

(19) 日本国特許庁(JP)

## (12) 公開特許公報(A)

(11) 特許出願公開番号

特開2005-292973  
(P2005-292973A)

(43) 公開日 平成17年10月20日(2005.10.20)

(51) Int.C1.<sup>7</sup>

G06F 17/21

F 1

G06F 17/21

530A

テーマコード(参考)

5B009

審査請求 未請求 請求項の数 11 O L (全 23 頁)

(21) 出願番号

特願2004-103993 (P2004-103993)

(22) 出願日

平成16年3月31日 (2004.3.31)

(71) 出願人 000002369

セイコーエプソン株式会社

東京都新宿区西新宿2丁目4番1号

(74) 代理人 100066980

弁理士 森 哲也

(74) 代理人 100075579

弁理士 内藤 嘉昭

(74) 代理人 100103850

弁理士 崔 秀▲てつ▼

(72) 発明者 三輪 真司

長野県諏訪市大和3丁目3番5号 セイコーエプソン株式会社内

(72) 発明者 永原 敦示

長野県諏訪市大和3丁目3番5号 セイコーエプソン株式会社内

F ターム(参考) 5B009 NA01 NA03 QA06

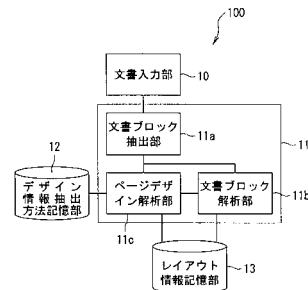
(54) 【発明の名称】 デザイン情報抽出装置、デザイン情報抽出プログラム及びデザイン情報抽出方法

## (57) 【要約】

【課題】既存の文書ファイルからデザイン情報を抽出することが可能なデザイン情報抽出装置、デザイン情報抽出プログラム及びデザイン情報抽出方法を提供する。

【解決手段】デザイン情報抽出装置100を、外部装置等から文書ファイルを入力する文書入力部10と、入力された文書ファイルからレイアウトテンプレート情報及びデザイン情報を抽出するデザイン情報抽出部11と、文書ファイルからデザイン情報を抽出するための抽出方法の情報が記憶されたデザイン情報抽出方法記憶部12と、デザイン情報抽出部11において抽出されたデザイン情報を記憶するレイアウト情報記憶部13とを含んだ構成とした。

【選択図】 図1



**【特許請求の範囲】****【請求項 1】**

文書ファイルから当該文書ファイルの出力内容のデザインに係るデザイン情報を抽出する装置であって、

前記文書ファイルを取得する文書ファイル取得手段と、

前記文書ファイル取得手段によって取得された前記文書ファイルから前記デザイン情報を抽出するデザイン情報抽出手段と、

前記デザイン情報抽出手段によって抽出されたデザイン情報を記憶するデザイン情報記憶手段と、を備えることを特徴とするデザイン情報抽出装置。

**【請求項 2】**

文書ファイルから当該文書ファイルの出力内容のデザインに係るレイアウトテンプレート情報及びデザイン情報を抽出する装置であって、

前記文書ファイルを取得する文書ファイル取得手段と、

前記文書ファイル取得手段によって取得された前記文書ファイルから前記レイアウトテンプレート情報及び前記デザイン情報を抽出するデザイン情報抽出手段と、

前記デザイン情報抽出手段によって抽出されたレイアウトテンプレート情報及びデザイン情報を記憶するデザイン情報記憶手段と、を備えることを特徴とするデザイン情報抽出装置。

**【請求項 3】**

前記文書ファイルの出力内容を、当該出力内容を構成する各構成要素毎に区分する文書ファイル区分手段と、

前記文書ファイル区分手段の区分結果に基づき、前記区分された各構成要素を解析する構成要素解析手段と、を備え、

前記デザイン情報抽出手段は、前記文書ファイル区分手段の区分結果及び前記構成要素解析手段の解析結果に基づき、前記文書ファイルから前記レイアウトテンプレート情報及び前記デザイン情報を抽出するようになっていることを特徴とする請求項 2 記載のデザイン情報抽出装置。

**【請求項 4】**

前記構成要素は、図形情報、文字情報及び線情報のうち少なくとも 1 つを含むことを特徴とする請求項 3 記載のデザイン情報抽出装置。

**【請求項 5】**

前記デザイン情報の抽出方法に係る情報を記憶するデザイン情報抽出方法記憶手段を備え、

前記デザイン情報抽出手段は、前記デザイン情報抽出方法記憶手段によって記憶された前記抽出方法に係る情報を用いて前記文書ファイルから前記デザイン情報を抽出するようになっていることを特徴とする請求項 2 乃至請求項 4 のいずれか 1 項に記載のデザイン情報抽出装置。

**【請求項 6】**

前記デザイン情報は、版面率、画線率、文字のジャンプ率、図形のジャンプ率、図版率、構成要素のグリッド拘束率及び前記各構成要素間の位置揃え情報のうち少なくとも 1 つを含むことを特徴とする請求項 2 乃至請求項 5 のいずれか 1 項に記載のデザイン情報抽出装置。

**【請求項 7】**

前記デザイン情報記憶手段は、前記抽出されたレイアウトテンプレート情報及びデザイン情報を、これらの情報に対応する文書ファイルを明示する明示情報と対応付けて記憶するようになっており、

利用者からの入力情報に適合する前記明示情報の前記レイアウトテンプレート情報及びデザイン情報を、前記デザイン情報記憶手段によって記憶されたレイアウトテンプレート情報及びデザイン情報の中から選択するデザイン情報選択手段を備えることを特徴とする請求項 2 乃至請求項 6 のいずれか 1 項に記載のデザイン情報抽出装置。

10

20

30

40

50

**【請求項 8】**

文書ファイルから当該文書ファイルの出力内容のデザインに係るデザイン情報を抽出するためのプログラムであって、

前記文書ファイルを取得する文書ファイル取得手段、

前記文書ファイル取得手段によって取得された前記文書ファイルから前記デザイン情報を抽出するデザイン情報抽出手段及び、

前記デザイン情報抽出手段によって抽出されたデザイン情報を記憶するデザイン情報記憶手段として実現される処理をコンピュータに実行させるためのプログラムであることを特徴とするデザイン情報抽出プログラム。

**【請求項 9】**

文書ファイルから当該文書ファイルの出力内容のデザインに係るレイアウトテンプレート情報及びデザイン情報を抽出するためのプログラムであって、

前記文書ファイルを取得する文書ファイル取得手段、

前記文書ファイル取得手段によって取得された前記文書ファイルから前記レイアウトテンプレート情報及び前記デザイン情報を抽出するデザイン情報抽出手段及び、

前記デザイン情報抽出手段によって抽出されたレイアウトテンプレート情報及びデザイン情報を記憶するデザイン情報記憶手段として実現される処理をコンピュータに実行させるためのプログラムであることを特徴とするデザイン情報抽出プログラム。

**【請求項 10】**

文書ファイルから当該文書ファイルの出力内容のデザインに係るデザイン情報を抽出するための方法であって、

前記文書ファイルを取得する文書ファイル取得ステップと、

前記文書ファイル取得ステップにおいて取得された前記文書ファイルから前記デザイン情報を抽出するデザイン情報抽出ステップと、

前記デザイン情報抽出ステップにおいて抽出されたデザイン情報を記憶するデザイン情報記憶ステップと、を含むことを特徴とするデザイン情報抽出方法。

**【請求項 11】**

文書ファイルから当該文書ファイルの出力内容のデザインに係るレイアウトテンプレート情報及びデザイン情報を抽出するための方法であって、

前記文書ファイルを取得する文書ファイル取得ステップと、

前記文書ファイル取得ステップにおいて取得された前記文書ファイルから前記レイアウトテンプレート情報及び前記デザイン情報を抽出するデザイン情報抽出ステップと、

前記デザイン情報抽出ステップにおいて抽出されたレイアウトテンプレート情報及びデザイン情報を記憶するデザイン情報記憶ステップと、を含むことを特徴とするデザイン情報抽出方法。

**【発明の詳細な説明】****【技術分野】****【0001】**

本発明は、文書のデザインを支援する装置に係り、特に、既存の文書ファイルからデザイン情報を抽出することが可能なデザイン情報抽出装置、デザイン情報抽出プログラム及びデザイン情報抽出方法に関する。

**【背景技術】****【0002】**

従来、PC等において、ワープロソフトなどを用いて文書ファイルを作成し、当該作成した文書ファイルをプリンタにより紙面上などにプリントアウトすることが可能である。しかし、印刷紙面上における文書の位置や文字サイズ等の単純な体裁の整った文書をプリントアウトすることは簡易にできても、一般の人がプロのグラフィックデザイナが作成するようなデザイン内容の文書を作成することは極めて困難である。このような課題に対して、予めデザインされたテンプレートを複数用意し、使用者はこの中から所望のデザインのテンプレートを選択し当該選択したテンプレート上に文書の内容を記入することで、一

般の人でも一定のデザインの文書を作成することを可能とする技術がある。このような技術としては、例えば、特許文献1に記載の文書作成装置及び特許文献2に記載の作品作成支援システムがある。

#### 【0003】

ここで、特許文献1の文書作成装置は、例文（体裁のサンプル）及び原文を論理解析し、例文と原文との論理構造が一致する場合に、原文の体裁を例文の体裁に自動的に割り付けるものである。

また、特許文献2の作品作成支援システムは、マルチメディアプレゼンテーションの作成において、プレゼンテーションの特徴情報（種類、対象者、雰囲気）に対応して使用する色、文字フォント、レイアウト規則などの知識ベースを持ち、特徴の指定により、知識ベースの知識に従ってテンプレートを作成するものである。

【特許文献1】特開平7-93329号公報

【特許文献2】特開平8-212258号公報

【発明の開示】

【発明が解決しようとする課題】

#### 【0004】

しかしながら、上記特許文献1の文書作成装置においては、例文と原文との論理構造が一致する場合にしか上記した例文の体裁への割り付けが行われないという問題がある。また、その実施例においては、体裁として章タイトルの字下げや改行等の文書の配置に関するものと、文字に付す下線や文字の倍率などの文字飾りに関するものが例示されているが、この程度では、プロのグラフィックデザイナが作成するようなデザイン内容の文書を作成することは困難である。

#### 【0005】

また、上記特許文献2の作品作成支援システムにおいては、知識ベースの構築が必要であることと、特徴情報に対応したテンプレート情報を作成することは専門的知識が必要であることからシステムを構成するのに手間がかかる。また、プレゼンテーションの特徴とデザイン知識の関係が直感的ではなく、デザインの選択が困難である。

そこで、本発明は、このような従来の技術の有する未解決の課題に着目してなされたものであって、既存の文書ファイルからデザイン情報を抽出することが可能なデザイン情報抽出装置、デザイン情報抽出プログラム及びデザイン情報抽出方法を提供することを目的としている。

【課題を解決するための手段】

#### 【0006】

〔発明1〕 上記目的を達成するために、発明1のデザイン情報抽出装置は、文書ファイルから当該文書ファイルの出力内容のデザインに係るデザイン情報を抽出する装置であって、

前記文書ファイルを取得する文書ファイル取得手段と、

前記文書ファイル取得手段によって取得された前記文書ファイルから前記デザイン情報を抽出するデザイン情報抽出手段と、

前記デザイン情報抽出手段によって抽出されたデザイン情報を記憶するデザイン情報記憶手段と、を備えることを特徴としている。

#### 【0007】

このような構成であれば、文書ファイル取得手段によって、前記文書ファイルを取得することが可能であり、デザイン情報抽出手段によって、前記文書ファイル取得手段によって取得された前記文書ファイルから前記デザイン情報を抽出することが可能であり、デザイン情報記憶手段によって、前記デザイン情報抽出手段によって抽出されたデザイン情報を記憶することが可能である。

#### 【0008】

従って、既存の文書ファイルから、その出力内容のデザインに係るデザイン情報を抽出することができるので、所望のデザインの文書ファイルからデザイン情報を抽出し、この

抽出したデザイン情報を用いて文書をデザインすることで、所望のデザインの文書に含まれるデザイン情報の反映された文書を容易に作成することができるという効果が得られる。なお、抽出されたデザイン情報は、文書の作成に用いるに限らず、テンプレートの作成に用いることも可能である。

#### 【0009】

ここで、文書ファイルの出力内容とは、文書ファイルの文書の画面への表示出力や紙面等への印刷出力などが含まれる。

また、文書ファイルは、文字情報、画像情報、文字や画像のパレット情報、文書や画像のレイアウト情報（レイアウトテンプレート情報）などから構成される。つまり、文書ファイルとしては、ワープロソフト等で作成された文書ファイル等が含まれる。また、文書ファイルの中には文字情報のみで構成されるものもある。以下、発明2のデザイン情報抽出装置、発明8及び9のデザイン情報抽出プログラム、並びに発明15及び16のデザイン情報抽出方法において同じである。

#### 【0010】

また、デザイン情報とは、文書ファイルの出力内容のデザインに係る情報であり、文書ファイルにおける、上記したレイアウトテンプレート情報に含まれない、文字や画像の配置位置の情報、文字や画像のサイズの情報、文字や画像（枠線や背景等を含む）の色情報などを含むものである。以下、発明2のデザイン情報抽出装置、発明8及び9のデザイン情報抽出プログラム、並びに発明15及び16のデザイン情報抽出方法において同じである。

#### 【0011】

また、レイアウトテンプレートとは、例えば、デザイナ等があらかじめ設計したレイアウトのひな形を基に生成されたもので、具体的には、所定のレイアウト領域に複数の情報格納枠を配置したレイアウトを規定したものであり、レイアウト時には、レイアウトテンプレートに従って、レイアウト対象となるレイアウト要素（例えば、文字情報、画像情報）を各情報格納枠に格納することによりレイアウト要素をレイアウト領域に配置していく。以下、発明2のデザイン情報抽出装置、発明8及び9のデザイン情報抽出プログラム、並びに発明15及び16のデザイン情報抽出方法において同じである。

#### 【0012】

〔発明2〕 一方、上記目的を達成するために、発明2のデザイン情報抽出装置は、文書ファイルから当該文書ファイルの出力内容のデザインに係るレイアウトテンプレート情報及びデザイン情報を抽出する装置であって、

前記文書ファイルを取得する文書ファイル取得手段と、

前記文書ファイル取得手段によって取得された前記文書ファイルから前記レイアウトテンプレート情報及び前記デザイン情報を抽出するデザイン情報抽出手段と、

前記デザイン情報抽出手段によって抽出されたレイアウトテンプレート情報及びデザイン情報を記憶するデザイン情報記憶手段と、を備えることを特徴としている。

#### 【0013】

このような構成であれば、文書ファイル取得手段によって、前記文書ファイルを取得することが可能であり、デザイン情報抽出手段によって、前記文書ファイル取得手段によって取得された前記文書ファイルから前記レイアウトテンプレート情報及び前記デザイン情報を抽出することが可能であり、デザイン情報記憶手段によって、前記デザイン情報抽出手段によって抽出されたレイアウトテンプレート情報及びデザイン情報を記憶することができる。

#### 【0014】

従って、既存の文書ファイルから、その出力内容のデザインに係るレイアウトテンプレート情報及びデザイン情報を抽出することができるので、所望のデザインの文書ファイルからレイアウトテンプレート情報及びデザイン情報を抽出し、この抽出したレイアウトテンプレート情報及びデザイン情報を用いて文書をデザインすることで、所望のデザインの文書のデザインが反映された文書を容易に作成することができるという効果が得られる。

10

20

30

40

50

ここで、抽出されたデザイン情報は、テンプレートの作成に用いたり、そのまま文書の作成に用いるなどすることが可能である。

【0015】

〔発明3〕更に、発明3のデザイン情報抽出装置は、発明2のデザイン情報抽出装置において、前記文書ファイルの出力内容を、当該出力内容を構成する各構成要素毎に区分する文書ファイル区分手段と、

前記文書ファイル区分手段の区分結果に基づき、前記区分された各構成要素を解析する構成要素解析手段と、を備え、

前記デザイン情報抽出手段は、前記文書ファイル区分手段の区分結果及び前記構成要素解析手段の解析結果に基づき、前記文書ファイルから前記レイアウトテンプレート情報及び前記デザイン情報を抽出するようになっていることを特徴としている。

【0016】

このような構成であれば、前記文書ファイル区分手段によって、前記文書ファイルの出力内容を、当該出力内容を構成する各構成要素毎に区分することが可能であり、構成要素解析手段によって、前記文書ファイル区分手段の区分結果に基づき、前記区分された各構成要素を解析することが可能であり、前記デザイン情報抽出手段は、前記文書ファイル区分手段の区分結果及び前記構成要素解析手段の解析結果に基づき、前記文書ファイルから前記レイアウトテンプレート情報及び前記デザイン情報を抽出することが可能である。

【0017】

従って、文書をスキャナで取り込んで生成した画像データから成る文書ファイルや、文書をデジタルカメラで撮影して生成した画像データから成る文書ファイル、ページ記述言語（PDL）で作成された文書ファイルなどの、文書を画像データ化して成る文書ファイルの出力内容を各構成要素に区分し且つ各構成要素を解析（例えば、文字認識等のパターン認識）することができるので、画像データから成る文書ファイルからもレイアウトテンプレート情報及びデザイン情報を抽出を行うことができるという効果が得られる。更に、ワープロ等で作成された文書ファイルに含まれる画像部分の解析等も可能となるので、画像内に含まれる文字や図形等の認識により、より正確にデザイン情報の抽出ができるという効果も得られる。

【0018】

〔発明4〕更に、発明4のデザイン情報抽出装置は、発明3のデザイン情報抽出装置において、前記構成要素は、図形情報、文字情報及び線情報のうち少なくとも1つを含むことを特徴としている。

このような構成であれば、文書をスキャナで取り込んで生成した画像データから成る文書ファイルや、デジタルカメラで撮影して生成した画像データから成る文書ファイル、ページ記述言語（PDL）で作成された文書ファイルなどを、図形情報、文字情報、線情報といった各構成要素に区分することができ、簡易にレイアウトテンプレート情報及びデザイン情報の抽出を行うことができるという効果が得られる。

ここで、図形情報は、文書ファイル中の写真、イラスト、模様等の画像情報を含む。以下、発明11のデザイン情報抽出プログラム及び発明18のデザイン情報抽出方法において同じである。

【0019】

〔発明5〕更に、発明5のデザイン情報抽出装置は、発明2乃至4のいずれか1のデザイン情報抽出装置において、前記デザイン情報の抽出方法に係る情報を記憶するデザイン情報抽出方法記憶手段を備え、

前記デザイン情報抽出手段は、前記デザイン情報抽出方法記憶手段によって記憶された前記抽出方法に係る情報を用いて前記文書ファイルから前記デザイン情報を抽出するようになっていることを特徴としている。

【0020】

このような構成であれば、デザイン情報抽出方法記憶手段によって、前記デザイン情報の抽出方法に係る情報を記憶することが可能であり、前記デザイン情報抽出手段は、前記

10

20

30

40

50

デザイン情報抽出方法記憶手段によって記憶された前記抽出方法に係る情報を用いて前記文書ファイルから前記デザイン情報を抽出することが可能である。

従って、例えば、デザイン情報の種類毎に抽出方法に係る情報（例えば、各抽出方法を実現するプログラム）を記憶することにより、個々の抽出方法の改良や新たな種類のデザイン情報に対応した抽出方法の追加等を簡易に行うことが可能である。

#### 【0021】

〔発明6〕更に、発明6のデザイン情報抽出装置は、発明2乃至5のいずれか1のデザイン情報抽出装置において、前記デザイン情報は、版面率、画線率、文字のジャンプ率、図形のジャンプ率、図版率、構成要素のグリッド拘束率及び前記各構成要素間の位置揃え情報のうち少なくとも1つを含むことを特徴としている。

このような構成であれば、版面率、画線率、文字のジャンプ率、図形のジャンプ率、図版率、構成要素のグリッド拘束率及び前記各構成要素間の位置揃え情報等の抽象的なデザイン情報を用いて文書を作成することが可能であるので、掲載情報に対するレイアウトの自由度を高くできるという効果が得られる。

#### 【0022】

ここで、版面率は、文書の全ての文書ブロック（文字情報により構成される領域）を含む外接四角形の面積の、紙面全体の面積に対する割合、あるいは、文書全ての文書ブロックの面積の合計の、紙面全体の面積に対する割合である。

また、画線率は、紙面全体、あるいは、文書ブロックの面積に対する、印字部分の面積の割合である。

#### 【0023】

また、文字のジャンプ率は、最大サイズの文字と最小サイズの文字との大きさの比率である。

また、画像のジャンプ率は、画像のサイズや色の明るさ等の画像に係る数値についての、最大値と最小値との比率である。

また、図版率は、文書中に占める画像部分の割合である。

#### 【0024】

また、グリッド拘束率は、文書ブロックの位置が文書中のグリッド（仮想的な基準線）に沿っている割合である。

また、前記各構成要素間の位置揃え情報は、複数の文書ブロックの対応する辺同士や中心線が同一直線上になるレイアウトの情報である。

上記、版面率、画線率、文字のジャンプ率、画像のジャンプ率、図版率、グリッド拘束率及び前記各構成要素間の位置揃え情報についての記載は、発明13のデザイン情報抽出プログラム及び発明20のデザイン情報抽出方法において同じである。

#### 【0025】

〔発明7〕更に、発明7のデザイン情報抽出装置は、発明2乃至6のいずれか1のデザイン情報抽出装置において、前記デザイン情報記憶手段は、前記抽出されたレイアウトテンプレート情報及びデザイン情報を、これらの情報に対応する文書ファイルを明示する明示情報と対応付けて記憶するようになっており、

利用者からの入力情報に適合する前記明示情報の前記レイアウトテンプレート情報及びデザイン情報を、前記デザイン情報記憶手段によって記憶されたレイアウトテンプレート情報及びデザイン情報の中から選択するデザイン情報選択手段を備えることを特徴としている。

#### 【0026】

このような構成であれば、前記デザイン情報記憶手段は、前記抽出されたレイアウトテンプレート情報及びデザイン情報を、これらの情報に対応する文書ファイルを明示する明示情報と対応付けて記憶することが可能であり、デザイン情報選択手段によって、利用者からの入力情報に適合する前記明示情報の前記レイアウトテンプレート情報及びデザイン情報を前記デザイン情報記憶手段によって記憶されたレイアウトテンプレート情報及びデザイン情報を前記デザイン情報の中から選択することが可能である。

## 【0027】

従って、抽出したテンプレート情報及びデザイン情報を用いて文書を作成するときに、抽出元の文書ファイルを明示した明示情報によってデザイン情報記憶手段によって記憶されたデザイン情報の中から所望のテンプレート情報及びデザイン情報を選択することができるので、記憶されたテンプレート情報及びデザイン情報の数が多くなっても所望の情報を容易に選択することができ、且つ、例えば、抽出元の文書ファイルのタイトルをテンプレート情報及びデザイン情報に対応付けることで、選択したテンプレート情報及びデザイン情報によって作成された文書ファイルの出力イメージが直感的に理解し易くなるという効果が得られる。

## 【0028】

〔発明8〕 一方、上記目的を達成するために、発明8のデザイン情報抽出プログラムは、文書ファイルから当該文書ファイルの出力内容のデザインに係るデザイン情報を抽出するためのプログラムであって、

前記文書ファイルを取得する文書ファイル取得手段、

前記文書ファイル取得手段によって取得された前記文書ファイルから前記デザイン情報を抽出するデザイン情報抽出手段及び、

前記デザイン情報抽出手段によって抽出されたデザイン情報を記憶するデザイン情報記憶手段として実現される処理をコンピュータに実行させるためのプログラムであることを特徴としている。

ここで、本発明は、発明1のデザイン情報抽出装置に適用可能なプログラムであり、これにより発明1のデザイン情報抽出装置と同等の効果が得られる。

## 【0029】

〔発明9〕 一方、上記目的を達成するために、発明9のデザイン情報抽出プログラムは、文書ファイルから当該文書ファイルの出力内容のデザインに係るレイアウトテンプレート情報及びデザイン情報を抽出するためのプログラムであって、

前記文書ファイルを取得する文書ファイル取得手段、

前記文書ファイル取得手段によって取得された前記文書ファイルから前記レイアウトテンプレート情報及び前記デザイン情報を抽出するデザイン情報抽出手段及び、

前記デザイン情報抽出手段によって抽出されたレイアウトテンプレート情報及びデザイン情報を記憶するデザイン情報記憶手段として実現される処理をコンピュータに実行させるためのプログラムであることを特徴としている。

ここで、本発明は、発明2のデザイン情報抽出装置に適用可能なプログラムであり、これにより発明2のデザイン情報抽出装置と同等の効果が得られる。

## 【0030】

〔発明10〕 更に、発明10のデザイン情報抽出プログラムは、発明9のデザイン情報抽出プログラムにおいて、前記文書ファイルの出力内容を、当該出力内容を構成する各構成要素毎に区分する文書ファイル区分手段及び、

前記文書ファイル区分手段の区分結果に基づき、前記区分された各構成要素を解析する構成要素解析手段として実現される処理をコンピュータに実行させるためのプログラムを更に含み、

前記デザイン情報抽出手段は、前記文書ファイル区分手段の区分結果及び前記構成要素解析手段の解析結果に基づき、前記文書ファイルから前記レイアウトテンプレート情報及び前記デザイン情報を抽出するようになっていることを特徴としている。

ここで、本発明は、発明3のデザイン情報抽出装置に適用可能なプログラムであり、これにより発明3のデザイン情報抽出装置と同等の効果が得られる。

## 【0031】

〔発明11〕 更に、発明11のデザイン情報抽出プログラムは、発明10のデザイン情報抽出プログラムにおいて、前記構成要素の種類は、図形情報、文字情報及び線情報のうち少なくとも1つを含むことを特徴としている。

ここで、本発明は、発明4のデザイン情報抽出装置に適用可能なプログラムであり、こ

10

20

30

40

50

れにより発明 4 のデザイン情報抽出装置と同等の効果が得られる。

【 0 0 3 2 】

〔発明 12〕 更に、発明 12 のデザイン情報抽出プログラムは、発明 9 乃至 11 のいずれか 1 のデザイン情報抽出プログラムにおいて、前記デザイン情報の抽出方法に係る情報を記憶するデザイン情報抽出方法記憶手段として実現される処理をコンピュータに実行させるためのプログラムを更に含み、

前記デザイン情報抽出手段は、前記デザイン情報抽出方法記憶手段によって記憶された前記抽出方法に係る情報を用いて前記文書ファイルから前記デザイン情報を抽出するようになっていることを特徴としている。

ここで、本発明は、発明 5 のデザイン情報抽出装置に適用可能なプログラムであり、これにより発明 5 のデザイン情報抽出装置と同等の効果が得られる。

【 0 0 3 3 】

〔発明 13〕 更に、発明 13 のデザイン情報抽出プログラムは、発明 9 乃至 12 のいずれか 1 のデザイン情報抽出プログラムにおいて、前記デザイン情報は、版面率、画線率、文字のジャンプ率、図形のジャンプ率、図版率、構成要素のグリッド拘束率及び前記各構成要素間の位置揃え情報のうち少なくとも 1 つを含むことを特徴としている。

ここで、本発明は、発明 6 のデザイン情報抽出装置に適用可能なプログラムであり、これにより発明 6 のデザイン情報抽出装置と同等の効果が得られる。

【 0 0 3 4 】

〔発明 14〕 更に、発明 14 のデザイン情報抽出プログラムは、発明 9 乃至 13 のいずれか 1 のデザイン情報抽出プログラムにおいて、前記デザイン情報記憶手段は、前記抽出されたレイアウトテンプレート情報及びデザイン情報を、これらの情報に対応する文書ファイルを明示する明示情報と対応付けて記憶するようになっており、

利用者からの入力情報に適合する前記明示情報の前記レイアウトテンプレート情報及びデザイン情報を、前記デザイン情報記憶手段によって記憶されたレイアウトテンプレート情報及びデザイン情報の中から選択するデザイン情報選択手段として実現される処理をコンピュータに実行させるためのプログラムを更に含むことを特徴としている。

ここで、本発明は、発明 7 のデザイン情報抽出装置に適用可能なプログラムであり、これにより発明 7 のデザイン情報抽出装置と同等の効果が得られる。

【 0 0 3 5 】

〔発明 15〕 一方、発明 15 のデザイン情報抽出方法は、文書ファイルから当該文書ファイルの出力内容のデザインに係るデザイン情報を抽出するための方法であって、

前記文書ファイルを取得する文書ファイル取得ステップと、

前記文書ファイル取得ステップにおいて取得された前記文書ファイルから前記デザイン情報を抽出するデザイン情報抽出ステップと、

前記デザイン情報抽出ステップにおいて抽出されたデザイン情報を記憶するデザイン情報記憶ステップと、を含むことを特徴としている。

これにより、発明 1 のデザイン情報抽出装置と同等の効果が得られる。

【 0 0 3 6 】

〔発明 16〕 一方、発明 16 のデザイン情報抽出方法は、文書ファイルから当該文書ファイルの出力内容のデザインに係るレイアウトテンプレート情報及びデザイン情報を抽出するための方法であって、

前記文書ファイルを取得する文書ファイル取得ステップと、

前記文書ファイル取得ステップにおいて取得された前記文書ファイルから前記レイアウトテンプレート情報及び前記デザイン情報を抽出するデザイン情報抽出ステップと、

前記デザイン情報抽出ステップにおいて抽出されたレイアウトテンプレート情報及びデザイン情報を記憶するデザイン情報記憶ステップと、を含むことを特徴としている。

これにより、発明 2 のデザイン情報抽出装置と同等の効果が得られる。

【 0 0 3 7 】

〔発明 17〕 更に、発明 17 のデザイン情報抽出方法は、発明 16 のデザイン情報抽

10

20

30

40

50

出方法において、前記文書ファイル取得ステップにおいては、前記文書ファイルの出力内容を、当該出力内容を構成する各構成要素毎に区分する文書ファイル区分ステップと、

前記文書ファイル区分ステップにおける区分結果に基づき、前記区分された各構成要素を解析する構成要素解析ステップとを更に含み、

前記デザイン情報抽出ステップにおいては、前記文書ファイル区分ステップにおける区分結果及び前記構成要素解析ステップにおける解析結果に基づき、前記文書ファイルから前記レイアウトテンプレート情報及び前記デザイン情報を抽出することを特徴としている。

これにより、発明3のデザイン情報抽出装置と同等の効果が得られる。

#### 【0038】

〔発明18〕更に、発明18のデザイン情報抽出方法は、発明17のデザイン情報抽出方法において、前記構成要素の種類は、図形情報、文字情報及び線情報のうち少なくとも1つを含むことを特徴としている。

これにより、発明4のデザイン情報抽出装置と同等の効果が得られる。

#### 【0039】

〔発明19〕更に、発明19のデザイン情報抽出方法は、発明16乃至18のいずれか1のデザイン情報抽出方法において、前記デザイン情報の抽出方法に係る情報を記憶するデザイン情報抽出方法記憶ステップを更に含み、

前記デザイン情報抽出ステップにおいては、前記デザイン情報抽出方法記憶ステップにおいて記憶された前記抽出方法に係る情報を用いて前記文書ファイルから前記デザイン情報を抽出することを特徴としている。

これにより、発明5のデザイン情報抽出装置と同等の効果が得られる。

#### 【0040】

〔発明20〕更に、発明20のデザイン情報抽出方法は、発明16乃至19のいずれか1のデザイン情報抽出方法において、前記デザイン情報は、版面率、画線率、文字のジャンプ率、図形のジャンプ率、図版率、構成要素のグリッド拘束率及び前記各構成要素間の位置揃え情報のうち少なくとも1つを含むことを特徴としている。

これにより、発明6のデザイン情報抽出装置と同等の効果が得られる。

#### 【0041】

〔発明21〕更に、発明21のデザイン情報抽出方法は、発明15乃至20のいずれか1のデザイン情報抽出方法において、前記デザイン情報記憶ステップにおいては、前記抽出されたレイアウトテンプレート情報及びデザイン情報を、これらの情報に対応する文書ファイルを明示する明示情報と対応付けて記憶し、

利用者からの入力情報に適合する前記明示情報の前記レイアウトテンプレート情報及びデザイン情報を、前記デザイン情報記憶ステップにおいて記憶されたレイアウトテンプレート情報及びデザイン情報の中から選択するデザイン情報選択ステップと、を更に含むことを特徴としている。

これにより、発明7のデザイン情報抽出装置と同等の効果が得られる。

#### 【発明を実施するための最良の形態】

#### 【0042】

以下、本発明の実施の形態を図面に基づいて説明する。図1～図10は、本発明に係るデザイン情報抽出装置、デザイン情報抽出プログラム及びデザイン情報抽出方法の実施の形態を示す図である。

#### 【0043】

##### 〔第1の実施の形態〕

まず、本発明の第1の実施の形態を図面に基づいて説明する。図1～図6は、本発明に係るデザイン情報抽出装置、デザイン情報抽出プログラム及びデザイン情報抽出方法の第1の実施の形態を示す図である。

#### 【0044】

まず、本発明の第1の実施の形態に係るデザイン情報抽出装置100の構成を図1に基

10

20

30

40

50

づいて説明する。図1は、本発明に係るデザイン情報抽出装置、デザイン情報抽出プログラム及びデザイン情報抽出方法を適用したデザイン情報抽出装置100の構成を示すプロック図である。

デザイン情報抽出装置100は、外部装置等から文書ファイルを入力する文書入力部10と、入力された文書ファイルからレイアウトテンプレート情報及びデザイン情報を抽出するデザイン情報抽出部11と、文書ファイルからデザイン情報を抽出するための抽出方法の情報が記憶されたデザイン情報抽出方法記憶部12と、デザイン情報抽出部11において抽出されたレイアウトテンプレート情報及びデザイン情報を記憶するレイアウト情報記憶部13とを含んだ構成となっている。

#### 【0045】

文書入力部10は、ワープロソフト等の文書エディタで作成された文書ファイル（以下、ワープロ文書ファイルと称す）や、イメージスキャナ、デジタルカメラ等により紙上に印字された文字や図形を画像データ化した文書ファイル（以下、画像文書ファイルと称す）、ページ記述言語（PDL）により作成された文書ファイル（以下、PDL文書ファイルと称す）等を入力でき、当該入力された文書ファイルをデザイン情報抽出部11に伝送する。

#### 【0046】

ここで、文書ファイルは、デザイナ等がその出力内容をデザインした既存の文書ファイルや、デザイナ等がデザインしたポスターやPOP（point-of-purchase advertising）等をスキャナやデジタルカメラ等で画像データ化したものを含む。

デザイン情報抽出部11は、文書ブロック抽出部11aと、文書ブロック解析部11bと、ページデザイン解析部11cとを含んだ構成となっている。

#### 【0047】

文書ブロック抽出部11aは、文書入力部10から伝送された文書ファイルから文字のブロックや図形のブロックなどの文書を構成する構成要素毎のブロックを抽出する。抽出結果は、文書ブロック解析部11bに伝送される。

ここで、PDL文書ファイル、及び、画像文書ファイルは、ワープロ文書ファイルとは異なり、背景、文字、図形、枠線や罫線等の線情報などの構成要素が明らかになっておらず、特に画像文書ファイルは、各構成要素の配置位置や大きさ等のレイアウトテンプレート情報が一切含まれていないため、構成要素の位置や大きさ等が解らない。従って、文書ブロック抽出部11aにおいて、まずPDL文書ファイル及び画像文書ファイルから各構成要素を抽出する必要がある。ここで、PDL文書ファイルについては、文書中のどの領域にテキストデータ、画像データ、ベクターデータが存在するかを解析する。

#### 【0048】

文書ブロック解析部11bは、文書ブロック抽出部11aにおいて抽出された各構成要素の画像を解析して、各画像に含まれる情報を認識する。例えば、構成要素の種類（文字、図形、表など）の認識に加え、画像に文字が含まれている場合は、その文字の認識や、文字の書体（明朝体、ゴシック体など）、文字サイズ等も認識する。他に、画像が表であった場合の表の構成の認識等も含む。解析結果は、ページデザイン解析部11cに伝送される。ここで、PDL形式のファイルについては、文書ブロック抽出部11aにおいて抽出された画像データの部分に対して文字認識を行う。

#### 【0049】

ページデザイン解析部11cは、入力された文書ファイルがワープロ文書ファイルである場合は、文書ファイルを直接解析してレイアウトテンプレート情報を抽出すると共に、デザイン情報抽出方法記憶部12に記憶されたデザイン情報の抽出方法の情報を用いて文書ファイルからデザイン情報を抽出する。一方、文書ファイルが画像文書ファイル及びPDL文書ファイルである場合は、文書ブロック解析部11bの解析結果（認識結果）に基づき文書ファイルのレイアウトテンプレート情報を抽出すると共に、デザイン情報抽出方法記憶部12に記憶されたデザイン情報の抽出方法の情報を用いて文書ファイルからデザイン情報を抽出する。これら抽出されたレイアウトテンプレート情報及びデザイン情報は

10

20

30

40

50

、文書ファイルの明示情報（例えば、文書ファイルのタイトル）と共にレイアウト情報記憶部13に伝送される。

【0050】

ここで、デザイン情報とは、版面率、画線率、文字のジャンプ率、図形のジャンプ率、図版率、ブロックのグリッド拘束率、前記各構成要素間の位置揃え情報などが含まれる。

また、レイアウトテンプレート情報は、情報格納枠の数、各情報格納枠ごとの大きさおよび配置位置、並びに情報格納枠の種別のほか、文字情報のフォントの大きさ、種別および色彩と、文字間隔や行ピッチと、写真情報または図形情報の品質、大きさ、アスペクト比およびトリミング等を含むものである。

【0051】

また、情報格納枠としては、レイアウト要素の各種別に対応し、文字情報を格納するための文字ブロック、画像情報を格納するための画像ブロック等がある。

デザイン情報抽出方法記憶部12は、文書ファイルから上記した各デザイン情報を抽出するためのルールが記憶されたものである。本実施の形態においては、各デザイン情報毎に文書ファイルからデザイン情報を抽出するためのプログラムが記憶されており、ページデザイン解析部11cで実行されるデザイン情報抽出用のプログラム（メインルーチン）においてサブルーチンとして呼び出され実行されることで、文書ファイルから各デザイン情報を抽出する。

【0052】

レイアウト情報記憶部13は、デザイン情報抽出部11において抽出されたレイアウトテンプレート情報及びデザイン情報を記憶するものである。本実施の形態においては、レイアウトテンプレート情報及びデザイン情報は、抽出元の文書ファイルを明示する明示情報（元文書のタイトル等）と対応付けて記憶される。

ここで、本実施の形態において、デザイン情報抽出装置100は、図示しないが、上記各部を制御するためのプログラムの記憶された記憶媒体と、これらのプログラムを実行するためのCPUと、プログラムの実行に必要なデータを記憶するRAMと、を備えている。そして、上記CPUにより前記プログラムを実行することによって上記各部の処理を実現するものである。

【0053】

また、記憶媒体とは、RAM、ROM等の半導体記憶媒体、FD、HD等の磁気記憶型記憶媒体、CD、CDV、LD、DVD等の光学的読み取り方式記憶媒体、MO等の磁気記憶型／光学的読み取り方式記憶媒体であって、電子的、磁気的、光学的等の読み取り方法のいかんにかかわらず、コンピュータで読み取り可能な記憶媒体であれば、あらゆる記憶媒体を含むものである。

【0054】

次に、図2に基づき、デザイン情報である、版面率、画線率、文字のジャンプ率、図形のジャンプ率、図版率、ブロックのグリッド拘束率、各文書ブロック間の位置揃え情報について具体的に説明する。ここで、図2(a)は、紙面と版面との関係を示す図であり、(b)は、文書に対するグリッド線の一例を示す図である。また、文書ブロックとは、文書を構成する各ブロック（構成要素）を示し、以下、当該文書ブロックのうち、文字情報によって構成されるブロックを文字ブロック、図形（写真や絵などを含む）情報によって構成されるブロックを画像ブロックと各々称することとする。

【0055】

まず、版面率について説明する。版面率は、文書を構成する全てのブロックを含む外接四角形の面積の、紙面全体の面積に対する割合、あるいは、文書を構成する全ての文書ブロックの合計面積の、紙面全体の面積に対する割合である。つまり、前者は、図2(a)に示すように、第1～第3文字ブロック22～24並びに第1及び第2画像ブロック25及び26を含む版面21（外接四角形）の面積の、紙面20の面積に対する割合となる。従って、紙面20の面積が $744 \text{ cm}^2$ で、版面21の面積が $690 \text{ cm}^2$ だとすると、版面率は、 $690 / 744 \times 100 = 92.7\%$ となる。一方、後者は、第1～第3文

10

20

30

40

50

字ブロック 22～24 並びに第1及び第2画像ブロック 25 及び 26 の各面積の合計の、紙面 20 の面積に対する割合となる。従って、第1～第3文字ブロック 22～24 の面積が、それぞれ  $174 \text{ cm}^2$ 、 $60 \text{ cm}^2$  及び  $24 \text{ cm}^2$  で、第1及び第2画像ブロック 25 及び 26 の各面積が、それぞれ  $49 \text{ cm}^2$  及び  $2 \text{ cm}^2$  だとすると、全ての文書ブロックの合計面積は、 $174 + 60 + 24 + 49 + 2 = 309 (\text{cm}^2)$  となるので、版面率は、 $309 / 744 \times 100 = 41.5\%$  となる。

#### 【0056】

次に、画線率について説明する。画線率は、紙面全体、あるいは、ある文書ブロックの面積に対する、印字部分の面積の割合である。ここで、印字部分とは、例えば、黒字で印字した文字であれば黒部分の面積となり、網点で表現した画像であれば網点の印字部分の面積の合計を示す。これら印字部分の面積を、紙面全体について計算したものは紙面の画線率となり、各文書ブロック毎に計算したものは各文書ブロックの画線率となる。

#### 【0057】

次に、文字のジャンプ率について説明する。文字のジャンプ率は、文書を構成する文字ブロック内に含まれる最大サイズ文字と最小サイズ文字との大きさの比率である。例えば、図2(a)に示す文字ブロック 21 の「秋の七草」の各文字サイズが 40 ポイントで最大であり、文字ブロック 24 の「秋の野に・・・」の各文字サイズが 14 ポイントで最小だとすると、文字のジャンプ率は、 $40 : 14 (2.86)$  となる。

#### 【0058】

次に、画像のジャンプ率について説明する。画像のジャンプ率は、文書中の複数の画像ブロック間における何らかの値（面積、縦横比、画線率等）の、最大値と最小値との比率である。

次に、図版率について説明する。図版率は、文書中に占める画像部分の割合である。例えば、文書ブロックの面積の合計に対する画像ブロックの面積の合計の割合、版面の面積に対する画像ブロックの面積の合計の割合、紙面全体の面積に対する画像ブロックの面積の合計の割合、紙面全体又は版面全体の面積に対する、文字ブロックの面積を除いた部分の面積の合計の割合、文字の画線部分の面積を除いた部分の割合などが含まれる。例えば、文書ブロックの面積の合計が  $309 \text{ cm}^2$  で、画像ブロックの面積の合計が  $51 \text{ cm}^2$  である場合は、図版率は、 $51 / 309 \times 100 = 16.5\%$  となる。

#### 【0059】

次に、グリッド拘束率について説明する。グリッド拘束率は、文書ブロックの位置が文書中のグリッド（仮想的な基準点）に沿っている割合である。例えば、図2(a)に示す文書においては、各文書ブロックの辺の位置は自由に配置されているが、図2(b)に示す文書においては、グリッドを一点鎖線で示すと、第2画像ブロック 26 及び第3文字ブロック 24 の上辺がそれぞれグリッドに沿う配置となっている。また、図2(b)に示す文書においては、第1文字ブロック 22 及び第1画像ブロック 25 は、縦方向の中心線が、グリッドと沿う（重なるのも含む）ように配置されている。ここで、文書ブロックがグリッドに沿っているかを判定する基準としては、各文書ブロックの、辺、中心線、中心点などがある。つまり、グリッド拘束率は、各文書ブロックにおける基準がグリッドにどれだけ沿っているかを示すものである。

#### 【0060】

次に、各構成要素間の位置揃え情報について説明する。各文書ブロック間の位置揃え情報は、複数の文書ブロック間の対応する辺や中心線などが同一直線上になるレイアウトの情報を言う。例えば、上下に並んでいる 2 つの文書ブロックの各々の左辺が同一直線上にあれば左揃え、各々の上辺が同一直線上にあれば上揃えといったようなレイアウト情報である。

#### 【0061】

更に、図3に基づき、デザイン情報抽出装置 100 のより具体的な動作を説明する。ここでは、デザイン情報として、文字のジャンプ率を抽出する例を説明する。ここで、図3(a)は、入力文書ファイルの一例を示す図であり、(b)は (a) の文書ファイルを各

10

20

30

40

50

構成要素（文書ブロック）毎に区分した結果を示す図である。

まず、図3（a）に示す出力内容の文書ファイル200が文書入力部10に入力される。ここで、入力された文書ファイル200は、スキャナ等により画像データ化された文書ファイルであるとする。文書ファイル200は、デザイン情報抽出部11に伝送され、文書ブロック抽出部11aに入力される。文書ブロック抽出部11aでは、まず入力された文書ファイルの種類が、ワープロ文書ファイルか、PDL文書ファイルか、画像文書ファイルかのいずれかを判別する。入力された文書ファイル200がワープロ文書ファイルである場合は、本実施の形態においては、ワープロソフトの種類に応じて予め文書ファイルが有している文書ブロック情報を抽出して、抽出情報を含む文書ファイルをページデザイン解析部11cに伝送し、一方、文書ファイル200がPDL文書ファイル又は画像文書ファイルである場合は、文書ブロックの抽出処理を行う。ここでは、入力された文書ファイルは画像文書ファイルであるので、図3（a）に示す文書ファイル200から文書ブロックの抽出を行う。図3（b）に示すように、第1～第3文字ブロック22～24並びに第1及び第2画像ブロック25及び26が文書ブロックとして抽出される。そして、これら抽出された文書ブロックの情報を含む文書ファイルは、文書ブロック解析部11bに伝送される。ここで、文書ブロック抽出部11aにおいては、まだ各文書ブロックが文字情報を含むのか図形情報を含むのか等の詳細な情報が解っていない状態である。

#### 【0062】

文書ブロック解析部11bでは、入力された文書ブロック情報を含む文書ファイルに基づき、各ブロックの解析を行う。解析は、上記したように各ブロックの画像に文字が含まれるか否か、あるいは、どのような図形情報（表など）が含まれるかを画像認識等により認識する。更に、文書ブロックに文字が含まれている場合は、その文字のサイズや書体等の詳細な情報を調べる。そして、これら解析結果は、ページデザイン解析部11cに伝送される。

#### 【0063】

ページデザイン解析部11cは、文書ブロック抽出部11aから入力された抽出結果及び文書ファイル、並びに文書ブロック解析部11bから入力された解析結果に基づき、まず、文書ファイル200における、各文書ブロックの大きさや配置位置、各文書ブロックが文字ブロックであるか画像ブロックであるかといったレイアウトテンプレート情報を抽出する。次に、デザイン情報抽出方法記憶部12に記憶された各デザイン情報を抽出するためのプログラムを順次実行することにより、文書ファイルから上記各デザイン情報を抽出する。各デザイン情報を抽出するプログラムは、ワープロ文書ファイルであれば、その文書ファイル及び上記抽出結果に基づきデザイン情報の抽出を行い、一方、PDL文書ファイル又は画像文書ファイルの場合は、上記解析結果に基づきデザイン情報の抽出を行う。そして、上記抽出されたレイアウトテンプレート情報及びデザイン情報は文書ファイル200のタイトル情報（秋の七草）と共に、レイアウト情報記憶部13に伝送される。

#### 【0064】

レイアウト情報記憶部13は、入力されたレイアウトテンプレート情報、デザイン情報及びタイトル情報を、タイトル情報とレイアウトテンプレート情報及びデザイン情報を対応付けて記憶する。

更に、図4に基づき、デザイン抽出装置100の動作処理の流れを説明する。図4は、デザイン抽出装置100の動作処理の流れを示す図である。

#### 【0065】

図4に示すように、まずステップS100に移行し、文書入力部10において、文書ファイルが入力されたか否かを判定し、入力されたと判定された場合(Yes)は入力された文書ファイルをデザイン情報抽出部11に伝送してステップS102に移行し、そうでない場合(No)は入力されるまで待機する。

ステップS102に移行した場合は、文書ブロック抽出部11aにおいて、文書ファイルはワープロ文書ファイルか否かを判定し、ワープロ文書ファイルであると判定された場合(Yes)はステップS104に移行し、そうでない場合(No)はステップS114に移行す

10

20

30

40

50

る。

#### 【0066】

ステップS104に移行した場合は、文書ブロック抽出部11aにおいて、ワープロ文書ファイルから文書ブロックを抽出してステップS106に移行する。

ステップS106では、文書ブロック抽出部11aにおいて、抽出した文書ブロック情報を含む文書ファイルをページデザイン解析部11cに伝送してステップS108に移行する。

#### 【0067】

ステップS108では、ページデザイン解析部11cにおいて、入力された情報に基づき、文書ファイルのレイアウトテンプレート情報を抽出してステップS110に移行する。  
10

ステップS110では、ページデザイン抽出部11cにおいて、デザイン情報抽出方法記憶部12に記憶された各デザイン情報抽出用のルール（例えば、プログラム）を、入力された情報に適用して、文書ファイルから各デザイン情報を抽出し、これら抽出したレイアウトテンプレート情報を文書ファイルの明示情報（例えば、抽出元の文書ファイルのタイトル）と共にレイアウト情報記憶部13に伝送してステップS112に移行する。

#### 【0068】

ステップS112では、レイアウト情報記憶部13において、入力されたレイアウトテンプレート情報を文書ファイルの明示情報に対応付けて記憶してステップS100に移行する。  
20

一方、ステップS102において、入力された文書ファイルがワープロ文書ファイルではなくステップS114に移行した場合は、文書ブロック抽出部11aにおいて、文書ファイルから文書ブロックを抽出し、当該抽出結果の情報を含む文書ファイルを文書ブロック解析部11bに伝送してステップS116に移行する。

#### 【0069】

ステップS116では、文書ブロック解析部11bにおいて、入力された抽出結果の情報を含む文書ファイルに基づき、文書ファイルにおける各文書ブロックを解析してステップS118に移行する。

ステップS118では、文書ブロック解析部11bにおいて、ステップS116の解析結果をページデザイン解析部11cに伝送してステップS108に移行する。  
30

#### 【0070】

更に、図5に基づき、図4のフローチャートにおけるステップS110におけるデザイン抽出処理を詳細に説明する。図5は、デザイン情報抽出処理を示すフローチャートである。

ステップS110でデザイン情報抽出処理が実行されると、図5に示すように、まずステップS200に移行し、デザイン情報抽出方法記憶部12からデザイン情報抽出用のプログラム（サブルーチン）を読み出してステップS202に移行する。

#### 【0071】

ステップS202では、ステップS200で読み出したプログラムを実行して、文書ファイルからデザイン情報を抽出してステップS204に移行する。  
40

ステップS204では、レイアウトテンプレート情報を文書ファイルの明示情報（例えば、抽出元の文書ファイルのタイトル）と共に、レイアウト情報記憶部13に伝送してステップS206に移行する。

#### 【0072】

ステップS206では、デザイン情報の抽出が全て終了したか否かを判定し、終了したと判定された場合(Yes)は処理を終了し、そうでない場合(No)はステップS200に移行する。

更に、図6に基づき、図5のフローチャートにおけるステップS202において、文字のジャンプ率を抽出するプログラムを実行した場合の処理の一例を説明する。図6は、文  
50

字のジャンプ率を抽出する処理を示すフローチャートである。

【0073】

ステップS202において文字のジャンプ率を抽出するプログラムが実行されると、図6に示すように、まずステップS300に移行し、最大文字サイズ格納用変数S<sub>max</sub>に0を代入し、最小文字サイズ格納用変数S<sub>min</sub>に値の十分大きい整数値(Large)を代入して変数の初期化を行いステップS302に移行する。ここで、十分大きい整数値とは、装置が表現可能な整数値の最大値や、文字サイズとしてはあり得ない大きさの数値等、文書ファイル中に含まれる最大文字サイズより大きい数値である。

【0074】

ステップS302では、未処理の文字ブロックを探索してステップS304に移行する 10。  
。

ステップS304では、ステップS302の探索結果に基づき、未処理の文字ブロックがあるか否かを判定し、あると判定された場合(Yes)はステップS306に移行し、そうでない場合(No)はステップS316に移行する。

【0075】

ステップS306に移行した場合は、対象の文字ブロック中に含まれる文字のサイズを比較用文字サイズ格納用変数Sに代入してステップS308に移行する。

ステップS308では、変数Sと変数S<sub>max</sub>とを比較して、Sに代入された数値がS<sub>max</sub>に代入された数値よりも大きいか否かを判定し、大きいと判定された場合(Yes)はステップS310に移行し、そうでない場合(No)はステップS312に移行する。 20

【0076】

ステップS310に移行した場合は、最大文字サイズ格納用変数S<sub>max</sub>に比較用文字サイズ格納用変数Sの数値を代入してS<sub>max</sub>を更新しステップS312に移行する。

ステップS312では、変数Sと変数S<sub>min</sub>とを比較して、Sに代入された数値がS<sub>min</sub>に代入された数値よりも小さいか否かを判定し、小さいと判定された場合(Yes)はステップS314に移行し、そうでない場合(No)はステップS302に移行する。

【0077】

ステップS314に移行した場合は、最小文字サイズ格納用変数S<sub>min</sub>に比較用文字サイズ格納用変数Sの数値を代入してS<sub>min</sub>を更新しステップS302に移行する。

一方、ステップS304で未処理の文字ブロックが無くステップS316に移行した場合は、文字ジャンプ率格納用変数Jcharに、S<sub>max</sub>の値をS<sub>min</sub>の値で除算した除算結果を代入して処理を終了する。 30

【0078】

以上、上記実施の形態におけるデザイン情報抽出装置100は、デザイン情報抽出部1及びデザイン情報抽出方法記憶部12によって、デザイナ等がデザインした既存の文書ファイルから、そのレイアウトテンプレート情報及びデザイン情報を抽出することが可能である。

また、文書ブロック抽出部11a及び文書ブロック解析部11bにより、PDL文書ファイルや画像文書ファイルからもレイアウトテンプレート情報及びデザイン情報を抽出することが可能である。 40

【0079】

[第2の実施の形態]

次に、本発明の第2の実施の形態を図面に基づいて説明する。図7～図10は、本発明に係るデザイン情報抽出装置、デザイン情報抽出プログラム及びデザイン情報抽出方法の第2の実施の形態を示す図である。

まず、本発明の第2の実施の形態に係るデザイン情報抽出装置300の構成を図7に基づいて説明する。図7は、本発明に係るデザイン情報抽出装置、デザイン情報抽出プログラム及びデザイン情報抽出方法を適用したデザイン情報抽出装置300の構成を示すブロック図である。

【0080】

デザイン情報抽出装置 300 は、外部装置等から文書ファイルを入力する文書入力部 10 と、入力された文書ファイルからレイアウトテンプレート情報及びデザイン情報を抽出するデザイン情報抽出部 11 と、文書ファイルからデザイン情報を抽出するための抽出方法の情報が記憶されたデザイン情報抽出方法記憶部 12 と、デザイン情報抽出部 11 において抽出されたデザイン情報を記憶するレイアウト情報記憶部 13 と、入力された明示情報に対応したレイアウトテンプレート情報及びデザイン情報をレイアウト情報記憶部 13 の記憶内容から選択するレイアウト情報選択部 14 と、レイアウト情報選択部 14 において選択されたレイアウトテンプレート情報及びデザイン情報を出力するレイアウト情報出力部 15 と、を含んだ構成となっている。ここで、本実施の形態におけるデザイン情報抽出装置 300 は、上記第 1 の実施の形態におけるデザイン情報抽出装置 100 に、レイアウト情報選択部 14 及びレイアウト情報出力部 15 を追加した構成であり、従って、重複する部分については詳細な説明を省略する。

10

#### 【0081】

レイアウト情報選択部 14 は、デザイン情報抽出装置 300 に付属の、図示しない、キーボードやマウス等の入力デバイスを用いて利用者等によって入力された明示情報に基づき、当該明示情報に対応するレイアウトテンプレート情報及びデザイン情報をレイアウト情報記憶部 13 から読み出してレイアウト情報出力部 15 に伝送する。ここで、本実施の形態において、明示情報は、レイアウトテンプレート情報及びデザイン情報の抽出元の文書ファイルのタイトルとする。

#### 【0082】

レイアウト情報出力部 15 は、レイアウト情報選択部 14 によって選択されたレイアウトテンプレート情報及びデザイン情報を出力する。ここで、出力される情報は、レイアウトテンプレート情報及びデザイン情報の反映された文書ファイルの離形等の所定形式のデータとして出力しても良い。

20

ここで、本実施の形態において、デザイン情報抽出装置 300 は、図示しないが、上記各部を制御するためのプログラムの記憶された記憶媒体と、これらのプログラムを実行するための C P U と、プログラムの実行に必要なデータを記憶する R A M と、を備えている。そして、上記 C P U により前記プログラムを実行することによって上記各部の処理を実現するものである。

#### 【0083】

また、記憶媒体とは、R A M 、R O M 等の半導体記憶媒体、F D 、H D 等の磁気記憶型記憶媒体、C D 、C D V 、L D 、D V D 等の光学的読み取り方式記憶媒体、M O 等の磁気記憶型 / 光学的読み取り方式記憶媒体であって、電子的、磁気的、光学的等の読み取り方法のいかんにかかわらず、コンピュータで読み取り可能な記憶媒体であれば、あらゆる記憶媒体を含むものである。

30

#### 【0084】

以下、図 8 及び図 9 に基づき、デザイン情報抽出装置 300 におけるレイアウトテンプレート情報及びデザイン情報の選択処理並びに選択したレイアウトテンプレート情報及びデザイン情報の出力処理の具体的な動作を説明する。図 8 ( a ) 、( b ) は、入力文書ファイル例の構成を示す図であり、図 9 は、レイアウト情報記憶部 13 の記憶内容の一例を示す図である。なお、レイアウトテンプレート情報及びデザイン情報の抽出処理については、上記第 1 の実施の形態におけるデザイン情報抽出装置 100 と同様となるので説明を省略する。

40

#### 【0085】

ここで、本実施の形態においては、デザイン情報抽出装置 300 は、図 3 に示す文書ファイル 200 に加え、図 8 ( a ) 及び ( b ) に示す文書ブロック構成の、文書ファイル 400 及び文書ファイル 500 を含めた 3 つの文書ファイルから、上記第 1 の実施の形態で説明した抽出処理によって、図 9 に示す、レイアウトテンプレート情報及びデザイン情報が既に抽出され、抽出元の文書のタイトルと対応付けられてレイアウト情報記憶部 13 に記憶された状態となっている。

50

## 【0086】

また、レイアウトテンプレート情報及びデザイン情報は、図9に示すように、明示情報である抽出元文書のタイトル、デザイン情報である版面率、文字ジャンプ率及びグリッド拘束率、並びにレイアウトテンプレート情報であるレイアウト情報1～5がデータテーブルとして記憶されている。

まず、利用者が入力デバイスを用いて、明示情報として「秋の七草」と入力（あるいは選択）すると、レイアウト情報選択部14は、レイアウト情報記憶部13に記憶されたデータテーブルの明示情報の中から「秋の七草」を検索する。検索により明示情報の中から「秋の七草」が見つかると、更に、「秋の七草」に対応するレイアウトテンプレート情報及びデザイン情報を読み出してレイアウト情報出力部15に伝送する。ここでは、図9に示すように、「秋の七草」に対応する、版面率「50%」、文字ジャンプ率「2.5」、グリッド拘束率「40%」、並びにレイアウト情報1～レイアウト情報5「第1～第3文字ブロックの配置位置・大きさ並びに第1及び第2画像ブロックの配置位置・大きさ」が読み出されて、レイアウト情報出力部15に伝送される。ここで、文字ブロックの配置位置・大きさは、紙面上における文字ブロックの配置位置の情報（ブロックの左上の点の2次元座標等）、文字ブロックの大きさの情報（ブロック形状、各辺の長さ等）である。同様に画像ブロックの配置位置・大きさは、紙面上における画像ブロックの配置位置の情報（ブロックの左上の点の2次元座標等）、画像ブロックの大きさの情報（ブロック形状、各辺の長さ等）である。上記「秋の七草」と同様に、明示情報として「春の七草」、「世界の草花」が入力された場合も、レイアウト情報記憶部13に記憶されたデータテーブルの明示情報の中から「春の七草」、「世界の草花」をそれぞれ検索し、見つかった場合は、図9に示すように、これら明示情報に対応するレイアウトテンプレート情報及びデザイン情報を読み出してレイアウト情報出力部15に伝送する。

## 【0087】

レイアウト情報出力部15は、レイアウト情報選択部14において選択され且つ伝送されたレイアウトテンプレート情報及びデザイン情報を取得すると、当該取得したレイアウトテンプレート情報及びデザイン情報を所定形式のファイル等として利用者に向けて出力する。

更に、図10に基づき、デザイン情報抽出装置300におけるレイアウトテンプレート情報及びデザイン情報の選択処理並びに選択したレイアウトテンプレート情報及びデザイン情報の出力処理の流れを説明する。図10は、レイアウトテンプレート情報及びデザイン情報の選択処理並びに選択したレイアウトテンプレート情報及びデザイン情報の出力処理を示すフローチャートである。

## 【0088】

図10に示すように、まずステップS400に移行し、レイアウト情報選択部14において、明示情報が入力されたか否かを判定し、入力されたと判定された場合(Yes)はステップS402に移行し、そうでない場合(No)は入力されるまで待機する。

ステップS402に移行した場合は、レイアウト情報選択部14において、レイアウト情報記憶部13の記憶内容から入力された明示情報を検索してステップS404に移行する。

## 【0089】

ステップS404では、レイアウト情報選択部14において、レイアウト情報記憶部13の記憶内容に入力された明示情報があるか否かを判定し、あると判定された場合(Yes)はステップS406に移行し、そうでない場合(No)はステップS410に移行する。

ステップS406に移行した場合は、レイアウト情報選択部14において、明示情報に対応するレイアウトテンプレート情報及びデザイン情報をレイアウト情報記憶部13から選択し（読み出し）、レイアウト情報出力部15に伝送してステップS408に移行する。

## 【0090】

ステップS408では、レイアウト情報出力部15において、レイアウト情報選択部1

10

20

30

40

50

4において選択されたレイアウトテンプレート情報及びデザイン情報を出力してステップS400に移行する。

一方、ステップS404において、明示情報が見つからずにステップS410に移行した場合は、明示情報が見つからない旨のメッセージを出力してステップS400に移行する。

#### 【0091】

以上、上記実施の形態におけるデザイン情報抽出装置300は、レイアウト情報選択部14によって、入力された明示情報に対応するレイアウトテンプレート情報及びデザイン情報をレイアウト情報記憶部13から選択することが可能であり、レイアウト情報出力部15によって、レイアウト情報選択部14において選択されたレイアウトテンプレート情報及びデザイン情報を出力することが可能である。10

#### 【0092】

上記実施の形態において、文書入力部10は、発明1又は2の文書ファイル取得手段に対応する。

上記実施の形態において、デザイン情報抽出部11は、発明1又は2のデザイン情報抽出手段に対応する。

上記実施の形態において、文書ブロック抽出部11aは、発明3の文書ファイル区分手段に対応し、文書ブロック解析部11bは、発明3の構成要素解析手段に対応する。

#### 【0093】

上記実施の形態において、ページデザイン解析部11cは、発明3のデザイン情報抽出手段に対応する。20

上記実施の形態において、デザイン情報抽出方法記憶部12は、発明5のデザイン情報抽出方法記憶手段に対応する。

上記実施の形態において、レイアウト情報記憶部13は、発明1、2及び7のいずれか1のデザイン情報記憶手段に対応する。

#### 【0094】

上記実施の形態において、レイアウト情報選択部14は、発明7のデザイン情報選択手段に対応する。

上記実施の形態において、ステップS100は、発明8又は9の文書ファイル取得手段、あるいは発明15又は16の文書ファイル取得ステップに対応する。30

上記実施の形態において、ステップS104及びステップS114は、発明10の文書ファイル区分手段又は発明17の文書ファイル区分ステップに対応する。

#### 【0095】

上記実施の形態において、ステップS116は、発明10の構成要素解析手段又は発明17の構成要素解析ステップに対応する。

上記実施の形態において、ステップS108及びステップS110は、発明10のデザイン情報抽出手段又は発明17のデザイン情報抽出ステップに対応する。

上記実施の形態において、ステップS112は、発明8、9及び14のデザイン情報記憶手段、並びに発明15、16及び21のいずれか1のデザイン情報記憶ステップに対応する。40

#### 【0096】

上記実施の形態において、ステップS400～ステップS406は、発明14のデザイン情報選択手段又は発明21のデザイン情報選択ステップに対応する。

なお、上記実施の形態においては、明示情報として、抽出元の文書ファイルのタイトルを例として説明したが、これに限らず、利用者が所望の明示情報を入力したり、抽出元の文書ファイルの出力内容中にある最大の文字サイズの文字列を明示情報としたりするなどしても良い。

#### 【図面の簡単な説明】

#### 【0097】

【図1】本発明に係るデザイン情報抽出装置、デザイン情報抽出プログラム及びデザイン

10

30

40

50

情報抽出方法を適用したデザイン情報抽出装置 100 の構成を示すブロック図である。

【図 2】(a) は、紙面と版面との関係を示す図であり、(b) は、文書に対するグリッド線の一例を示す図である。

【図 3】(a) は、入力文書ファイルの一例を示す図であり、(b) は(a)の文書ファイルを各構成要素(文書ブロック)毎に区分した結果を示す図である。

【図 4】デザイン抽出装置 100 の動作処理の流れを示す図である。

【図 5】デザイン情報抽出処理を示すフローチャートである。

【図 6】文字のジャンプ率を抽出する処理を示すフローチャートである。

【図 7】本発明に係るデザイン情報抽出装置、デザイン情報抽出プログラム及びデザイン情報抽出方法を適用したデザイン情報抽出装置 300 の構成を示すブロック図である。 10

【図 8】(a)、(b) は、入力文書ファイル例の構成を示す図である。

【図 9】レイアウト情報記憶部 13 の記憶内容の一例を示す図である。

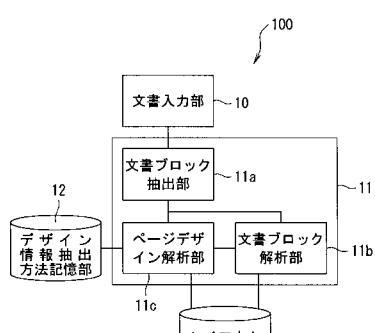
【図 10】レイアウトテンプレート情報及びデザイン情報の選択処理並びに選択したレイアウトテンプレート情報及びデザイン情報の出力処理を示すフローチャートである。 10

【符号の説明】

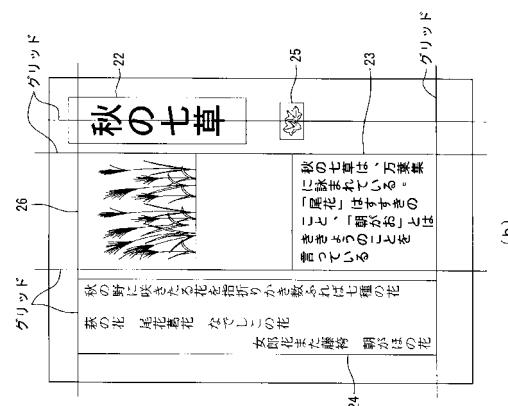
【0098】

100...文書入力部、11...デザイン情報抽出部、11a...文書ブロック抽出部、11b...文書ブロック解析部、11c...ページデザイン解析部、12...デザイン情報抽出方法記憶部、13...レイアウト情報記憶部、14...レイアウト情報選択部、15...レイアウト情報出力部、20...紙面、21...版面、22...第1文字ブロック、23...第2文字ブロック、24...第3文字ブロック、25...第1画像ブロック、26...第2画像ブロック、100, 300...デザイン情報抽出装置、200, 400, 500...文書ファイル 20

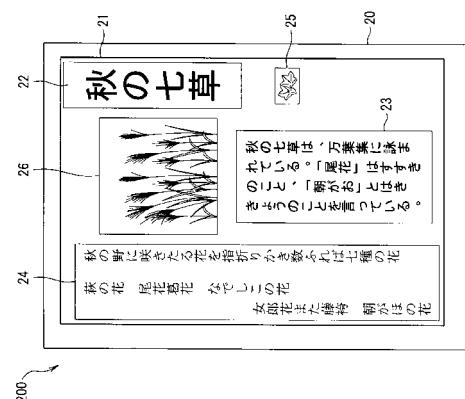
【図 1】



【図 2】

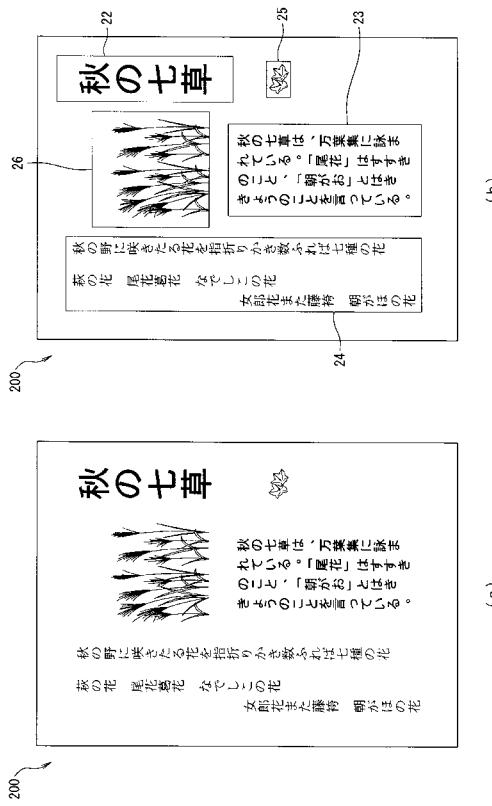


(b)

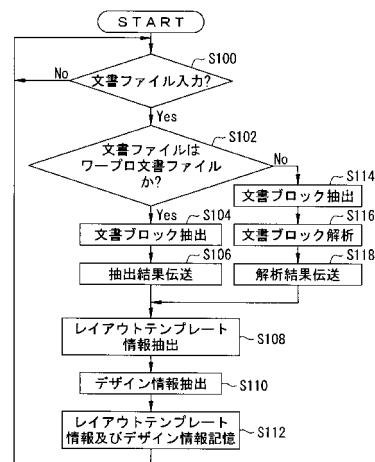


(a)

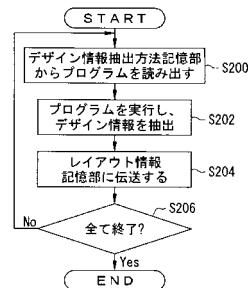
【図3】



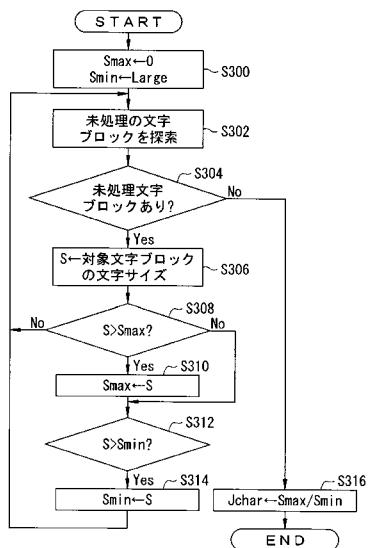
【図4】



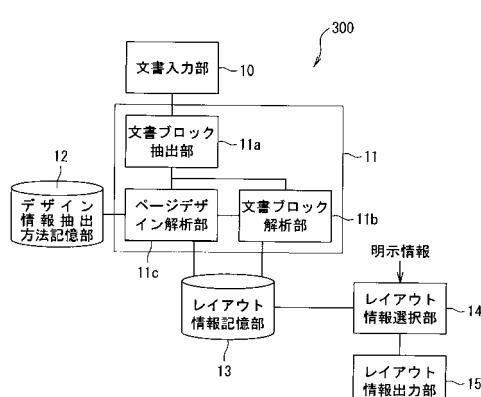
【図5】



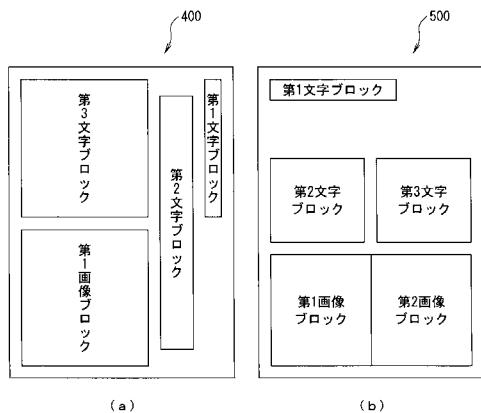
【図6】



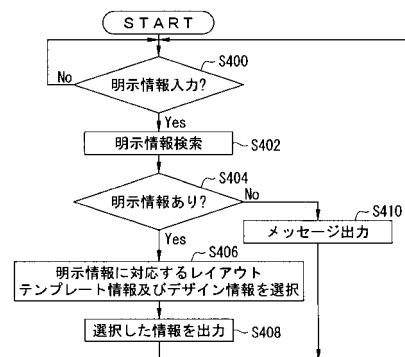
【図7】



【図8】



【図10】



【図9】

入力文書	文書ファイル200	文書ファイル400	文書ファイル500
元文書	秋の七草	春の七草	世界の草花
版面率	50%	60%	70%
文字ジャンプ率	2.5	1.7	1.5
グリット拘束率	40%	50%	75%
レイアウト情報1	第1文字ブロックの 配置位置・大きさ	第1文字ブロックの 配置位置・大きさ	第1文字ブロックの 配置位置・大きさ
レイアウト情報2	第2文字ブロックの 配置位置・大きさ	第2文字ブロックの 配置位置・大きさ	第2文字ブロックの 配置位置・大きさ
レイアウト情報3	第3文字ブロックの 配置位置・大きさ	第3文字ブロックの 配置位置・大きさ	第3文字ブロックの 配置位置・大きさ
レイアウト情報4	第1画像ブロックの 配置位置・大きさ	第1画像ブロックの 配置位置・大きさ	第1画像ブロックの 配置位置・大きさ
レイアウト情報5	第2画像ブロックの 配置位置・大きさ	なし	第2画像ブロックの 配置位置・大きさ

## 【手続補正書】

【提出日】平成16年4月21日(2004.4.21)

## 【手続補正1】

【補正対象書類名】図面

【補正対象項目名】図6

【補正方法】変更

【補正の内容】

【図6】

