

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6171476号
(P6171476)

(45) 発行日 平成29年8月2日(2017.8.2)

(24) 登録日 平成29年7月14日(2017.7.14)

(51) Int.Cl.

F I

G 0 6 F 17/22 (2006.01)

G 0 6 F 17/22 6 7 6

請求項の数 3 (全 16 頁)

(21) 出願番号	特願2013-69223 (P2013-69223)	(73) 特許権者	000005223
(22) 出願日	平成25年3月28日 (2013. 3. 28)		富士通株式会社
(65) 公開番号	特開2014-191778 (P2014-191778A)		神奈川県川崎市中原区上小田中4丁目1番1号
(43) 公開日	平成26年10月6日 (2014. 10. 6)	(74) 代理人	100094525
審査請求日	平成27年12月4日 (2015. 12. 4)		弁理士 土井 健二
		(74) 代理人	100094514
			弁理士 林 恒徳
		(72) 発明者	向山 芳孝
			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
		(72) 発明者	山岡 大亮
			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

最終頁に続く

(54) 【発明の名称】 ドキュメント作成支援装置、ドキュメント作成支援プログラム及びドキュメント作成支援方法

(57) 【特許請求の範囲】

【請求項 1】

入力される文字列についてドキュメント作成支援処理を行うドキュメント作成支援装置であって、

入力された1つまたは複数の文字を有する入力文字列の仮名漢字変換が確定したときに、前記確定した入力文字列を単語に分解する単語分解手段と、

前記入力文字列に含まれる第1の単語が、秘密ワードデータベース内の特定の秘密ワードと前方一致するか否かの第1の検索を行い、前方一致した秘密ワードを置換対象候補として登録する置換対象候補検出処理手段と、

前記第1の単語に当該第1の単語に後続する入力文字列に含まれる第2の単語を、前記置換対象候補の秘密ワードの単語構成数に達する数まで結合した結合単語が、前記置換対象候補の秘密ワードと一致するか否かの第2の検索を行い、一致した置換対象候補の秘密ワードを置換対象の秘密ワードとして登録する置換対象検出処理手段と、

前記入力文字列をそのまま有するオリジナルドキュメントファイル内の前記置換対象の秘密ワードとして登録された秘密ワードを、前記秘密ワードに対応する置換ワードに置換してマスキングドキュメントファイルを生成するマスキングドキュメント作成手段とを有するドキュメント作成支援装置。

【請求項 2】

入力される文字列についてドキュメント作成支援処理を行うドキュメント作成支援プログラムであって、

10

20

前記ドキュメント作成支援処理は、

入力された１つまたは複数の文字を有する入力文字列の仮名漢字変換が確定したときに、前記確定した入力文字列を単語に分解する単語分解工程と、

前記入力文字列に含まれる第１の単語が、秘密ワードデータベース内の特定の秘密ワードと前方一致するか否かの第１の検索を行い、前方一致した秘密ワードを置換対象候補として登録する置換対象候補検出処理工程と、

前記第１の単語に当該第１の単語に後続する入力文字列に含まれる第２の単語を、前記置換対象候補の秘密ワードの単語構成数に達する数まで結合した結合単語が、前記置換対象候補の秘密ワードと一致するか否かの第２の検索を行い、一致した置換対象候補の秘密ワードを置換対象の秘密ワードとして登録する置換対象検出処理工程と、

10

前記入力文字列をそのまま有するオリジナルドキュメントファイル内の前記置換対象の秘密ワードとして登録された秘密ワードを、前記秘密ワードに対応する置換ワードに置換してマスキングドキュメントファイルを生成するマスキングドキュメント作成工程とを有するドキュメント作成支援プログラム。

【請求項３】

入力される文字列についてドキュメント作成支援処理を行うドキュメント作成支援方法であって、

入力された１つまたは複数の文字を有する入力文字列の仮名漢字変換が確定したときに、前記確定した入力文字列を単語に分解する単語分解工程と、

前記入力文字列に含まれる第１の単語が、秘密ワードデータベース内の特定の秘密ワードと前方一致するか否かの第１の検索を行い、前方一致した秘密ワードを置換対象候補として登録する置換対象候補検出処理工程と、

20

前記第１の単語に当該第１の単語に後続する入力文字列に含まれる第２の単語を、前記置換対象候補の秘密ワードの単語構成数に達する数まで結合した結合単語が、前記置換対象候補の秘密ワードと一致するか否かの第２の検索を行い、一致した置換対象候補の秘密ワードを置換対象の秘密ワードとして登録する置換対象検出処理工程と、

前記入力文字列をそのまま有するオリジナルドキュメントファイル内の前記置換対象の秘密ワードとして登録された秘密ワードを、前記秘密ワードに対応する置換ワードに置換してマスキングドキュメントファイルを生成するマスキングドキュメント作成工程とを有するドキュメント作成支援方法。

30

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、ドキュメント作成支援装置、ドキュメント作成支援プログラム及びドキュメント作成支援方法に関する。

【背景技術】

【０００２】

ドキュメント作成支援プログラムは、ワードプロセッサやエディタなど、文字列であるテキストデータが含まれているドキュメントファイルの作成を支援する。企業や官公庁内で作成されるドキュメントには、多くの個人情報や秘密情報（以下まとめて秘密情報）が含まれている。したがって、そのような秘密情報が含まれているドキュメントファイルは、外部者に開示することはできないし、セキュリティ性が低い個人の端末内にコピーすることも好ましくない。

40

【０００３】

そこで、ドキュメント内の秘密情報をマスキングしたり、秘密情報を別の単語に置換する書類管理システムが提案されている。たとえば、以下の特許文献などである。

【先行技術文献】

【特許文献】

【０００４】

【特許文献１】特開２００１－３２５２５８号公報

50

【特許文献2】特開2007-65778号公報
【特許文献3】特開2005-165737号公報
【特許文献4】特開2005-108006号公報
【特許文献5】特開2002-259363号公報
【特許文献6】特開2009-199385号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

作成済みのドキュメントファイルを外部者に開示する時に、ドキュメントファイル内の文字列を検索して秘密情報と一致する文字列を抽出し、その秘密情報の文字列をアスタリスクなどのマスキング文字や他の単語に置換してマスキングドキュメントを生成する方法では、指定されている秘密情報をキーにして作成済みドキュメントファイルの文字列を検索しドキュメント内の秘密情報を検出する必要があるという問題がある。

10

【0006】

特に、大きな組織の場合には、秘密情報データベース内の秘密情報の数が膨大になり、処理工数の増大が顕著になる。

【0007】

そこで、1つの側面では、本発明の目的は、マスキングドキュメントの作成を容易に行うドキュメント作成支援装置、ドキュメント作成支援プログラム及びドキュメント作成支援方法を提供することにある。

20

【課題を解決するための手段】

【0008】

本実施の形態の第1の側面は、入力される文字列についてドキュメント作成支援処理を行うドキュメント作成支援装置であって、

入力された入力文字列に含まれる第1の単語が、秘密ワードデータベース内の秘密ワードと前方一致するか否かの第1の検索を行い、一致した秘密ワードを置換対象候補に登録する置換対象候補検出処理手段と、

前記第1の単語に当該第1の単語に後続する入力文字列に含まれる第2の単語を、前記置換対象候補の秘密ワードの単語構成数に達する数まで結合した結合単語が、前記置換対象候補の秘密ワードと一致するか否かの第2の検索を行い、一致した置換対象候補の秘密ワードを置換対象に登録する置換対象検出処理手段と、

30

前記入力文字列をそのまま有するオリジナルドキュメントファイル内の前記置換対象の秘密ワードを、前記秘密ワードに対応する置換ワードに置換してマスキングドキュメントファイルを生成するマスキングドキュメント作成手段とを有する。

【発明の効果】

【0009】

第1の側面によれば、入力される入力文字列について秘密ワードに一致するか否かを効率的に検出するので、マスキングドキュメントの作成工数が少なくなる。

40

【図面の簡単な説明】

【0010】

【図1】本実施の形態におけるドキュメント作成支援装置を有するドキュメント作成サーバとクライアント装置の構成図である。

【図2】本実施の形態におけるドキュメント作成支援装置の構成図である。

【図3】秘密ワードデータベースの一例を示す図である。

【図4】本実施の形態における入力文字列処理とファイル保存処理の概略を説明する図である。

【図5】図4の入力文字列INが分解された単語群の例が示されている。

【図6】本実施の形態におけるドキュメント作成支援装置の処理を示すフローチャート図

50

である。

【図 7】入力文字列処理S13のフローチャート図である。

【図 8】置換対象候補検出処理S20のフローチャート図である。

【図 9】文字列入力処理S201のフローチャート図である。

【図 10】置換対象検出処理S30のフローチャート図である。

【図 11】ドキュメント保存処理のフローチャート図である。

【発明を実施するための形態】

【0011】

図 1 は、本実施の形態におけるドキュメント作成支援装置を有するドキュメント作成サーバとクライアント装置の構成図である。クライアント装置 10 は、ドキュメントを作成する作成者により文字入力が行われドキュメントデータを表示する端末 12 と、必要に応じてドキュメントファイルを保存するローカルディスク 14 とを有する。端末 12 は、少なくともエディタの機能を有し、ドキュメントを作成する。

10

【0012】

クライアント装置 10 は、インターネットやイントラネットなどのネットワークNETを介してドキュメント作成サーバ 20 と通信可能に接続される。そして、ドキュメント作成者は、クライアント装置 10 をドキュメント作成サーバ 20 にネットワークNETを介して接続し、ドキュメント作成サーバ 20 内のドキュメント作成支援装置 21 を使用してマスキングドキュメントを作成する。

【0013】

20

ドキュメント作成サーバ 20 は、クライアント装置 10 から入力される入力文字列についてドキュメント作成支援処理を行うドキュメント作成支援装置 21 と、作成されるドキュメント内の秘密にすべき秘密ワードとその秘密ワードに置き換えられる置換ワードとを有する秘密ワードデータベース 22 とを有する。ドキュメント作成支援装置 21 は、図示しないプロセッサとドキュメント作成支援プログラムを格納する記憶媒体とを有する。さらに、ドキュメント作成サーバ 20 は、メモリ 23 と、ディスクユニット 26 とを有し、メモリ 23 内には、ドキュメント作成支援処理中に入力文字を一時的に格納する入力文字バッファ 24 と、置換対象の秘密ワードを判定するキーとなる結合単語を格納する結合単語バッファ 25 とを有する。これらのバッファの利用については後述する。

【0014】

30

そして、ドキュメント作成サーバ 20 は、作成中のドキュメントに対応して生成されるオリジナルドキュメントファイル 28 と、オリジナルドキュメントファイル内の秘密ワードを置換ワードに置換したマスキングドキュメントファイル 29 とを記憶するディスクユニット 26 とを有する。ディスクユニット 26 は、メモリディスク以外の揮発性または不揮発性の記憶媒体でもよい。

【0015】

図 2 は、本実施の形態におけるドキュメント作成支援装置の構成図である。ドキュメント作成支援装置 21 は、図示しないプロセッサとドキュメント作成支援プログラムで実現され、そのドキュメント作成支援プログラムがドキュメント作成サーバ 20 内の図示しないプロセッサにより実行されて、ドキュメント作成支援処理を行う。

40

【0016】

ドキュメント作成支援装置 21 は、入力文字列処理部 210 と、ファイル保存処理部 214 とを有する。そして、入力文字列処理部 210 は、置換対象候補検出処理部 211 と置換対象検出処理部 212 とを有し、置換対象候補検出部 211 及び置換対象検出処理部 212 はそれぞれ文字列入力処理部 212 を利用する。

【0017】

入力文字列処理部 210 は、クライアント装置 10 の端末 12 から入力される入力文字列に秘密ワードデータベース 22 内の秘密ワードが含まれるか否かをチェックし、該当する秘密ワードを置換ワードへの置換対象である置換対象秘密ワードとして登録する。これにより、秘密ワードデータベース内の膨大な数の秘密ワードのうち、現在作成中のドキュ

50

メントファイルに含まれる秘密ワードを抽出することができ、その後ドキュメントファイル内の秘密ワードを置換ワードに置換してマスキングドキュメントファイルを作成する処理の処理量を少なくすることができる。

【0018】

この秘密ワードデータベース22内のどの秘密ワードが置換対象秘密ワードかの検出は、(1)入力文字列に含まれる第1の単語が、秘密ワードデータベース22内の秘密ワードと前方一致するか否かの第1の検索を行い、一致した秘密ワードを置換対象候補に登録する置換対象候補検出処理と、(2)第1の単語に、それに後続する入力文字列に含まれる第2の単語を、置換対象候補の秘密ワードの単語構成数に達する数まで結合した結合単語が、置換対象候補の秘密ワードと一致するか否かの第2の検索を行い、一致した置換対象候補の秘密ワードを置換対象の秘密ワードに登録する置換対象検出処理とにより行われる。

10

【0019】

上記(1)の処理は置換対象候補検出処理部211が行い、上記(2)の処理は置換対象検出処理部212が行う。前述のとおり、置換対象候補検出処理部211及び置換対象検出処理部212は、文字列入力処理部213を利用する。文字列入力処理部213は、クライアント装置10の端末12で入力される文字列が仮名漢字変換で確定したことを検出して、検索対象の文字列として処理する。または、クライアント装置10の端末12が仮名漢字変換などの機能を有しない場合は、文字列入力処理部213は、クライアント装置10の端末12で入力された文字列について仮名漢字変換などの入力処理も行う。

20

【0020】

ファイル保存処理部214は、ドキュメント作成完了後にドキュメント作成者による作成中のドキュメントを保存するドキュメント保存要求に応答して、入力文字列をそのまま有するオリジナルドキュメントファイルと、オリジナルドキュメントファイル内の秘密ワードをその秘密ワードに対応する置換ワードに置換したマスキングドキュメントファイル29とを生成し、記憶媒体内に保存する。

【0021】

上記のドキュメント作成支援装置21の各処理部については後で詳述する。

【0022】

図3は、秘密ワードデータベースの一例を示す図である。秘密ワードデータベース22には、企業や官公庁などの組織体内で秘密にすべき文字列が秘密ワードとして登録され、その秘密ワードを置換する文字列が置換ワードとして各秘密ワードに対応して登録される。

30

【0023】

秘密ワードは、単一のまたは複数の単語で構成される。図3中の秘密ワードデータベース22には、各秘密ワードに対して、それを単語に分解した秘密ワード構成と、その単語構成数とが登録されている。例えば、秘密ワードである「川崎市川崎区B町1丁目1番地」には、置換ワードとして「AA市BB区CCCCCCCC」が登録されている。したがって、作成中のドキュメント内にこの秘密ワードが含まれている場合は、それに対する置換ワードに置き換えられてマスキングドキュメントファイルが作成される。また、秘密ワードの「川崎市川崎区B町1丁目1番地」は、「川崎市」「川崎区」「B町」「1丁目」「1番地」のように単語構成数が5となっている。

40

【0024】

さらに、秘密ワードデータベース22には、ドキュメント作成中に入力文字列内の単語が秘密ワードの先頭単語に該当する場合に、その該当した秘密ワードを置換対象候補として登録するための置換対象候補フラグの領域と、置換対象候補の秘密ワードのうち入力文字列内の結合単語と一致する秘密ワードを置換対象として登録するための置換対象フラグの領域とを有する。

【0025】

そして、ドキュメント作成中の入力文字列に含まれる単語について、後述する方法で、

50

置換対象候補の秘密ワードを検出して置換対象候補フラグを「1」にし、その置換対象候補フラグが「1」の置換対象候補の秘密ワードのうち、それぞれの秘密ワードの単語構成数の入力文字列である結合単語と一致する置換対象の秘密ワードを検出して置換対象フラグを「1」にする。図3に示された秘密ワードデータベース22の例では、既に、置換対象候補フラグと置換対象フラグの領域に「1」が登録されている。本実施の形態では、上記の置換候補の秘密ワードの検出を少ない処理工数で行うことができる。

【0026】

〔本実施の形態におけるドキュメント作成支援装置の処理の概略〕

図4は、本実施の形態における入力文字列処理とファイル保存処理の概略を説明する図である。図4には、ドキュメント作成中に入力される入力文字列INの一例として図示される以下の文字列が入力されるものとする。

「2000年神奈川県川崎市川崎区B町1丁目1番地（電話044-999-9999）に山田太郎さんと鈴木花子さんが住んでいました。」

この入力文字列について、入力文字列処理部210がどのような処理をするのかについて説明する。まず、入力文字列処理部210では、入力された文字列であって仮名漢字変換で確定した文字列を単語に分解する。

【0027】

図5は、図4の入力文字列INが分解された単語群の例が示されている。例えば、端末12で「2000年」という入力文字列が入力され仮名漢字変換され確定したとすると、「2000年」は1つの単語であるので、1つの単語として取り扱われる。また、次に、端末12で「神奈川県川崎市川崎区B町1丁目1番地」という入力文字列が入力され仮名漢字変換され確定したとすると、所定の構文解析アルゴリズムにより6つの単語「神奈川県」「川崎市」「川崎区」「B町」「1丁目」「1番地」に分解される。分解された単語に付された番号は入力の順番を示している。

【0028】

ただし、端末12での入力が「神奈川」「県」「川崎」「市」などのように4つに細切れに入力され仮名漢字変換され確定される場合もある。その場合は、入力され確定した文字列をそれぞれ単語として認識するようにしてもよい。または、その4つの確定した文字列「神奈川県川崎市」を上記と同じ構文解析アルゴリズムにより2つの単語「神奈川県」「川崎市」に分解しても良い。ドキュメント作成支援装置21がこの単語に分解する構文解析機能を持つことにより、端末12のエディタの仮名漢字変換処理機能が異なっても、共通のアルゴリズムにより単語に分解することができる。ただし、前者のように入力文字列が細切れの単語に分解されても、後述する候補対象検出処理部212は対応可能である。

【0029】

図4に戻り、入力文字列処理部210の置換対象候補検出処理部211は、入力文字列を分解した複数の単語のうち先頭の単語（第1の単語）をキーにして、秘密ワードデータベース22内の秘密ワードを前方一致検索する。図4に示した例では、図5の「2000年」「神奈川県」についての検索が終了した後、「川崎市」をキーにして秘密ワードを前方一致検索している。この検索の結果、以下の4つの秘密ワードが前方一致している。

「川崎市川崎区A町1丁目1番地」

「川崎市川崎区B町1丁目1番地」

「川崎市中原区B町1丁目」

「川崎市高津区B町」

その結果、置換対象候補検出処理部211は、秘密ワードデータベース22内の上記4つの秘密ワードについて置換対象候補フラグを「1」にし、それら4つの文字列を置換対象候補の秘密ワードとして登録する。

【0030】

次に、入力文字列処理部210の置換対象検出処理部212は、置換対象候補フラグが「1」の置換対象候補の秘密ワードの単語構成数に注目し、上記の第1の単語「川崎市」

10

20

30

40

50

に後続する入力文字列内の単語（第2の単語）を，単語構成数に達するまで結合した結合単語が，置換対象候補の秘密ワードと一致するか否かまたは少なくとも前方一致するか否かの検索を行う。

【0031】

例えば，上記4つの置換対象候補の秘密ワードのうち，「川崎市高津区B町」は単語構成数が3個であるので，「川崎市」に「川崎区」「B町」を結合して単語構成数が3個の結合単語「川崎市川崎区B町」が，単語構成数3個の置換対象候補の秘密ワード「川崎市高津区B町」と一致するか否かを検索する。この例では，一致しないし前方一致もしていない。

【0032】

次に，上記4つの置換対象候補の秘密ワードのうち，「川崎市中原区B町1丁目」は単語構成数が4個であるので，「川崎市」に「川崎区」「B町」「1丁目」を結合して単語構成数が4個の結合単語「川崎市川崎区B町1丁目」が，単語構成数4個の置換対象候補の秘密ワード「川崎市中原区B町1丁目」と一致するか否かを検索する。この例でも，一致しないし前方一致もしていない。

【0033】

さらに，上記4つの置換対象候補の秘密ワードのうち，「川崎市川崎区A町1丁目1番地」と「川崎市川崎区B町1丁目1番地」は単語構成数が5個であるので，「川崎市」に「川崎区」「B町」「1丁目」「1番地」を結合して単語構成数が5個の結合単語「川崎市川崎区B町1丁目1番地」が，単語構成数5個の置換対象候補の秘密ワード「川崎市川崎区A町1丁目1番地」と「川崎市川崎区B町1丁目1番地」のいずれかと一致するか否かを検索する。この例では，「川崎市川崎区B町1丁目1番地」と一致している。

【0034】

その結果，候補対象検出処理部212は，4つの置換対象候補の秘密ワードのうち，結合単語で一致した「川崎市川崎区B町1丁目1番地」を，置換対象の秘密ワードとして置換対象フラグを「1」にして登録する。この結果，秘密ワードデータベース22内の膨大な数の秘密ワード群のうち「川崎市川崎区B町1丁目1番地」が，作成中のドキュメントに含まれていることが判明する。

【0035】

上記の置換対象の秘密ワードを検出するアルゴリズムによれば，入力される入力文字列の単語全てについて秘密ワードデータベース22内の秘密ワードと一致するか否かの複雑な検索処理を行う必要はない。入力文字列に含まれる単語が秘密ワードと前方一致した場合に，その置換対象候補の秘密ワードについて，その単語構成数まで単語の入力を待って，結合した結合単語と一致するかを検索を行えばよい。図4の例では，少なくとも「川崎市」に続く2番目の単語「川崎区」について置換対象候補の秘密ワードに対する検索処理は行われていない。秘密ワードの単語構成数が大きいほど，検索処理の効率が高くなることが理解できる。

【0036】

図4において，ファイル保存処理部214は，入力文字列を秘密ワード一致する文字列も含めてそのまま有するオリジナルドキュメントファイル28を保存すると共に，オリジナルドキュメント内の秘密ワードを置換ワードに置換して秘密ワードをマスクしたマスキングドキュメントファイル29を作成し保存する。

【0037】

オリジナルドキュメントファイルは，入力文字列をそのまま有するので，文字列の入力が終了した時点で作成済みであり，ディスクユニット26などに保存すればよい。一方，マスキングドキュメントファイルは，オリジナルドキュメント内の置換対象の秘密ワードと一致する文字列を検出し，検出した文字列を置換ワードに置換することで生成され，ディスクユニット26などに保存される。

【0038】

[本実施の形態におけるドキュメント作成支援装置の詳細]

10

20

30

40

50

図 6 は、本実施の形態におけるドキュメント作成支援装置の処理を示すフローチャート図である。既存のドキュメントファイルを編集するか、新規のドキュメントファイルを編集するかによって (S10)、ドキュメント作成支援装置は、作成対象の既存のドキュメントファイルを読み込むか (S11)、新規にドキュメントファイルを作成する (S12)。

【 0 0 3 9 】

そして、ドキュメント作成支援装置は、そのドキュメントファイルについて、次の入力文字列がある限り (S14のYES)、入力文字列処理 (S13) を繰り返す。そして、全ての入力文字列について入力文字列処理S13が完了すると (S14のNO)、作成したドキュメントファイルの保存処理 (S15) を実行する。

【 0 0 4 0 】

上記の入力文字列処理S13は、図 4 で説明した入力文字列処理部 2 1 0 による処理であり、ドキュメントファイル保存処理S15は、図 4 で説明したファイル保存処理部 2 1 4 による処理である。入力文字列処理S13は、サブルーチンであり、図 7 (図 8, 9, 10 を含む) に詳述されている。また、ドキュメントファイル保存処理S15も、サブルーチンであり、図 1 1 に詳述されている。

【 0 0 4 1 】

図 7 は、入力文字列処理S13のフローチャート図である。入力文字列処理S13では、入力され仮名漢字変換が確定した入力文字列に含まれる先頭の単語をキーにして、秘密ワードを前方一致検索し、置換対象候補の秘密ワードを検出する置換対象候補検出処理 (S20) が実行される。その結果検出された置換対象候補の秘密ワードには候補フラグ「1」が付与される。

【 0 0 4 2 】

そして、置換対象候補の秘密ワードが検出されない場合は (S21のNO)、ドキュメント作成支援装置は、今回の先頭の単語に対する入力文字列処理S13を終了し、図 6 に示したとおり、次の入力文字列に含まれる先頭の単語について入力文字列処理S13を実行する。

【 0 0 4 3 】

一方、置換対象候補の秘密ワードが検出された場合は (S21のYES)、ドキュメント作成支援装置は、今回の処理対象の先頭の単語に後続する単語を連結した結合単語が、置換対象候補の秘密ワードと一致するかの検索を行い、一致した秘密ワードを置換対象として登録する置換対象検出処理 (S30) を実行する。

【 0 0 4 4 】

図 8 は、置換対象候補検出処理S20のフローチャート図である。置換対象候補検出処理S20では、入力文字バッファ 2 4 内に単語が格納されていなければ (S200のNO)、文字列入力処理 (S201) が実行され、入力が確定した入力文字列が入力文字バッファに保存される。ドキュメント作成処理開始時は入力文字バッファ内に入力文字列は保存されていない。一方、入力文字バッファ 2 4 内に単語が格納されていれば (S200のYES)、その先頭の単語について処理S203が実行される。

【 0 0 4 5 】

図 9 は、文字列入力処理S201のフローチャート図である。文字列入力処理S201では、端末 1 2 で文字列が入力され (S2011)、仮名漢字変換処理が行われ (S2012)、変換が確定し (S2013)、次の文字列入力が行われる直前で (S2014)、入力文字列が確定する。そして、確定した入力文字列が入力文字バッファ 2 4 内に保存される (S2015)。この文字列入力処理S201は、ドキュメント作成支援装置 2 1 による入力文字列処理S13と並行してバックグラウンドで行われるのが好ましい。

【 0 0 4 6 】

図 8 に戻り、入力文字バッファ 2 4 に保存された入力文字列は、複数の単語に分解される (S202)。入力文字列を複数の単語に分解することは、図 5 で説明したとおりである。

【 0 0 4 7 】

そして、ドキュメント作成支援装置 2 1 は、入力文字バッファ内の先頭の単語をキーにして、秘密ワードデータベース 2 2 内の秘密ワードを前方一致検索する (S203)。

10

20

30

40

50

【 0 0 4 8 】

上記の前方一致検索の結果，前方一致した秘密ワードが検出されると（S204のYES），一致した秘密ワードに置換対象候補フラグが付与される（S205）。これにより置換対象候補の秘密ワードが登録される。この先頭の文字をキーにした秘密ワードの前方一致検索と置換対象候補フラグの付与については，図4の置換候補検出処理部の処理の例で説明したとおりである。

【 0 0 4 9 】

一方，前方一致した秘密ワードが検出されないと（S204のNO），分解された全単語についての検索が行われるまで（S206のNO），次の処理S207が繰り返される。この処理S207では，入力文字バッファ24から先に検索済みの単語が削除され，その次の単語をキーにして秘密ワードについて前方一致検索を行う。したがって，処理S203,S207では，入力文字バッファ24内の先頭の単語をキーにして，秘密ワードへの前方一致検索が行われる。

10

【 0 0 5 0 】

図7に戻り，置換対象候補検出処理S20が完了して，先頭の単語と前方一致して置換対象候補フラグ「1」を付与された秘密ワードが存在する場合は，置換対象検出処理（S30）が実行される。

【 0 0 5 1 】

図10は，置換対象検出処理S30のフローチャート図である。置換対象検出処理では，先頭の単語（第1の単語）に後続する単語（第2の単語）を置換対象候補の秘密ワードの単語構成数に達するまで結合した結合単語が，置換対象候補の秘密ワードと一致するか否かを検索する。そのために，入力文字バッファ24内の単語を，置換対象候補の秘密ワードの単語構成数と同じ数まで結合し，その結合単語をキーにして，置換対象候補の秘密ワードを一致検索する。

20

【 0 0 5 2 】

そこで，前処理的に，入力文字バッファ内に置換対象候補の秘密ワードの単語構成数と同じ数の単語があるか否かをチェックし（S301），なければ（S301のNO），文字入力処理S201と入力文字列を単語に分解する処理S202とを行う。この文字入力処理S201と分解処理S202は，図8，9で説明したとおりである。

【 0 0 5 3 】

そして，置換対象候補の秘密ワードの単語構成数と同じ数の単語を有する結合単語の文字列をキーにして，置換対象候補フラグ「1」を付与された秘密ワードのうち結合単語と同じ単語構成数の秘密ワードと一致するか否かの検索が行われる（S302）。この検索処理（S302）は，置換対象候補の秘密ワードの単語構成数が複数種類ある場合は，その全ての単語構成数について行われる。すなわち，図4で説明したとおり，単語構成数3，4，5の全てについて結合単語をキーにして，同じ単語構成数3，4，5の秘密ワードに対して一致検索が行われる。

30

【 0 0 5 4 】

ここでの一致検索は，少なくとも前方一致したか否かを行うことが望ましい。その理由は次のとおりである。

【 0 0 5 5 】

40

入力文字列は一定の構文解析アルゴリズムにより単語に分解されている。しかし，前述したとおり，単語に分解する構文解析アルゴリズムによっては，端末12での入力「神奈川県」「川崎」「市」などのように細切れに入力され仮名漢字変換され確定される場合に，入力され確定した文字列をそれぞれ単語として認識するようにしてもよい。その場合は，置換対象候補の秘密ワードの単語構成数と同じ数の単語を有する結合単語の文字列は，置換対象候補の秘密ワードとは前方一致はするものの，完全に一致しない場合がある。そこで，上記の検索処理S302では，そのような単語分解アルゴリズムの場合に対応して，置換対象候補の秘密ワードと少なくとも前方一致するか否かの処理をする。そのようにして検出された置換対象の秘密ワードは，ドキュメントファイル保存時において，オリジナルドキュメント内にその置換対象の秘密ワードと一致するか否かのチェックが再度行

50

われるので、前方一致はするものの完全一致しない文字列は置換対象から外されるので、問題はない。

【 0 0 5 6 】

図 1 0 において、結合単語が、置換対象候補フラグ「 1 」が付与された秘密ワードと一致または少なくとも前方一致した場合は (S303 の YES) , その一致した秘密ワードに置換対象フラグ「 1 」が付与され置換対象の秘密ワードとして登録される (S304) 。また、この時、入力文字バッファから一致した結合単語の文字列が削除される。

【 0 0 5 7 】

一方、結合単語が、置換対象候補フラグ「 1 」が付与された秘密ワードと一致せず、前方一致もしない場合は (S303 の NO) , 置換対象候補フラグ「 1 」を削除し、入力文字バッファから先頭単語を削除する (S305) 。すなわち、図 4 の例において、もし結合単語が「川崎市、の、山田太郎、さん、と」の場合は、先頭単語「川崎市」を入力文字バッファから削除して、次の先頭単語の「の」、「山田太郎」について、図 6 の処理 S14 の YES にしたがつて、再度入力文字列処理 S13 が実行される必要がある。

【 0 0 5 8 】

図 6 に戻り、入力文字バッファ 2 4 内の全ての入力文字列について入力文字列処理 S13 が完了すると、最後にドキュメントファイル保存処理 S15 が実行される。

【 0 0 5 9 】

図 1 1 は、ドキュメント保存処理のフローチャート図である。端末 1 2 から作成中のドキュメントファイルの保存要求があると (S151 の YES) , ドキュメント作成支援装置 2 1 は、まず、作成中のオリジナルドキュメントファイルを保存する (S152) 。図 4 で説明したとおり、オリジナルドキュメントファイル 2 8 は、入力文字列をそのまま有するドキュメントファイルである。

【 0 0 6 0 】

次に、マスキングドキュメントファイルを作成するために、オリジナルドキュメントファイルをコピーしてマスキングドキュメントファイルを作成し (S153) , そのマスキングドキュメントファイル内の文字列について置換対象の秘密ワードと一致する文字列を検索する (S154) 。この検索は、置換対象フラグ「 1 」が付与された秘密ワードについて順番に行う。そして、検出した置換対象の秘密ワードを置換ワードに置き換える (S155) 。置換対象の秘密ワードと一致する文字列は必ずドキュメント内に存在することは確認されているので、置換対象の秘密ワードをキーにしてドキュメント内の文字列を検索すると必ず一致する文字列が検索される。

【 0 0 6 1 】

ただし、結合単語の文字列が、置換対象候補の秘密ワードとは前方一致はするものの、完全に一致しなかった場合に置換対象の秘密ワードとして登録されている場合は、S154での検索で置換対象の秘密ワードと完全一致する文字列が検出されないことが、例外的にある。その場合は、置換ワードへの置き換えは行われない。

【 0 0 6 2 】

上記の検索 S154 と置換ワードへの置換 S155 は、全ての置換対象の秘密ワードについて実行される (S156 の NO) 。ただし、入力文字列処理で作成中のドキュメントに含まれる置換対象の秘密ワードは検出済みであるので、秘密ワードデータベース内の全ての秘密ワードについて検索と置換を行うよりも、短時間でこの処理を完了することができる。

【 0 0 6 3 】

全ての置換対象の秘密ワードについて検索 S154 と置換ワードへの置換 S155 が実行されると (S156 の YES) , 秘密ワードを置換ワードに置き換えられたマスキングドキュメントファイル 2 9 が保存される (S157) 。マスキングドキュメントファイル 2 9 の例は、図 4 に示したとおりである。

【 0 0 6 4 】

以上のとおり、本実施の形態によれば、ドキュメント作成中に入力される文字列について、秘密ワードデータベース内の秘密ワードと一致するか否かの処理を効率的に行うこと

10

20

30

40

50

ができる。

【 0 0 6 5 】

以上の実施の形態をまとめると、次の付記のとおりである。

【 0 0 6 6 】

(付記 1)

入力される文字列についてドキュメント作成支援処理を行うドキュメント作成支援装置であって、

入力された入力文字列に含まれる第 1 の単語が、秘密ワードデータベース内の秘密ワードと前方一致するか否かの第 1 の検索を行い、一致した秘密ワードを置換対象候補に登録する置換対象候補検出処理手段と、

前記第 1 の単語に当該第 1 の単語に後続する入力文字列に含まれる第 2 の単語を、前記置換対象候補の秘密ワードの単語構成数に達する数まで結合した結合単語が、前記置換対象候補の秘密ワードと一致するか否かの第 2 の検索を行い、一致した置換対象候補の秘密ワードを置換対象の秘密ワードに登録する置換対象検出処理手段と、

前記入力文字列をそのまま有するオリジナルドキュメントファイル内の前記置換対象の秘密ワードを、前記秘密ワードに対応する置換ワードに置換してマスキングドキュメントファイルを生成するマスキングドキュメント作成手段と

を有するドキュメント作成支援装置。

【 0 0 6 7 】

(付記 2)

付記 1 において、

さらに、入力された 1 つまたは複数の文字を有する入力文字列の仮名漢字変換が確定したときに、前記確定した入力文字列を単語に分解する単語分解手段を有するドキュメント作成支援装置。

【 0 0 6 8 】

(付記 3)

付記 2 において、

前記置換対象検出処理手段は、前記第 2 の検索において、前記結合単語が前記置換対象候補の秘密ワードと少なくとも前方一致するか否かを行うドキュメント作成支援装置。

【 0 0 6 9 】

(付記 4)

入力される文字列についてドキュメント作成支援処理を行うドキュメント作成支援プログラムであって、

前記ドキュメント作成支援処理は、

入力された入力文字列に含まれる第 1 の単語が、秘密ワードデータベース内の秘密ワードと前方一致するか否かの第 1 の検索を行い、一致した秘密ワードを置換対象候補に登録する置換対象候補検出処理工程と、

前記第 1 の単語に当該第 1 の単語に後続する入力文字列に含まれる第 2 の単語を、前記置換対象候補の秘密ワードの単語構成数に達する数まで結合した結合単語が、前記置換対象候補の秘密ワードと一致するか否かの第 2 の検索を行い、一致した置換対象候補の秘密ワードを置換対象の秘密ワードに登録する置換対象検出処理工程と、

前記入力文字列をそのまま有するオリジナルドキュメントファイル内の前記置換対象の秘密ワードを、前記秘密ワードに対応する置換ワードに置換してマスキングドキュメントファイルを生成するマスキングドキュメント作成工程と

を有するドキュメント作成支援プログラム。

【 0 0 7 0 】

(付記 5)

入力される文字列についてドキュメント作成支援処理を行うドキュメント作成支援方法であって、

入力された入力文字列に含まれる第 1 の単語が、秘密ワードデータベース内の秘密ワー

10

20

30

40

50

ドと前方一致するか否かの第１の検索を行い、一致した秘密ワードを置換対象候補に登録する置換対象候補検出処理工程と、

前記第１の単語に当該第１の単語に後続する入力文字列に含まれる第２の単語を、前記置換対象候補の秘密ワードの単語構成数に達する数まで結合した結合単語が、前記置換対象候補の秘密ワードと一致するか否かの第２の検索を行い、一致した置換対象候補の秘密ワードを置換対象の秘密ワードに登録する置換対象検出処理工程と、

前記入力文字列をそのまま有するオリジナルドキュメントファイル内の前記置換対象の秘密ワードを、前記秘密ワードに対応する置換ワードに置換してマスキングドキュメントファイルを生成するマスキングドキュメント作成工程とを有するドキュメント作成支援方法。

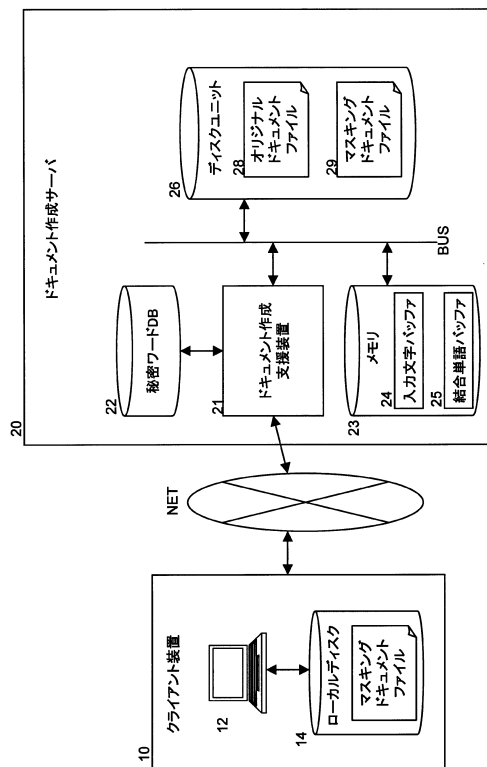
10

【符号の説明】

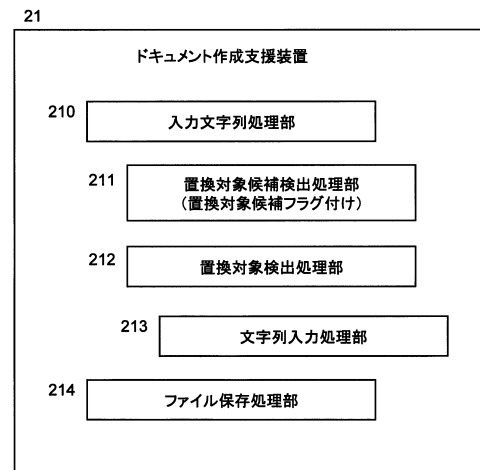
【 0 0 7 1 】

- 2 1 0 : 入力文字列処理部
- 2 1 1 : 置換対象候補検出処理部
- 2 1 2 : 置換対象検出処理部
- 2 1 4 : ファイル保存処理部

【図 1】



【図 2】

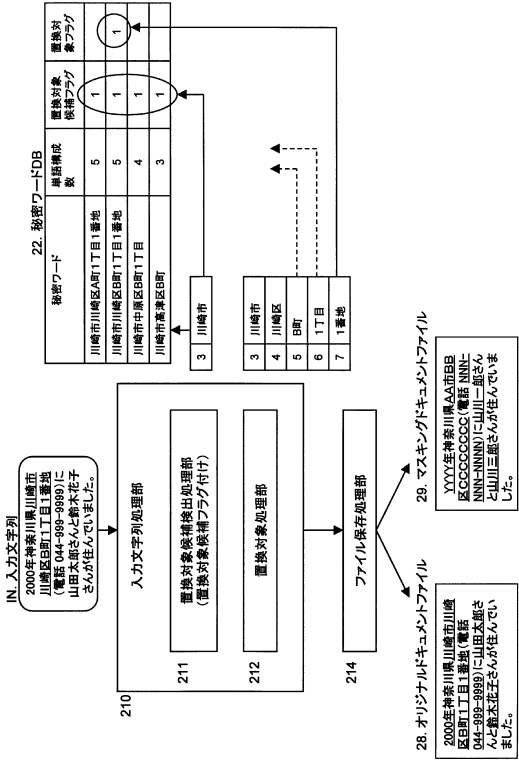


【図 3】

22. 秘密ワードDB

秘密ワード	秘密ワード構成	単語構成数	置換対象候補フラグ	置換対象フラグ	置換ワード
2000年1月1日	2000年1月1日	2	1	1	YYY年MM月DD日
2000年	2000年	1	1	1	YYY年
2001年	2001年	1	1	1	YYY年
...
川崎市川崎区A町1丁目1番地	川崎市/川崎区/A町/1丁目/1番地	5	1	...	AA市BB区CCGGGGGG
川崎市川崎区B町1丁目1番地	川崎市/川崎区/B町/1丁目/1番地	5	1	1	AA市BB区CCGGGGGG
川崎市中原区B町1丁目	川崎市/中原区/B町/1丁目	4	1	...	AA市BB区CCGGGGGG
川崎市高津区B町	川崎市/高津区/B町	3	1	...	AA市BB区CCGGGGGG
...
043-999-9997	043-/-/999-/-/9997	5	NNN-NNN-NNN
044-999-9998	044-/-/999-/-/9998	5	1	...	NNN-NNN-NNN
044-999-9999	044-/-/999-/-/9999	5	1	1	NNN-NNN-NNN
...
山田太郎	山田/太郎	2	1	1	山川一郎
山田一郎	山田/一郎	2	1	1	山川二郎
鈴木花子	鈴木/花子	2	1	1	山川三郎
鈴木花子	鈴木/花子	2	1	1	山川三郎
...

【図 4】

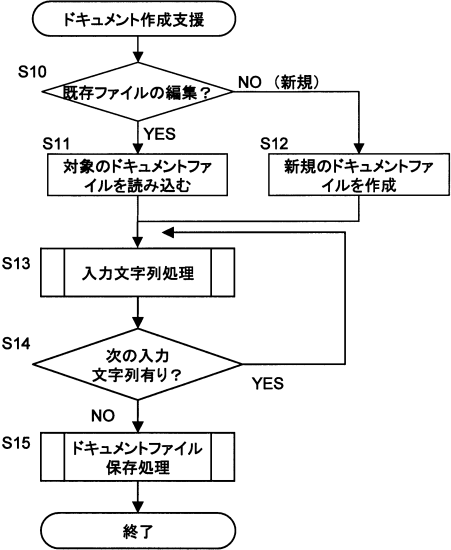


【図 5】

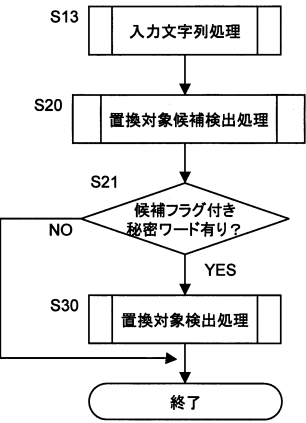
入力文字列を分類した単語例

1 2000年	17 さん
2 神奈川県	18 と
3 川崎市	19 鈴木花子
4 川崎区	20 さんが
5 B町	21 住んで
6 1丁目	22 いました。
7 1	
8 番地	
9 (
10 電話	
11 044-999-	
12 9999	
13)	
14 に	
15 山田	
16 太郎	

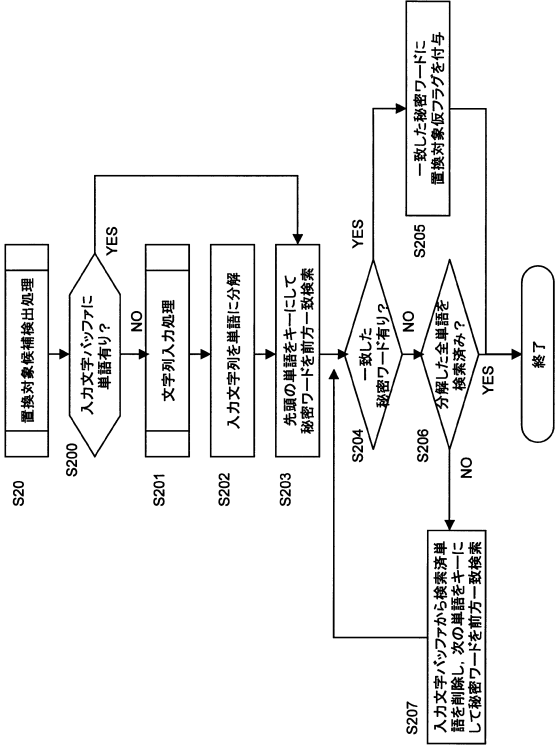
【図 6】



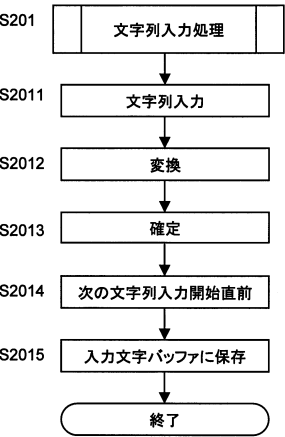
【図 7】



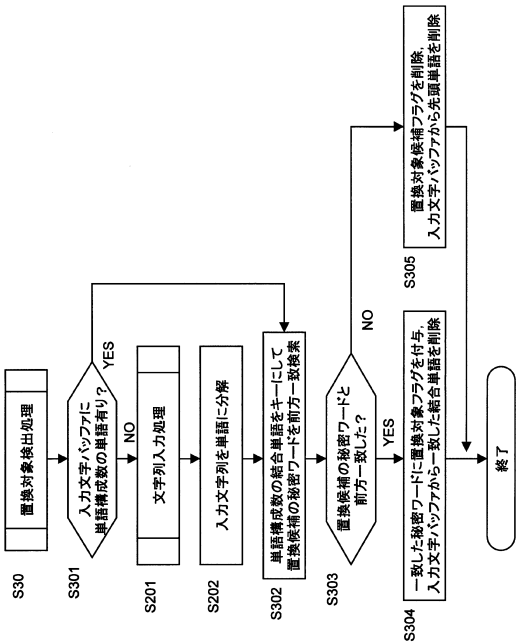
【図 8】



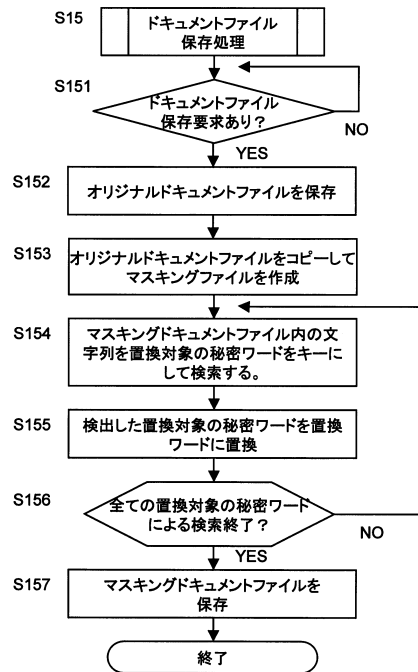
【図 9】



【図 10】



【図 11】



フロントページの続き

- (72)発明者 池村 長之
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
- (72)発明者 ゴーハー ヨン
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

審査官 長 由紀子

- (56)参考文献 特開平09-204434(JP,A)
特開2002-259368(JP,A)
特開2009-205647(JP,A)
特開2006-227425(JP,A)
特開2010-026689(JP,A)
特開2011-138365(JP,A)
特開平06-149787(JP,A)

- (58)調査した分野(Int.Cl., DB名)
G06F 17/20-28