

(19)日本国特許庁(JP)

**(12)特許公報(B2)**

(11)特許番号  
**特許第7671513号**  
**(P7671513)**

(45)発行日 令和7年5月2日(2025.5.2)

(24)登録日 令和7年4月23日(2025.4.23)

(51)国際特許分類  
**G 0 6 F 40/279 (2020.01)**

F I  
**G 0 6 F 40/279**

請求項の数 10 (全15頁)

(21)出願番号	特願2022-541356(P2022-541356)	(73)特許権者	717005132 株式会社LegalOn Technologies 東京都渋谷区桜丘町1番1号 渋谷サクラステージSHIBUYAタワー
(86)(22)出願日	令和2年8月4日(2020.8.4)	(74)代理人	100079108 弁理士 稲葉 良幸
(86)国際出願番号	PCT/JP2020/029747	(74)代理人	100109346 弁理士 大貫 敏史
(87)国際公開番号	WO2022/029863	(74)代理人	100117189 弁理士 江口 昭彦
(87)国際公開日	令和4年2月10日(2022.2.10)	(74)代理人	100134120 弁理士 内藤 和彦
審査請求日	令和5年3月28日(2023.3.28)	(72)発明者	川戸 崇志 東京都江東区豊洲三丁目2番20号 豊最終頁に続く
特許法第30条第2項適用 令和2年5月18日にhttps://legalforce-cloud.com/marshall/index.htmlにて公開			
令和2年5月18日にhttps://jp.techcrunch.com/2020/05/18/legalforce-marshall/にて公開			
前置審査			

(54)【発明の名称】 文書処理プログラム、情報処理装置及び文書処理方法

**(57)【特許請求の範囲】****【請求項1】**

コンピュータが、

利用者の端末から文書情報を取得することと、

前記文書情報から、固有表現を表す文字列を当該文書情報中の位置情報と共に取得することと、

前記文字列の少なくとも一つに対して、キーワードの統一、記載形式の統一、他の情報の参照による補足、他の情報に基づく推論、及び名寄せ処理の少なくとも一つを含む正規化を行うことにより正規化抽出情報を取得することと、

前記正規化抽出情報を前記端末において指定する操作を可能にすることと、

前記利用者の操作により指定された前記正規化抽出情報に対応する文字列の前記文書情報中の位置を前記端末に表示させることと、

を含む文書処理方法。

**【請求項2】**

前記コンピュータが、前記利用者により指定された前記正規化抽出情報に対応する文字列の位置の情報を含む前記文書情報と、前記正規化抽出情報を前記端末において同一の画面上に表示されること、をさらに含む、請求項1に記載の文書処理方法。

**【請求項3】**

前記コンピュータが、前記正規化抽出情報を、当該正規化抽出情報が属する項目の名称と共に前記端末に表示させること、をさらに含む、請求項1に記載の文書処理方法。

**【請求項 4】**

前記コンピュータが、前記文書情報中の指定された文字列を抽出情報として登録すること、をさらに含む、請求項 1～3 のいずれか一項に記載の文書処理方法。

**【請求項 5】**

前記コンピュータが、前記端末とネットワークを通じて通信すること、をさらに含む、請求項 1～4 のいずれか一項に記載の文書処理方法。

**【請求項 6】**

前記コンピュータが、前記文書情報を条、項、又は号に基づいて分割することにより単位情報を取得すること、をさらに含む、請求項 1～5 のいずれか一項に記載の文書処理方法。

10

**【請求項 7】**

前記コンピュータが、前記正規化抽出情報と、前記単位情報と、前記文書情報とを関連付けて表示するための表示情報を生成すること、をさらに含む、請求項 6 に記載の文書処理方法。

**【請求項 8】**

前記文書情報は契約書である、請求項 1～7 のいずれか一項に記載の文書処理方法。

**【請求項 9】**

プロセッサが、

利用者の端末から文書情報を取得し、

前記文書情報から、固有表現を表す文字列を当該文書情報中の位置情報と共に取得し、

20

前記文字列の少なくとも一つに対して、キーワードの統一、記載形式の統一、他の情報の参照による補足、他の情報に基づく推論、及び名寄せ処理の少なくとも一つを含む正規化を行うことにより正規化抽出情報を取得し、

前記正規化抽出情報を前記端末において指定する操作を可能とし、

前記利用者の操作により指定された前記正規化抽出情報に対応する文字列の前記文書情報中の位置を前記端末に表示する、

ように構成されている、情報処理装置。

**【請求項 10】**

所定の指示命令に加えて、文書情報を記憶するように構成されたメモリと、

前記メモリに記憶された指示命令に基づいて、

30

利用者の端末から前記文書情報を取得し、

前記文書情報から、固有表現を表す文字列を当該文書情報中の位置情報と共に取得し、

前記文字列の少なくとも一つに対して、キーワードの統一、記載形式の統一、他の情報の参照による補足、他の情報に基づく推論、及び名寄せ処理の少なくとも一つを含む正規化を行うことにより正規化抽出情報を取得し、

前記正規化抽出情報を前記端末において指定する操作を可能とし、

前記利用者の操作により指定された前記正規化抽出情報に対応する文字列の前記文書情報中の位置を前記端末に表示する、

ための処理を実行するように構成されたプロセッサと、

を含む情報処理装置。

40

**【発明の詳細な説明】****【技術分野】****【0001】**

本開示は、文書処理プログラム、情報処理装置及び文書処理方法に関する。

**【背景技術】****【0002】**

従来の技術として、契約書からキーワードを検出して契約者にリスクを事前に知らせる情報処理装置が提案されている（例えば、特許文献 1 参照）。

**【0003】**

特許文献 1 に開示された情報処理装置は、契約書の文書中から所定のキーワードを検出

50

し、検出されたキーワードの中から契約する利用者にとっての重要箇所を認識して、重要箇所をリスクとともに提示して締結するか否かを利用者に判断させる。また、情報処理装置は、契約書の文書中から契約の更新期限等を検出すると、更新の設定画面を提示し、更新を通知するか否かを利用者に判断させる。

【先行技術文献】

【特許文献】

【0004】

【文献】国際公開第2018/042548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、上記した特許文献1に開示された情報処理装置によると、契約書からキーワードを検出して契約者にリスクを事前に知らせるとともに、契約書の更新通知の要否を判断させるものの、検出されるキーワードに表記上のゆれがある場合には対応できない、という問題があった。また、検出されたキーワードから対応する契約書内の記載が知りたい場合に、検出されたキーワードに表記上のゆれがある場合、すべての該当箇所を一元して管理することができない、という問題があった。

【0006】

本開示の目的は、上記従来技術を踏まえ、より使い勝手のよい文書処理プログラム、情報処理装置及び文書処理方法を提供することにある。

【課題を解決するための手段】

【0007】

本開示に係る発明の一態様は、上記目的を達成するため、以下の文書処理プログラム、情報処理装置及び文書処理方法を提供する。

【0008】

[1] コンピュータを、

文書情報から文書の内容を示す文字列を当該文書情報中の位置情報とともに抽出する抽出手段と、

前記抽出手段によって抽出された前記文字列を正規化して正規化抽出情報とする正規化手段と、

前記正規化抽出情報の内容を、前記位置情報に基づいて前記文書情報の該当位置を示して表示する表示制御手段として機能させるための文書処理プログラム。

[2] 文書情報を予め定めた単位で分割して単位情報とする分割手段としてさらに機能させ、

前記抽出手段は、前記文書の内容を示す文字列を前記単位情報の位置情報とともに抽出し、

前記表示制御手段は、前記正規化抽出情報の内容を、前記単位情報の位置情報に基づいて前記単位情報の該当位置を示して表示する前記[1]に記載の文書処理プログラム。

[3] 前記抽出手段は、前記文書情報中の指定された文字列を抽出情報として登録する請求項1又は2に記載の文書処理プログラム。

[4] 前記表示制御手段は、前記正規化抽出情報と、前記単位情報と、前記文書情報とを関連付けて表示する前記[1]から[3]のいずれかに記載の文書処理プログラム。

[5] 前記コンピュータは、ネットワークを通じて通信可能に1又は複数の端末と接続される、前記[1]から[4]のいずれかに記載の文書処理プログラム。

[6] 前記コンピュータは、無線通信ネットワークを通じて1又は複数の端末と接続される、前記[1]から[5]のいずれかに記載の文書処理プログラム。

[7] 文書情報から文書の内容を示す文字列を当該文書情報中の位置情報とともに抽出する抽出手段と、

前記抽出手段によって抽出された前記文字列を正規化して正規化抽出情報とする正規化手段と、

10

20

30

40

50

前記正規化抽出情報の内容を、前記位置情報に基づいて前記文書情報の該当位置を示して表示する表示制御手段とを有する情報処理装置。

- [ 8 ] 所定の指示命令に加えて、文書情報を記憶するように構成されたメモリと、前記メモリに記憶された指示命令に基づいて、前記文書情報から文書の内容を示す文字列を当該文書情報中の位置情報とともに抽出し、抽出された前記文字列を正規化して正規化抽出情報とし、前記正規化抽出情報の内容を、前記位置情報に基づいて前記文書情報の該当位置を示して表示する、ための処理を実行するように構成されたプロセッサと、を含む情報処理装置。

- [ 9 ] 文書情報から文書の内容を示す文字列を当該文書情報中の位置情報とともに抽出する抽出ステップと、抽出された前記文字列を正規化して正規化抽出情報とする正規化ステップと、前記正規化抽出情報の内容を、前記位置情報に基づいて前記文書情報の該当位置を示して表示する表示制御ステップとを有する文書処理方法。

**【発明の効果】**

**【0009】**

本開示の様々な実施形態によれば、より使い勝手のよい文書処理プログラム、情報処理装置及び文書処理方法を提供することが可能となる。

**【図面の簡単な説明】**

**【0010】**

- 【図1】図1は、実施の形態に係る文書処理システムの構成の一例を示す概略図である。
- 【図2】図2は、実施の形態に係る文書処理サーバ装置の構成例を示すブロック図である。
- 【図3】図3は、文書情報の構成例を示す概略図である。
- 【図4】図4は、抽出情報の構成例を示す概略図である。
- 【図5】図5は、正規化抽出情報の構成例を示す概略図である。
- 【図6】図6は、情報抽出動作の処理の一例を説明するための概略図である。
- 【図7】図7は、正規化手段の動作例を説明するための概略図である。
- 【図8】図8は、表示制御手段の表示例を示す概略図である。
- 【図9】図9は、操作に応じて抽出情報を登録する場合に表示される画面の表示例を示す概略図である。
- 【図10】図10は、文書処理サーバ装置の情報抽出動作を示すフローチャートである。
- 【図11】図11は、表示処理動作を説明するためのフローチャートである。

**【発明を実施するための形態】**

**【0011】**

**【実施の形態】**

(文書処理システムの構成)

図1は、実施の形態に係る文書処理システムの構成の一例を示す概略図である。

**【0012】**

この文書処理システム5は、文書処理サーバ装置1と、端末2と、端末3とをネットワーク4によって互いに通信可能に接続することで構成される。端末2は、例えば、文書管理、文書作成、文書内容確認及び／又は文書レビューを希望する利用者によって操作され、端末3は他の利用者によって操作されるが、これに限られるものではない。端末2の利用者と、端末3の利用者とは文書情報として、例えば、契約書を扱うものであり、主に、一方又は双方の利用者の作成した文書情報を管理するとともに、締結後の文書情報を管理し、当該文書情報の内容を確認、把握するために文書処理システム5を利用する。

**【0013】**

文書処理サーバ装置1は、サーバ型の情報処理装置であり、端末2及び端末3の要求に応じて動作するものであって、本体内に情報を処理するための機能を有するCPU(Central Processing Unit)やHDD(Hard Disk Drive)

10

20

30

40

50

e)、フラッシュメモリ等の電子部品を備える。なお、文書処理サーバ装置1は、必ずしも単体の情報処理装置で構成される必要はなく、複数の情報処理装置が協働して動作するものであってもよいし、任意のクラウドサービスによって動作するものであってもよい。また、文書処理サーバ装置1の機能を端末2及び/又は端末3内で実現するものであってもよい。

#### 【0014】

端末2及び端末3は、PC(Personal Computer)やタブレット端末等の情報処理装置であって、本体内に情報を処理するための機能を有するCPUやフラッシュメモリ等の電子部品を備える。

#### 【0015】

ネットワーク4は、高速通信が可能な通信ネットワークであり、例えば、インターネット、イントラネットやLAN(Local Area Network)等の有線又は無線の通信網である。

#### 【0016】

上記構成において、一例として、文書処理サーバ装置1が処理する文書は契約書等の法律分野の文書であり、利用者の方又は双方は法律の専門家ではないが契約書の作成を必要とする人物、又は弁護士等の法律の専門家であって契約書の作成の知識を有する人物である。また、例えば、一方又は双方の利用者は社内の営業部の社員、又は社内法務部の社員である。

#### 【0017】

基本動作としては、端末2又は端末3から文書情報を文書処理サーバ装置1にアップロードした後、文書処理サーバ装置1において文書情報を管理し、端末2又は端末3から文書処理サーバ装置1にアクセスすることで文書情報の内容の確認等を行うが、その際に利用者の確認作業負担を減らすべく、又は利用者の契約書の作成を支援するべく、文書処理サーバ装置1は、文書情報から情報を抽出し、抽出した情報を内容の把握を助ける形式で表示する。表示方法の具体例については後述する。

#### 【0018】

本実施の形態では主に、文書処理サーバ装置1は、文書情報中の契約状況を示す情報を抽出し、抽出した情報を一方又は双方の利用者に提示する。なお、以下において、条項のことを「条文」と言うことがある。実施の形態を以下で説明する。

#### 【0019】

また、端末2及び端末3はそれぞれ単数を図示しているが、複数台がネットワーク4に接続されるものであってもよく、同様にこれらを操作する利用者は複数人であってもよい。

#### 【0020】

(文書処理サーバ装置の構成)

図2は、実施の形態に係る文書処理サーバ装置1の構成例を示すプロック図である。

#### 【0021】

文書処理サーバ装置1は、CPU等から構成され、各部を制御するとともに、各種のプログラムを実行する制御部10と、フラッシュメモリ等の記憶媒体から構成され情報を記憶する記憶部11と、ネットワーク4を介して外部と通信するための通信インターフェイスとして機能する通信部12とを備える。

#### 【0022】

制御部10は、CPU等のプロセッサから構成され、メモリから構成される記憶部11及び通信インターフェイスとして機能する通信部12に電気的に接続される。当該制御部10は、後述する文書処理プログラム110を実行することで、契約書受付手段100、契約書分割手段101、情報抽出手段102、正規化手段103及び表示制御手段104等として機能する。

#### 【0023】

契約書受付手段100は、端末2又は端末3から契約書を文書情報111として受け付けて記憶部11に格納する。文書情報111は、文章がレイアウトされたPDF等の画像

10

20

30

40

50

情報であってもよいし、テキストデータを含むテキストファイル、ワードファイル等の情報であってもよい。

#### 【0024】

契約書分割手段101は、文書情報111がテキスト以外の情報の場合はO C R ( O p t i c a l C h a r a c t e r R e c o g n i t i o n ) 等を行ってテキスト化した後、文書情報111をタイトル、序文、条文単位等の契約書の構成要素毎に分割し、単位情報112として記憶部11に格納する。なお、条文単位は、分割の単位が条文に限定されるものではなく、項、号であってもよいし、意味のまとまりのある複数の条、項、号のまとまりであってもよい。また、異なる条、項に渡って意味のまとまりのある項、号であってもよい。

10

#### 【0025】

情報抽出手段102は、単位情報112から契約内容を表す情報を抽出して抽出情報113として記憶部11に格納する。

#### 【0026】

正規化手段103は、情報抽出手段102が抽出した抽出情報113の内容を、それぞれキーワードの統一、記載形式の統一、他の情報の参照による補足、他の情報に基づく推論、名寄せ処理等を行うことで正規化して正規化抽出情報114として記憶部11に格納する。なお、他の情報の参照による補足については、契約開始日と有効期限から契約終了日を算出する。また、他の情報に基づく推論については、例えば、社名、日付、住所等に基づき法人番号を推定する。また、名寄せ処理は、レーベンシュタイン距離やキーワード類似度を用いて辞書(キーワードリスト)のキーワードを比較し、近いものを正規化文字列とすることで行う。

20

#### 【0027】

表示制御手段104は、記憶部11の文書情報111、単位情報112、抽出情報113及び正規化抽出情報114並びに各手段100～103の出力結果を予め定めた方法で端末2及び端末3の表示部に表示制御する。なお、表示方法の詳細は後述する。

#### 【0028】

記憶部11は、フラッシュメモリ等のメモリから構成され、プロセッサ等から構成される制御部10及通信インターフェイスとして機能する通信部12に電気的に接続される。当該記憶部11は、制御部10を上述した各手段100～104として動作させる文書処理プログラム110、文書情報111、単位情報112、抽出情報113及び正規化抽出情報114等を記憶する。

30

#### 【0029】

図3は、文書情報111の構成例を示す概略図である。

#### 【0030】

文書情報111aは、一例として、契約書であり、契約書分割手段101によって分割される単位情報112として、契約書のタイトル112a<sub>1</sub>と、序文112a<sub>2</sub>と、これらに続く複数の条文(条文単位)112a<sub>3</sub>、112a<sub>4</sub>、112a<sub>5</sub>…とを有する。また、条文(条文単位)112a<sub>4</sub>は、複数の項(条文単位)112a<sub>41</sub>、112a<sub>42</sub>を含むものとする。

40

#### 【0031】

図4は、抽出情報113の構成例を示す概略図である。

#### 【0032】

抽出情報113は、情報抽出手段102により抽出された情報であり、抽出情報を識別するための抽出IDと、単位情報112から抽出された抽出情報と、抽出された情報の属する項目である抽出項目と、単位情報112中の位置情報である参照先位置とを有する。

#### 【0033】

図5は、正規化抽出情報114の構成例を示す概略図である。

#### 【0034】

正規化抽出情報114は、正規化手段103により抽出情報113を正規化した情報で

50

あり、抽出 ID と、抽出情報と、当該抽出情報を正規化した正規化抽出情報とを有する。

#### 【0035】

(文書処理システムの動作)

次に、第1の実施の形態の作用を、(1)基本動作、(2)情報抽出動作、(3)抽出情報表示動作に分けて説明する。以降、端末2を操作対象として動作を説明するが、端末3に置き換えた場合も同様の動作となる場合は説明を省略する。

#### 【0036】

(1) 基本動作

まず、利用者は、文書処理サーバ装置1の提供するサービスへログインを行うため端末2を操作する。端末2は、利用者から利用者ID及びパスワード等の情報の入力を受け付けると、当該情報とともに認証要求を文書処理サーバ装置1に送信する。

10

#### 【0037】

文書処理サーバ装置1は、端末2から利用者ID及びパスワード等の情報とともに認証要求を受信すると、予め登録された利用者ID及びパスワード等を含む図示しない利用者情報を参照し、利用者としての依頼者の認証を行う。

#### 【0038】

次に、利用者は、サービスへのログインが完了すると、契約書の文書情報を文書処理サーバ装置1へアップロードするため端末2を操作する。端末2は、当該文書情報を文書処理サーバ装置1へアップロードする。

20

#### 【0039】

(2) 情報抽出動作

図10は、文書処理サーバ装置1の情報抽出動作を示すフローチャートである。図6は、情報抽出動作の処理の一例を説明するための概略図である。

#### 【0040】

文書処理サーバ装置1の契約書受付手段100は、依頼者の操作する端末2から文書情報111を受け付けて記憶部11に格納する(S1)。

#### 【0041】

次に、文書処理サーバ装置1の契約書分割手段101は、図3に示すように文書情報111aを、契約書の構成要素毎に分割して単位情報112aとし、構造化する(S2)。一例として、文書情報111aは、契約書原本をスキャンしたものであってPDF(Portable Document Format)ファイル等の画像情報をあり、契約書分割手段101は、まず、文書情報111aをOCR(Optical Character Reader)等の方法を用いてテキスト化する。契約書分割手段101は、テキスト化された情報を、タイトル112a1、序文112a2、条文単位112a3~112a7...、項単位112a41、112a42...のように分割する。なお、契約書分割手段101は、対象とする文書情報111aをさらに号の単位で分割するものであってもよく、文書情報111aの構造に適した単位を用いて単位情報112aとする。なお、上記の分割は機械学習、正規表現等の技術を用いて行う。また、分割は、必須の動作項目ではなく、分割を行わずに以降の動作を行うものであってもよい。

30

#### 【0042】

次に、情報抽出手段102は、構造化された文書である単位情報112aから契約内容を表す情報を単位情報112aの参照先の位置情報とともに抽出情報113a1~113a5...(抽出情報113a)として抽出して記憶部11に格納する(S3)。なお、上記の情報の抽出は条件付き確率場(Conditional Random Field)による固有表現認識(Named Entity Recognition)等の技術を用いて行う。

40

#### 【0043】

次に、正規化手段103は、情報抽出手段102が抽出した抽出情報113aの内容を、それぞれキーワードの統一、他の情報の参照による補足等及び以下の図7に示すような処理を行うことで正規化して正規化抽出情報114a1~114a5...(正規化抽出情報

50

114a)として記憶部11に格納する(S4)。なお、正規化抽出情報114aは、タイトル、当事者1、当事者2、締結日、開始日、終了日、全文等の項目で管理される。また、上記の情報の正規化はレーベンシュタイン距離やキーワード類似度等の技術を用いて辞書(キーワードリスト)のキーワードを比較することで行う。

#### 【0044】

図7は、正規化手段103の動作例を説明するための概略図である。

#### 【0045】

正規化手段103は、抽出情報113として日付に関する抽出項目において「平成29年7月1日」のように和暦で記載された日付を受け付けた場合、正規化抽出情報114として「2017年7月1日」のように西暦で記載された日付に正規化する。なお、抽出情報113が西暦で記載されていても「2017/7/1」や「7/1/2017」、「July 1, 2017」のように記載順が異なる場合にも同様に正規化する。

10

#### 【0046】

また、正規化手段103は、抽出情報113として日付に関する抽出項目において「平成29年7月1日から1年間」のように期間を伴う記載を受け付けた場合、正規化抽出情報114として「2018年6月30日」のように終了日に該当する日付に正規化する。

20

#### 【0047】

また、正規化手段103は、抽出情報113として更新に関する抽出項目において「...本契約は同条件で更新されるものとし、以降も同様とする。」のように具体的な記載を受け付けた場合、正規化抽出情報114として「自動更新アリ」のように端的な内容の記載に正規化する。

20

#### 【0048】

また、正規化手段103は、抽出情報113として当事者に関する抽出項目において「甲(借主)株式会社LegalForce」のように当事者立場と当事者名を受け付けた場合、正規化抽出情報114として「株式会社LegalForce」のように当事者に正規化する。

#### 【0049】

また、正規化手段103は、抽出情報113として当事者に関する抽出項目において「本契約書の末尾に記載される者(中略)株式会社LegalForce」のように記載位置とその記載位置に記載された当事者名を受け付けた場合、正規化抽出情報114として「株式会社LegalForce」のように当事者名に正規化する。

30

#### 【0050】

また、正規化手段103は、抽出情報113として当事者に関する抽出項目において「株式会社LegalForce(以下、「甲」という。)と弁護士角田望(以下、「乙」という。)は、以下のとおり、顧問契約(以下、「本契約」という。)を締結する。」のように当事者の対を受け付けた場合、正規化抽出情報114として「株式会社LegalForce/角田望」のように当事者の対に正規化する。

#### 【0051】

また、正規化手段103は、抽出情報113として期間に関する抽出項目において「本契約の有効期限は、本契約の締結日より1年間とする。...締結日:2020年1月1日」のように有効期限の開始日と期間を受け付けた場合、正規化抽出情報114として「2020年1月1日」のように締結開始日に正規化する。

40

#### 【0052】

また、情報抽出手段102及び正規化手段103は、上記したように自動で抽出と正規化を行うが、下記の図9に示すように、固有名詞の他、日付、期間等については利用者の操作に応じて抽出と正規化を行うものであってもよい。

#### 【0053】

図9は、操作に応じて抽出情報を登録する場合に表示される画面の表示例を示す概略図である。

#### 【0054】

50

画面 103 b は、利用者が検索を所望する文字列を入力する入力欄 103 b<sub>1</sub>と、すべての検索結果を登録対象とするためのチェック欄 103 b<sub>2</sub>と、検索結果毎に登録対象とするためのチェック欄 103 b<sub>24</sub>、103 b<sub>25</sub>、103 b<sub>26</sub>…と、チェック欄でチェックした検索結果を登録するための登録ボタン 103 b<sub>3</sub>と、検索結果 103 b<sub>4</sub>、103 b<sub>5</sub>、103 b<sub>6</sub>…とを有する。各検索結果 103 b<sub>4</sub>、103 b<sub>5</sub>、103 b<sub>6</sub>…は同様の構成を有するため検索結果 103 b<sub>4</sub>の構成について代表して説明する。検索結果 103 b<sub>4</sub>は、契約書のタイトルとして検索結果を登録するためのボタン 103 b<sub>41</sub>と、当事者名として検索結果を登録するためのボタン 103 b<sub>42</sub>と、開始日として検索結果を登録するためのボタン 103 b<sub>43</sub>と、終了日として検索結果を登録するためのボタン 103 b<sub>44</sub>と、検索結果の文字列を表示する表示欄 103 b<sub>45</sub>とを有する。

10

#### 【0055】

利用者は、画面 103 bにおいて検索を所望する文字列を入力欄 103 b<sub>1</sub>に入力して検索された検索結果 103 b<sub>4</sub>、103 b<sub>5</sub>、103 b<sub>6</sub>…の内容を確認し、登録を所望する検査結果についてチェック欄 103 b<sub>2</sub>、103 b<sub>24</sub>、103 b<sub>25</sub>、103 b<sub>26</sub>…にチェックして、ボタン 103 b<sub>41</sub>～103 b<sub>44</sub>を適宜選択状態とし、登録ボタン 103 b<sub>3</sub>を押下操作することで登録操作を行う。

#### 【0056】

情報抽出手段 102 及び正規化手段 103 は、チェックされた文字列を抽出情報 113 及び正規化抽出情報 114 に登録する。

20

#### 【0057】

##### (3) 抽出情報表示動作

次に、利用者は、所望の契約書の内容参照を文書処理サーバ装置 1へ要求するため端末 2を操作して契約書を選択する。端末 2は、契約書の選択及び選択した契約書の内容参照を文書処理サーバ装置 1へ要求する。

#### 【0058】

文書処理サーバ装置 1の表示制御手段 104 は、契約書の選択及び契約書の内容参照の要求を受け付けると、正規化抽出情報 114 a を文書情報 111 a 及び単位情報 112 a とともに表示処理する。

#### 【0059】

図 8 は、表示制御手段 104 の表示例を示す概略図である。

30

#### 【0060】

画面 104 a は、表示制御手段 104 により表示される画面であり、契約書原本である文書情報 111 a を表示する文書情報表示欄 104 a<sub>1</sub>と、構造化された文書である単位情報 112 a を表示する構造化文書表示欄 104 a<sub>2</sub>と、正規化された抽出情報である正規化抽出情報 114 a を項目毎に表示する正規化抽出情報表示欄 104 a<sub>3</sub>とを有する。

#### 【0061】

次に、利用者は、正規化抽出情報表示欄 104 a<sub>3</sub>のうち所望の項目を選択するため端末 2を操作する。端末 2は、正規化抽出情報 114 a の項目の選択を文書処理サーバ装置 1へ要求する。

#### 【0062】

図 11 は、表示処理動作を説明するためのフローチャートである。

40

#### 【0063】

表示制御手段 104 は、正規化抽出情報 114 a の項目選択を受け付けると (S10) 、選択された項目である正規化抽出情報 104 a<sub>32</sub>を選択状態とし、抽出情報 113 a の参照先位置を参照して位置情報を取得し (S11) 、構造化された文書である構造化文書表示欄 104 a<sub>2</sub>の単位情報 104 a<sub>22</sub>のうち該当位置の文字列 104 b<sub>2</sub>を指示して表示する (S12) 。

#### 【0064】

利用者は、該当位置の文字列 104 b<sub>2</sub>を確認し、該当位置の単位情報 112 a 中の位置、文書情報 111 a における位置等を確認する。

50

**【 0 0 6 5 】**

(実施の形態の効果)

上記した実施の形態によれば、文書情報 1 1 1 をテキスト化して構造化し、構造化した単位情報 1 1 2 から文書の内容を示す文字列を抽出して抽出情報 1 1 3 とした後、抽出情報 1 1 3 の各文字列を正規化して正規化抽出情報 1 1 4 とともに、正規化抽出情報 1 1 4 の内容を単位情報 1 1 2 の該当位置とともに表示するようにしたため、当該抽出された文字列に表記上のゆれがある場合であっても、該当箇所を一元して管理することができる。

**【 0 0 6 6 】**

また、文書情報 1 1 1 、単位情報 1 1 2 、正規化抽出情報 1 1 4 を関連付けて表示制御するようにしたため、抽出された文字列の単位情報 1 1 2 中の位置情報を確認させることができるとともに、原本である文書情報 1 1 1 で確実に記載されていることを確認させることができる。

10

**【 0 0 6 7 】**

[他の実施の形態]

なお、本発明は、上記実施の形態に限定されず、本発明の趣旨を逸脱しない範囲で種々な変形が可能である。

**【 0 0 6 8 】**

例えば、文書情報 1 1 1 は契約書に限らず、情報が抽出できるものであれば、法律そのものの文書や取り扱い説明書等の法律以外の分野の文書であってよいし、同様に本発明を適用可能である。また、構成要素は、条文、項、号に限らず、単語や文字・記号であってもよいし、段落、文章であってもよい。また、文書情報 1 1 1 の言語は日本語、英語に限らず、情報を抽出可能な文を構成できる言語であれば他の言語にも適用可能である。

20

**【 0 0 6 9 】**

上記実施の形態では制御部 1 0 の各手段 1 0 0 ～ 1 0 4 の機能をプログラムで実現したが、各手段の全て又は一部を ASIC 等のハードウェアによって実現してもよい。また、上記実施の形態で用いたプログラムを CD-ROM 等の記録媒体に記憶して提供することもできる。また、上記実施の形態で説明した上記ステップの入れ替え、削除、追加等は本発明の要旨を変更しない範囲内で可能である。

**【 産業上の利用可能性】**

30

**【 0 0 7 0 】**

文書情報から文書の内容を示す文字列が抽出され、当該抽出された文字列に表記上のゆれがある場合であっても、該当箇所を一元して管理する文書処理プログラム、情報処理装置及び文書処理方法を提供する。

**【 符号の説明】****【 0 0 7 1 】**

- |       |             |
|-------|-------------|
| 1     | : 文書処理サーバ装置 |
| 2、 3  | : 端末        |
| 4     | : ネットワーク    |
| 5     | : 文書処理システム  |
| 1 0   | : 制御部       |
| 1 1   | : 記憶部       |
| 1 2   | : 通信部       |
| 1 0 0 | : 契約書受付手段   |
| 1 0 1 | : 契約書分割手段   |
| 1 0 2 | : 情報抽出手段    |
| 1 0 3 | : 正規化手段     |
| 1 0 4 | : 表示制御手段    |
| 1 1 0 | : 文書処理プログラム |
| 1 1 1 | : 文書情報      |

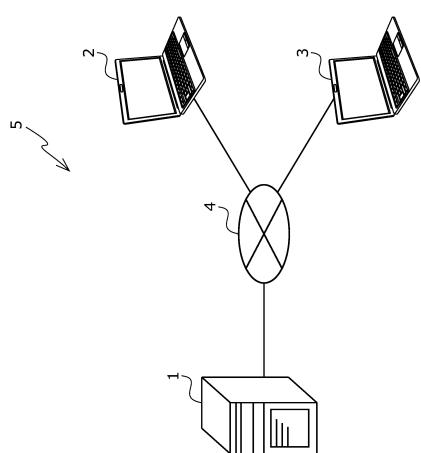
40

50

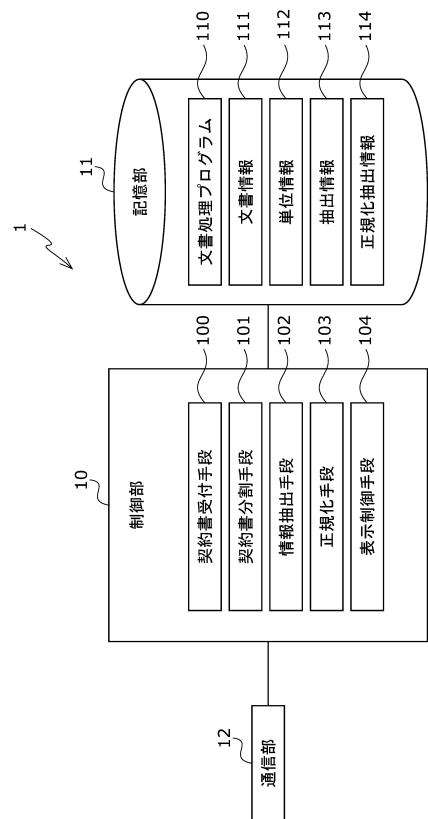
- 1 1 2 : 単位情報  
 1 1 3 : 抽出情報  
 1 1 4 : 正規化抽出情報

【図面】

【図 1】



【図 2】



10

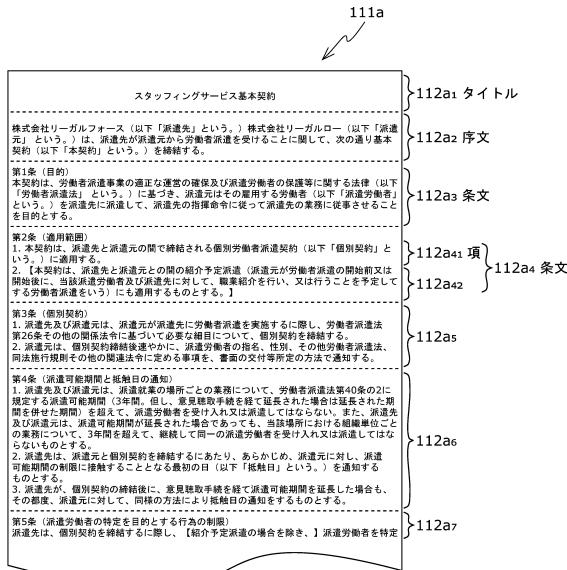
20

30

40

50

【図3】



【図4】

抽出ID	抽出情報	抽出項目	参照先位置
001	スタッフサービス基本契約	契約類型	タイトル
002	本契約の締結日	本契約の発効日	条文(2);1-7w
003	令和元年2月1日	締結日	条文(n);1-8w
004	株式会社リーガルフォース	当事者名1	序文;1-12w
⋮	⋮	⋮	⋮

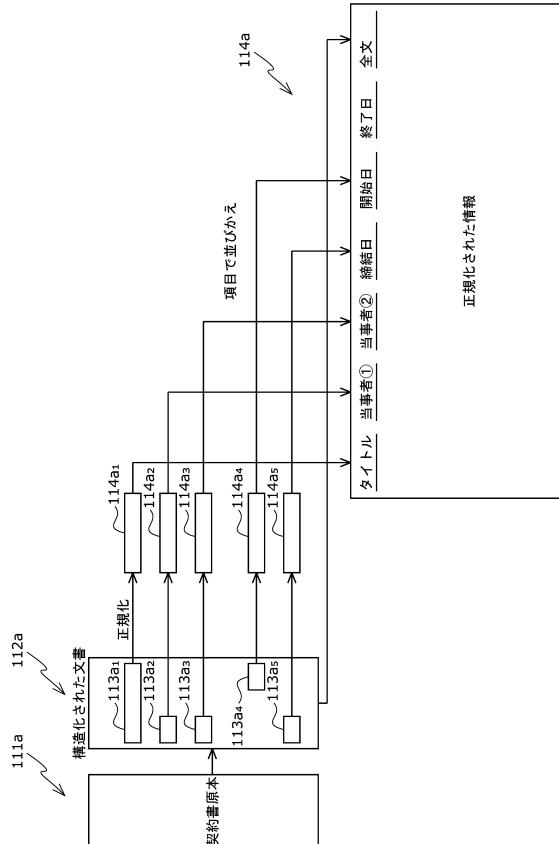
10

20

【図5】

抽出ID	抽出情報	正規化抽出情報
001	スタッフサービス基本契約	労働者派遣基本契約
002	本契約の締結日	2019/2/1 (ID:003 参照)
003	令和元年2月1日	2019/2/1
004	株式会社リーガルフォース	株式会社LegalForce
⋮	⋮	⋮

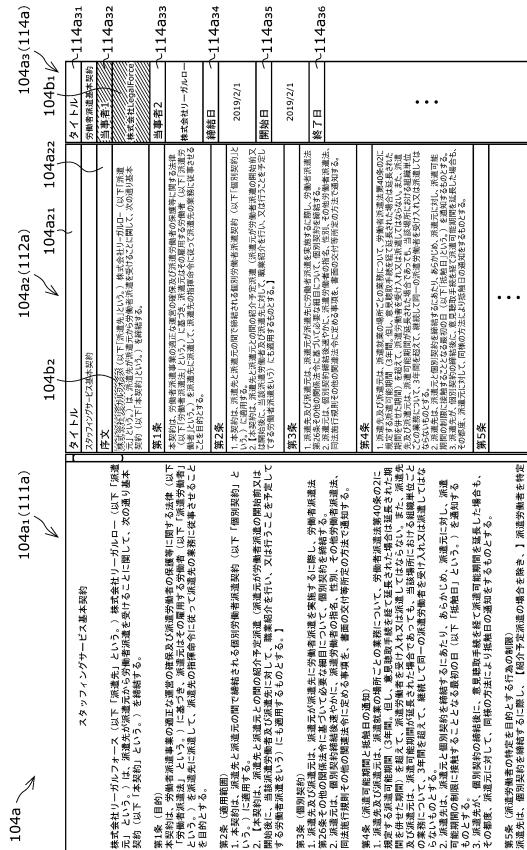
【図6】



〔 図 7 〕

抽出情報	抽出項目	正規化抽出情報
平成29年7月1日	契約締結日	2017年7月1日
平成29年7月1日から1年間	契約終了日	2018年6月30日
本契約の有効期間は、harami導入要綱第1項「1.契約期間」に掲げるとおりとし、甲又は乙のいずれかから3ヶ月前までに本契約を更新しない旨の書面による通知がなされない限り、本契約は自動的に同条件で更新されるものとし、以後も同様とする。	自動更新の有無	自動更新アリ
甲(借主)株式会社LegalForce	当事者	株式会社LegalForce
本契約書の末尾に記載される者 (中略) 株式会社LegalForce	当事者	株式会社LegalForce
株式会社LegalForce(以下、「甲」という。)と弁護士 角田望(以下、「乙」という。)は、以下のとおり、顧問契約(以下、「本契約」という。)を締結する。	当事者	株式会社LegalForce / 角田望
本契約の有効期間は、 <u>本契約の締結日</u> より1年間とする。 ... <u>締結 : 2020年1月1日</u>	契約開始日	2020年1月1日
⋮	⋮	⋮

【四】 8

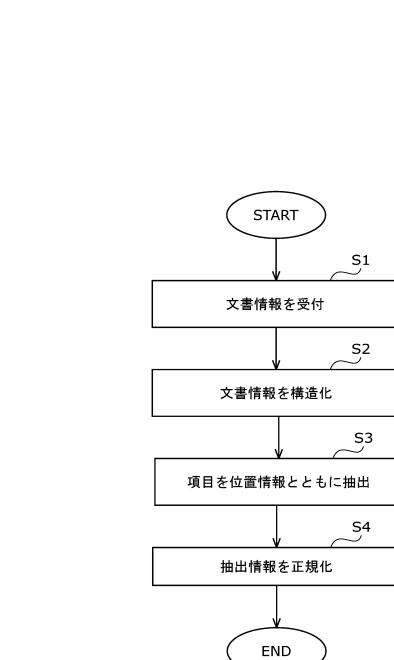


【 义 9 】

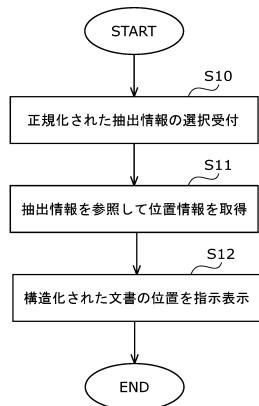
The diagram illustrates the structure of a document with several sections, each labeled with a specific 103b identifier:

- Section 1:** A search icon is followed by the text "株式会社LegalForce". This section includes a checkbox labeled "103b2" and a link labeled "103b1".
- Section 2:** A link labeled "情報の登録" (Information Registration) is shown, which leads to a form with fields: "Document title" (103b24), "Parties" (103b41), "Effective Date" (103b42), and "Expiration Date" (103b43). A checkmark is present next to the "Document title" field.
- Section 3:** A large text box contains the sentence: "株式会社LegalForceと、株式会社XXは、以下通り契約を締結する。".
- Section 4:** A checkbox labeled "103b25" is shown, leading to another form with fields: "Document title" (103bs), "Parties" (103bs), "Effective Date" (103bs), and "Expiration Date" (103bs).
- Section 5:** A text box contains the sentence: "株式会社YYと株式会社ZZは、株式会社LegalForceの株式の取得について下記の通り合意する。".
- Section 6:** A checkbox labeled "103b26" is shown, leading to a final form with fields: "Document title" (103b6), "Parties" (103b6), "Effective Date" (103b6), and "Expiration Date" (103b6).
- Section 7:** A text box contains the sentence: "株式会社YY（以下甲という）と、株式会社LegalForce（以下乙という）は、甲乙間で交換する秘密情報の取り扱いについて下記の通り合意する。".

【図 10】



【図 1 1】



10

20

30

40

50

---

フロントページの続き

洲フロント6階 株式会社L e g a l F o r c e内

(72)発明者 舟木 類佳

東京都江東区豊洲三丁目2番20号 豊洲フロント6階 株式会社L e g a l F o r c e内

審査官 成瀬 博之

(56)参考文献 国際公開第2018/042548 (WO, A1)

特開2019-045953 (JP, A)

特開2006-236262 (JP, A)

(58)調査した分野 (Int.Cl., DB名)

G 06 F 40/00 - 40/58