聞紙面を製作するユーザ（新聞社の社員等）が利用する端末である。ユーザ端末 60 では、情報処理装置 10 に対して、ユーザが製作しようとしている新聞紙面に関する情報を提供し、情報処理装置 10 から提示されたひな形の一覧画面を表示する。また、ユーザ端末 60 は、一覧画面においてユーザにより選択されたひな形の情報を情報処理装置 10 に送信し、情報処理装置 10 から送信されてくるひな形及び / 又は情報処理装置 10 にあるひな形を表示部 193 上に展開する。また、ユーザ端末 60 は、ユーザが製作した新聞紙面のレイアウトデータをレイアウトデータ格納装置 70 に記憶する。この場合、ユーザ端末 60 は、レイアウトデータに関連付けて、ユーザが製作した新聞紙面やモジュールに関する情報（掲載年月日、掲載曜日、左右ページ、面名、版名、縦大きさ、横大きさ、記事部段数、記事数、記事行数、画像数、画像の縦横など）をレイアウトデータ格納装置 70 に記憶する。

**【 0028 】**

図 2 (b) には、ユーザ端末 60 のハードウェア構成が示されている。図 2 (b) に示すように、ユーザ端末 60 は、CPU 190、ROM 192、RAM 194、記憶部 (HDD) 196、ネットワークインタフェース 197、表示部 193、入力部 195、及び可搬型記憶媒体用ドライブ 199 等を備えている。これらユーザ端末 60 の構成各部は、バス 198 に接続されている。表示部 193 は液晶ディスプレイ等を含み、入力部 195 は、キーボードやマウスタッチパネル等を含む。ユーザ端末 60 では、ROM 192 あるいは HDD 196 に格納されているプログラム、或いは可搬型記憶媒体用ドライブ 199 が可搬型記憶媒体 191 から読み取ったプログラムを CPU 190 が実行することにより、図 3 に示す、紙面製作部 62、紙面送信部 64 の機能が実現される。紙面製作部 62 は、ユーザの操作に従って、新聞紙面を製作する処理を実行する。例えば、紙面製作部 62 は、ユーザが配置した記事や見出し、区切り線に基づいて、新聞紙面を製作する。また、紙面製作部 62 は、例えば、ユーザにより選択されたひな形に対するユーザの操作に応じて紙面を製作する。紙面送信部 64 は、紙面製作部 62 が製作した紙面をレイアウトデータ格納装置 70 に格納する。

**【 0029 】**

レイアウトデータ格納装置 70 は、データサーバ等であり、ユーザ端末 60 において製作された新聞紙面のレイアウトデータや、紙面及びモジュールの情報を管理する。

**【 0030 】**

( 情報処理装置 10 のひな形生成処理 )

次に、情報処理装置 10 によるひな形生成処理について、図 6 のフローチャートに沿って詳細に説明する。なお、本処理が行われるタイミングは、ユーザ端末 60 において新規の新聞紙面が製作され、新規新聞紙面のデータ (レイアウトデータ) がレイアウトデータ格納装置 70 に格納されたタイミングであるものとする。ただし、これに限らず、例えば

10

20

30

40

50

、新規の新聞紙面が所定数だけ製作された場合や、所定期間経過した場合（例えば1日1回）など、所定の条件を満たしたタイミングで図6の処理を実行することとしてもよい。また、ユーザや情報処理装置10の管理者等が手動で図6の処理の実行開始を指示してもよい。

【0031】

図6の処理では、まず、ステップS10において、情報処理装置10のひな形生成部20は、新たにレイアウトデータ格納装置70に格納された新聞紙面のレイアウトデータを取得する。

【0032】

次いで、ステップS11では、ひな形生成部20が、ユーザが製作した新聞紙面の中から蓄積対象面を決定する。ひな形生成部20は、蓄積対象テーブル30を参照して、1日分の新聞紙面から、ひな形を生成する面を決定する。図4の蓄積対象テーブル30からは、「生活」、「くらし」、「文化」の面がひな形を生成する面として決定される。なお、以下においては、「文化」の面のひな形生成を例にとり説明する。

【0033】

次いで、ステップS12では、ひな形生成部20が、ユーザが製作した新聞紙面の解析処理を実行する。ステップS12の解析処理においては、具体的には、ひな形生成部20は、図7のフローチャートに沿った処理を実行する。

【0034】

ひな形生成部20は、図7の処理において、配置領域の特定処理（S120）と、配置領域の種別判定処理（S122）と、配置領域のカテゴリ判定処理（S124）とを実行する。

【0035】

配置領域の特定（S120）においては、ひな形生成部20は、区切り線、文字のフォント、フォントサイズ、文字の色の少なくとも1つに基づいて、配置領域を抽出する。例えば、ひな形生成部20は、区切り線を用いて、新聞紙面の領域を区画するとともに、区画した範囲内で、文字のフォント、フォントサイズや色に基づいて、更に領域を区画することで、配置領域を抽出する。また、画像が配置されている領域についても、配置領域として抽出する。

【0036】

次いで、配置領域の種別判定処理（S122）においては、まず、ステップS1221において、ひな形生成部20が、記事領域判定を実行する。具体的には、ひな形生成部20は、割り付いた文字のフォントが記事本文用のフォントである配置領域を記事領域として認識する。

【0037】

次いで、ステップS1222では、ひな形生成部20は、見出し領域判定処理を実行する。具体的には、ひな形生成部20は、割り付いた文字のフォントが見出し用のフォントである配置領域、記事フォントよりもサイズが大きい配置領域、フォントの色が見出し用の色である配置領域、例えば白抜き文字など装飾されている文字が割り付いた配置領域、などを見出し領域と認識する。

【0038】

次いで、ステップS1223では、ひな形生成部20は、画像領域判定処理を実行する。具体的には、ひな形生成部20は、文字が割り付いておらず、画像のみが割り付いている配置領域または画像と写真説明文の領域とを合わせた領域を画像領域と認識する。

【0039】

以上のようにして配置領域の種別判定処理（S122）が行われた後は、ひな形生成部20は、ステップS124において、配置領域のカテゴリ判定を実行する。具体的には、ひな形生成部20は、次のようにして、配置領域のカテゴリ判定を行う。

【0040】

（1）まず、ひな形生成部20は、見出し領域の文字（見出し文）を形態素解析し、名詞

10

20

30

40

50

を抽出する。例えば、図 8 に示すように、見出し文が「日本総人口 初の減少」であったとする。この場合、ひな形生成部 20 は、形態素解析により、名詞「日本」「人口」「初」「減少」を抽出する。

(2) 次いで、ひな形生成部 20 は、記事領域の文字(記事本文)の中から、(1)で抽出された名詞「日本」「人口」「初」「減少」を検索する。

(3) 次いで、ひな形生成部 20 は、見出し文と記事本文の文字が所定の類似関係を満たす場合に、同一のカテゴリとする。具体的には、ひな形生成部 20 は、一致する名詞が最も多い記事本文と見出しを関連付け、同一のカテゴリ名「日本/人口/初/減少」を付与する。

#### 【0041】

以上の処理により、図 7 の処理(ステップ S 12 の処理)が終了する。この図 7 の処理により、各配置領域の種別が判定できるとともに、配置領域のカテゴリが判定できるので、以降においては、カテゴリ名が共通する配置領域を 1 つのまとまり(モジュール)として扱うことが可能になる。この場合、共通するカテゴリ名がモジュール名として扱われる。なお、ひな形生成部 20 は、ステップ S 12 の判定結果を、配置領域のレイアウトデータとして管理する。なお、配置領域のレイアウトデータには、配置領域の位置や大きさも含まれている。

#### 【0042】

図 6 に戻り、ステップ S 13 では、ひな形生成部 20 が、モジュール分割を実行する。例えば、ユーザ端末 60 において、新聞紙面の「文化」の面が、図 9 に示すように製作されたものとする。この場合、ひな形生成部 20 は、文化の面をカテゴリ(=記事、話題)ごとに分割する。なお、本実施形態においては、新聞紙面のレイアウトデータでは、前述のように配置領域の位置や大きさと種別(見出し、記事、画像、区切り線の別)とが管理されている。すなわち、レイアウトデータからは、図 10 に示すように、配置領域の位置や大きさと種別とが得られる。また、レイアウトデータにおいては、配置領域内に配置された記事や見出し、画像のカテゴリ名(何に関する記事であるかを示す名称)が管理されている。したがって、ひな形生成部 20 は、カテゴリ名ごとのまとまりで、配置領域を分割する。図 11 には、モジュール分割により新聞紙面の配置領域が分割された状態が太線枠にて示されている。

#### 【0043】

次いで、ステップ S 14 では、ひな形生成部 20 が、書き割り化を実行する。具体的には、ひな形生成部 20 は、新聞紙面のレイアウトデータを編集し、文字や画像を削除して配置領域だけの状態にする。図 12 には、書き割り化後の新聞レイアウトが示されている。なお、書き割り化後の新聞レイアウトは、新聞紙面から抽出された形状データであるといえる。

#### 【0044】

次いで、ステップ S 16 では、ひな形生成部 20 が、レイアウトデータの書き出しを行う。すなわち、ひな形生成部 20 は、書き割り化した新聞紙面のレイアウトデータを書き出し、ストレージ 96 に格納するとともに、格納先のパスを取得する。この場合、ひな形生成部 20 は、面全体のレイアウトデータを書き出すとともに、面の一部であるモジュールごとのレイアウトデータ(図 11 の太線枠ごとのレイアウトデータ)についても書き出す。

#### 【0045】

次いで、ステップ S 18 では、ひな形生成部 20 が、管理情報の収集を行う。この場合、ひな形生成部 20 は、レイアウトデータ格納装置 70 に格納されているユーザが製作した新聞紙面の情報を参照し、蓄積する必要のある管理情報を収集する。具体的には、ひな形生成部 20 は、レイアウトデータ格納装置 70 から、新聞紙面の管理情報として、図 5 のひな形 DB 32 の「掲載年月日」、「掲載曜日」、「左右ページ」、「面名」、「版名」、「モジュール名」、「縦大きさ」、「横大きさ」、「記事部段数」のフィールドに格納する情報などを収集する。

10

20

30

40

50



## 【 0 0 4 6 】

次いで、ステップ S 2 0 では、ひな形生成部 2 0 が、表示画を作成する。この場合、図 1 3 ( a )、図 1 3 ( b ) に示すような、モジュールごとの表示画を書き出す。より具体的には、ひな形生成部 2 0 は、図 1 3 ( a ) に示すような組み上がり状態（新聞発行時の状態）を示す表示画と、図 1 3 ( b ) に示すような書き割り状態（配置領域のみの状態）を示す表示画の 2 種類の表示画を作成する。また、ひな形生成部 2 0 は、図 9 の新聞紙面そのものの表示画についても作成する。なお、ひな形生成部 2 0 は、作成した表示画をストレージ 9 6 に格納し、格納先のパスを取得する。

## 【 0 0 4 7 】

次いで、ステップ S 2 2 では、ひな形生成部 2 0 が、ひな形 D B 3 2 ヘデータを格納する。すなわち、ひな形生成部 2 0 は、これまでの処理において作成、収集したデータをひな形 D B 3 2 ヘ格納する。なお、図 9 の新聞紙面そのものの表示画の格納パスについては、図 5 のひな形 D B 3 2 のデータ種別が「面」のデータ（行）の「表示画格納パス（組み上がり）」のフィールドに格納される。

## 【 0 0 4 8 】

（情報処理装置 1 0 によるひな形提示処理）

次に、図 1 4 のフローチャートに沿って、情報処理装置 1 0 により実行されるひな形提示処理について説明する。図 1 4 の処理は、例えば、ユーザがユーザ端末 6 0 において新規の新聞紙面の製作を開始するタイミングで実行される処理である。

## 【 0 0 4 9 】

図 1 4 の処理では、まず、ステップ S 5 0 において、ひな形提示部 2 2 は、ユーザが製作しようとしている紙面の情報を取得する。この場合、紙面編集開始時に自動で紙面の情報を取得する場合と、ユーザが手入力した紙面の情報を取得する場合とがある。

## 【 0 0 5 0 】

紙面編集開始時に自動で紙面の情報を取得する場合、ひな形提示部 2 2 は、ユーザ端末 6 0 から、製作しようとしている紙面の情報と作成済みで新聞紙面に配置しようとしているコンテンツ（記事）の属性を自動的に取得する。なお、ユーザ端末 6 0 において紙面編集を開始する際に表示される画面は、図 1 5 ( a ) に示すような画面である。ユーザは、紙面を製作するとき、図 1 5 ( a ) の画面の入力欄 7 2 において日付や、製作する面の情報などを入力、選択等するため、ひな形提示部 2 2 は、図 1 5 ( a ) の画面に入力された情報を取得する。また、ひな形提示部 2 2 は、ユーザ端末 6 0 において作成されたコンテンツ（記事）の情報（モジュール名や、モジュールの大きさ、画像数、画像の縦横の情報など）を取得する。なお、コンテンツ（記事）の情報は、文字、文字の構成情報、写真の構成情報を含んでいる。

## 【 0 0 5 1 】

一方、ユーザが手動で紙面の情報を入力する画面は、図 1 5 ( b ) に示すような画面である。ユーザは、ひな形を呼び出す際に、図 1 5 ( b ) の画面において、掲載年月日や掲載曜日、記事数、画像数、モジュール名、面名、版名、左右ページ、記事部段数、記事行数、画像縦横などの情報を入力し、確定ボタンを押す。ユーザによって確定ボタンが押されると、ユーザ端末 6 0 からひな形提示部 2 2 に図 1 5 ( b ) の画面において入力された情報が送信される。

## 【 0 0 5 2 】

図 1 4 に戻り、次のステップ S 5 2 では、ひな形提示部 2 2 が、紙面の属性（掲載曜日、左右ページ、面名、版名など）でひな形を絞り込む。この場合、ひな形提示部 2 2 は、ひな形 D B 3 2 を参照して、ユーザが製作しようとしている紙面の属性と一致するひな形を抽出する。なお、ステップ S 5 2 では、ユーザが製作しようとしている紙面の属性と完全一致するひな形を抽出する場合に限らず、属性がある程度類似するひな形を抽出することとしてもよい。

## 【 0 0 5 3 】

次いで、ステップ S 5 4 では、ひな形提示部 2 2 が、記事（コンテンツ）の属性でひな

10

20

30

40

50

形を絞り込む。この場合、ひな形提示部 2 2 は、ステップ S 5 2 で抽出したひな形のうち、ステップ S 5 0 で取得した記事の属性（例えば、モジュール名、記事数、行数、画像数、画像の縦横など）と一致するひな形を抽出する。なお、ステップ S 5 2 では、記事の属性と完全一致するひな形を抽出する場合に限らず、属性がある程度類似するひな形を抽出することとしてもよい。

#### 【 0 0 5 4 】

次いで、ステップ S 5 6 では、ひな形提示部 2 2 が、ひな形の候補を一覧表示する。具体的には、ひな形提示部 2 2 は、図 1 6 や図 1 7 に示すような一覧画面を情報処理装置 1 0 の紙面製作部 6 2 に送信する。これにより、紙面製作部 6 2 が、表示部 1 9 3 上を一覧表示を行う。図 1 6 の一覧画面では、表示画として「組み上がり」、「書き割り」が表示される。一方、図 1 7 の一覧画面では、表示画として「組み上がり」、「書き割り」に加え、「出来上がり」が表示される。表示画（出来上がり）は、ひな形を生成した際に用いた新聞紙面そのものの画像（図 9 の新聞紙面から生成した表示画）である。表示画（出来上がり）は、ひな形 D B 3 2 のうちデータ種別が「面」のデータの「表示画格納パス（組み上がり）」のフィールドに格納されているパスで管理されている表示画である。なお、図 1 6 と図 1 7 のいずれの一覧画面を表示するかは、ユーザにより予め設定されているものとする。なお、ユーザからの入力により、一覧画面を図 1 6 から図 1 7 へ、又は図 1 7 から図 1 6 へ切り替えられるようにしてもよい。ユーザは、図 1 6 や図 1 7 の一覧画面を参照することで、製作しようとしている新聞紙面において記事や画像がどのように配置されるかを確認することができるので、一覧画面からのひな形選択がしやすくなっている。

#### 【 0 0 5 5 】

次いで、ステップ S 5 8 では、ひな形提示部 2 2 が、一覧画面の中からユーザがいずれかのひな形を選択するまで待機する。ユーザが入力部 1 9 5 を介していずれかのひな形を選択し、選択したひな形の情報が紙面製作部 6 2 からひな形提示部 2 2 に送信されると、ステップ S 6 0 に移行する。

#### 【 0 0 5 6 】

ステップ S 6 0 に移行すると、ひな形提示部 2 2 は、ひな形をユーザ端末 6 0 の表示部 1 9 3 上に展開する。図 1 8 には、選択されたひな形を表示部 1 9 3 上に展開した状態が示されている。ユーザは、展開されたひな形の各配置領域に対して、見出しや記事を配置したり、画像を配置することが可能となっている。なお、展開されたひな形の配置領域の大きさは、配置時に更新またはユーザが適宜変更できるようになっている。すなわち、ステップ S 6 0 では、ひな形提示部 2 2 は、ユーザによって選択されたひな形を、ユーザが製作しようとしている新聞紙面の構成要素として提示しているといえる。

#### 【 0 0 5 7 】

以上のように、図 1 4 の処理が実行されることで、本実施形態では、ユーザが、適切なひな形を利用することができるので、簡易に新聞紙面を製作することが可能となっている。

#### 【 0 0 5 8 】

これまでの説明からわかるように、本実施形態では、ひな形生成部 2 0 は、製作した新聞紙面から、区切り線、文字のフォント、フォントサイズ、文字の色の少なくとも 1 つに基づいて配置領域を特定する特定部、及び特定した配置領域から文字を除いた形状を抽出する抽出部として機能する。また、ひな形生成部 2 0 は、特定した配置領域に配置された文字に基づいて配置領域の属性を生成し、生成した属性を抽出した形状に対応づけてひな形 D B 3 2 に記憶させる生成部として機能する。また、ひな形生成部 2 0 は、新聞紙面から、区切り線で区切られた領域（図 1 1 の太枠参照）を特定する特定部、及び特定した領域のデータと、該領域から文字を除いた形状とをひな形 D B 3 2 やストレージ 9 6 に記憶させる記憶制御部として機能する。

#### 【 0 0 5 9 】

以上、詳細に説明したように、本実施形態によると、ひな形生成部 2 0 は、ユーザがユーザ端末 6 0 において製作した新聞紙面から、区切り線、文字のフォント、フォントサイ

ズ、色等に基づいて配置領域を特定し（S 1 2 0）、特定した配置領域から文字を除いた形状を抽出する（S 1 3）。これにより、ユーザが製作した新聞紙面において配置領域が指定されていなくても、自動的に配置領域を特定し、ひな形に利用可能な配置領域の形状を簡易に抽出することができる。

【 0 0 6 0 】

また、本実施形態によると、ひな形生成部 2 0 は、特定した配置領域に配置された文字に基づいて属性情報（見出し、記事、画像の別や、カテゴリ（モジュール名））を生成し（S 1 2 2）、配置領域の形状等に対応付けてレイアウトデータとしてストレージ 9 3 に記憶する。これにより、ユーザが製作した新聞紙面において配置領域の属性が定められていない場合であっても、新聞紙面から配置領域の属性情報と配置領域の形状等に対応付けたレイアウトデータを自動的に生成することが可能である。

10

【 0 0 6 1 】

また、本実施形態によると、2つの配置領域のそれぞれに含まれる文字が所定の類似関係を満たす場合（記事本文が、見出し文に含まれる名詞を多く含む場合）に、2つの配置領域を同じグループ（モジュール）としてストレージ 9 6 及びひな形 DB 3 2 に記憶する。これにより、ユーザが製作した新聞紙面において配置領域間の関係が定められていない場合であっても、新聞紙面から配置領域のグループ（モジュール）を自動的に生成することが可能である。

【 0 0 6 2 】

また、本実施形態では、新聞紙面から、区切り線で区切られた領域を特定し（図 1 1）、特定した領域のデータ（例えばモジュール名や記事数、画像数など）と、特定した領域から文字を除いた形状データ（ひな形）とをストレージ 9 6 及びひな形 DB 3 2 に記憶している。これにより、モジュール名や記事数、画像数などに基づいて、ストレージ 9 6 からユーザが製作しようとしている新聞紙面に適したひな形をモジュール単位で読み出すことが可能である。また、本実施形態では、ひな形提示部 2 2 がひな形をユーザに提示する際に、ひな形に対応する特定した領域のデータ（モジュール名や記事数、画像数など）を併せて表示する（図 1 6、図 1 7）ので、ユーザは、ひな形が複数提示された場合でも、適切なひな形を選択することが可能である。

20

【 0 0 6 3 】

また、本実施形態では、ひな形 DB 3 2 において、モジュールが配置された面の面名、面の記事部段数、モジュールの行数情報をモジュールのひな形に対応づけて記憶している。これにより、ひな形に対応付けて記憶されている情報に基づいて、ひな形 DB 3 2 からユーザが製作しようとしている新聞紙面に適したひな形を抽出し、ユーザに提供することが可能である。

30

【 0 0 6 4 】

また、本実施形態では、ひな形 DB 3 2 において、モジュールが配置された面が新聞の左右いずれの面であるかの情報を、モジュールのひな形に対応づけて記憶している。新聞の場合、左右いずれの面であるかによってレイアウトが異なるため、上記のように左右いずれの面であるかの情報をひな形に対応付けて記憶しておくことで、ユーザが製作しようとしている新聞紙面に適したひな形を抽出し、ユーザに提供することが可能である。

40

【 0 0 6 5 】

また、本実施形態では、ひな形 DB 3 2 において、モジュールが掲載された新聞紙面が発行された曜日や日付を、モジュールのひな形に対応づけて記憶している。新聞の場合、曜日や日付によってレイアウトが異なることが多いため、上記のように曜日や日付の情報をひな形に対応付けて記憶しておくことで、ユーザが製作しようとしている新聞紙面に適したひな形を抽出し、ユーザに提供することが可能である。

【 0 0 6 6 】

また、本実施形態によれば、ユーザ端末 6 0 で生成された新聞紙面のレイアウトデータを取得し（S 1 0）、取得したデータのうち、区切り線と、各記事の配置領域を抽出し（S 1 3 ~ S 1 6）、抽出した区切り線と、各記事の配置領域の形状とをひな形（レイアウト

50

トデータ)として、読み出し可能にひな形DB32及びストレージ96に記憶させる(S22)。これにより、ユーザが新聞紙面を製作する際に利用可能なひな形を自動的に生成して、ひな形DB32及びストレージ96に記憶させることができる。この場合、ひな形を一から手作業で生成したり、製作済みの新聞紙面から文字や画像を除外してひな形を生成したりする場合に比べて、手間や時間をかけずにひな形を生成することができ、新聞紙面の製作作業を適切に支援することが可能である。

【0067】

また、本実施形態では、ひな形DB32において、記事のカテゴリ(モジュール名)をひな形に対応付けて記憶している。これにより、ユーザが製作しようとしている記事のカテゴリに応じて、適切なひな形をひな形DB32から読み出して、ユーザに対して提供することができる。

10

【0068】

また、本実施形態では、ひな形DB32において、ひな形を生成するときに用いた新聞紙面から生成した画像(表示画)をひな形に対応付けて記憶している。これにより、ユーザに対してひな形を提示するときに、表示画を表示することができるため、ユーザは、表示画を参照して、提示されたひな形の中から利用するひな形を選択することが可能である。

【0069】

また、本実施形態では、新たに製作された新聞紙面のうち、蓄積対象テーブル30において対象とされている面を用いて、ひな形(レイアウトデータ)を生成し、ひな形DB32及びストレージ96に読み出し可能に記憶することとしている。これにより、ユーザが利用する可能性の高い、利便性の高いひな形を生成し、ひな形DB32及びストレージ96に記憶しておくことができる。

20

【0070】

また、本実施形態によると、ひな形提示部22は、ユーザが製作しようとしている(編集対象の)紙面や記事の属性(掲載曜日、左右ページ、面名、版名、記事数、記事行数、画像数、画像の縦横など)を取得し、ひな形DB32を参照して、取得した紙面の属性に対応するひな形を絞り込み(S52、S54)、絞り込んだひな形をユーザが製作しようとしている紙面のひな形として提示する(S56)。これにより、ユーザが製作しようとしている紙面の製作に適したひな形を自動的に特定し、提示することができる。したがって、多数のひな形がひな形DB32及びストレージ96に格納されている場合であっても、新聞の製作に適したひな形を容易に絞り込み、ユーザに提示することができる。

30

【0071】

また、本実施形態では、ひな形提示部22は、絞り込んだひな形を提示する際に、ひな形のデータとともに、文字や写真の配置候補(表示画(組み上がり)や表示画(出来上がり))を提示する。これにより、ユーザは、提示されたひな形から利用するひな形を選択する際に、組み上がりや出来上がり後の紙面を参考にすることができる。

【0072】

また、本実施形態では、ひな形提示部22は、ひな形を提示する際に、製作しようとしている面全体または面の一部を提示する。これにより、ユーザは、紙面作成の際に利用しやすい大きさのひな形を選択することが可能である。

40

【0073】

なお、上記実施形態では、ひな形提示部22は、ステップS52及びS54において、面及びコンテンツの属性でひな形を絞り込む場合について説明したが、これに限られるものではない。すなわち、ステップS52とステップS54のいずれか一方を実行することとしてもよい。

【0074】

なお、上記実施形態では、情報処理装置10において生成されたひな形を情報処理装置10内のひな形DB32で管理する場合について説明したが、これに限らず、外部のデータサーバ等で管理してもよい。この場合、ひな形を生成する装置と、ひな形を提示する装

50

置とが異なってもよい。

【0075】

なお、上記実施形態では、新聞製作時に利用するひな形の生成、提示において、情報処理装置10が上記処理を実行する場合について説明したが、これに限られるものではない。例えば、雑誌などの書籍の製作時に利用するひな形の生成、提示において、情報処理装置10が上記処理を実行することとしてもよい。

【0076】

なお、上記の処理機能は、コンピュータによって実現することができる。その場合、処理装置が有すべき機能の処理内容を記述したプログラムが提供される。そのプログラムをコンピュータで実行することにより、上記処理機能がコンピュータ上で実現される。処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体（ただし、搬送波は除く）に記録しておくことができる。

【0077】

プログラムを流通させる場合には、例えば、そのプログラムが記録されたDVD（Digital Versatile Disc）、CD-ROM（Compact Disc Read Only Memory）などの可搬型記録媒体の形態で販売される。また、プログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することもできる。

【0078】

プログラムを実行するコンピュータは、例えば、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、自己の記憶装置に格納する。そして、コンピュータは、自己の記憶装置からプログラムを読み取り、プログラムに従った処理を実行する。なお、コンピュータは、可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することもできる。また、コンピュータは、サーバコンピュータからプログラムが転送されるごとに、逐次、受け取ったプログラムに従った処理を実行することもできる。

【0079】

上述した実施形態は本発明の好適な実施の例である。但し、これに限定されるものではなく、本発明の要旨を逸脱しない範囲内において種々変形実施可能である。

【0080】

なお、以上の実施形態の説明に関して、更に以下の付記を開示する。

（付記1） 文字を含む文書情報から、区切り線、前記文字のフォント、フォントサイズ、前記文字の色 of 1 又は複数の組み合わせに基づいて部分領域を特定し、特定した前記部分領域から文字を除いた形状を抽出する、処理をコンピュータに実行させることを特徴とする形状抽出プログラム。

（付記2） 特定した前記部分領域に配置された文字に基づいて属性情報を生成し、生成した前記属性情報を抽出した前記形状に対応づけて記憶部に記憶させる、ことを特徴とする付記1に記載の形状抽出プログラム。

（付記3） 前記部分領域が複数特定され、2つの部分領域のそれぞれに含まれる文字が所定の類似関係を満たす場合に、前記2つの部分領域を同じグループとして記憶部に記憶させる、ことを特徴とする付記1に記載の形状抽出プログラム。

（付記4） 前記属性情報の生成に用いられる文字は、フォントサイズが所定のサイズより大きい文字、又は、見出し文字、又は、特定の色に対応する文字である、ことを特徴とする付記2に記載の形状抽出プログラム。

（付記5） 前記文書情報は複数ページの文書情報を含み、特定した前記部分領域が配置されたページに対応するカテゴリ情報、前記部分領域が配置されたページに配置された広告の段数情報、特定した前記部分領域の文字の行数情報の少なくともいずれかが、特定した前記部分領域に対応づけて前記記憶部に記憶される、ことを特徴とする付記2～4のいずれかに記載の形状抽出プログラム。

10

20

30

40

50

(付記 6) 前記文書情報は見開きのページ構成を有し、特定した前記部分領域が配置されたページが見開きの右側か、左側かの別を示す情報を前記部分領域に対応づけて前記記憶部に記憶させる、ことを特徴とする付記 2 ~ 5 のいずれかに記載の形状抽出プログラム。

(付記 7) 前記文書情報の曜日又は日付情報が、前記部分領域に対応づけて前記記憶部に記憶される、ことを特徴とする付記 2 ~ 6 のいずれかに記載の形状抽出プログラム。

(付記 8) 編集対象の文書ページの構成要素として前記記憶部に記憶された情報を選択し、表示することを特徴とする付記 2 ~ 7 のいずれかに記載の形状抽出プログラム。

(付記 9) 区切り線で区切られて配置された文字を含む文書情報から、区切り線で区切られた部分領域を特定し、

特定した前記部分領域のデータと、該部分領域から文字を除いた形状とを記憶部に記憶させる、

処理をコンピュータに実行させることを特徴とする形状抽出プログラム。

(付記 10) 編集対象の文書ページの構成要素として前記記憶部に記憶された前記部分領域から文字を除いた形状を表示する際に、前記部分領域のデータも併せて表示させる、処理をコンピュータに実行させることを特徴とする付記 9 に記載の形状抽出プログラム。

(付記 11) 文字を含む文書情報から、区切り線、前記文字のフォント、フォントサイズ、前記文字の色の 1 又は複数の組み合わせに基づいて部分領域を特定し、

特定した前記部分領域から文字を除いた形状を抽出する、

処理をコンピュータが実行することを特徴とする形状抽出方法。

(付記 12) 区切り線で区切られて配置された文字を含む文書情報から、区切り線で区切られた部分領域を特定し、

特定した前記部分領域のデータと、該部分領域から文字を除いた形状とを記憶部に記憶させる、

処理をコンピュータが実行することを特徴とする形状抽出方法。

(付記 13) 文字を含む文書情報から、区切り線、前記文字のフォント、フォントサイズ、前記文字の色の 1 又は複数の組み合わせに基づいて部分領域を特定する特定部と、

特定した前記部分領域から文字を除いた形状を抽出する抽出部と、

を備える形状抽出装置。

(付記 14) 特定した前記部分領域に配置された文字に基づいて属性情報を生成し、生成した前記属性情報を抽出した前記形状に対応づけて記憶部に記憶させる生成部、を備える付記 13 に記載の形状抽出装置。

(付記 15) 前記部分領域が複数特定され、2つの部分領域のそれぞれに含まれる文字が所定の類似関係を満たす場合に、前記2つの部分領域を同じグループとして記憶部に記憶させる生成部、を備える付記 13 に記載の形状抽出装置。

(付記 16) 前記属性情報の生成に用いられる文字は、フォントサイズが所定のサイズより大きい文字、又は、見出し文字、又は、特定の色に対応する文字である、ことを特徴とする付記 14 に記載の形状抽出装置。

(付記 17) 区切り線で区切られて配置された文字を含む文書情報から、区切り線で区切られた部分領域を特定する特定部と、

特定した前記部分領域のデータと、該部分領域から文字を除いた形状とを記憶部に記憶させる記憶制御部と、を備える形状抽出装置。

【符号の説明】

【0081】

- 10 情報処理装置(形状抽出装置)
- 20 ひな形生成部(特定部、抽出部、生成部、記憶制御部)
- 32 ひな形DB(記憶部)
- 96 ストレージ(記憶部)

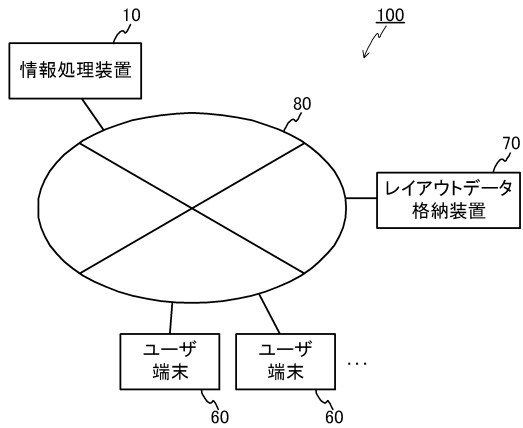
10

20

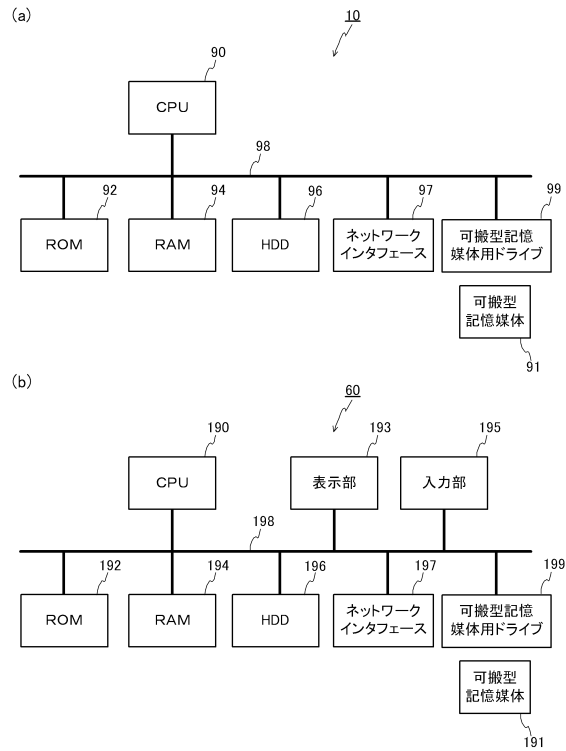
30

40

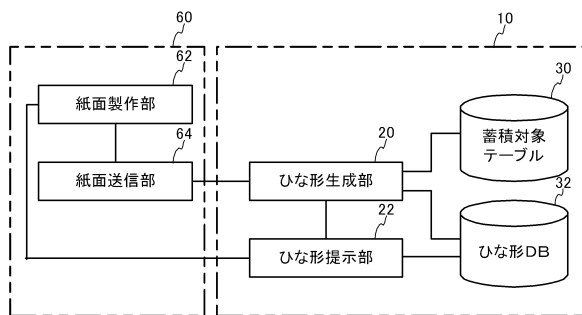
【図1】



【図2】



【図3】



【図4】

＜蓄積対象テーブル30＞

面名	蓄積対象／対象外
一面	対象外
二面	対象外
社会	対象外
生活	対象
くらし	対象
文化	対象
...	...

【図5】

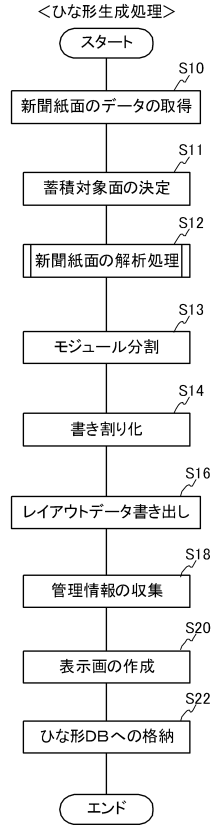
＜ひな形DB32＞

データ種別	作成日時	最終使用日時	使用回数	掲載年月日	掲載曜日	掲載ページ	面名	版名	モジュール名	縦向き	横向き	記事部行数
面	2016/08/01 23:15:30	2016/08/30 18:03:15	3	2016/07/30	土曜日	左	くらし	10版	くらし	2323200	1615680	10
モジュール	2016/08/01 23:15:35		0	2016/07/30	土曜日	左	くらし	10版	金曜アート	3300	6000	
モジュール	2016/08/01 23:15:40		0	2016/07/30	土曜日	左	くらし	10版	連続小販	12800	82450	
モジュール	2016/08/01 23:15:45		0	2016/07/30	土曜日	左	くらし	10版	制作の扉	24800	45680	
モジュール	2016/08/01 23:15:50	2016/10/01 17:01:03	1	2016/07/30	土曜日	左	くらし	10版	夏のストレッチ	12800	162830	
モジュール	2016/08/01 23:15:55		0	2016/07/30	土曜日	左	くらし	10版	投稿募集	35000	213670	

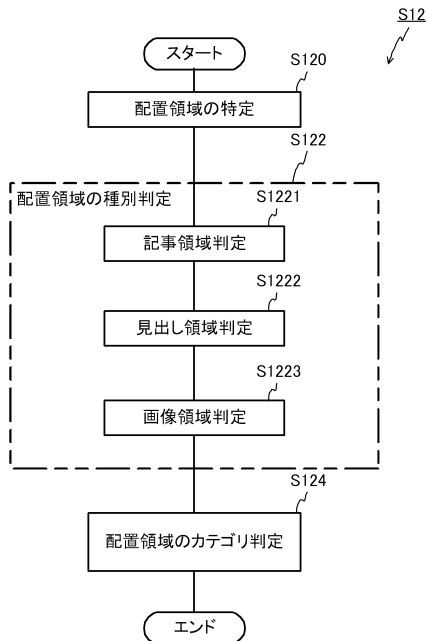
  

レイアウト格納ハス	表示面格納ハス (書き割り)	表示面格納ハス (組み上がり)	モジュール構成情報 格納ハス	記事数	記事行数	画像数	画像の縦横
20160730MM0730 AA10	20160730MM0730AA1 0.T	20160730MM0730AA1 0.J	20160730MM0730AA1 0.001	1	33	1	横
20160730MM0730 AA10	20160730MM0730AA1 0.001.T	20160730MM0730AA1 0.001.J	20160730MM0730AA1 0.002	2	60.30	1	縦
20160730MM0730 AA10	20160730MM0730AA1 0.002.T	20160730MM0730AA1 0.002.J	20160730MM0730AA1 0.003	1	48	1	縦
20160730MM0730 AA10	20160730MM0730AA1 0.003.T	20160730MM0730AA1 0.003.J	20160730MM0730AA1 0.004	3	30.30.40	2	横、縦
20160730MM0730 AA10	20160730MM0730AA1 0.004.T	20160730MM0730AA1 0.004.J	20160730MM0730AA1 0.005	1	52	1	横

【図6】



【図7】

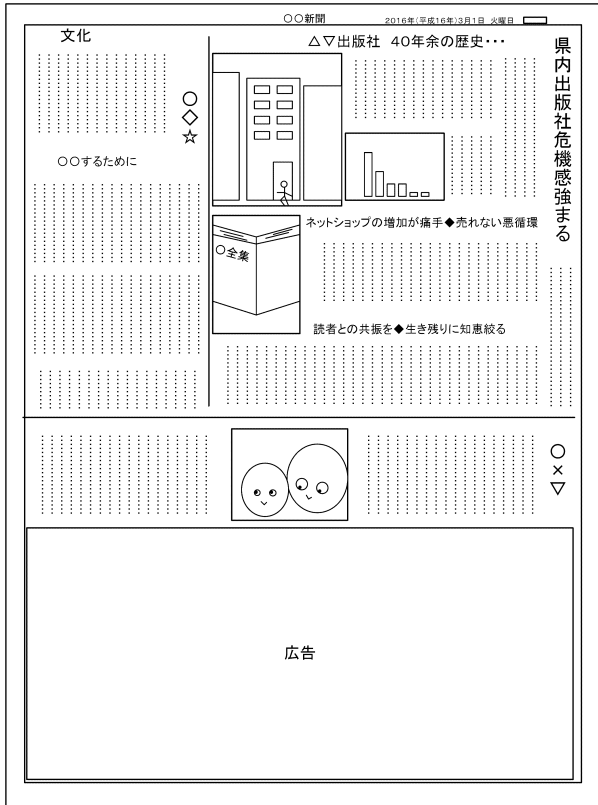


【図8】

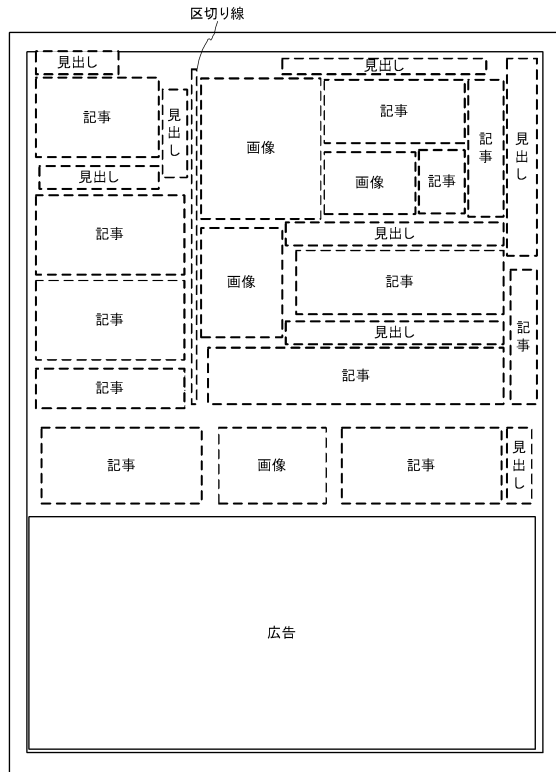
見出し文	単語1	単語2	単語3	単語4	単語5	単語6
日本総人口 初の減少	日本	総	人口	初	の	減少
	名詞	接頭詞	名詞	名詞	助詞	名詞
固有名詞	固有名詞	名詞接続	一般	一般	連体化	サ変接続



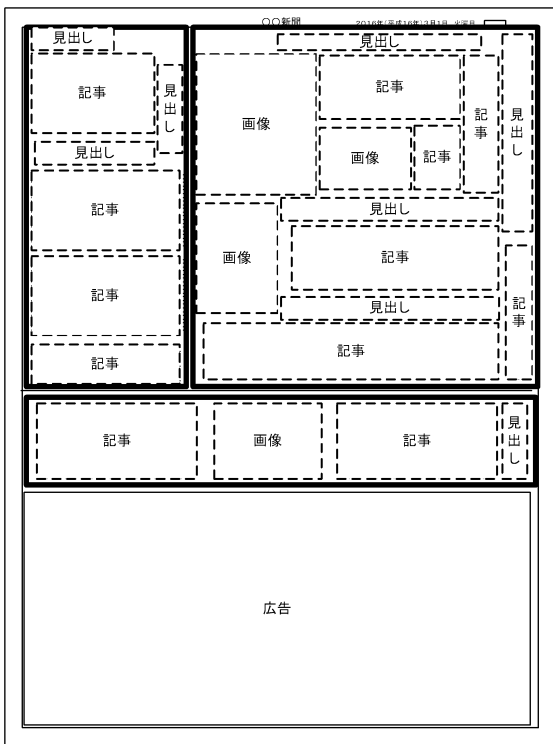
【図 9】



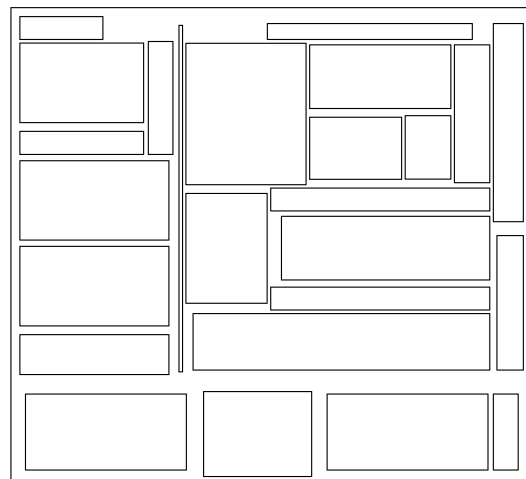
【図 10】



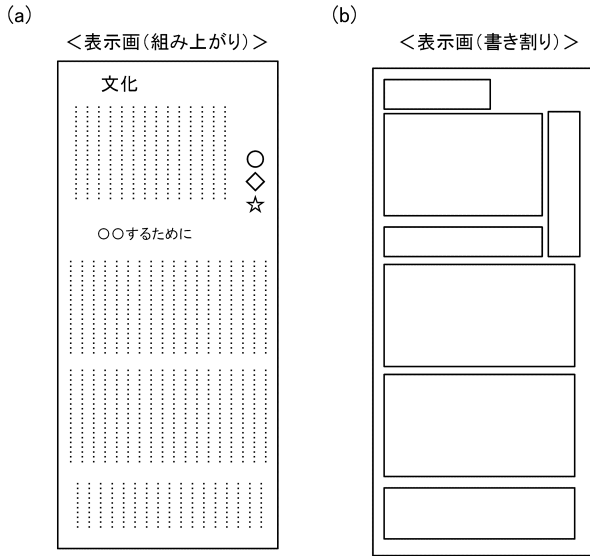
【図 11】



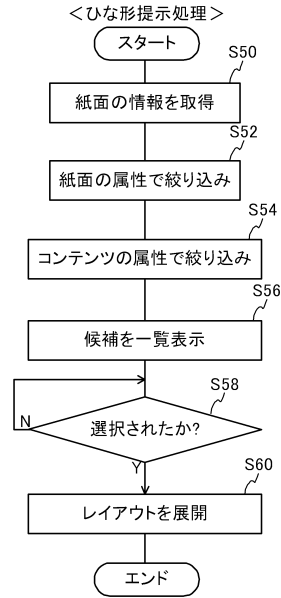
【図 12】



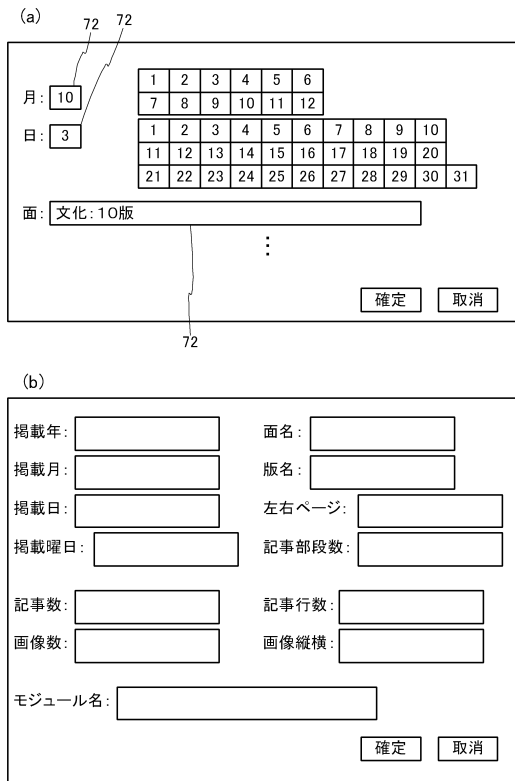
【図13】



【図14】



【図15】



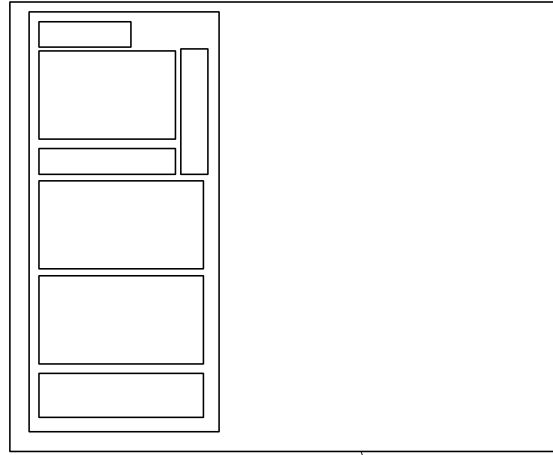
【図16】

画像の縦横	横	横	横
画像数	1	1	1
行数	80/20	100	100
記事数	2	1	1
段数	10	12	10
左右	左	右	左
版名	10版	10版	10版
面名	文化	文化	文化
曜日	1月	8月	15月
月	8	8	8
年	2016	2016	2016
モジュール名	総備名備	総備名備	総備名備
書き割り			
組み上がり			

【図17】

組み上がり		...
書き割り		...
出来上がり		...
モデル名	総論各論	...
年	2016	...
月	8	...
日	1	...
曜日	文化	...
面名	10版	...
左右	左	...
段数	10	...
記事数	2	...
行数	8020	...
画像数	1	...
画像の縦横	横	...

【図18】



---

フロントページの続き

- (56)参考文献 特開2000-251067(JP,A)  
特開昭57-199066(JP,A)  
特表2012-500428(JP,A)  
特開2002-041501(JP,A)  
特開2008-022159(JP,A)  
特開平08-194829(JP,A)  
駱 琴、他2名、ルールベースの適用による日本語新聞紙紙面の構造認識、電子情報通信学会論文誌、日本、社団法人電子情報通信学会、1992年 9月25日、第J75-D-II巻 第9号、第1514~1525頁

(58)調査した分野(Int.Cl., DB名)

G06T 11/60  
G06F 40/00 - 40/197  
G06K 9/18 - 9/44  
G06K 9/54 - 9/60