

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4936650号  
(P4936650)

(45) 発行日 平成24年5月23日 (2012.5.23)

(24) 登録日 平成24年3月2日 (2012.3.2)

(51) Int.Cl.

F I

G O 6 F 17/30 (2006.01)

G O 6 F 17/30 3 2 O C

G O 6 F 17/22 (2006.01)

G O 6 F 17/30 3 5 O C

G O 6 F 17/22 5 2 O S

請求項の数 18 (全 26 頁)

(21) 出願番号 特願2004-217628 (P2004-217628)  
 (22) 出願日 平成16年7月26日 (2004.7.26)  
 (65) 公開番号 特開2006-39866 (P2006-39866A)  
 (43) 公開日 平成18年2月9日 (2006.2.9)  
 審査請求日 平成19年6月22日 (2007.6.22)  
 審判番号 不服2009-24809 (P2009-24809/J1)  
 審判請求日 平成21年12月15日 (2009.12.15)

(73) 特許権者 500257300  
 ヤフー株式会社  
 東京都港区赤坂9丁目7番1号  
 (72) 発明者 藤田 澄男  
 東京都江東区塩浜2丁目4番29号 株式  
 会社パトリス内

合議体  
 審判長 小曳 満昭  
 審判官 本郷 彰  
 審判官 飯田 清司

最終頁に続く

(54) 【発明の名称】 類似単語検索装置、その方法、そのプログラム、および、情報検索装置

(57) 【特許請求の範囲】

【請求項 1】

端末装置の入力プログラムによって変換された単語に関する単語情報を取得する単語情報取得手段と、

前記取得した単語情報から、その単語に変換する前の入力ローマ字列を出力するローマ字入力キーボードの操作キーの入力操作毎に出力される入力ローマ字列を生成する信号列情報生成手段と、

この信号列情報生成手段にて生成された前記単語の入力ローマ字列と単語情報記憶手段に複数記憶された単語における前記入力操作に応じた入力ローマ字列との類似状態に関する類似度情報を演算する類似度演算手段と、

前記類似度情報に基づいて前記入力ローマ字列が類似する前記単語情報記憶手段に記憶された単語を選出する単語候補選出手段と、

を具備したことを特徴とした類似単語検索装置。

【請求項 2】

請求項 1 に記載の類似単語検索装置であって、

前記信号列情報生成手段は、前記端末装置から取得した単語情報から、その単語に携帯通信端末装置に設けられた入力プログラムで変換する前の前記携帯通信端末装置のローマ字入力キーボードの操作キーの入力操作毎に出力される入力ローマ字列を生成する

ことを特徴とした類似単語検索装置。

【請求項 3】

請求項 1 又は請求項 2 に記載の類似単語検索装置であって、  
前記類似度演算手段は、前記単語情報記憶手段に記憶された前記単語における前記入力ローマ字列を前記信号列情報生成手段にて生成させ、この生成した入力ローマ字列との類似度情報を演算する

ことを特徴とした類似単語検索装置。

【請求項 4】

請求項 1 又は請求項 2 に記載の類似単語検索装置であって、  
前記類似度演算手段は、前記単語情報記憶手段に記憶された前記単語に関連付けられた前記入力ローマ字列を取得し、この取得した入力ローマ字列との類似度情報を演算する

ことを特徴とした類似単語検索装置。

10

【請求項 5】

請求項 1 ないし請求項 4 のいずれかに記載の類似単語検索装置であって、  
前記類似度演算手段は、前記入力ローマ字列における前記ローマ字入力キーボードの操作キーの入力操作毎に出力される入力ローマ字列の単位信号およびこれら単位信号を組み合わせた部分信号のうちの少なくともいずれか 1 つを有する分割信号情報を生成する分割信号情報生成手段と、この生成された分割信号情報と前記単語情報記憶手段に記憶された前記単語における前記分割信号情報との同一性に応じた重み付けに関する重み度情報を演算する重み度情報演算手段と、を備え、

前記重み度情報に基づいて前記単語情報記憶手段に記憶された前記単語の類似状態に関する類似度情報を演算する

20

ことを特徴とした類似単語検索装置。

【請求項 6】

請求項 5 に記載の類似単語検索装置であって、  
前記分割信号情報生成手段は、前記部分信号の前記入力ローマ字列を 1 字分ずつずらした文字列の並びとして生成する

ことを特徴とした類似単語検索装置。

【請求項 7】

請求項 1 ないし請求項 6 のいずれかに記載の類似単語検索装置であって、  
前記単語候補選出手段にて選出された単語を、入力操作により選択可能に表示手段に表示させる表示制御手段を具備した

30

ことを特徴とした類似単語検索装置。

【請求項 8】

請求項 1 ないし請求項 7 のいずれかに記載の類似単語検索装置であって、  
所定の入力操作により前記単語候補選出手段にて選出された単語を前記取得した単語に代える単語修正手段を具備した

ことを特徴とした類似単語検索装置。

【請求項 9】

請求項 1 ないし請求項 8 のいずれかに記載の類似単語検索装置であって、  
前記類似度演算手段は、前記信号列情報生成手段にて生成された前記単語の入力ローマ字列と、前記単語情報記憶手段に記憶された前記単語における前記入力操作に応じた入力ローマ字列との距離に関する距離情報を前記類似度情報として演算する

40

ことを特徴とした類似単語検索装置。

【請求項 10】

請求項 1 ないし請求項 8 のいずれかに記載の類似単語検索装置であって、  
前記単語候補選出手段は、前記選出した単語のうち、前記取得した単語との距離に関する距離情報を演算し、この演算した距離情報に基づいて前記選出した単語のいずれかを選出する

ことを特徴とした類似単語検索装置。

【請求項 11】

請求項 10 に記載の類似単語検索装置であって、

50

前記単語候補選出手段は、前記距離情報として、前記選出した単語の入力ローマ字列と、前記取得した単語における前記入力ローマ字列との距離に関する距離情報を演算することを特徴とした類似単語検索装置。

【請求項 1 2】

請求項 1 ないし請求項 1 1 のいずれかに記載の類似単語検索装置であって、

前記単語候補選出手段は、前記選出する単語のうち、前記単語を含む文献に関する文献情報を複数記憶する文献記憶手段における前記選出する単語の利用頻度を演算し、この利用頻度に基づいて前記単語情報記憶手段に記憶された単語を選出する

ことを特徴とした類似単語検索装置。

【請求項 1 3】

請求項 1 2 に記載の類似単語検索装置であって、

前記単語候補選出手段は、前記文献記憶手段に記憶された文献情報での出現頻度、および前記選出する単語を含む文献頻度のうちの少なくともいずれか一方を利用頻度として演算する

ことを特徴とした類似単語検索装置。

【請求項 1 4】

請求項 1 2 又は請求項 1 3 に記載の類似単語検索装置であって、

前記単語候補選出手段は、前記単語記憶手段に記憶された単語情報に、この単語情報の単語を含む文献に対応し文献を特定する文献固有情報が関連付けられた単語出現一覧テーブルに基づいて前記利用頻度を演算する

ことを特徴とした類似単語検索装置。

【請求項 1 5】

請求項 1 ないし請求項 1 4 のいずれかに記載の類似単語検索装置と、

単語を含む文献に関する文献情報を複数記憶する文献記憶手段を利用して、前記類似単語検索装置にて選出された単語を含む前記文献情報を検索する検索手段と、

を具備したことを特徴とした情報検索装置。

【請求項 1 6】

コンピュータにより、取得した単語に類似する単語を選出する類似単語検索方法であって、

前記コンピュータは、

端末装置の入力プログラムによって変換された単語を取得し、この取得した単語に変換する前の入力ローマ字列を出力するローマ字入力キーボードの操作キーの入力操作毎に出力される入力ローマ字列を生成する処理をし、

この生成した単語の入力ローマ字列と単語情報記憶手段に複数記憶された単語における前記入力操作に応じた入力ローマ字列との類似状態に関する類似度情報を演算する処理をし、

この演算した類似度情報に基づいて前記入力ローマ字列が類似する前記単語情報記憶手段に記憶された単語を選出する処理をする

ことを特徴とする類似単語検索方法。

【請求項 1 7】

コンピュータを、請求項 1 ないし請求項 1 4 のいずれかに記載の類似単語検索装置、または請求項 1 5 に記載の情報検索装置として機能させる

ことを特徴とした類似単語検索プログラム。

【請求項 1 8】

請求項 1 6 に記載の類似単語検索方法をコンピュータに実行させる

ことを特徴とした類似単語検索プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、入力される単語に類似する単語を検索する類似単語検索装置、その方法、そ

10

20

30

40

50

のプログラム、および、情報検索装置に関する。

【背景技術】

【0002】

従来、入力される単語を正しい単語に修正する装置として、取得した単語をスペルチェックにより正しい単語に修正する構成が知られている（例えば、特許文献1または特許文献2参照）。

【0003】

特許文献1に記載のものは、光学的文字読取装置や画像読取装置などにより入力され切り取った欧文文字列の単語を、単語辞書に記憶された単語と比較して正誤判定するスペルチェックを実施する。誤りであると判断した場合、誤認文字データベースを参照して入力された単語中の文字を間違いやすい文字と置換して新たな単語を生成し、再度スペルチェックする。このスペルチェックにより正しいと判定した単語が複数ある場合、表示装置に表示し、オペレータが指定した単語に変更する構成が採られている。

【0004】

一方、特許文献2に記載のものは、取り出した単語のスペルを、辞書記憶部に記憶された正しいスペルの単語と比較し、辞書記憶部にないと判断するとスペルミスと判定する。そして、ミス癖ルール表からルールを取り出し、このルールのスペル変換によりスペルミスと判定した単語を変換する。そして、変換した後のスペルが辞書記憶部にあれば候補として表示し、ユーザが表示される候補を選択することで正しく修正する構成が採られている。

【0005】

【特許文献1】特開平5 - 298495号公報（第4頁左欄～第6頁左欄）

【特許文献2】特開平10 - 207888号公報（第3頁左欄～第4頁右欄）

【発明の開示】

【発明が解決しようとする課題】

【0006】

しかしながら、上述した特許文献1または特許文献2に記載のような従来の単語を修正する構成では、判定する単語と辞書中の単語との字面の類似度合いに基づいて正誤判断しているため、正しい単語を適切に選出できないおそれがある。例えば、正誤判断する単語が「ケミッタ」である場合、「エミッタ」か「リミッタ」かいずれが最も確からしいか判定できない。

【0007】

本発明の目的は、このような点に鑑みて、類似する単語が適切に検索される類似単語検索装置、その方法、そのプログラム、および、情報検索装置を提供することにある。

【課題を解決するための手段】

【0008】

本発明の類似単語検索装置は、端末装置の入力プログラムによって変換された単語に関する単語情報を取得する単語情報取得手段と、前記取得した単語情報から、その単語に変換する前の入力ローマ字列を出力するローマ字入力キーボードの操作キーの入力操作毎に出力される入力ローマ字列を生成する信号列情報生成手段と、この信号列情報生成手段にて生成された前記単語の入力ローマ字列と単語情報記憶手段に複数記憶された単語における前記入力操作に応じた入力ローマ字列との類似状態に関する類似度情報を演算する類似度演算手段と、前記類似度情報に基づいて前記入力ローマ字列が類似する前記単語情報記憶手段に記憶された単語を選出する単語候補選出手段と、を具備したことを特徴とする。

【0009】

この発明では、単語情報取得手段で端末装置の入力プログラムによって変換された単語に関する単語情報を取得し、この単語情報に基づいて、その単語に変換する前の入力ローマ字列を出力するローマ字入力キーボードの操作キーの入力操作毎に出力される入力ローマ字列を信号列情報生成手段で生成させる。この生成した単語の入力ローマ字列と、単語情報記憶手段に複数記憶された単語における同様の入力ローマ字列との類似状態に関する

類似度情報を類似度演算手段にて演算させる。この演算した類似度情報に基づいて、入力ローマ字列が類似する単語情報記憶手段に記憶された単語を単語候補選出手段にて選出させる。このことにより、例えば単語の文字列での編集距離などにて数学的に定義される類似度による判断では、確からしさの差別化ができない場合でも類比判別が得られ、より適切な類似する単語の選出が得られる。

【 0 0 1 1 】

また、本発明では、請求項 1 に記載の類似単語検索装置であって、前記信号列情報生成手段は、前記端末装置から取得した単語情報から、その単語に携帯通信端末装置に設けられた入力プログラムで変換する前の前記携帯通信端末装置の入力ローマ字列を出力するローマ字入力キーボードの操作キーの入力操作毎に出力される入力ローマ字列を生成する構成とすることが好ましい。この構成では、例えば携帯電話や P H S (Personal Handyphone System)、P D A (Personal Digital Assistant) などの携帯通信端末装置に設けられたローマ字入力キーボードの操作キーにおける入力操作に応じた入力ローマ字列を生成するので、数字キーなどの入力操作が限られたローマ字入力キーボードの操作キーを備えた携帯通信端末装置でも適切な類似する単語の選出が得られ、利用の拡大が容易に図れる。

10

【 0 0 1 2 】

さらに、本発明では、請求項 1 又は請求項 2 に記載の類似単語検索装置であって、前記類似度演算手段は、前記単語情報記憶手段に記憶された前記単語における前記入力ローマ字列を前記信号列情報生成手段にて生成させ、この生成した入力ローマ字列との類似度情報を演算する構成とすることが好ましい。この構成では、類似度演算手段は、取得した単語の入力ローマ字列との類似度情報を演算するための単語情報記憶手段に記憶された単語の入力ローマ字列を信号列情報生成手段にて生成させるので、単語情報記憶手段に単語の入力ローマ字列を記憶させておく必要がなく、単語情報記憶手段の構成が簡略化する。

20

【 0 0 1 3 】

さらに、本発明では、請求項 1 又は請求項 2 に記載の類似単語検索装置であって、前記類似度演算手段は、前記単語情報記憶手段に記憶された前記単語に関連付けられた前記入力ローマ字列を取得し、この取得した入力ローマ字列との類似度情報を演算する構成とすることが好ましい。この構成では、類似度演算手段は、取得した単語の入力ローマ字列との類似度情報を演算するための単語情報記憶手段に記憶された単語の入力ローマ字列を、単語譲歩うき置く手段に記憶された単語に関連付けられた入力ローマ字列を用いるので、単語情報記憶手段に記憶された単語から入力ローマ字列を生成させる処理が不要で、処理負荷が低減し、処理の高速化が容易に得られる。

30

【 0 0 1 4 】

そして、本発明では、請求項 1 ないし請求項 4 のいずれかに記載の類似単語検索装置であって、前記類似度演算手段は、前記入力ローマ字列における前記ローマ字入力キーボードの操作キーの入力操作毎に出力される入力ローマ字列の単位信号およびこれら単位信号を組み合わせた部分信号のうちの少なくともいずれか 1 つを有する分割信号情報を生成する分割信号情報生成手段と、この生成された分割信号情報と前記単語情報記憶手段に記憶された前記単語における前記分割信号情報との同一性に応じた重み付けに関する重み度情報を演算する重み度情報演算手段と、を備え、前記重み度情報に基づいて前記単語情報記憶手段に記憶された前記単語の類似状態に関する類似度情報を演算する構成とすることが好ましい。この構成では、類似度演算手段は、分割信号情報生成手段にて入力ローマ字列におけるローマ字入力キーボードの操作キーでの入力操作毎に出力される入力ローマ字列の単位信号およびこれら単位信号を組み合わせた部分信号のうちの少なくともいずれか 1 つを有する分割信号情報を生成させ、重み度情報演算手段にて単位情報記憶手段に記憶された単語の分割信号情報と取得した単語の分割信号情報との同一性に応じた重み付けに関する重み度情報を演算させ、重み度情報に基づいて単語情報記憶手段に記憶された単語における取得した単語に対する類似度情報を演算するので、信号列に基づいた適切な類比判別が得られる。

40

50

## 【 0 0 1 5 】

また、本発明では、請求項5に記載の類似単語検索装置であって、前記分割信号情報生成手段は、前記部分信号の前記入力ローマ字列を1字分ずつずらした文字列の並びとして生成する構成とすることが好ましい。この構成では、分割信号情報生成手段は、部分信号として入力ローマ字列を1文字分ずつずらした文字列の並びとして生成させるので、検索漏れが減少し、適切な類似する単語が適切に検出される。また、形態素解析などの分析をする必要が無く、簡単な処理で容易に適切な類似する単語の選出が得られる。なお、文字列の文字数は、特に限定されない。例えば、1字ずつずらした3文字の部分文字列、2文字の部分文字列、または1文字の単位文字列としてもよい。また、これら3文字の部分文字列、2文字の部分文字列、および1文字の単位文字列を全て分割文字列として分割信号情報として抽出する構成とするなどしてもよい。

10

## 【 0 0 1 6 】

さらに、本発明では、請求項1ないし請求項6のいずれかに記載の類似単語検索装置であって、前記単語候補選出手段にて選出された単語を、入力操作により選択可能に表示手段に表示させる表示制御手段を具備した構成とすることが好ましい。この構成では、表示制御手段により、単語候補選出手段で選出した単語が入力操作により選択可能に表示手段で表示されるので、選択された類似する単語が利用者により確認可能となり、例えば類似する単語に置換するなどの入力支援処理や類似する単語での検索処理など、適切な処理が確実に得られる。

20

## 【 0 0 1 7 】

そして、本発明では、請求項1ないし請求項7のいずれかに記載の類似単語検索装置であって、所定の入力操作により前記単語候補選出手段にて選出された単語を前記取得した単語に代える単語修正手段を具備した構成とすることが好ましい。この構成では、単語修正手段により、所定の入力操作に基づいて、単語候補選出手段にて選出された単語を取得した単語に代えることにより、例えば誤入力された単語が適切に類似する単語にて修正され、良好な入力支援処理が得られる。

## 【 0 0 1 8 】

また、本発明では、請求項1ないし請求項8のいずれかに記載の類似単語検索装置であって、前記類似度演算手段は、前記信号列情報生成手段にて生成された前記単語の入力ローマ字列と、前記単語情報記憶手段に記憶された前記単語における前記入力操作に応じた入力ローマ字列との距離に関する距離情報を前記類似度情報として演算する構成とすることが好ましい。この構成では、類似度演算手段により、信号列情報生成手段にて生成された単語の入力ローマ字列と、単語情報記憶手段に記憶された単語の入力ローマ字列との例えば編集距離またはリーベンシュタイン距離などの距離に関する距離情報を類似度情報として演算するので、例えば入力ローマ字列から分割信号情報を生成させて重み度を演算させる処理に代えて距離情報を演算して類似状態を判別する構成としても入力操作に応じた信号列を利用するため、適切な類似判別が得られ、分割信号情報により重み度を演算させてさらに距離情報を演算することにより、より適切な類似判別が得られ、適切な類似する単語の選出が得られる。

30

## 【 0 0 1 9 】

さらに、本発明では、請求項1ないし請求項8のいずれかに記載の類似単語検索装置であって、前記単語候補選出手段は、前記選出した単語のうち、前記取得した単語との距離に関する距離情報を演算し、この演算した距離情報に基づいて前記選出した単語のいずれかを選出する構成とすることが好ましい。この構成では、単語候補選出手段により、選出した単語のうち、取得した単語との例えば編集距離またはリーベンシュタイン距離などの距離に関する距離情報を演算し、選出した単語のいずれかをさらに選出することとなるので、より適切に類似する単語の絞り込みが得られ、適切な類似する単語の選出が得られる。

40

## 【 0 0 2 0 】

そして、本発明では、請求項10に記載の類似単語検索装置であって、前記単語候補選出手段は、前記距離情報として、前記選出した単語の入力ローマ字列と、前記取得した単語

50

における前記入力ローマ字列との距離に関する距離情報を演算する構成とすることが好ましい。この構成では、単語候補選出手段により、距離情報として選出した単語の入力ローマ字列と、取得した単語の入力ローマ字列との距離に関する距離情報を演算するので、例えば単語の文字列での距離を演算するなどの処理に比してより適切な類似判別が可能となり、より適切な類似する単語の選出が得られる。

【0021】

また、本発明では、請求項1ないし請求項1 1のいずれかに記載の類似単語検索装置であって、前記単語候補選出手段は、前記選出する単語のうち、前記単語を含む文献に関する文献情報を複数記憶する文献記憶手段における前記選出する単語の利用頻度を演算し、この利用頻度に基づいて前記単語情報記憶手段に記憶された単語を選出する構成とすることが好ましい。この構成では、単語候補選出手段により、選出する単語のうち、単語を含む文献に関する文献情報を複数記憶する文献手段における選出する単語の利用頻度を演算し、この利用頻度に基づいて単語情報記憶手段に記憶された単語を選出するので、例えば類似する単語の文献における利用頻度に応じた類似判別が得られ、より適切に類似する単語の絞り込みが得られ、適切な類似する単語の選出が得られる。

10

【0022】

さらに、本発明では、請求項1 2に記載の類似単語検索装置であって、前記単語候補選出手段は、前記文献記憶手段に記憶された文献情報での出現頻度、および前記選出する単語を含む文献頻度のうちの少なくともいずれか一方を利用頻度として演算する構成とすることが好ましい。この構成では、単語候補選出手段により、文献記憶手段に記憶された文献情報での出現頻度、および選出する単語を含む文献頻度のうちの少なくともいずれか一方を利用頻度として演算するので、類似する単語の文献における利用頻度に応じた適切な類似判別が得られ、より適切に類似する単語を選出する処理が容易に得られる。

20

【0023】

そして、本発明では、請求項1 2又は請求項1 3に記載の類似単語検索装置であって、前記単語候補選出手段は、前記単語記憶手段に記憶された単語情報に、この単語情報の単語を含む文献に対応し文献を特定する文献固有情報が関連付けられた単語出現一覧テーブルに基づいて前記利用頻度を演算する構成とすることが好ましい。この構成では、単語候補選出手段により、単語記憶手段に記憶された単語情報に、この単語情報の単語を含む文献に対応し文献を特定する文献固有情報が関連付けられた単語出現一覧テーブルに基づいて利用頻度を演算するので、類似する単語における適切な類似判別のための文献における利用頻度が簡単なデータ構造で演算可能となり、処理効率の向上が容易に得られる。

30

【0024】

また、本発明の情報検索装置は、請求項1ないし請求項1 4のいずれかに記載の類似単語検索装置と、単語を含む文献に関する文献情報を複数記憶する文献記憶手段を利用して、前記類似単語検索装置にて選出された単語を含む前記文献情報を検索する検索手段と、を具備したことを特徴とする。

【0025】

この発明では、検索手段により、単語を含む文献に関する文献情報を複数記憶する文献記憶手段を利用して、適切な類似する単語を選出する請求項1ないし請求項1 4のいずれかに記載の類似単語検索装置にて選出された単語を含む文献情報を検索する。このことにより、文献情報の検索のための単語に適切に類似する単語での検索が得られ、検索漏れが減少し、適切な文献検索が得られる。

40

【0026】

本発明の類似単語検索方法は、コンピュータにより、取得した単語に類似する単語を選出する類似単語検索方法であって、前記コンピュータは、端末装置の入力プログラムによって変換された単語を取得し、この取得した単語に変換する入力ローマ字列を出力するローマ字入力キーボードの操作キーの入力操作毎に出力される入力ローマ字列を生成する処理をし、この生成した単語の入力ローマ字列と単語情報記憶手段に複数記憶された単語における前記入力操作に応じた入力ローマ字列との類似状態に関する類似度情報を演算する

50

処理をし、この演算した類似度情報に基づいて前記入力ローマ字列が類似する前記単語情報記憶手段に記憶された単語を選出する処理をすることを特徴とする。

【0027】

この発明では、請求項1に記載の類似単語検索装置の動作を、演算手段により実施する方法に展開したもので、請求項1に記載の発明と同様の作用効果を奏する。

【0028】

本発明の類似単語検索プログラムは、演算手段を、請求項1ないし請求項14のいずれかに記載の類似単語検索装置、または請求項15に記載の情報検索装置として機能させることを特徴とする。

【0029】

この発明では、例えば汎用のコンピュータなどを演算手段として利用しインストールすることにより請求項1ないし請求項14のいずれかに記載の類似単語検索装置、または請求項15に記載の情報検索装置として機能させて実施させることが可能となり、本発明の利用促進が大幅に図れる。

【0030】

本発明の類似単語検索プログラムは、請求項16に記載の類似単語検索方法をコンピュータに実行させることを特徴とする。

【0031】

この発明では、例えば汎用のコンピュータなどを利用しコンピュータにインストールすることにより請求項16に記載の類似単語検索方法がコンピュータに実行されるので、本発明の利用促進が大幅に図れる。

【0034】

そして、本発明において、コンピュータは、例えば1つのパーソナルコンピュータ、複数のコンピュータをネットワーク状に組み合わせた構成、マイクロコンピュータなどの素子、あるいは複数の電子部品が搭載された回路基板などをも含むものである。

【発明を実施するための最良の形態】

【0035】

〔文献情報検索システムの構成〕

以下、本発明の一実施の形態について図面を参照して説明する。本実施の形態では、本発明の類似単語検索装置を備え、産業財産権として、特許、実用新案および意匠に関する文献情報を検索するための情報検索装置としての文献情報検索システムの構成について例示するが、例えば商標などの他の産業財産権に関わる文献情報や、技術文献などいずれの文献情報をも対象とすることができる。さらには、文献情報検索システムに適用する構成に限らず、例えばワードプロセッサなどの文字入力支援装置として類似単語検索装置を適用するなどしてもよい。図1は、本実施の形態における文献情報検索システムの概略構成を示す概念図である。図2は、文献情報検索システムを構成する端末装置の概略構成を示すブロック図である。図3は、文献情報検索システムを構成するサーバ装置の概略構成を示すブロック図である。図4は、文献情報記憶領域のテーブル構造を概念的に示す模式図である。図5は、単語出現一覧記憶領域のテーブル構造を概念的な単語出現一覧テーブルを示す模式図である。図6は、入力されたキーワードの入力ローマ字列を分割して生成された分割文字列に関する分割信号情報を概念的に示す模式図である。図7は、単語出現一覧テーブルの単語の入力ローマ字列を分割して生成された分割文字列に関する分割信号情報を概念的に示す模式図である。

【0036】

図1において、100は文献情報検索システムで、この文献情報検索システム100は、例えば特許、実用新案および意匠などの産業財産権に関わる文献に関する文献情報を検索するためのシステムである。この文献情報検索システム100は、端末装置200と、この端末装置200がネットワーク110を介して各種情報を送受信可能に接続する文献情報検索装置としても機能するサーバ装置300と、を備えている。

【0037】



ここで、ネットワーク 110 としては、サーバ装置 300 と端末装置 200 との間を、各種データを送受信可能に接続する。このネットワーク 110 は、TCP/IP (Transmission Control Protocol/Internet Protocol) などの汎用のプロトコルに基づくインターネット、LAN (Local Area Network) などのイントラネット、無線媒体により情報が送受信可能な複数の基地局がネットワークを構成する通信回線網や放送網などのネットワーク、さらには、データを直接送受信するための媒体となる無線媒体自体など、データを送受信させるいずれの構成が利用できる。ここで、無線媒体としては、電波、光、音波、電磁波などのいずれの媒体をも適用できる。

#### 【0038】

端末装置 200 は、サーバ装置 300 から単語である所定のキーワードに対応する文献情報を取得するための装置で、例えばパーソナルコンピュータやオフィスコンピュータなどが例示できる。そして、端末装置 200 は、例えば図 2 に示すように、端末通信手段 210 と、操作手段としての端末入力手段 220 と、表示手段としての端末表示手段 230 と、端末記憶手段 240 と、端末処理手段 250 と、を備えている。

#### 【0039】

端末通信手段 210 は、ネットワーク 110 を介してサーバ装置 300 に接続される。この端末通信手段 210 は、ネットワーク 110 を介してサーバ装置 300 から端末信号を受信可能で、この端末信号の取得によりあらかじめ設定されている入力インターフェース処理を実施し、処理端末信号として端末処理手段 250 に出力する。また、端末通信手段 210 は、端末処理手段 250 から処理端末信号が入力可能で、この入力される処理端末信号の取得によりあらかじめ設定されている出力インターフェース処理を実施し、端末信号としてネットワーク 110 を介してサーバ装置 300 に送信する。

#### 【0040】

端末入力手段 220 は、例えばキーボードやマウスなどで、検索を実施する利用者により入力操作される図示しない各種操作ボタンや操作つまみなどを有している。これら操作ボタンや操作つまみなどの入力操作としては、端末装置 200 の動作内容の設定の他、文献情報の検索のための条件項目、例えばキーワードなどである。そして、端末入力手段 220 は、入力操作に対応した所定の信号を端末処理手段 250 へ適宜出力して設定させる。なお、この端末入力手段 220 としては、操作ボタンや操作つまみなどの入力操作に限らず、例えば端末表示手段 230 に設けられたタッチパネルによる入力操作や、音声による入力操作など、さらには光学的文字読取装置 (Optical Character Reader: OCR) や画像読取装置などにて読み取るなど、各種入力事項を設定入力できないいずれの構成が適用できる。

#### 【0041】

端末表示手段 230 は、端末処理手段 250 にて制御され端末処理手段 250 からの画像データを画面表示させる。画像データとしては、例えば文献情報に関する各種検索のための条件項目の入力を促す画面表示、検索した文献情報に関する情報を表示する画面表示などの他、端末装置 200 の動作内容や他のアプリケーションソフトウェアに関する画像データなど、各種画像データが表示可能である。この端末表示手段 230 としては、種々のディスプレイが挙げられ、例えば液晶表示パネルや有機 EL (Electro Luminescence) パネル、PDP (Plasma Display Panel)、CRT (Cathode-Ray Tube)、FED (Field Emission Display)、電気泳動ディスプレイパネルなどが例示できる。

#### 【0042】

端末記憶手段 240 は、ネットワーク 110 を介してサーバ装置 300 から取得した各種情報や、端末入力手段 220 で入力操作される入力事項の他、外部から取得した各種情報をも適宜記憶する。また、端末記憶手段 240 には、端末装置 200 全体を動作制御する OS (Operating System) 上に展開される各種プログラムなどを記憶する。なお、端末記憶手段 240 としては、HD (Hard Disk) や光ディスクなどの記録媒体に読み出し可能に記憶するドライブやドライバなどを備えた構成の他、メモリなど、各種情報を記憶できないいずれの構成が利用でき、ドライブとメモリとの双方さらには複数を備えた構成など

10

20

30

40

50

としてもよい。

【 0 0 4 3 】

端末処理手段 2 5 0 は、例えば C P U (Central Processing Unit) を備え、図示しない各種入出力ポート、例えば端末通信手段 2 1 0 が接続される通信制御ポート、端末入力手段 2 2 0 が接続されるキー入力ポート、端末表示手段 2 3 0 が接続される表示部制御ポート、印刷出力する図示しない出力手段としての印刷装置が接続可能な印刷制御ポート、端末記憶手段 2 4 0 が接続される記憶ポートなどを有する。そして、端末処理手段 2 5 0 は、各種プログラムとして、入力支援手段としても機能するキーワード設定手段 2 5 1 と、検索要求手段 2 5 2 と、検索結果取得手段 2 5 3 と、表示制御手段としても機能する出力制御手段 2 5 4 と、などを備えている。

10

【 0 0 4 4 】

キーワード設定手段 2 5 1 は、端末入力手段 2 2 0 による入力操作にて入力された情報に基づいて、文献情報を検索するための単語であるキーワードを設定する。具体的には、所定の入力操作にて端末入力手段 2 2 0 で入力されたキーワードに関する単語情報を生成する。なお、設定されるキーワードは、例えば入力支援手段としての入力プログラム (Input Method Editor: I M E) などにより、いわゆるローマ字入力やかな入力などの入力に対応して変換された、ひらがな、カタカナ、漢字などの日本語に限らず、アルファベット入力された英単語、中国語、ハングル文字など、いずれの単語が対象となる。

【 0 0 4 5 】

検索要求手段 2 5 2 は、キーワード設定手段 2 5 1 で取得した単語情報のキーワードに基づいて、このキーワードを含む文献情報の検索の実行をサーバ装置 3 0 0 に要求する旨の検索要求情報を生成する。すなわち、検索要求情報には、キーワードと、このキーワードを含む文献情報を検索する条件である検索条件情報と、検索を要求する利用者や端末装置 2 0 0 を特定する顧客情報などを有した端末特定情報とが、1 つのデータ構造として関連付けられている。そして、検索要求手段 2 5 2 は、生成した検索要求情報を端末通信手段 2 1 0 によりネットワーク 1 1 0 を介してサーバ装置 3 0 0 へ送信する。

20

【 0 0 4 6 】

検索結果取得手段 2 5 3 は、ネットワーク 1 1 0 を介して端末通信手段 2 1 0 によりサーバ装置 3 0 0 から取得した検索結果情報を取得する。この取得した検索結果情報は、端末記憶手段 2 4 0 や図示しないメモリなどに適宜記憶される。この検索結果情報としては、例えば文献情報を特定する文献固有情報、例えば出願番号や登録番号などの他、文献情報自体の情報をも含む。

30

【 0 0 4 7 】

出力制御手段 2 5 4 は、端末入力手段 2 2 0 による入力操作にて設定された出力状況に基づいて、検索結果取得手段 2 5 3 にて取得した検索結果情報を適宜処理し、端末表示手段 2 3 0 で表示させたり、図示しない接続された印刷装置にて印刷出力させたりする検索結果情報の出力制御をする。また、出力制御手段 2 5 4 は、サーバ装置 3 0 0 から出力される単語候補を、端末入力手段 2 2 0 による入力操作にて選択可能に端末表示手段 2 3 0 で複数表示させる制御をする。

【 0 0 4 8 】

サーバ装置 3 0 0 は、端末装置 2 0 0 からのキーワードを含む文献情報を検索して端末装置 2 0 0 へ提供する装置で、ネットワーク 1 1 0 を介して端末装置 2 0 0 と各種情報の送受信が可能となっている。このサーバ装置 3 0 0 は、例えば図 3 に示すように、サーバ通信手段 3 1 0 と、サーバ入力手段 3 2 0 と、サーバ表示手段 3 3 0 と、記憶手段として機能するサーバ記憶手段 3 4 0 と、演算手段としての類似単語検索装置であるサーバ処理手段 3 5 0 と、などを備えている。

40

【 0 0 4 9 】

サーバ通信手段 3 1 0 は、ネットワーク 1 1 0 を介して入力されるサーバ信号に対してあらかじめ設定されている入力インターフェース処理を実行し、処理サーバ信号としてサーバ処理手段 3 5 0 へ出力する。また、サーバ通信手段 3 1 0 は、サーバ処理手段 3 5 0

50

から端末装置 200 に対して送信すべき処理サーバ信号が入力されると、入力された処理サーバ信号に対してあらかじめ設定されている出力インターフェース処理を実行し、サーバ信号としてネットワーク 110 を介して端末装置 200 へ出力する。なお、サーバ信号は、処理サーバ信号に記載された情報に基づいて、適宜所定の端末装置 200 のみに出力させたり、ネットワーク 110 を介して接続される全ての端末装置 200 に出力させたりすることも可能である。

#### 【0050】

サーバ入力手段 320 は、例えば端末入力手段 220 と同様に、サーバ管理者などにて入力操作される図示しない各種操作ボタンや操作つまみなどを有している。これら操作ボタンや操作つまみなどの入力操作は、サーバ装置 300 の動作内容の設定や、サーバ記憶手段 340 に記憶する情報の設定入力、サーバ記憶手段 340 に記憶された情報の更新など、各種の設定事項である。そして、サーバ入力手段 320 は、設定事項の入力操作により、設定事項に対応する信号をサーバ処理手段 350 へ適宜出力して設定入力させる。なお、入力操作としては、同様に、操作ボタンや操作つまみなどの操作に限られない。

#### 【0051】

サーバ表示手段 330 は、例えば端末表示手段 230 と同様に、各種表示装置が用いられる。このサーバ表示手段 330 は、サーバ処理手段 350 に接続され、サーバ処理手段 350 の制御によりこのサーバ処理手段 350 から出力される画像データを適宜表示する。

#### 【0052】

サーバ記憶手段 340 は、端末記憶手段 240 と同様に、各種情報を記憶可能ないずれの構成が用いられ、文献情報記憶領域 341 と、単語出現一覧記憶領域 342 と、を備えている。文献情報記憶領域 341 は、複数の文献情報 341A を記憶するテーブル構造に構成されている。文献情報 341A は、例えば図 4 に示すように、テキスト形式の本文や画像データ形式の図面などが 1 つのデータ構造に構成された本文情報 341A1 と、文献情報 341A を特定する固有情報である例えば ID (identification) である識別情報 341A2 と、を 1 つのデータ構造として関連付けられて構成されている。なお、文献情報 341A には、出願番号や出願人、著者などの付帯情報などをも適宜関連付けられている。単語出現一覧記憶領域 342 は、例えば図 5 に示すように、単語に関する単語情報 342A1 と、その単語情報 342A1 を本文情報 341A1 に含む文献情報 341A の識別情報 341A2 とを 1 つのレコードとした単語出現情報 342A を複数記録したテーブル構造に構成されている。この単語出現一覧記憶領域 342 の単語情報 342A1 は、例えば各文献情報 341A の本文情報 341A1 に基づいて形態素解析などにより本文情報 341A1 に含まれる単語を抜き出して設定される。すなわち、各文献情報 341A に含まれる単語について、その単語を含む文献情報 341A の識別情報 341A2 が関連付けられたテーブル構造に構成されている。

#### 【0053】

また、サーバ記憶手段 340 は、サーバ装置 300 全体および文献情報検索システム 100 全体を動作制御する OS 上に展開される各種プログラムなどをも記憶している。さらに、サーバ記憶手段 340 は、端末装置 200 などから受信した各種情報や、サーバ入力手段 320 にて入力された各種情報をも適宜記憶可能となっている。

#### 【0054】

サーバ処理手段 350 は、端末処理手段 250 と同様に CPU を備え、図示しない各種入出力ポート、例えばサーバ通信手段 310 が接続される通信制御ポート、サーバ入力手段 320 が接続される入力ポート、サーバ表示手段 330 が接続される表示制御ポート、サーバ記憶手段 340 が接続される記憶ポートなどを有している。そして、サーバ処理手段 350 は、図 3 に示すように、サーバ記憶手段 340 に記憶された各種プログラムとして、単語情報取得手段としても機能する検索要求情報取得手段 351 と、単語変換手段 352 と、類似度演算手段 353 と、単語候補選出手段 354 と、検索手段 355 と、検索結果生成手段 356 と、報知手段 357 と、課金演算手段 358 と、決済手段 359 と、

10

20

30

40

50

などを備えている。

【0055】

検索要求情報取得手段351は、端末装置200から送信されサーバ通信手段310を介して検索要求情報を取得する。そして、検索要求情報取得手段351は、検索要求情報に含まれるキーワードを取得し、単語変換手段352へ出力する。

【0056】

単語変換手段352は、検索要求情報取得手段351から出力されるキーワードを所定の信号列の信号列情報に変換する。このキーワードの変換処理は、例えば入力支援手段としての入力プログラム(Input Method Editor:IME)などによりそのキーワードを入力するために端末入力手段220やサーバ入力手段320で実行されるキー操作などの入力操作に対応した信号列であるローマ字読み文字列となる入力ローマ字列に変換する。具体的には、図6にも示すように、キーワードが「ケミッタ」であれば信号列である入力ローマ字列は「kemitta」、「信号」であれば「shingou」の文字列となる信号列の信号列情報に変更する。なお、この入力ローマ字列への変更は、例えば入力プログラムを利用して変更処理される。そして、この変換された信号列情報は、サーバ記憶手段340に適宜記憶される。

【0057】

類似度演算手段353は、単語変換手段352で変換された信号列情報に基づいて、キーワードに類似した単語を選出するためにキーワードと単語出現一覧記憶領域342に記憶された単語との類似状態に関する類似度を演算する。すなわち、類似度演算手段353は、例えば、分割信号情報生成手段353Aと、重み度情報演算手段353Bと、を備えている。

【0058】

分割信号情報生成手段353Aは、単語変換手段352でキーワードから変換された文字列の信号列情報に基づいて、分割信号情報を作成する。例えば、変換した文字列を構成する文字や文字列の文字を1字ずつずらした際の部分的な文字列となる分割文字列に分割してこれら分割した文字列に関する信号の集合体である分割信号情報を生成する。具体的には、図6に示すように、キーワードが「ケミッタ」の場合、変換された入力ローマ字列である「kemitta」から、「kem」、「emi」、「mit」、「itt」、「tta」の3文字で構成された分割文字列である部分文字列と、「ke」、「em」、「mi」、「it」、「tt」、「ta」の2文字で構成された分割文字列である部分文字列と、「k」、「e」、「m」、「i」、「t」、「t」、「a」の1文字の分割文字列である単位文字列とに分割され、これら部分文字列に関する信号である部分信号と、単位文字列に関する単位信号とを有する集合体である分割信号情報を生成する。そして、生成された分割信号情報は、サーバ記憶手段340に適宜記憶される。なお、本実施の形態では、1字ずつずらした3文字の部分文字列、2文字の部分文字列、および1文字の単位文字列を全て分割文字列として抽出して説明するが、これに限られない。すなわち、部分文字列の文字数は4文字以上でもよく、特に限定されるものではなく、また、部分文字列または単位文字列を分割文字列としてもよく、いずれの組み合わせが適用できる。

【0059】

重み度情報演算手段353Bは、分割信号情報生成手段353Aにて生成された分割信号情報と、単語出現一覧記憶領域342に記憶された単語における分割信号情報との同一性に関する度合い情報を演算する。ここで、単語出現一覧記憶領域342に記憶された単語における分割信号情報は、例えば図7に示すように、上述した分割信号情報と同様に、単語出現一覧記憶領域342の各単語における部分文字列や単位文字列などの分割文字列に関する部分信号や単位信号の集合である。なお、図7は、説明の都合上、単語情報342A1として、「エミッタ」(emitta)と「リミッタ」(limitta)とについて例示した図である。そして、あらかじめ単語出現一覧記憶領域342の各単語情報342A1にそれぞれ関連付けられた上述したような分割信号情報を利用したり、別途単語出現一覧記憶領域342に記憶された各単語について単語変換手段352でそれぞれ分割信号情報を生

10

20

30

40

50

成させ、これら生成した分割信号情報を利用したりするなどが例示できる。なお、処理の高速化や処理負荷の低減などの点であらかじめ単語情報 3 4 2 A 1 に分割信号情報を関連付けておく構成とすることが好ましいことから、本実施の形態では分割信号情報が関連付けられた構成について説明する。

#### 【0060】

そして、分割信号情報の同一性に関する度合い情報の演算として、例えばキーワードが「ケミッタ」の場合、図 6 に示す分割した各部分文字列や単位文字列などの各分割文字列と一致する図 7 に示す分割した部分文字列や単位文字列などの各分割文字列における重み度を演算する。すなわち、キーワードの各部分文字列や単位文字列の分割文字列を  $q_1, \dots, q_n$  (上記「ケミッタ」の場合、 $n$  は 18)、単語出現一覧記憶領域 3 4 2 の各単語における部分文字列や単位文字列の分割文字列を  $d_1, \dots, d_m$  とすると、重み度  $W(q_i, d_j, q, d)$  は、 $q_i = d_j$  について、以下の数 1 により演算される。そして、演算された重み度  $W(q_i, d_j, q, d)$  は、サーバ記憶手段 3 4 0 に適宜記憶される。

#### 【0061】

〔数 1〕

$$W(q_i, d_j, q, d) = \text{Log}(\text{総単語数} / (\text{ } q_i = d_j \text{ の出現する単語数})) \\ \times (\text{ } q_i \text{ のキーワード内での出現数}) \\ \times (\text{ } d_j \text{ の単語内での出現数})$$

#### 【0062】

さらに、類似度演算手段 3 5 3 は、演算した重み度に基づいて、単語出現一覧記憶領域 3 4 2 の各単語の類似状態に関する類似度を演算して類似度情報を生成する。例えば、各単語の分割信号情報についての各分割文字列において、キーワードの分割信号情報についての各分割文字列との共通部分の重み度の総和を類似度  $S$  として演算する。具体的には、類似度  $S$  の演算としては、以下の数 2 により演算される。

#### 【0063】

〔数 2〕

$$S(q, d) = g(W(q_i, d_j), q, d)$$

#### 【0064】

なお、類似度演算手段 3 5 3 は、各単語の類似度  $S$  の演算として、高速かつ多数の各単語について効率よく演算するために、以下の加算方法が例示できる。すなわち、各単語  $I_n$  ( $N$  は自然数) のうち、所定の単語  $I_{n1}$  を特定する。そして、所定の単語  $I_{n1}$  の分割信号情報を構成する各分割文字列  $d_1, \dots, d_t$  のうち、キーワードの分割文字列  $q_1$  と一致する分割文字列  $d_1, \dots, d_t$  の重み度  $W(q_1, I_{n1})$  を演算し、サーバ記憶手段 3 4 0 にあらかじめ設けた単語毎の類似度  $S$  の記憶領域  $s(I_{n1})$  番地に加算する。同様にキーワードの分割文字列  $q_2 \sim q_n$  についても上述した処理を繰り返し、記憶領域  $s(I_{n1})$  番地に加算する処理をする。さらに、他の単語  $I_{n2} \sim I_{nN}$  についても同様に重み度  $W(q_i, I_{ni})$  を順次計算し、サーバ記憶手段 3 4 0 の記憶領域  $s(I_{n2})$  番地 ~ 記憶領域  $s(I_{nN})$  番地に加算する処理をする。これらにより、各単語の類似度  $S$  がそれぞれ演算される。

#### 【0065】

単語候補選出手段 3 5 4 は、類似度演算手段 3 5 3 にて演算した類似度  $S$  に基づいて、キーワードに類似する単語を選出する。すなわち、類似度  $S$  の値が多い順からあらかじめ設定された数の単語、あるいは所定の閾値以上の類似度  $S$  となる単語を類似単語として 1 つあるいは複数選出する。なお、選出される類似単語の数は、端末装置 2 0 0 における入力設定により設定してもよい。また、単語候補選出手段 3 5 4 は、選出した類似単語を端末装置 2 0 0 で確認させて設定する。そして、単語候補選出手段 3 5 4 は、複数の類似単語からさらに絞り込みを実施する距離情報算出手段 3 5 4 A と、選出した類似単語を設定する単語修正手段としても機能する検索単語設定手段 3 5 4 B と、を備えている。

#### 【0066】

距離情報算出手段 3 5 4 A は、キーワードと、選出した類似単語との距離に関する距離

情報を演算し、距離情報に基づいて類似単語のいずれかをさらに選出する絞り込み処理をする。この距離に関する距離情報としては、例えばキーワードと類似単語との編集距離またはリーベンシュタイン距離を演算する。ここで、編集距離は、キーワードと所定の類似単語とのそれぞれ文字列の異なり度合いを、1文字ごとの置換・削除・挿入などの操作を繰り返すにより、それぞれの文字列を同一にするために必要な最小の操作数である。具体的には、編集距離を動的計画法に基づいて演算する。

【0067】

すなわち、 $x_1, \dots, x_n$ の文字の集合であるキーワードを $X$ とし、 $y_1, \dots, y_n$ の文字の集合である類似単語を $Y$ とする。そして、 $X[i] = x_i$ 、 $Y[j] = y_j$ である。また、キーワード $X$ と類似単語 $Y$ との編集距離 $\text{Dist}(X, Y)$ を演算するため、 $x_1, \dots, x_i$ と、 $y_1, \dots, y_j$ との距離を距離行列 $D[i, j]$ とする。そして、距離情報算出手段354Aは、距離行列 $D[i, j]$ において、以下の数3に示すように、 $i$ および $j$ のうちの少なくともいずれか一方が「0」に関しての距離を設定する。

【0068】

〔数3〕

$$\begin{aligned} d[0, 0] &= 0, \\ d[i, 0] &= i, \\ d[0, j] &= j \end{aligned}$$

【0069】

ここで、編集距離の計算として、挿入および削除の操作数を「1」とし、置換の操作数を同一文字の場合には「0」、異なる文字の場合には「1」とする。そして、距離行列 $D[0, 0]$ は空文字列と空文字列との距離であり、編集距離の計算としては $D[0, 0] = 0$ と設定される。また、距離行列 $D[i, 0]$ は $x_1, \dots, x_i$ と $y_1$ の前の空文字列とを一致させる操作であり、この操作では $x_1, \dots, x_i$ を削除する操作となるので、編集距離の計算としては $D[i, 0] = i$ と設定される。さらに、距離行列 $D[0, j]$ は $x_1$ の前の空文字列と $y_1, \dots, y_j$ とを一致させる操作であり、この操作では $y_1, \dots, y_j$ を $x_1$ の前に挿入する操作となるので、編集距離の計算としては $D[0, j] = j$ と設定される。そして、距離情報算出手段354Aは、以下の数4および数5に基づいて、 $i$ が $1 \sim n$ 、 $j$ が $1 \sim m$ の距離行列 $D[i, j]$ をそれぞれ計算する。

【0070】

〔数4〕

$$\begin{aligned} D[i, j] &= \text{MIN} ( D[i-1, j] + 1, \\ &\quad D[i, j-1] + 1, \\ &\quad D[i-1, j-1] + \text{Cost} ) \end{aligned}$$

【0071】

〔数5〕

$$\begin{aligned} \text{Cost} &= 0 && \text{if } X[i] = Y[j], \\ \text{Cost} &= 1 && \text{if } X[i] \neq Y[j] \end{aligned}$$

【0072】

ここで、数4において、距離行列 $D[i, j]$ を計算する際、 $x_1, \dots, x_{i-1}$ と $y_1, \dots, y_j$ との距離行列 $D[i-1, j]$ の計算、 $x_1, \dots, x_i$ と $y_1, \dots, y_{j-1}$ との距離行列 $D[i, j-1]$ の計算、および、 $x_1, \dots, x_{i-1}$ と、 $y_1, \dots, y_{j-1}$ との距離行列 $D[i-1, j-1]$ の計算が済んでいるとする。この後の距離計算を続けていくとして、 $x_i$ と $y_j$ とについて可能な操作は、以下のような3つの操作のみである。

【0073】

すなわち、1つ目の可能な操作としては、 $x_1, \dots, x_{i-1}$ と $y_1, \dots, y_j$ との距離行列 $D[i-1, j]$ の計算から、 $x_1, \dots, x_i$ と $y_1, \dots, y_j$ との距離行列 $D[i, j]$ の計算を続ける場合、 $x_i$ を削除する操作となる。このため、1つ目の距離行列 $D[i, j]$ の計算として、 $D[i-1, j] + 1$ となる。また、2つ目の可能な操作としては、 $x_1, \dots, x_i$ と $y_1, \dots, y_{j-1}$ との距離行列 $D[i, j-1]$ の計算から、 $x_1, \dots, x_i$ と

$y_1, \dots, y_j$ との距離行列  $D[i, j]$  の計算を続ける場合、 $y_j$ を  $x_i$ の後に挿入する操作となる。このため、2つ目の距離行列  $D[i, j]$  の計算として、 $D[i, j-1] + 1$ となる。さらに、3つ目の可能な操作としては、 $x_1, \dots, x_{i-1}$ と  $y_1, \dots, y_{j-1}$ との距離行列  $D[i-1, j-1]$  の計算から、 $x_1, \dots, x_i$ と  $y_1, \dots, y_j$ との距離行列  $D[i, j]$  の計算を続ける場合、 $x_i$ と  $y_j$ とを置換する操作となる。このため、3つ目の距離行列  $D[i, j]$  の計算として、 $D[i-1, j-1] + \text{Cost}$ となる。ここで、 $\text{Cost}$ は、 $x_i$ と  $y_j$ とを置換する操作数を示し、数5に示すように、 $X[i]$ と  $Y[j]$ とが同一か否か、すなわち  $x_i$ と  $y_j$ とが同一か否かに応じて設定される。そして、距離行列  $D[i, j]$  を演算するため、数4に示すように、これら3つの可能な操作の中で計算値が最小となる操作が選択される。

10

#### 【0074】

そして、距離情報算出手段354Aは、計算した距離行列  $D[i, j]$  のうち、距離行列  $D[n, m]$  をキーワードXと類似単語Yとの編集距離  $\text{Dist}(X, Y)$  として算出する。この編集距離  $\text{Dist}(X, Y)$  を、キーワードXと全ての選出した類似単語Yについて算出する。これら各類似単語Yとの各編集距離  $\text{Dist}(X, Y)$  は、サーバ記憶手段340に適宜記憶される。そして、単語候補選出手段354は、選出した類似単語のうち、編集距離  $\text{Dist}(X, Y)$  の値が小さい類似単語を1つあるいは所定の数で選出する。

#### 【0075】

また、単語候補選出手段354の検索単語設定手段354Bは、選出した類似単語を端末装置200で出力可能に送信させる処理をする。例えば、端末表示手段230で表示させるとともに端末入力手段220で表示された類似単語を選択さらには修正など入力操作可能に、サーバ通信手段310にてネットワーク110を介して端末装置200へ送信する。そして、端末装置200で類似単語が選択さらには修正されることにより、選択された類似単語や修正された類似単語をサーバ通信手段310にて読み取らせ、確定する処理をする。この確定処理は、文献情報341Aの検索要求のキーワードと合わせ、設定した類似単語をキーワードとして含む文献情報341Aを検索する対象とする処理である。なお、端末装置200での入力操作により、検索要求時のキーワードに代えて類似単語をキーワードとして文献検索させるようにしたり、類似単語が選択されずにキーワードのみで文献検索させたりするなどし、利用者の要求する検索条件に対応可能としてもよい。

20

30

#### 【0076】

検索手段355は、検索要求情報取得手段351にて取得した検索要求情報に基づいて、キーワードおよび単語候補選出手段354で選出されて設定された類似単語を含む文献情報341Aを検索する。この検索は、単語出現一覧記憶領域342の単語出現一覧テーブルにより、キーワードおよび類似単語に対応する単語情報342A1に関連付けられた識別情報341A2を読み取る。この読み取った識別情報341A2に対応する文献情報341Aを文献情報記憶領域341から読み出し、その本文情報341A1を読み取る。そして、検索手段355は、読み取った文献情報341Aを検索結果生成手段356へ適宜出力する。なお、検索の結果、キーワードおよび類似単語を本文情報341A1に含む文献情報341Aを検出できなかった場合、検出できなかった旨の信号を検索結果生成手段356へ適宜出力する。

40

#### 【0077】

検索結果生成手段356は、検索手段355で取得した文献情報341Aを端末装置200へ報知するために適宜処理して検索結果情報を生成する。すなわち、端末装置200で文献情報341Aの本文情報341A1を確認可能な例えばサーバ記憶手段340にあらかじめ記憶された所定のフォームに基づいて検索結果情報を生成する。また、検出できなかった旨の信号を取得した場合には、その旨を端末装置200で出力可能な所定のフォームの検索結果情報を生成、例えばサーバ記憶手段340にあらかじめ記憶された所定のフォームを読み出す。そして、生成された検索結果情報は、サーバ記憶手段340に適宜記憶されるとともに、報知手段357へ適宜出力される。

50

## 【 0 0 7 8 】

報知手段 3 5 7 は、検索結果生成手段 3 5 6 で生成した検索結果情報を取得して適宜処理し、サーバ通信手段 3 1 0 によりネットワーク 1 1 0 を介して端末装置 2 0 0 へ送信する。このことにより、端末装置 2 0 0 は、検索結果取得手段 2 5 3 にて取得した検索結果情報を適宜処理し、文献情報 3 4 1 A の本文情報 3 4 1 A 1 を端末表示手段 2 3 0 で表示させたり、図示しない接続された印刷装置にて印刷出力させたりし、検索結果が端末装置 2 0 0 で報知される。

## 【 0 0 7 9 】

課金演算手段 3 5 8 は、検索要求情報や検索手段 3 5 5 にて検索し報知手段 3 5 7 にて送信する検索結果情報などに基づいて、検索要求をした利用者に対して検索結果の報知に関する対価を演算する。この対価の演算としては、例えば端末装置 2 0 0 がネットワーク 1 1 0 を介してサーバ装置 3 0 0 に接続した時間、検索を実行するための検索料金、検索結果の情報量に対応する回答出力料金などに基づいて演算される。なお、この対価の額は、検索結果情報とともに報知手段 3 5 7 により端末装置 2 0 0 で確認可能に送信される。

10

## 【 0 0 8 0 】

決済手段 3 5 9 は、課金演算手段 3 5 8 にて演算した対価を回収するための処理をする。例えば、サーバ記憶手段 3 4 0 にあらかじめ記憶されている端末装置 2 0 0 を用いて文献検索を利用する利用者を特定する顧客情報などに基づいて、請求書を発行したり、ネットワーク 1 1 0 を介して各金融機関が管理する付加価値通信網 (Value-Added Network: V A N) であるいわゆる金融 V A N に構築されたファームバンキング (Firm Banking: F B ) を利用して決済処理したり、決済処理する旨をサーバ表示手段 3 3 0 に表示させて管理者に決済処理を促す案内を報知するなどの処理をする。

20

## 【 0 0 8 1 】

## 〔 文献情報検索システムの動作 〕

次に、上述した文献情報検索システム 1 0 0 における所定の文献情報 3 4 1 A を検索する検索処理の動作について図面を参照して説明する。図 8 は、文献情報を検索する検索処理の動作を示すフローチャートである。

## 【 0 0 8 2 】

文献情報検索システム 1 0 0 を利用した文献の検索に際しては、まず利用者が端末装置 2 0 0 の端末入力手段 2 2 0 を適宜入力操作することで、サーバ装置 3 0 0 とネットワーク 1 1 0 を介して各種情報を送受信可能に端末装置 2 0 0 を接続させる。この接続としては、例えばあらかじめサーバ装置用のアプリケーションソフトウェアがインストールされて構築されたサーバ装置 3 0 0 が常駐するホームページなどにアクセスするなどが例示できる。そして、利用者が端末表示手段 2 3 0 に表示させる画面表示にしたがって端末入力手段 2 2 0 を適宜入力操作し、文献を検索するための文献に含まれるキーワードを入力する。この入力操作により、端末処理手段 2 5 0 のキーワード設定手段 2 5 1 が、端末入力手段 2 2 0 からの入力操作に対応、例えばキーボードの操作キーの操作毎に出力される信号列に対応して適宜変換されたキーワードを設定する (ステップ S 1 0 1 )。

30

## 【 0 0 8 3 】

この後、利用者が画面表示に従って端末入力手段 2 2 0 を適宜操作し、設定したキーワードを含む文献の検索を要求する旨の入力操作により、文献検索を実行させる。この文献検索の実行を要求する旨の入力操作としては、例えば表示画面中に設けられ文献検索の実行を促すコマンドボタンの入力操作などが例示できる。この入力操作により、端末処理手段 2 5 0 の検索要求手段 2 5 2 は、設定させたキーワードを含む文献検索の実行をサーバ装置 3 0 0 に要求する旨の検索要求情報を生成し (ステップ S 1 0 2 )、端末通信手段 2 1 0 によりネットワーク 1 1 0 を介してサーバ装置 3 0 0 へ送信させる処理をする (ステップ S 1 0 3 )。

40

## 【 0 0 8 4 】

そして、ステップ S 1 0 3 で端末装置 2 0 0 から送信された検索要求情報を、サーバ通信手段 3 1 0 でサーバ装置 3 0 0 が受信すると (ステップ S 1 0 4 )、サーバ処理手段 3

50



50の検索要求情報取得手段351が検索要求情報に含まれるキーワードを取得する(ステップS105)。このステップS105で取得したキーワードに基づいて、単語変換手段352により信号列情報を生成する信号列情報処理工程が実施する。すなわち、単語変換手段352は、取得したキーワードから、入力操作に応じて入力プログラムが変換する際のキー操作などの入力操作に対応した信号列、例えばローマ字読み文字列となる入力ローマ字列に変換して信号列情報を生成する(ステップS106)。

#### 【0085】

この後、類似度演算手段353により、生成した信号列情報に基づいて、サーバ記憶手段340の単語出現一覧記憶領域342に記憶された単語との類似状態に関する類似度を演算して類似度情報を生成させる類似度情報演算工程を実施する。すなわち、類似度演算手段353の分割信号情報生成手段353Aにより、キーワードから変換された入力ローマ字列の信号列情報に基づいて、例えば図6に示すように、例えば1字ずつずらした際の部分的な文字列となる分割文字列に分割して分割信号情報を生成する(ステップS107)。さらに、類似度演算手段353の重み度情報演算手段353Bは、単語出現一覧記憶領域342に記憶された単語における同様の入力ローマ字列から分割した例えば図7に示すような複数の分割文字列に関する分割信号情報を取得する。そして、重み度情報演算手段353Bは、キーワードの入力ローマ字列から分割した複数の分割文字列に関する分割信号情報と、単語出現一覧記憶領域342の単語の分割信号情報とに基づいて、上述した数1によりキーワードの各分割文字列と一致する単語の各分割文字列における重み度を演算する(ステップS108)。さらに、類似度演算手段353は、ステップS108で演算した重み度に基づいて、上述した数2により、各単語の分割信号情報についての各分割文字列において、キーワードの分割信号情報についての各分割文字列との共通部分の重み度の総和を演算し、キーワードと単語出現一覧記憶領域342の各単語の類似状態に関する類似度を演算し、類似度情報を生成する(ステップS109)。

#### 【0086】

そして、サーバ処理手段350は、ステップS107~S109の類似度情報演算工程で演算された類似度に基づいて、類似する単語を選出する類似単語選出工程を実施する。すなわち、単語候補選出手段354により、ステップS109で演算した類似度に基づいて、キーワードに類似する単語を、例えば類似度の値が多い順からあらかじめ設定された数の単語を類似単語として複数選出する(ステップS110)。なお、このステップS110における類似単語の選出の際、例えば選出する類似単語の数が多いなどの場合、距離情報算出手段354Aにより、キーワードと類似単語との編集距離またはリーベンシュタイン距離などの距離に関する距離情報を演算する処理などを適宜実施して、さらに絞り込み処理を実施してもよい。このステップS110の後、単語候補選出手段354の検索単語設定手段354Bは、選出した類似単語を端末装置200へ送信する処理をする(ステップS111)。この選出した類似単語に関する情報を送信する先の端末装置200は、例えばステップS104で受信した検索要求情報に基づいて特定される。

#### 【0087】

このステップS111により、サーバ装置300で選出された類似単語に関する情報を端末通信手段210で端末装置200が受信すると(ステップS112)、端末処理手段250の出力制御手段254により、端末表示手段230を適宜制御して類似単語を端末入力手段220での入力操作により選択さらには修正などが可能に画面表示させる。そして、画面表示にしたがって所定の類似単語が選択あるいは修正されると(ステップS113)、出力制御手段254は入力操作で選択あるいは修正された類似単語をサーバ装置300へ送信させる(ステップS114)。この類似単語の選択あるいは修正に関する情報を端末装置200からサーバ通信手段310で受信すると(ステップS115)、単語候補選出手段354の検索単語設定手段354Bが選択あるいは修正された類似単語を検索要求のキーワードとして確定する(ステップS116)。なお、例えば、検索要求時のキーワードに代えて類似単語をキーワードとして文献検索させるようにしたり、類似単語が選択されずにキーワードのみで文献検索させたり、キーワードとともに類似単語をキーワ

10

20

30

40

50

ードとして文献検索させたりするなど、端末装置 200 での入力操作に対応して検索要求のキーワードが適宜確定される。このような一連の処理により、キーワードに類似する類似単語が選出される。

#### 【0088】

このステップ S110 ~ S116 の類似単語選出工程の後、選出された類似単語およびキーワードを含む文献を検索する検索工程が実施される。すなわち、ステップ S116 の後、サーバ処理手段 350 の検索手段 355 は、ステップ S104 で受信した検索要求情報に基づいて、ステップ S116 で確定された検索要求のキーワードおよび類似単語を含む文献情報 341A を、単語出現一覧記憶領域 342 の単語出現一覧テーブルを用いて検索する（ステップ S117）。すなわち、キーワードおよび類似単語に対応する単語情報 342A1 に関連付けられた識別情報 341A2 を読み取り、識別情報 341A2 に対応する文献情報 341A を文献情報記憶領域 341 から読み出し、その本文情報 341A1 を読み取る。そして、検索結果生成手段 356 により、検索手段 355 で取得した文献情報 341A を端末装置 200 へ報知するために、例えばサーバ記憶手段 340 にあらかじめ記憶された所定のフォームに基づいて検索結果情報を生成する。さらに、課金演算手段 358 により、例えば端末装置 200 がネットワーク 110 を介してサーバ装置 300 に接続した時間、検索を実行するための検索料金、検索結果の情報量に対応する回答出力料金などに基づいて、検索要求をした利用者に対して検索結果の報知に関する対価を演算する（ステップ S118）。この検索工程の後、サーバ処理手段 350 は、検索結果を報知する検索結果報知工程を実施する。すなわち、報知手段 357 により、検索結果生成手段 356 で生成した検索結果情報を、課金演算手段 358 で演算した文献検索の対価に関する情報とともに、サーバ通信手段 310 によりネットワーク 110 を介して端末装置 200 へ送信する処理をする（ステップ S119）。

#### 【0089】

このステップ S119 で送信される検索結果情報および文献検索の対価に関する情報を端末装置 200 が端末通信手段 210 にて受信すると（ステップ S120）、端末処理手段 250 の検索結果取得手段 253 が取得し、出力制御手段 254 にて例えば所定のフォームに基づいて文献情報 341A を端末表示手段 230 で表示させたり、図示しない接続された印刷装置にて印刷出力させたりする検索結果情報の出力制御をし（ステップ S121）、端末装置 200 における文献のキーワード検索の処理が終了する。

#### 【0090】

一方、ステップ S119 の後、サーバ装置 300 は、決済手段 359 により、課金演算手段 358 で演算した対価を回収するための処理をする。例えば、サーバ記憶手段 340 にあらかじめ記憶されている端末装置 200 を用いて文献検索を利用する利用者を特定する顧客情報などに基づいて、請求書を発行したり、ネットワーク 110 を介していわゆる金融 V A N に構築されたファームバンキングを利用して決済処理したり、決済処理する旨をサーバ表示手段 330 に表示させて管理者に決済処理を促す案内を報知するなどの処理をし（ステップ S122）、サーバ装置 300 における文献のキーワード検索の処理が終了する。

#### 【0091】

ここで、検索要求のキーワードが上述した「ケミッタ」の場合、「エミッタ」と「リミッタ」とは編集距離またはリーベンシュタイン距離において数学的に定義される類似度が同一となる。ところで、入力操作に応じた文字列、例えば入力ローマ字列で比較した場合、図 6 および図 7 に示すように、「kemitta」と、「emitta」および「limitta」とでは、字面が似ていることで「エミッタ」が「ケミッタ」により似ていることがわかる。すなわち、入力操作に応じた入力文字列においてより高い類似性が認められることは、そのキーワードを入力しようとする意図の基、入力操作上の誤りによりキーワードが誤入力された蓋然性が高いことが判断できる。

#### 【0092】

〔実施の形態の作用効果〕

10

20

30

40

50

上述したように、上記実施の形態では、端末装置 200 からネットワーク 110 を介して取得した検索要求情報に含まれるキーワードを取得し、所定の入力操作に対応して所定の単語に変換する入力支援手段でキーワードに変換するための入力操作に応じた入力ローマ字列の信号列情報を単語変換手段 352 で生成させる。この生成したキーワードの入力ローマ字列と、単語出現一覧記憶領域 342 に複数記憶された単語における同様の信号列情報の入力ローマ字列との類似状態に関する類似度情報を類似度演算手段 353 にて演算させる。この演算した類似度情報に基づいて、信号列情報の入力ローマ字列が類似する単語出現一覧記憶領域 342 に記憶された単語を単語候補選出手段 354 にて選出させる。このため、例えば上述したように、単語の文字列での編集距離などにて数学的に定義される類似度による判断では確からしさの差別化ができない場合でも、キーワードを入力しようとする意図で入力操作上の誤りによりキーワードが誤入力された蓋然性が高いことによる高い類似性が認められる入力操作に応じた信号列である文字列すなわち入力ローマ字列に変換することで、類似状態に差が生じて類似判別が可能となり、より適切な類似する単語を選出できる。

10

#### 【0093】

そして、単語変換手段 352 により、入力操作に応じた信号列として、ローマ字読みした場合における入力ローマ字列に変換して信号列情報を生成する構成としている。このため、特に日本語入力されたキーワードに類似する単語をより適切に判別でき、より適切な類似単語を選出できる。

#### 【0094】

20

また、類似度演算手段 353 により、単語出現一覧記憶領域 342 に記憶された単語情報 342A1 に関連付けられた入力ローマ字列の信号列情報、さらには分割信号情報を読み取って類似度情報の演算を実施している。このため、単語出現一覧記憶領域 342 に記憶された全単語から信号列情報さらには分割信号情報を生成させる処理が不要で、処理負荷を低減でき、処理の高速化を容易に得ることができる。

#### 【0095】

そして、類似度演算手段 353 により、分割信号情報生成手段 353A にてキーワードの入力ローマ字列の入力操作毎に対応した部分文字列や単位文字列などの分割文字列の分割信号情報を生成させ、重み度情報演算手段 353B にて単語出現一覧記憶領域 342 に記憶された単語の分割信号情報と、キーワードの分割信号情報との同一性に応じた重み付けに関する重み度を演算させ、重み度に基づいて単語出現一覧記憶領域 342 に記憶された単語におけるキーワードに対する類似度を演算している。このため、入力操作に応じた入力ローマ字列に基づいた適切な類似判別が容易に得られる。

30

#### 【0096】

さらに、分割信号情報生成手段 353A で分割文字列を入力ローマ字列を 1 字分ずつずらした文字列の並びとして生成している。このため、入力ローマ字列に基づいてキーワードに対して類似性を有する類似単語を検索する際に、検索漏れを減少させることができ、適切な類似単語を選出できる。検索漏れが減少し、適切な類似する単語が適切に検出される。さらには、キーワードおよび単語を構成する複数の部分的な分割文字列としているので、キーワードおよび単語を形態素解析などの分析を実施する必要がなく、簡単な処理で容易にキーワードおよび単語の特徴となる情報を抽出でき、簡単な処理で容易に適切な類似する単語の選出が得られる。

40

#### 【0097】

そして、出力制御手段 254 により、単語候補選出手段 354 で選出した単語が入力操作により選択可能に端末表示手段 230 で表示している。このため、選択された類似する単語が利用者にて確認でき、例えば類似する単語に置換するなどの入力支援処理や類似する単語での検索処理など、適切な処理が確実に得られる。

#### 【0098】

また、単語候補選出手段 354 により、選出した類似単語のうち、キーワードとの例えば編集距離またはリーベンシュタイン距離などの距離に関する距離情報を演算し、選出し

50

た類似単語のいずれかをさらに選出して絞り込み処理をしている。このため、より適切に類似する類似単語の絞り込みが得られ、また重み度と異なる距離による類比判別により、類比判断が多面的となり、より適切な類似する類似単語を選出できる。さらに、このキーワードと類似単語との距離を演算する絞り込み処理の際、キーワードを入力しようとする意図で入力操作上の誤りによりキーワードが誤入力された蓋然性が高いことによる高い類似性が認められるそれぞれの入力ローマ字列における距離を演算することで、より適切な類似単語の絞り込みが得られる。

【 0 0 9 9 】

そして、文献検索のキーワードに類似する類似単語を選出する構成に適用している。このため、より漏れのない文献検索ができる。

10

【 0 1 0 0 】

また、ネットワーク 1 1 0 を介して接続する端末装置 2 0 0 からの検索要求によりサーバ装置 3 0 0 で文献検索して報知するシステム構成としている。このため、簡単な端末装置 2 0 0 の構成でも、文献情報 3 4 1 A の関連に関する検索が容易にでき、利用の拡大を容易に図ることができる。さらには、サーバ装置 3 0 0 にて統括的に容量の大きな文献情報 3 4 1 A を管理でき、データの更新などが容易で、運用管理性を容易に向上できる。したがって、携帯電話などの処理能力が比較的にかさいものを用いても容易に検索でき、利用の拡大が容易に図れる。

【 0 1 0 1 】

そして、パーソナルコンピュータなどにプログラムとして検索処理が実行可能に構成している。このため、プログラムを組み込む、さらにはプログラムを記録した記録媒体を用いることで、上記処理が容易に得られ、利用の拡大が容易に図れる。

20

【 0 1 0 2 】

〔実施形態の変形〕

なお、本発明は、上述した一実施の形態に限定されるものではなく、本発明の目的を達成できる範囲で以下に示される変形をも含むものである。

【 0 1 0 3 】

例えば、システム構成として説明したが、1 台のパーソナルコンピュータを用い、文献情報記憶領域 3 4 1 である記録媒体から文献情報 3 4 1 A を適宜検索する装置構成としたり、文献情報記憶領域 3 4 1 をネットワーク 1 1 0 を介して読み取る構成としたり、端末装置 2 0 0 に類似単語を選出させる構成、すなわちキーワードを取得する手段、信号列に変換する手段、類似度を演算する手段、類似単語を選出する手段を設け、サーバ装置 3 0 0 では検索処理するのみの構成としたり、類似単語を選出させる構成と検索する構成とをそれぞれ異なるサーバ装置 3 0 0 で実施するシステム構成としたり、サーバ記憶手段 3 4 0 の構成をサーバ装置 3 0 0 と切り離してネットワーク 1 1 0 を介して別のサーバ装置として構成したりするなど、いずれの形態とすることができる。

30

【 0 1 0 4 】

また、文献を検索するためのキーワードに類似する単語を選出する構成について説明したが、例えば上述したように、ワードプロセッサなどの文字入力支援装置として類似単語検索装置を独立して構成するなどしてもよい。このような文字入力支援装置として構成する場合には、入力操作に応じて類似単語を正しい入力文字として置換する構成、すなわち単語修正手段を設けると、より入力操作性が向上するので好ましい。

40

【 0 1 0 5 】

そして、単語変換手段 3 5 2 にてキーワードをローマ字読み文字列となる入力ローマ字列に変換して説明したが、入力ローマ字列に限らず、そのキーワードを入力するために入力操作する操作手段で入力操作することにより出力されるいずれの入力信号列に変換してもよい。例えば携帯電話による「ケミッタ」の入力の際には、「2」の操作ボタンを4回操作し、「7」を2回操作し、「4」を3回操作して小文字に変換する操作ボタンを操作し、「4」を1回操作することによる信号列に変換するなどが例示できる。このような例えば携帯電話や P H S、P D A などの携帯通信端末装置に設けられた操作手段における入

50

力操作に応じた信号列の信号列情報を生成することで、数字キーなどの入力操作が限られた操作手段を備えた携帯通信端末装置でも適切な類似する単語の選出が得られ、利用の拡大を容易に図ることができる。さらには、入力文字列として、入力ローマ字列に変換して説明したが、例えば入力かな文字列に変換するなどして類似度を演算してもよい。

【0106】

また、類似度の演算として、分割文字列での重み付けに基づいて実施したが、例えば単語変換手段352で変換した信号列を用いて、単語出現一覧テーブルの単語の入力ローマ字列との距離を演算して類似判断するなどしてもよい。

【0107】

さらに、選出した類似単語の絞り込みの処理として、距離を演算する構成としては、入力ローマ字列での距離の他、キーワードおよび類似単語の文字列での距離を演算してもよい。さらには、距離を演算する構成に代えて、例えば単語出現一覧テーブルを利用して関連付けられた文献が多いすなわち文献頻度が高い方、あるいは全文献中での出現頻度が多い方などのその単語の文献での利用頻度に基づいて判断し、利用頻度が高い方が確からしい単語であるとして重み付け（例えばスコア付け）をし、絞り込みの類似判断をするなどしてもよい。なお、単語出現一覧テーブルを用いて文献頻度に応じて類似単語のスコア付けをしてスコアの大きい順が類似度がより高いとして絞り込みを実施する構成では、単語出現一覧テーブルの共用が図れ、情報の有効利用による構成の簡略化や処理の高速化などが容易に得られる。そして、出現頻度を利用する構成では、例えば以下の数6に示す指揮に基づいてスコアV付けを実施することが例示できる。

【0108】

〔数6〕

$$V = \text{Log}(\text{Freq}(w) + C) / d$$

Freq(w)：単語wの出現頻度

C：定数

d：リーベンシュタイン距離

【0109】

これらのように、文献における利用頻度、例えば全文献での単語の出現頻度やその単語を含む文献頻度が高い単語は、頻繁に文献中にキーワードとして利用されていることから、文献検索のキーワードとしての蓋然性が高いこととなる。このため、類似単語のうち、その類似単語を含む文献における利用頻度を演算し、この利用頻度に基づいてさらに絞り込みを実施することで、より適切な類似単語を選出できる。

【0110】

また、入力されたキーワードを入力ローマ字列に変換して説明したが、例えば文字入力支援装置として利用する構成において入力された単語に類似する単語を選出する際、単語を入力操作に応じた信号列に変換する処理に代えて、例えば入力された当初の信号列を取得し、この取得した信号列で類似度を演算して類似する単語を選出するなどしてもよい。

【0111】

そして、類似度演算手段353により、キーワードの入力ローマ字列との類似度情報を演算するための単語出現一覧テーブルにおける単語の入力ローマ字列を、あらかじめ単語出現一覧テーブルの単語情報342A1に関連付けた信号列情報を読み取らせて説明したが、例えば単語出現一覧テーブルの単語に基づいて単語変換手段352により入力ローマ字列を生成させ、この生成した入力ローマ字列を利用するなどしてもよい。このことにより、単語出現一覧テーブルの単語情報342A1に入力ローマ字列の信号列情報を記憶させておく必要がなく、サーバ記憶手段340の構成の簡略化を容易に得ることができる。

【0112】

また、分割信号情報生成手段353Aで分割する分割文字列としては、1字ずつずらした3文字の部分文字列、2文字の部分文字列、および1文字の単位文字列を全て分割文字列として抽出する構成に限られない。すなわち、上述したように、部分文字列の文字数は4文字以上でもよく、特に限定されるものではなく、また、1字ずつずらした3文字の部

10

20

30

40

50

分文字列、2文字の部分文字列、または1文字の単位文字列、あるいはこれらのいずれの組み合わせが適用できる。

#### 【0113】

さらに、キーワードおよび類似単語での文献検索の結果を一括して報知する構成としたが、この場合に限られない。例えばキーワードで文献検索し、その文献結果を端末装置200に報知する際に類似単語での文献検索を続行してよいか否かの設定を促す案内を実施させ、類似単語での文献検索の続行の設定を認識することで初めて類似単語での文献検索を実施しその結果を報知するなどしてもよい。このことにより、キーワードで十分に文献が検出されたと利用者が判断した場合には、類似単語での文献検索は不要となるので、不要な処理を省略することでの処理効率を向上でき、また文献検索の適切な対価請求を実施でき、利用の拡大が容易に図れる。また、課金処理さらには決済処理などを実施しない構成としてもよい。

10

#### 【0114】

その他、本発明は、上述した一実施の形態および実施形態の変形のものに限ることなく、その他、本発明の目的を逸脱しない範囲で、例えばデータ構造を変更したり、データ構造の項目を設定したり、処理のための構成を変更するなど、様々な応用が可能である。

#### 【産業上の利用可能性】

#### 【0115】

本発明は、特許や実用新案、意匠、商標などの他、商号など広義の意味の産業財産権に関わる文献情報に関連する他の文献情報を検索する構成に利用できる。また、文献情報としては、公報以外の刊行物などをも対象とすることができる。さらには、文献情報を検索する構成に限らず、ワードプロセッサなどの文字入力支援装置として誤入力された単語に類似する単語を選出、さらには置換して修正する構成にも適用できる。

20

#### 【図面の簡単な説明】

#### 【0116】

【図1】本発明の一実施の形態に係る文献情報検索システムの概略構成を示す概念図である。

【図2】前記一実施の形態における端末装置の概略構成を示すブロック図である。

【図3】前記一実施の形態におけるサーバ装置の概略構成を示すブロック図である。

【図4】前記一実施の形態における文献情報記憶領域のテーブル構造を概念的に示す模式図である。

30

【図5】前記一実施の形態における単語出現一覧記憶領域のテーブル構造を概念的な単語出現一覧テーブルを示す模式図である。

【図6】前記一実施の形態における入力されたキーワードの入力ローマ字列を分割して生成された分割文字列に関する分割信号情報を概念的に示す模式図である。

【図7】前記一実施の形態における単語出現一覧テーブルの単語の入力ローマ字列を分割して生成された分割文字列に関する分割信号情報を概念的に示す模式図である。

【図8】前記一実施の形態における文献情報を検索する検索処理の動作を示すフローチャートである。

#### 【符号の説明】

40

#### 【0117】

100 ... 情報検索装置としての文献情報検索システム

110 ... ネットワーク

220 ... 操作手段としての端末入力手段

230 ... 表示手段としての端末表示手段

251 ... 入力支援手段としても機能するキーワード設定手段

254 ... 表示制御手段としても機能する出力制御手段

341 ... 文献記憶手段としての文献情報記憶領域

341 A ... 文献情報

341 A 2 ... 文献固有情報としての識別情報

50

3 4 2 ... 単語情報記憶手段としての単語出現一覧テーブルを構成する単語出現一覧記憶領域

3 4 2 A 1 ... 単語情報

3 5 0 ... 演算手段としての類似単語検索装置であるサーバ処理手段

3 5 1 ... 単語情報取得手段としても機能する検索要求情報取得手段

3 5 2 ... 単語変換手段

3 5 3 ... 類似度演算手段

3 5 3 A ... 分割信号情報生成手段

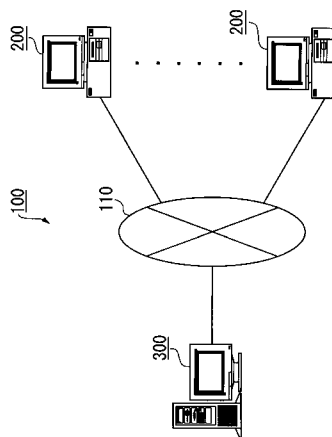
3 5 3 B ... 重み度情報演算手段

3 5 4 ... 単語候補選出手段

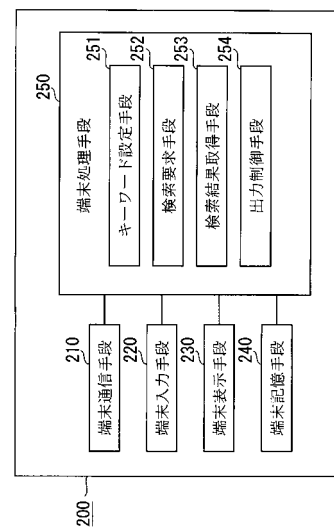
3 5 6 ... 検索手段

10

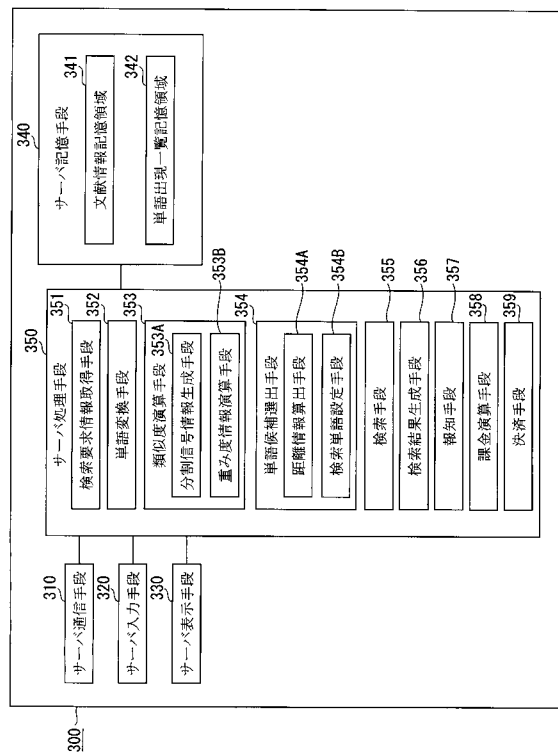
【図 1】



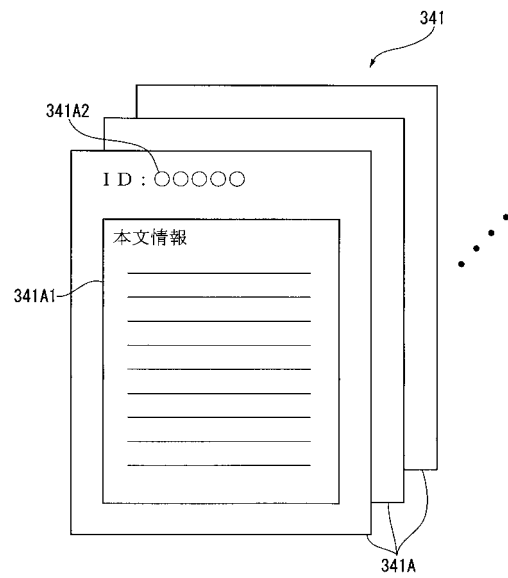
【図 2】



【図 3】



【図 4】



【図 5】

単語一覧	342A1	341A2	341A2	342
	出現文献の識別番号			
AABA	○○○△○	○○△×○	.....	
:	.....	.....	.....	
エミッタ	○○△△×	△○××○	.....	
:	.....	.....	.....	
リミッタ	×○△○○	×△△○×	.....	
.	.....	.....	.....	

【図 6】

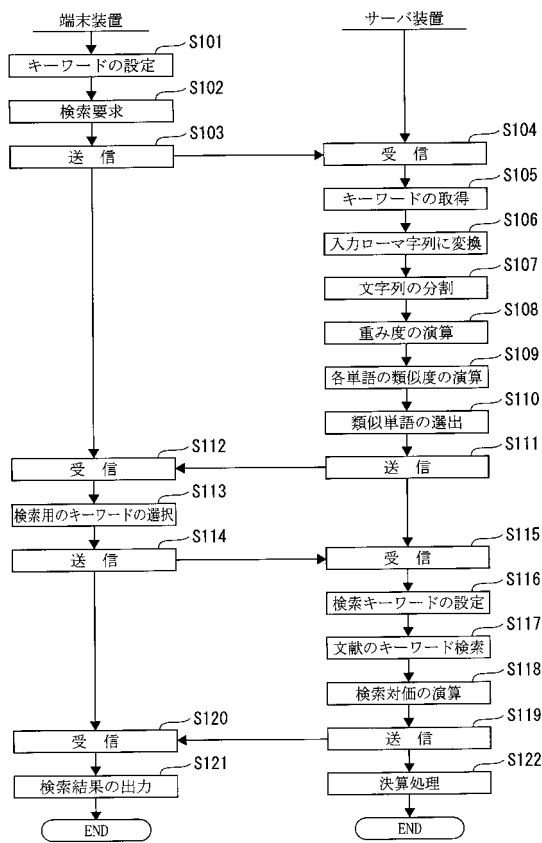
ケミッタ	[kem], [emi], [mit], [itt], [tta]
(kemitta)	[ke], [em], [mi], [it], [tt], [ta]
	[k], [e], [m], [i], [t], [t], [a]



【図 7】

エミッタ (emitta)	[emi], [mit], [itt], [tta] [em], [mi], [it], [tt], [ta] [e], [m], [i], [t], [t], [a]
リミッタ (limitta)	[lim], [imi], [mit], [itt], [tta] [li], [im], [mi], [it], [tt], [ta] [l], [i], [m], [i], [t], [t], [a]
⋮	⋮
⋮	⋮

【図 8】



---

フロントページの続き

- (56)参考文献 特開平 5 - 3 1 4 1 1 4 ( J P , A )  
特開 2 0 0 2 - 2 9 7 6 6 0 ( J P , A )  
特開平 1 1 - 3 2 8 3 1 7 ( J P , A )  
特開平 1 0 - 2 7 5 0 4 6 ( J P , A )  
特開平 7 - 1 8 2 3 4 3 ( J P , A )  
特開 2 0 0 1 - 5 2 0 2 9 ( J P , A )  
特開平 4 - 2 5 7 9 6 5 ( J P , A )

- (58)調査した分野(Int.Cl. , D B 名)  
G06F 17/30, G06F 17/22