(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7598005号
(P7598005)

(45)発行日　令和6年12月11日(2024.12.11)

(24)登録日　令和6年12月3日(2024.12.3)

(51)国際特許分類		FI		
G06Q 10/107 (2023.01)		G06Q	10/107	
G06F 40/279 (2020.01)		G06F	40/279	
G06F 21/62 (2013.01)		G06F	21/62	345
G06F 13/00 (2006.01)		G06F	13/00	

請求項の数　14　（全18頁）

(21)出願番号　特願2020-188393(P2020-188393)
(22)出願日　令和2年11月12日(2020.11.12)
(65)公開番号　特開2022-77566(P2022-77566A)
(43)公開日　令和4年5月24日(2022.5.24)
審査請求日　令和5年11月9日(2023.11.9)

(73)特許権者　390002761
キヤノンマーケティングジャパン株式会社
東京都港区港南２丁目１６番６号
(73)特許権者　592135203
キヤノンＩＴソリューションズ株式会社
東京都港区港南２丁目１６番６号
(74)代理人　100189751
弁理士　木村 友輔
(72)発明者　田中 靖大
東京都港区港南２丁目１６番６号 キヤノンＩＴソリューションズ株式会社内
審査官　宮地 匡人

最終頁に続く

(54)【発明の名称】　情報処理装置、情報処理方法、プログラム

(57)【特許請求の範囲】
【請求項１】
電子メールからテキストデータを取得する取得手段と、
　テキストデータと、当該テキストデータに個人情報が含まれる旨の情報とを対応付けたデータにより学習された学習済みモデルにより、前記取得手段で取得したテキストデータにおける個人情報指数を出力する出力手段と、
電子メールの送信指示を受け付ける受付手段と、
電子メールから取得されたテキストデータに係る個人情報指数が所定の条件を満たす場合、前記受付手段により当該電子メールの送信指示を受け付けたことに応じて当該電子メールから取得したテキストデータに個人情報が含まれるか否かの選択を受け付ける選択受付手段と、
　を備え、
　前記選択受付手段により個人情報が含まれる旨の選択を受け付けた場合、当該テキストデータにより前記学習済みモデルを学習させることを特徴とする情報処理システム。
【請求項２】
電子メールから取得されたテキストデータに係る個人情報指数が所定の条件を満たす場合、当該電子メールの送信を保留する送信制御手段と、
　を備えることを特徴とする請求項１に記載の情報処理システム。
【請求項３】
　前記選択受付手段は、さらに、前記送信制御手段により送信が保留された電子メールに

ついて、送信を許可するか否かの選択を受け付けることを特徴とする請求項２に記載の情報処理システム。

【請求項４】

前記送信制御手段は、電子メールから取得されたテキストデータに係る個人情報指数が所定の条件を満たすことにより送信が保留された電子メールについて、前記選択受付手段により当該電子メールから取得したテキストデータに個人情報が含まれるか否かの選択および送信を許可する旨の選択を受け付けた場合に、当該電子メールを送信するよう制御することを特徴とする請求項３に記載の情報処理システム。

【請求項５】

前記送信制御手段は、電子メールから取得されたテキストデータに係る個人情報指数が所定の条件を満たさない場合、前記選択受付手段による個人情報が含まれるか否かの選択を受け付けることなく、当該電子メールを送信することを特徴とする請求項２乃至４のいずれか１に記載の情報処理システム。

【請求項６】

テキストデータから個人情報形成情報を抽出する抽出手段と、

前記抽出手段により抽出された個人情報形成情報と、当該個人情報形成情報の抽出元であるテキストデータに対して設定された個人情報が含まれるか否かの情報と、を対応付けたデータにより、学習済みモデルを生成する生成手段をさらに備えることを特徴とする請求項１乃至５のいずれか１項に記載の情報処理システム。

【請求項７】

前記生成手段は、前記抽出手段により抽出された個人情報形成情報の種別毎の抽出数と、当該個人情報形成情報の抽出元であるテキストデータに対して設定された個人情報が含まれるか否かの情報と、を対応付けたデータにより、学習済みモデルを生成することを特徴とする請求項６に記載の情報処理システム。

【請求項８】

前記選択受付手段により受け付けた個人情報が含まれるか否かの選択と、当該選択を行ったユーザまたはユーザの所属に係る情報とを対応付けて記憶し、

前記生成手段は、前記選択を行ったユーザまたはユーザの所属ごとに、学習済みモデルを生成することを特徴とする請求項６または７に記載の情報処理システム。

【請求項９】

前記所定の条件を満たす場合とは、出力手段により出力された個人情報指数が、既定の値以上である場合であることを特徴とする請求項１乃至８のいずれか１項に記載の情報処理システム。

【請求項１０】

テキストデータとは、電子メールの本文、および電子メールに添付されたファイルに含まれるテキストの少なくとも１つを含むことを特徴とする請求項１乃至９のいずれか１項に記載の情報処理システム。

【請求項１１】

コンピュータを、請求項１乃至１０のいずれか１項に記載の各手段として機能させるためのプログラム。

【請求項１２】

情報処理システムの取得手段が、電子メールからテキストデータを取得する取得工程と、

前記情報処理システムの出力手段が、テキストデータと、当該テキストデータに個人情報が含まれる旨の情報とを対応付けたデータにより学習された学習済みモデルにより、前記取得手段で取得したテキストデータにおける個人情報指数を出力する出力工程と、

前記情報処理システムの受付手段が、電子メールの送信指示を受け付ける受付工程と、

前記情報処理システムの選択受付手段が、電子メールから取得されたテキストデータに係る個人情報指数が所定の条件を満たす場合、前記受付工程により当該電子メールの送信指示を受け付けたことに応じて当該電子メールから取得したテキストデータに個人情報が含まれるか否かの選択を受け付ける選択受付工程と、

を備え、
前記選択受付工程により個人情報が含まれる旨の選択を受け付けた場合、当該テキストデータにより前記学習済みモデルを学習させることを特徴とする情報処理方法。
【請求項１３】
テキストデータを取得する取得手段と、テキストデータと、当該テキストデータに個人情報が含まれる旨の情報とを対応付けたデータにより学習された学習済みモデルにより、前記取得手段で取得したテキストデータにおける個人情報指数を出力する出力手段と、を備える装置と通信可能な情報処理装置であって、
送信指示された電子メールから取得したテキストデータを、前記装置に対して送信する送信手段と、
前記装置の出力手段により出力された個人情報指数を取得する取得手段と、
前記取得手段において取得した個人情報指数に基づき個人情報が所定の条件を満たす場合、当該送信指示された電子メールの送信を制御する送信制御手段と、
前記送信制御手段により送信が制御された電子メールから取得したテキストデータに個人情報が含まれるか否かの選択を受け付ける選択受付手段と、
前記選択受付手段により個人情報が含まれる旨の選択を受け付けた場合、当該テキストデータにより前記学習済みモデルを学習させるよう前記装置を制御する制御手段と、
を備えることを特徴とする情報処理装置。
【請求項１４】
テキストデータを取得する取得手段と、テキストデータと、当該テキストデータに個人情報が含まれる旨の情報とを対応付けたデータにより学習された学習済みモデルにより、前記取得手段で取得したテキストデータにおける個人情報指数を出力する出力手段と、を備える装置と通信可能な情報処理装置を、
送信指示された電子メールから取得したテキストデータを、前記装置に対して送信する送信手段と、
前記装置の出力手段により出力された個人情報指数を取得する取得手段と、
前記取得手段において取得した個人情報指数に基づき個人情報が所定の条件を満たす場合、当該送信指示された電子メールの送信を制御する送信制御手段と、
前記送信制御手段により送信が制御された電子メールから取得したテキストデータに個人情報が含まれるか否かの選択を受け付ける選択受付手段と、
前記選択受付手段により個人情報が含まれる旨の選択を受け付けた場合、当該テキストデータにより前記学習済みモデルを学習させるよう前記装置を制御する制御手段と、
して機能させるためのプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、情報処理装置、情報処理方法、プログラムに関する。
【背景技術】
【０００２】
個人情報保護法の改正に伴い適用範囲が拡大されたことにより、事業者における情報漏洩時のリスクは高まっており、個人情報漏洩防止対策の重要性は増してきている。
【０００３】
個人情報を扱う場合は、個人データの安全管理のために必要な措置を講じる必要があるが、電子化された個人データは複製が容易なため、個人の端末において作業を行った際に複製したファイルが作業後も放置されているなど、管理困難な状態に陥りやすい。
【０００４】
更に、電子メールなどによってデータを外部に送信する場合は、誤送信のチェックだけではなく、誤って個人情報を含んだファイルを添付していないかの確認が必要となる。
【０００５】
しかしながら、組織において、各個人が管理する端末に保存されているファイルを全て

確認したり、電子メールに添付されているファイルを都度開いて確認したりするには膨大なコストが必要となる。

【０００６】

膨大な作業コストを抑えながら、組織における個人情報を適切に管理するために、個人情報を含むファイルを高い確度で判断できる技術が望まれている。

【０００７】

特許文献１には、個人情報を形成し得る個人情報形成情報を検出し、検出した個人情報形成情報の位置関係から個人情報を特定し、個人情報の文字数や項目数などを計測し、計測値から個人情報を含む度合を示す指標値として示す技術が記載されている。

【先行技術文献】

【特許文献】

【０００８】

【文献】特開２００８－１１７２８７号公報

【発明の開示】

【発明が解決しようとする課題】

【０００９】

しかし、特許文献１においては、個人情報を含む度合いを示す指標値の算出式を人手で調整する必要があり、適切な指標値を算出するには手間がかかってしまう。

【００１０】

そこで、本発明は、より効率的に、高い精度で、個人情報を含む文書を判定できる仕組み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】

【００１１】

本発明の情報処理システムは、テキストデータを取得する取得手段と、テキストデータと、当該テキストデータに個人情報が含まれる旨の情報とを対応付けたデータにより学習された学習済みモデルにより、前記取得手段で取得したテキストデータに個人情報が含まれるかを判定する判定手段と、を備えることを特徴とする。

【発明の効果】

【００１２】

本発明によれば、より効率的に、高い精度で、個人情報を含む文書の判定が可能となる。

【図面の簡単な説明】

【００１３】

【図１】本発明の実施形態における、個人情報検査装置を用いたＷｅｂメールサービスのシステム構成の一例を示す図である。

【図２】本発明の実施形態における、個人情報検査装置、Ｗｅｂメールサービス、管理者端末、クライアント端末のハードウェア構成の一例を示すブロック図である。

【図３】本発明の実施形態における、個人情報検査装置およびＷｅｂメールサービスの機能構成の一例を示す図である。

【図４】本発明の実施形態における、個人情報を学習する処理の一例を示すフローチャートである。

【図５】本発明の実施形態における、判定知識保存領域の構成の一例を示す図である。

【図６】本発明の実施形態における、文書に対する特徴ベクトルを生成する処理の一例を示すフローチャートである。

【図７】本発明の実施形態における、学習データの一例を示す図である。

【図８】本発明の実施形態における、文書から取得した個人情報形成情報の一例を示す図である。

【図９】本発明の実施形態における、特徴ベクトルの一例を示す図である。

【図１０】本発明の実施形態における、特徴量保存テーブルの一例を示す図である。

【図１１】本発明の実施形態における、学習の結果獲得した機械学習のモデルの一例を示す図である。

【図１２】本発明の実施形態における、個人情報指数算出処理の一例を示すフローチャートである。
【図１３】本発明の実施形態における、個人情報指数算出式の一例を示す図である。
【図１４】本発明の実施形態における、予測値算出の過程の一例を示す図である。
【図１５】本発明の実施形態における、Ｗｅｂメールサービスにおける電子メールの検査処理の一例を示すフローチャートである。
【図１６】本発明の実施形態における、Ｗｅｂメールサービスにおけるメール編集画面の一例を示す図である。
【図１７】本発明の実施形態における、Ｗｅｂメールサービスにおける送信確認画面の一例を示す図である。
【発明を実施するための形態】
【００１４】
以下、図面を参照して、本発明の実施形態を詳細に説明する。
【００１５】
図１は、本発明の実施形態における個人情報検査装置を用いたＷｅｂメールサービスのシステム構成の一例を示す図である。
【００１６】
個人情報検査装置１００と、Ｗｅｂメールサービス１１０および管理者端末１２０はローカルエリアネットワーク１３０を介して接続される構成となっている。また、Ｗｅｂメールサービス１１０は、外部ネットワーク１４０を経由して、クライアント１５０に接続可能な構成となっている。
【００１７】
Ｗｅｂメールサービス１１０は、一般的な電子メールクライアントの機能、即ち電子メールの編集機能とメールの送受信機能を、クライアント端末１５０に提供する。
【００１８】
また、Ｗｅｂメールサービス１１０は、クライアント端末１５０においてユーザがメールの送信または明示的は個人情報の検査を指示すると、個人情報検査装置１００に個人情報の有無を判定する処理の実行を要求する。
【００１９】
個人情報検査装置１００が送信しようとする電子メールに個人情報を含む可能性があると判断した場合、Ｗｅｂメールサービス１１０は、ユーザに対し個人情報有無の確認と送信可否の判断を促すメッセージを表示し、ユーザの送信可否の判断に基づき送信を制御する。
【００２０】
本発明の実施形態においては、個人情報検査装置１００は、クライアント端末１５０に対し、外部ネットワーク１４０を経由するサービスとして構成してもよいし、Ｗｅｂメールサービス１１０の機能の一部として構成してもよい。
【００２１】
また、クライアント端末１５０上で動作するメールクライアントプログラムの機能として構成してもよい。
【００２２】
図２は、本発明の実施形態における個人情報検査装置１００、Ｗｅｂメールサービス１１０、管理者端末１２０、クライアント端末１５０のハードウェア構成の一例を示すブロック図である。
【００２３】
図２に示すように、情報処理装置は、システムバス２００を介してＣＰＵ（Ｃｅｎｔｒａｌ Ｐｒｏｃｅｓｓｉｎｇ Ｕｎｉｔ）２０１、ＲＯＭ（Ｒｅａｄ Ｏｎｌｙ Ｍｅｍｏｒｙ）２０２、ＲＡＭ（Ｒａｎｄｏｍ Ａｃｃｅｓｓ Ｍｅｍｏｒｙ）２０３、記憶装置２０４、入力コントローラ２０５、音声コントローラ２０６、ビデオコントローラ２０７、メモリコントローラ２０８、および通信Ｉ／Ｆコントローラ２０９が接続される。

【００２４】
ＣＰＵ２０１は、システムバス２００に接続される各デバイスやコントローラを統括的に制御する。
【００２５】
ＲＯＭ２０２あるいは外部メモリ２１３は、ＣＰＵ２０１が実行する制御プログラムであるＢＩＯＳ（Ｂａｓｉｃ　Ｉｎｐｕｔ／Ｏｕｔｐｕｔ　Ｓｙｓｔｅｍ）やＯＳ（Ｏｐｅｒａｔｉｎｇ　Ｓｙｓｔｅｍ）や、本情報処理方法を実現するためのコンピュータ読み取り実行可能なプログラムおよび必要な各種データ（データテーブルを含む）を保持している。
【００２６】
ＲＡＭ２０３は、ＣＰＵ２０１の主メモリ、ワークエリア等として機能する。ＣＰＵ２０１は、処理の実行に際して必要なプログラム等をＲＯＭ２０２あるいは外部メモリ２１３からＲＡＭ２０３にロードし、ロードしたプログラムを実行することで各種動作を実現する。
【００２７】
入力コントローラ２０５は、キーボード２１０や不図示のマウス等のポインティングデバイス等の入力装置からの入力を制御する。入力装置がタッチパネルの場合、ユーザがタッチパネルに表示されたアイコンやカーソルやボタンに合わせて押下（指等でタッチ）することにより、各種の指示を行うことができることとする。
【００２８】
また、タッチパネルは、マルチタッチスクリーンなどの、複数の指でタッチされた位置を検出することが可能なタッチパネルであってもよい。
【００２９】
ビデオコントローラ２０７は、ディスプレイ２１２などの外部出力装置への表示を制御する。ディスプレイは本体と一体になったノート型パソコンのディスプレイも含まれるものとする。なお、外部出力装置はディスプレイに限ったものははく、例えばプロジェクタであってもよい。また、前述のタッチ操作を受け付け可能な装置については、入力装置も提供する。
【００３０】
なおビデオコントローラ２０７は、表示制御を行うためのビデオメモリ（ＶＲＡＭ）を制御することが可能で、ビデオメモリ領域としてＲＡＭ２０３の一部を利用することもできるし、別途専用のビデオメモリを設けることも可能である。
【００３１】
メモリコントローラ２０８は、外部メモリ２１３へのアクセスを制御する。外部メモリとしては、ブートプログラム、各種アプリケーション、フォントデータ、ユーザファイル、編集ファイル、および各種データ等を記憶する外部記憶装置（ハードディスク）、フレキシブルディスク（ＦＤ）、或いはＰＣＭＣＩＡカードスロットにアダプタを介して接続されるコンパクトフラッシュ（登録商標）メモリ等を利用可能である。
【００３２】
通信Ｉ／Ｆコントローラ２０９は、ネットワークを介して外部機器と接続・通信するものであり、ネットワークでの通信制御処理を実行する。例えば、ＴＣＰ／ＩＰを用いた通信やＩＳＤＮなどの電話回線、および携帯電話の４Ｇ回線、５Ｇ回線等を用いた通信が可能である。
【００３３】
尚、ＣＰＵ２０１は、例えばＲＡＭ２０３内の表示情報用領域へアウトラインフォントの展開（ラスタライズ）処理を実行することにより、ディスプレイ２１２上での表示を可能としている。また、ＣＰＵ２０１は、ディスプレイ２１２上の不図示のマウスカーソル等でのユーザ指示を可能とする。
【００３４】
図３は、個人情報検査装置およびＷｅｂメールサービスの機能構成の一例を示す図である。

【００３５】
メール編集処理部３１１は、クライアント端末１５０のＷｅｂブラウザ３２０に表示された画面を介して受け付けたユーザからの操作・指示に基づき、電子メールの作成・編集を行う機能を備える。
【００３６】
送信制御処理部３１２は、ユーザから電子メールの送信指示を受け付けると、当該電子メールからテキスト（本文や添付ファイル等）を抽出し、個人情報検査装置１００の各処理部による個人情報指数の算出処理を制御する。また、個人情報指数の算出処理の結果に基づき、図１７に示す確認ダイアログを表示したり、ユーザから送信の可否を受け付けることで電子メールの送信処理を制御する機能を備える。
【００３７】
送信処理部３１３は、送信制御処理部３１２において送信することが決定された電子メールを外部ネットワーク１４０に送出する機能を備える。
【００３８】
個人情報学習処理部３０１は、管理者端末１２０からアップロードされた情報に基づき、図４のフローチャートに示す個人情報の学習処理を実行する機能を備える。
【００３９】
特徴ベクトル生成処理部３０２は、処理対象の文書に対して特徴ベクトルを生成する機能を備える。
【００４０】
判定知識保存領域３０３は、特徴ベクトル生成処理部３０２において生成された特徴ベクトルと、学習データ学習データにおいて文書ごとに付与されている個人情報・非個人情報を示すラベルとの組を保存する領域である。
【００４１】
個人情報検査処理部３０４は、送信制御処理部３１２から取得したテキストに対して、図１２に示すテキスト内容が個人情報である可能性を算出する処理を実行する機能を備える
【００４２】
（個人情報学習処理）
次に図４のフローチャートを用いて、本発明の実施形態における個人情報学習処理部３０１が実行する個人情報の学習処理について説明する。
図４は、個人情報検査装置１００のＣＰＵ２０１が所定の制御プログラムを読み出して実行する処理であり、管理者端末１２０からアップロードされた個人情報・非個人情報を予め分類された文書に基づき、個人情報の特徴を学習する処理を示すフローチャートである。
【００４３】
ステップＳ４０１では、個人情報学習処理部３０１は、管理者端末１３０におけるＷｅｂブラウザ３３０からアップロードされた文書を学習データとして取得する。
【００４４】
ステップＳ４０２では、個人情報学習処理部３０１は、ステップＳ４０１で取得した学習データにおける文書に対して、ステップＳ４０５までの繰り返し処理を開始する。
【００４５】
ステップＳ４０３では、個人情報学習処理部３０１は、処理対象の文書に対して、特徴ベクトル生成処理部３０２を用いて特徴ベクトルを生成する。特徴ベクトルの生成方法については後述する。
【００４６】
ステップＳ４０４では、個人情報学習処理部３０１は、ステップＳ４０３で生成した特徴ベクトルと、学習データにおいて文書ごとに付与されている個人情報・非個人情報を示すラベルとの組を、図５に示す判定知識保存領域３０３における特徴ベクトル保存テーブル５０１に保存する。
【００４７】
ステップＳ４０５では、個人情報学習処理部３０１は、処理対象となる文書がまだあれ

ば、ステップＳ４０２からの繰り返し処理を実施する。処理対象となる文書がなければ、ステップＳ４０６に処理を移す。

【００４８】

ステップＳ４０６では、個人情報学習処理部３０１は、特徴ベクトル保存テーブル５０１の特徴ベクトルとラベル（個人情報・非個人情報）に対して機械学習を用いて学習を行い、学習の結果としてモデル（学習済みモデル）を獲得する。

【００４９】

ステップＳ４０７では、個人情報学習処理部３０１は、ステップＳ４０６で獲得したモデルを判定知識保存領域３０３におけるモデル保存テーブル５０２に保存する。

【００５０】

本発明の実施形態においては、個人情報学習処理を個人情報検査処理と同一のハードウェア上で実行するように構成しているが、異なるハードウェア上で実行するように構成しても構わない。

また、学習データをアップロードするように構成しているが、個人情報学習処理を行うハードウェア上で管理するように構成してもよい

【００５１】

（特徴ベクトル生成処理）

次に図６のフローチャートを用いて、本発明の実施形態における特徴ベクトル生成処理部３０２が実行する特徴ベクトルの生成処理について説明する。

【００５２】

ステップＳ６０１では、特徴ベクトル生成処理部３０２は、入力として受け取った文書のテキストから、予め個人情報項目や要配慮情報項目として定義された単語やパターンに基づき、形態素解析やパターマッチングを用いて、出現位置および文字数とともに個人情報形成情報として抽出する。

【００５３】

ステップＳ６０２では、特徴ベクトル生成処理部３０２は、文書および抽出した個人情報形成情報から、種別や文字数に基づいた統計値を求める。

【００５４】

ステップＳ６０３では、特徴ベクトル生成処理部３０２は、ステップＳ６０２で求めた統計値を特徴ベクトルとしてまとめる。

【００５５】

（個人情報学習処理具体例）

次に個人情報学習処理の具体例として、図７に示す学習データ７００に対して、図４に示す処理が実施された場合について説明する。

【００５６】

ステップＳ４０１では、個人情報学習処理部３０１は、管理者端末１３０におけるＷｅｂブラウザ３３０からアップロードされた文書７００を学習データとして取得する。

【００５７】

ステップＳ４０２では、個人情報学習処理部３０１は、ステップＳ４０１で取得した学習データ７００における文書７０１に対して、ステップＳ４０５までの繰り返し処理を開始する。

【００５８】

ステップＳ４０３では、個人情報学習処理部３０１は、処理対象の文書７０１に対して、特徴ベクトル生成処理部３０２を用いて特徴ベクトルを生成する。

【００５９】

ステップＳ６０１では、特徴ベクトル生成処理部３０２は、入力として受け取った文書７０１のテキストから、予め個人情報項目や要配慮情報項目として定義された単語やパターンに基づき、形態素解析やパターマッチングを用いて、出現位置および文字数とともに図８に示す個人情報形成情報８０１を抽出する。

【００６０】

ステップＳ６０２では、特徴ベクトル生成処理部３０２は、文書および抽出した個人情報形成情報８０１から、図９に示す種別や文字数に基づいた統計値を求める。

【００６１】

ステップＳ６０３では、特徴ベクトル生成処理部３０２は、ステップＳ６０２で求めた統計値を特徴ベクトル９０１としてまとめる。本発明の実施例においては、説明を容易にするために正規化等の処理は省略してるが、正規化を行うように構成してもよい。

【００６２】

ステップＳ４０４では、個人情報学習処理部３０１は、ステップＳ４０３で生成した特徴ベクトル９０１と、学習データにおいて文書ごとに付与されている個人情報・非個人情報を示すラベルＴＲＵＥの組１００１を、図１０に示す特徴ベクトル保存テーブル５０１に保存する。

【００６３】

ステップＳ４０５では、個人情報学習処理部３０１は、処理対象となる文書７０２がまだあるので、ステップＳ４０２からの繰り返し処理を実施する。

【００６４】

以下、学習データ７００の文書全てに対して同様のステップを実施する。

【００６５】

ステップＳ４０６では、個人情報学習処理部３０１は、特徴ベクトル保存テーブル５０１に対して機械学習により学習を行い、モデルを獲得する。

【００６６】

本発明の実施例においては、線形のＳＶＭ（Ｓｕｐｐｏｒｔ　Ｖｅｃｔｏｒ　Ｍａｃｈｉｎｅ）を用いる。

【００６７】

ステップＳ４０７では、個人情報学習処理部３０１は、学習結果として獲得した図１１に示すモデルを判定知識保存領域３０３におけるモデル保存テーブル５０２に保存する。獲得したモデルは予測値（計算結果）が正の値となる場合に個人情報と判断する。

【００６８】

本発明の実施例においては、説明を容易にするために機械学習の手法として線形のＳＶＭを用いて説明したが、正否の二値に分類することができ、その度合いを数値として表現できるなら、いずれの手法を用いてもよい

【００６９】

（個人情報検査処理）

次に図１２のフローチャートを用いて、本発明の実施形態における個人情報検査処理部３０４が実行する個人情報の検査処理について説明する。

【００７０】

図１２のフローチャートは、個人情報検査装置１００のＣＰＵ２０１が所定の制御プログラムを読み出して実行する処理であり、入力テキストに対して、テキストの内容が個人情報である可能性を示す度合を数値として算出する処理を示すフローチャートである。

【００７１】

ステップＳ１２０１では、個人情報検査処理部３０４は、処理対象となる文書のテキストを取得する。

【００７２】

ステップＳ１２０２では、個人情報検査処理部３０４は、ステップＳ１２０１で取得したテキストに対して、前述した図６のフローチャートを用いて、入力文書に対する特徴ベクトルを取得する。

【００７３】

ステップＳ１２０３では、個人情報検査処理部３０４は、ステップＳ１２０２で取得した特徴ベクトルに対して、判定知識保存領域３０３におけるモデル保存テーブル５０２に保存されているモデルを用いて予測値を算出する。

【００７４】

ステップＳ１２０４では、個人情報検査処理部３０４は、ステップＳ１２０３で算出した予測値に対して、感覚的に把握が容易な値となるように変換し、個人情報指数として出力する。
本発明の実施例においては、図１３に示すような個人情報指数算出式を用いるが、他の算出式を用いてもよいし、予測値をそのまま用いてもよい
【００７５】
（個人情報検査処理具体例）
次に個人情報検査処理の具体例として、図１４に示す検査対象文書１４０１に対して、図１２に示す処理が実施された場合について説明する。
【００７６】
ステップＳ１２０１では、個人情報検査処理部３０４は、処理対象となる文書１４０１からテキスト１４０２を取得する。
【００７７】
ステップＳ１２０２では、個人情報検査処理部３０４は、ステップＳ１２０１で取得したテキストに対して、前述した図６のフローチャートを用いて、入力文書に対する特徴ベクトル１４０３を取得する。
【００７８】
ステップＳ１２０３では、個人情報検査処理部３０４は、ステップＳ１２０２で取得した特徴ベクトル１４０３に対して、判定知識保存領域３０３におけるモデル保存テーブル５０２に保存されているモデルを用いて予測値として１４６０．９６０を算出する（１４０４）。
【００７９】
ステップＳ１２０４では、個人情報検査処理部３０４は、ステップＳ１２０３で算出した予測値１４６０．９６０に対して、図１３に示す個人情報指数算出式を用いて個人情報指数を算出する。ここではＲｍｉｎ＝－１７００、Ｒｍａｘ＝２０００であったとする。結果として５０＋５０×１４６０．９６０÷ｍａｘ（２０００，－（－１７００））＝８６．５２４を個人情報指数として得る。
【００８０】
このように、個人情報を含むか否かを判定された文書集合を用意するだけで、人手による調整不要で、個人情報である確度となる個人情報指数を算出すことが可能となる。
また、個人情報指数は、統計情報に基づいて算出されるので、個人情報形成情報の抽出に多少の誤りがあったとしても、高い精度で個人情報である確度を示すことができる。
算出した個人情報指数を用いれば、確度に応じた処理を実現することができ、効率的な個人情報の管理が可能となる
【００８１】
次に図１５に示すフローチャートを用いて、本発明の具体的な適用例として、電子メールの送信時に個人情報の送信を確認し、更にユーザからのフィードバックを取得する例を説明する。
なお、本実施例においては電子メールの送信を適用例として説明するが、これ以外にも、文書を印刷する際に個人情報が含まれる文書かを判定したり、文書をWebサービスにアップロードする際に判定するなどの適用例がある。
【００８２】
図１５は、Ｗｅｂメールサービス１１０のＣＰＵ２０１が所定の制御プログラムを読み出して実行する処理であり、クライアント端末１５０のＷｅｂブラウザ３２０に表示された図１６のような電子メール編集画面１６０１で、送信ボタン１６０２の押下によって送信指示がなされた場合に実施される処理を示すフローチャートである。
【００８３】
ステップＳ１５０１では、送信制御処理部３１２は、電子メールの送信指示を受け取る。
【００８４】
ステップＳ１５０２では、送信制御処理部３１２は、送信指示された電子メールに添付

されているファイルに対し、ステップＳ１５０９までの繰り返し処理を開始する。なお、添付ファイルだけでなく、電子メール本文なども対象としても良い。

【００８５】

ステップＳ１５０３では、送信制御処理部３１２は、処理対象の添付ファイルからテキストを抽出する。添付ファイルが画像の場合には、当該画像に対してＯＣＲ処理することでテキストを取得する。

【００８６】

ステップＳ１５０４では、送信制御処理部３１２は、ステップＳ１５０３で抽出したテキストに対して、個人情報検査装置１００を用いて個人情報指数を算出する。

【００８７】

ステップＳ１５０５では、送信制御処理部３１２は、ステップＳ１５０４で算出した個人情報指数が規定値以上の場合、処理をステップＳ１５０６に移す。規定値未満であった場合、ステップＳ１５０９に処理を移す。

【００８８】

ステップＳ１５０６では、送信制御処理部３１２は、図１７に示す確認ダイアログを表示し、ユーザに送信の可否を取得するとともに個人情報か否かの判断を取得する。

【００８９】

ステップＳ１５０７では、送信制御処理部３１２は、ステップＳ１５０５で取得した個人情報か否かのユーザによる判断を、個人情報検査装置１００に送り、個人情報検査装置１００で処理中の特徴ベクトルと一緒に特徴ベクトル保存テーブル５０１に保存するよう指示する。

【００９０】

ステップＳ１５０８では、送信制御処理部３１２は、ステップＳ１５０５で取得した送信可否が「送信」であった場合、ステップＳ１５０９に処理を移す。送信可否が「キャンセル」であった場合、処理を終了する。

【００９１】

ステップＳ１５０９では、送信制御処理部３１２は、処理対象となる添付ファイルがまだあれば、ステップＳ１５０２からの繰り返し処理を実施する。処理対象となる添付ファイルがなければ、ステップＳ１５１０に処理を移す。

【００９２】

ステップＳ１５１０では、送信制御処理部３１２は、電子メールを送信処理部３１３に送り電子メールを送出する。

【００９３】

このように、ユーザによる個人情報か否かの判断を収集すれば、個人情報検査装置１００において、定期的に学習処理を実施することで、人手の調整を必要とせずに、ユーザの判断に適合した個人情報指数の算出が可能となる。

【００９４】

本発明の実施例においては、保存した特徴ベクトル全てから学習するように構成しているが、特徴ベクトルの保存する際に、ユーザの所属する部門やユーザを識別する情報を付与し、部門やユーザごとに学習を行うように構成してもよい。そして、当該ユーザや当該ユーザの所属部門に係る文書に対しては、当該学習により生成されたモデルを用いて、個人情報が含まれるか否かを判定してもよい。

【００９５】

本発明は、例えば、システム、装置、方法、プログラムもしくは記録媒体等としての実施態様をとることが可能である。具体的には、複数の機器から構成されるシステムに適用しても良いし、また、一つの機器からなる装置に適用しても良い。

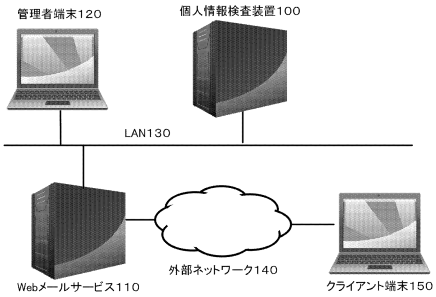
【００９６】

また、本発明におけるプログラムは、図４、図１２に示すフローチャートの処理方法をコンピュータが実行可能なプログラムであり、本発明の記憶媒体は図４、図１２の処理方法をコンピュータが実行可能なプログラムが記憶されている。なお、本発明におけるプロ

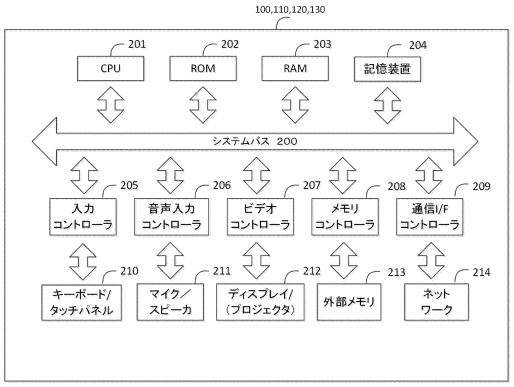
グラムは図４、図１２の各装置の処理方法ごとのプログラムであってもよい。
【００９７】
以上のように、前述した実施形態の機能を実現するプログラムを記録した記録媒体を、システムあるいは装置に供給し、そのシステムあるいは装置のコンピュータ（またはＣＰＵやＭＰＵ）が記録媒体に格納されたプログラムを読み出し、実行することによっても本発明の目的が達成されることは言うまでもない。
【００９８】
この場合、記録媒体から読み出されたプログラム自体が本発明の新規な機能を実現することになり、そのプログラムを記録した記録媒体は本発明を構成することになる。
【００９９】
プログラムを供給するための記録媒体としては、例えば、フレキシブルディスク、ハードディスク、光ディスク、光磁気ディスク、ＣＤ－ＲＯＭ、ＣＤ－Ｒ、ＤＶＤ－ＲＯＭ、磁気テープ、不揮発性のメモリカード、ＲＯＭ、ＥＥＰＲＯＭ、シリコンディスク等を用いることが出来る。
【０１００】
また、コンピュータが読み出したプログラムを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムの指示に基づき、コンピュータ上で稼働しているＯＳ（オペレーティングシステム）等が実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。
【０１０１】
さらに、記録媒体から読み出されたプログラムが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書き込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるＣＰＵ等が実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。
【０１０２】
また、本発明は、複数の機器から構成されるシステムに適用しても、ひとつの機器から成る装置に適用しても良い。また、本発明は、システムあるいは装置にプログラムを供給することによって達成される場合にも適応できることは言うまでもない。この場合、本発明を達成するためのプログラムを格納した記録媒体を該システムあるいは装置に読み出すことによって、そのシステムあるいは装置が、本発明の効果を享受することが可能となる。
【０１０３】
さらに、本発明を達成するためのプログラムをネットワーク上のサーバ、データベース等から通信プログラムによりダウンロードして読み出すことによって、そのシステムあるいは装置が、本発明の効果を享受することが可能となる。なお、上述した各実施形態およびその変形例を組み合わせた構成も全て本発明に含まれるものである。
【符号の説明】
【０１０４】
１００　個人情報検査装置
１１０　Ｗｅｂメールサービス
１２０　管理者端末
１３０　ＬＡＮ
１４０　外部ネットワーク
１５０　クライアント端末１５０

【図面】

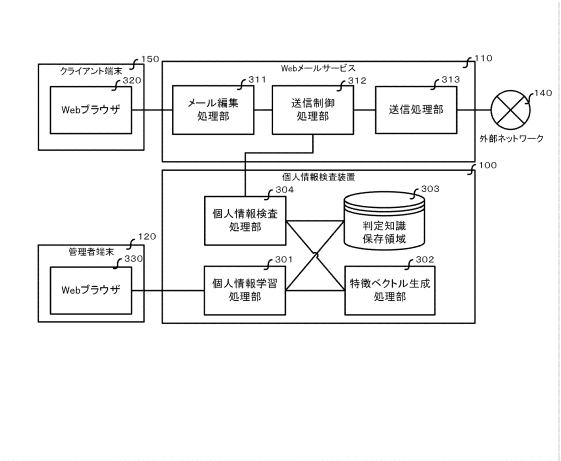
【図１】


管理者端末120
個人情報検査装置100
LAN130
Webメールサービス110
外部ネットワーク140
クライアント端末150

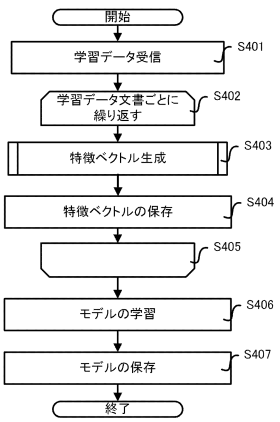
【図２】


100,110,120,130
201 CPU
202 ROM
203 RAM
204 記憶装置
システムバス 200
205 入力コントローラ
206 音声入力コントローラ
207 ビデオコントローラ
208 メモリコントローラ
209 通信I/Fコントローラ
210 キーボード/タッチパネル
211 マイク/スピーカ
212 ディスプレイ/(プロジェクタ)
213 外部メモリ
214 ネットワーク

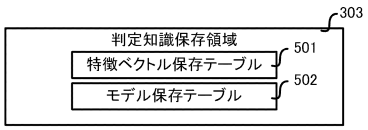
【図３】


クライアント端末 150
320 Webブラウザ
Webメールサービス 110
311 メール編集処理部
312 送信制御処理部
313 送信処理部
140 外部ネットワーク
100 個人情報検査装置
304 個人情報検査処理部
303 判定知識保存領域
管理者端末 120
330 Webブラウザ
301 個人情報学習処理部
302 特徴ベクトル生成処理部

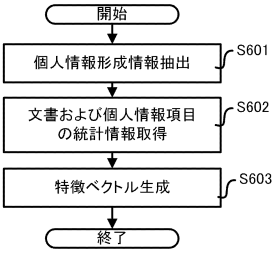
【図４】


開始
学習データ受信 S401
学習データ文書ごとに繰り返す S402
特徴ベクトル生成 S403
特徴ベクトルの保存 S404
S405
モデルの学習 S406
モデルの保存 S407
終了

【図５】



【図６】



【図７】

700

文章	個人情報
履歴書　西暦 2019 年 10 月 7 日現在 写真を貼る位置　1.縦36～40mm 横24～30mm 2.本人単身胸から上 3.裏面にのりづけ 4.裏面に氏名記入 姓 名 大迫 裕也 1993年03月21日生 満２７歳...	TRUE
事業レビュー議事録 日時：平成2５年４月５日 場所：６階６０２会議室 参加者：森保部長、吉田課長、大迫、鈴木、久保 目的：毎年、事業レビュー（次に掲げる取組をいう。以下同じ。）を実施することにより、所掌する事業のより効果的かつ効率的な実 施並びに説明責任及び透明性の確保を図り...	FALSE
検証を行うに当たっては、次に掲げる取組を行うこと。 事業に係る予算の執行状況等を分かりやすい形で公表すること...	FALSE
姓　名　郵便番号　住所　生年月日　年齢 4614　北田　清一　849-1323　佐賀県鹿島... 1936年09月02日 83 3753　石本　利昭　277-0823　千葉県柏市... 1947年01月10日 73 184　飯塚　英瑠　904-2422　沖縄県...	TRUE
領収書 No. 発行日 2020/8/9 第１電気工事株式会社御中 〒123-4567 滋賀県甲賀市１２３番地 農事組合法人ＡＢＣ　営農組合 ＸＹＺ TEL:123-456-7890 合計金額(税込) ￥但し 上記正に領収いたしました。 収入 印紙 単位：円 品名 数量 単位 単価 金額...	FALSE
:	:

701

702

【図８】

抽出値	種別	位置	文字数
大迫 裕也	名前	94	4
1993年03月21日生	年齢	107	12
満27歳	年齢	120	4
080-1234-5678	電話	139	13
yuya0321@daihyo.jp	Email	160	18
0123-45-6789	電話	216	12
649-2523	郵便	233	8
和歌山県西牟婁郡白浜町田野井	住所	242	14
:	:	:	:

801

【図９】

個人情報項目種別数(f1)	文書文字数(f2)	個人情報項目比率(f3)	個人情報項目長(f4)	個人情報項目数(f5)	個人情報項目間距離(f6)	住所数(f7)	人名数(f8)
7	277	0.462	128	16	128	3	1

901

【図１０】

個人情報項目種別数(f1)	文書文字数(f2)	個人情報項目比率(f3)	個人情報項目長(f4)	個人情報項目数(f5)	個人情報項目間距離(f6)	住所数(f7)	人名数(f8)	個人情報
7	277	0.462	128	16	128	3	1	TRUE
2	551	0.067	37	4	113	0	3	FALSE
				:				:

501

1001

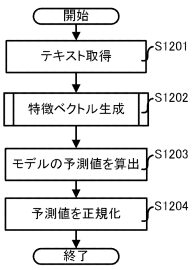
【図１１】

$$\omega_0 + \sum_{i=\text{特徴}} \omega_i \times f_i$$

f_i: i番目の特徴の特徴量

ω0	ω1	ω2	ω3	ω4	ω5	ω6	ω7	ω8
-1.777	5.067	-1.639	0.634	7.915	2.585	2.604	-9.928	10.750

【図１２】



【図１３】

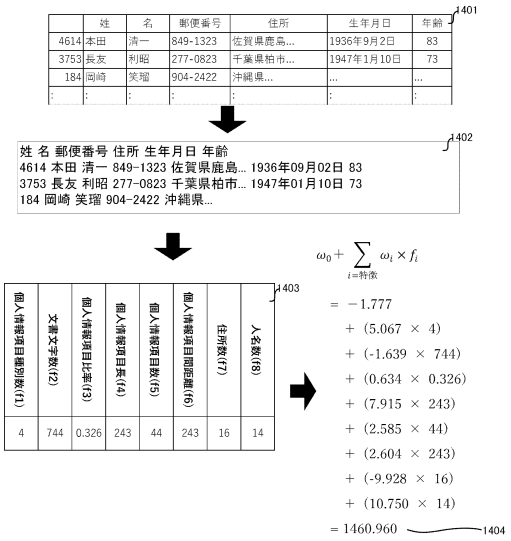
$$(\text{個人情報指数}) = \begin{cases} 0 & (r < R_{min}) \\ 50 + 50 \times \dfrac{r}{\max(R_{max}, -R_{min})} & (R_{min} \leq r \leq R_{max}) \\ 100 & (r > R_{max}) \end{cases}$$

r: 検査対象文書に対する予測値

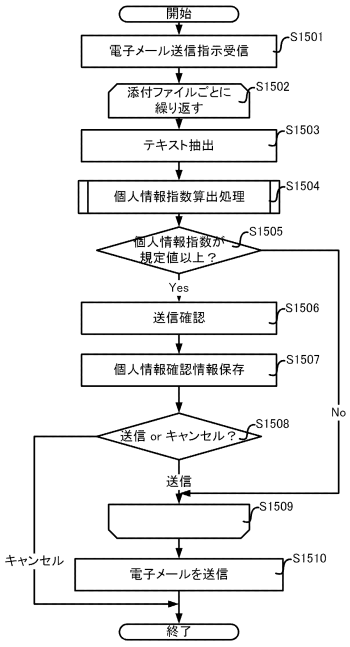
R_{min}: 学習データにおける予測値の最小値

R_{max}: 学習データにおける予測値の最大値

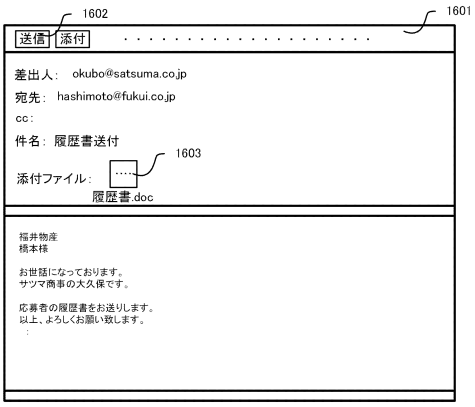
【図１４】



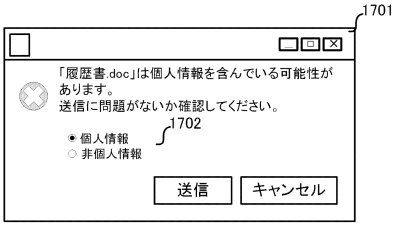
【図１５】



【図１６】



【図１７】


1701
「履歴書.doc」は個人情報を含んでいる可能性があります。
送信に問題がないか確認してください。
1702
個人情報
非個人情報
送信
キャンセル

フロントページの続き

(56)参考文献　特開２０１０－０７２７７９（ＪＰ，Ａ）
国際公開第２０１２／０９５９７１（WO，Ａ１）
特開２０１９－０６１４１９（ＪＰ，Ａ）

(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｑ　１０／００－９９／００
Ｇ０６Ｆ　４０／２７９
Ｇ０６Ｆ　２１／６２
Ｇ０６Ｆ　１３／００