

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4511892号
(P4511892)

(45) 発行日 平成22年7月28日(2010.7.28)

(24) 登録日 平成22年5月14日(2010.5.14)

(51) Int.Cl.

G06F 17/30 (2006.01)
G06F 17/28 (2006.01)

F 1

G06F 17/30 170Z
G06F 17/30 320D
G06F 17/28 C

請求項の数 20 (全 29 頁)

(21) 出願番号 特願2004-217744 (P2004-217744)
 (22) 出願日 平成16年7月26日 (2004.7.26)
 (65) 公開番号 特開2006-39871 (P2006-39871A)
 (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)

(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
 (74) 代理人 110000637
 特許業務法人樹之下知的財産事務所
 (74) 代理人 100079083
 弁理士 木下 實三
 (74) 代理人 100094075
 弁理士 中山 寛二
 (74) 代理人 100106390
 弁理士 石崎 剛
 (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

最終頁に続く

(54) 【発明の名称】類義語検索装置、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】

所定の単語に対する類義語を検索する類義語検索装置であって、複数の単語にて構成された文章に関する文章情報を複数記憶する記憶手段から抽出した前記単語に関する単語情報を取得する単語情報取得手段と、

前記単語に関する単語情報に、その単語の予め定められた範囲内に出現する前記文章中の他の単語を前記単語情報の単語に関連する関連単語として、該関連単語に関する関連単語情報を複数有する関連単語集合情報が、1つのデータ構造として関連付けられた単語関連情報が複数格納されるテーブル構造の関連単語テーブルに基づき、前記取得した単語情報の所定の単語に対応する単語関連情報を取得する単語関連情報取得手段と、

前記関連単語テーブルにおける前記取得した単語関連情報の関連単語集合情報に類似する関連単語集合情報を備えた単語関連情報を検索する関連単語検索手段と、

この関連単語検索手段にて検索された前記単語関連情報の単語情報を取得して前記所定の単語に対する類義語として選出する類義語選出手段と、を具備し、

前記関連単語検索手段は、前記所定の単語に対応する単語関連情報の関連単語集合情報に対する前記関連単語テーブルにおける単語関連情報の関連単語集合情報の類似状態に関する類似度情報を演算する際、前記所定の単語に対応する単語関連情報の関連単語集合情報を構成する各関連単語情報について、該関連単語情報と文字列が一致する関連単語情報を関連単語集合情報に有する、前記関連単語テーブルにおける前記単語関連情報の単語情報の数を一致単語数として、前記関連単語テーブルに複数記憶された単語関連情報の単語

10

20

情報についての総単語数を前記一致単語数で除算することで重み度情報を演算し、この重み度情報に基づいて前記類似度情報を演算し、この演算した類似度情報に基づいて前記単語関連情報を検出する

ことを特徴とした類義語検索装置。

【請求項 2】

請求項1に記載の類義語検索装置であって、

前記関連単語テーブルは、文章に関する文章情報を複数記憶する文章記憶手段に基づいて生成された

ことを特徴とした類義語検索装置。

【請求項 3】

請求項2に記載の類義語検索装置であって、

前記文章記憶手段は、前記文章を複数含む文献毎の集合とした前記文章情報を複数有する文献情報を複数記憶するテーブル構造に構築された

ことを特徴とした類義語検索装置。

【請求項 4】

請求項3に記載の類義語検索装置であって、

前記単語情報には、その単語が出現する文章を含む文献に関する前記文献情報が関連付けられ、

前記類義語選出手段にて類義語として選出した単語情報に関連付けられた文献の数を計数し、その計数した数に基づいて演算した文献頻度を単語情報にスコア付けする頻度認識手段を備える

ことを特徴とする類義語検索装置。

【請求項 5】

請求項3または請求項4に記載の類義語検索装置であって、

前記関連単語検索手段は、前記所定の単語が含まれる前記文献の文献情報における前記文章情報の文章に含まれる前記他の単語に対応する単語関連情報を検索する

ことを特徴とした類義語検索装置。

【請求項 6】

請求項3ないし請求項5のいずれかに記載の類義語検索装置であって、

前記文献情報は、前記文献の分類に関する文献分類情報を有し、

前記関連単語検索手段は、前記所定の単語が含まれる前記文献の文献情報の文献分類情報と同一の文献分類情報を有した前記文献情報における前記文章情報の文章に含まれ前記所定の単語の近傍に位置する前記他の単語に対応する単語関連情報を検索する

ことを特徴とした類義語検索装置。

【請求項 7】

請求項3ないし請求項6のいずれかに記載の類義語検索装置であって、

前記文献情報は、前記文献に関する書誌情報を有し、

前記関連単語検索手段は、前記所定の単語が含まれる前記文献の文献情報に関連付けられた書誌情報を有した前記文献情報における前記文章情報の文章に含まれ前記所定の単語の近傍に位置する前記他の単語に対応する単語関連情報を検索する

ことを特徴とした類義語検索装置。

【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義語検索装置であって、

前記単語関連情報の単語情報は、形態素解析により前記文章から抽出した単語に関する情報である

ことを特徴とした類義語検索装置。

【請求項 9】

請求項1に記載の類義語検索装置であって、

前記単語関連情報取得手段は、単語に関する単語情報に、その単語に関する説明文に出現する単語に関する関連単語情報を複数有する関連単語集合情報を 1 つのデータ構造とし

10

20

30

40

50

て関連付けられた単語関連情報が単語毎に複数格納されるテーブル構造の関連単語テーブルに基づいて、前記単語関連情報を取得する
ことを特徴とした類義語検索装置。

【請求項 10】

請求項 1 に記載の類義語検索装置であって、

前記単語関連情報取得手段は、単語に関する単語情報に、その単語のシソーラス項目となる単語群を構成する各単語に関する関連単語情報を複数有する関連単語集合情報が 1 つのデータ構造として関連付けられ、単語毎に複数格納されるテーブル構造の関連単語テーブルに基づいて、前記単語関連情報を取得する
ことを特徴とした類義語検索装置。

10

【請求項 11】

請求項 1 ないし請求項 10 のいずれかに記載の類義語検索装置であって、

前記単語情報取得手段にて前記所定の単語に関する単語情報を取得すると、前記関連単語テーブルを作成する関連単語テーブル生成手段を具備した
ことを特徴とした類義語検索装置。

【請求項 12】

請求項 1 ないし請求項 11 のいずれかに記載の類義語検索装置であって、

前記関連単語テーブルを記憶する関連単語記憶手段を具備した
ことを特徴とした類義語検索装置。

【請求項 13】

請求項 1 ないし請求項 12 のいずれかに記載の類義語検索装置であって、

前記関連単語テーブルの単語関連情報は、前記単語情報の単語の分類に関する単語分類情報を有し、

前記関連単語検索手段は、前記所定の単語に対応する単語関連情報の前記単語分類情報と同一の単語分類情報を有した前記関連単語テーブルにおける単語関連情報から検索する
ことを特徴とした類義語検索装置。

20

【請求項 14】

請求項 1 ないし請求項 13 のいずれかに記載の類義語検索装置であって、

前記類義語選出手段にて類義語として選出した単語情報を、前記所定の単語に対応する単語情報に関連付けて 1 つのデータ構造とした類義語情報を生成し、各種情報を記憶する記憶手段に複数記憶させる類義語情報生成手段を具備した
ことを特徴とした類義語検索装置。

30

【請求項 15】

請求項 1 ないし請求項 14 のいずれかに記載の類義語検索装置と、

この類義語検索装置にて類義語として選出した前記単語情報および前記単語情報取得手段で取得した単語情報のうちの少なくともいずれか 1 つを含む文章に関する文章情報の検索を実行する検索手段と、
を具備したことを特徴とした情報検索装置。

【請求項 16】

請求項 3 ないし請求項 7 のいずれかに記載の類義語検索装置と、

この類義語検索装置にて類義語として選出した前記単語情報および前記単語情報取得手段で取得した単語情報のうちの少なくともいずれか 1 つを含む文献情報の検索を実行する検索手段と、
を具備したことを特徴とした情報検索装置。

40

【請求項 17】

請求項 15 または請求項 16 に記載の情報検索装置であって、

前記検索手段は、前記類義語検索装置にて選出した前記単語情報および前記単語情報取得手段で取得した単語情報の単語を、入力操作可能な操作手段における入力操作に応じて選択および変更のうちの少なくともいずれか一方を実施可能に前記表示手段に表示させ、前記入力操作に応じて前記単語の選択および変更のうちの少なくともいずれか一方が実施
50

されて選択された単語および変更された単語のうちの少なくともいずれか一方の単語に関する単語情報を基づいて検索を実行する

ことを特徴とした情報検索装置。

【請求項 18】

コンピュータにより、複数の単語にて構成された文章に関する文章情報を複数記憶する記憶手段を利用して所定の単語に対する類義語を検索する類義語検索方法であって、

前記コンピュータは、

所定の単語に関する単語情報を取得すると、単語に関する単語情報に、その単語の予め定められた範囲内に出現する前記文章中の他の単語を前記単語情報の単語に関する関連単語として、該関連単語に関する関連単語情報を複数有する関連単語集合情報が、1つのデータ構造として関連付けられた単語関連情報を複数記憶するテーブル構造に構築された関連単語記憶領域に基づき、前記取得した所定の単語に対応する前記単語関連情報を取得し、

この取得した単語関連情報の関連単語集合情報に類似する関連単語集合情報を備えた単語関連情報を検索し、

この検索した単語関連情報の単語情報を前記所定の単語に対する類義語として選出する際、

前記所定の単語に対応する単語関連情報の関連単語集合情報を前記関連単語テーブルにおける単語関連情報の関連単語集合情報の類似状態に関する類似度情報を演算する時、前記所定の単語に対応する単語関連情報の関連単語集合情報を構成する各関連単語情報について、該関連単語情報と文字列が一致する関連単語情報を関連単語集合情報に有する、前記関連単語テーブルにおける前記単語関連情報の単語情報の数を一致単語数として、前記関連単語テーブルに複数記憶された単語関連情報の単語情報についての総単語数を前記一致単語数で除算することで重み度情報を演算し、この重み度情報に基づいて前記類似度情報を演算し、この演算した類似度情報に基づいて前記単語関連情報を検出する

ことを特徴とする類義語検索方法。

【請求項 19】

コンピュータを、請求項1ないし請求項14のいずれかに記載の類義語検索装置、または、請求項15ないし請求項17のいずれかに記載の情報検索装置として機能させる

ことを特徴とした検索プログラム。

【請求項 20】

請求項18に記載の類義語検索方法をコンピュータに実行させる

ことを特徴とした検索プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、文章に関する文章情報を利用して類義語を検索する類義語検索装置、その方法、そのプログラム、および、情報検索装置に関する。

【背景技術】

【0002】

従来、所定の単語に関する関連語を検索する装置として、文書ファイルから抽出した単語を利用する構成が知られている（例えば、特許文献1参照）。この特許文献1に記載のものは、文書ファイルを参照して文書中で前後に共通した単語が連なる2つの異なる単語を関連語の候補対として選定し、二連単語リストを作成する。注目単語に対して、前置単語のリストと、後置単語のリストを生成する。そして、二連単語リストを参照して、各前置単語に対する後置単語の集合と、各後置単語に対する前置単語の集合とを生成する。この語、これらの集合の共通要素から、関連語の候補対を抽出する構成が採られている。

【0003】

【特許文献1】特開2003-256447号公報（第3頁左欄～第4頁右欄）

【発明の開示】

10

20

30

40

50

【発明が解決しようとする課題】**【0004】**

しかしながら、上述した特許文献1に記載のような従来の関連語を検索する構成では、二連単語の候補対から前置単語のリストおよび後置単語のリストを生成して、それぞれに対する集合を生成する演算をしているため、演算負荷が比較的に大きい問題がある。また、二連単語に基づいて関連語を検索しているため、所定の単語に対する類義語の検索漏れが生じるおそれがある。

【0005】

本発明の目的は、このような点に鑑みて、類義語を簡単で適切に検索する類義語検索装置、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。

10

【課題を解決するための手段】**【0006】**

本発明の類義語検索装置は、所定の単語に対する類義語を検索する類義語検索装置であつて、複数の単語にて構成された文章に関する文章情報を複数記憶する記憶手段から抽出した前記単語に関する単語情報を取得する単語情報取得手段と、前記単語に関する単語情報に、その単語の予め定められた範囲内に出現する前記文章中の他の単語を前記単語情報の単語に関連する関連単語として、該関連単語に関する関連単語情報を複数有する関連単語集合情報が、1つのデータ構造として関連付けられた単語関連情報が複数格納されるテーブル構造の関連単語テーブルに基づき、前記取得した単語情報の所定の単語に対応する単語関連情報を取得する単語関連情報取得手段と、前記関連単語テーブルにおける前記取得した単語関連情報の関連単語集合情報に類似する関連単語集合情報を備えた単語関連情報を検索する関連単語検索手段と、この関連単語検索手段にて検索された前記単語関連情報の単語情報を取得して前記所定の単語に対する類義語として選出する類義語選出手段と、を具備し、前記関連単語検索手段は、前記所定の単語に対応する単語関連情報の関連単語集合情報に対する前記関連単語テーブルにおける単語関連情報の関連単語集合情報の類似状態に関する類似度情報を演算する際、前記所定の単語に対応する単語関連情報の関連単語集合情報を構成する各関連単語情報について、該関連単語情報と文字列が一致する関連単語情報を関連単語集合情報に有する、前記関連単語テーブルにおける前記単語関連情報の単語情報の数を一致単語数として、前記関連単語テーブルに複数記憶された単語関連情報の単語情報についての総単語数を前記一致単語数で除算することで重み度情報を演算し、この重み度情報を基づいて前記類似度情報を演算し、この演算した類似度情報に基づいて前記単語関連情報を検出することを特徴とする。

20

【0007】

この発明では、単語に関する単語情報をその単語に関連しその単語を特徴付ける単語に関する関連単語情報を複数有する関連単語集合情報を1つのデータ構造として関連付けた単語関連情報を複数格納するテーブル構造の関連単語テーブルを構築する。この関連単語テーブルに基づいて、単語情報取得手段にて取得した所定の単語に関する単語情報を対応する単語関連情報を取得する。取得した単語関連情報の関連単語集合情報に類似する関連単語集合情報を備えた単語関連情報を、関連単語テーブルに基づいて関連単語検索手段にて検索する。検出した単語関連情報の単語情報を、類義語選出手段により所定の単語に対する類義語として選出させる。この構成では、単語に関連し特徴付ける単語の集合が類似することで、単語同士の特徴が類似することとなり類義性が裏付けられるので、関連単語テーブルを利用して関連単語集合情報の類似性の判断により、単語にその単語を特徴付ける単語群を関連付けた簡単なテーブル構造を利用して容易に適切な類義語が検索可能となる。

30

そして、本発明では、関連単語検索手段により、所定の単語に対応する単語関連情報の関連単語集合情報に対して、関連単語テーブルにおける各単語関連情報の関連単語集合情報の類似状態に関する類似度情報を演算し、この類似度情報に基づいて単語関連情報を検出しているため、関連単語テーブルを利用した簡単な演算により、類似状態に応じた類義語が検出される。

40

50

さらに、本発明では、関連単語検索手段における類似度情報の演算として、所定の単語に対応する単語関連情報の関連単語集合情報を構成する関連単語情報と、関連単語テーブルにおける単語関連情報の関連単語集合情報を構成する関連単語情報との文字列が一致する数を一致単語数として、前記関連単語テーブルに複数記憶された単語関連情報の単語情報についての総単語数を前記一致単語数で除算することで重み度情報を演算し、この重み度情報に基づいて演算しているため、関連単語テーブルを利用した類似状態に応じた類義語の検索のための類似度情報の演算が容易で、処理負荷が低減する。

【0008】

そして、本発明では、前記単語関連情報取得手段は、複数の単語にて構成された文章に出現する単語情報に、その単語の文章における近傍に位置する他の単語、すなわち予め定められた範囲内に出現する文章中の他の単語に関する関連単語情報を複数有する関連単語集合情報が1つのデータ構造として関連付けられ、前記文章に出現する各単語についてそれぞれ構成された複数の単語関連情報が格納されるテーブル構造の関連単語テーブルに基づいて、前記単語関連情報を取得する構成とすることが好ましい。この構成では、文章における単語の近傍に位置する他の単語の集合は、その単語の機能や効能または作用あるいは効果、これら機能などを奏すものなどの比較的に関連する単語であるので、他の単語の集合をその単語に関連付けた単語関連情報を複数格納するテーブル構造の関連単語テーブルを利用することで、所定の単語に対して関連する単語が検索されることとなる。このため、他の単語の集合に関連付く単語を類義語としているため、単語にその単語に関連する単語を関連付けた簡単なテーブル構造を利用して容易に適切な類義語の検索が可能となる。

【0009】

さらに、本発明では、前記関連単語テーブルは、文章に関する文章情報を複数記憶する文章記憶手段に基づいて生成された構成とすることが好ましい。この構成では、文章に関する文章情報を複数記憶する文章記憶手段に基づいて関連単語テーブルを生成しているので、例えば用語辞典などの言語資源を利用する必要がなく、検索対象の文献を蓄積するデータベースを利用するなどにより、関連単語テーブルが容易に構築される。

【0010】

そして、本発明では、前記単語関連情報の単語情報は、形態素解析により前記文章から抽出した単語に関する情報である構成とすることが好ましい。この構成では、単語関連情報の単語情報として、形態素解析により文章から抽出した単語としているため、簡単な処理で関連単語テーブルが構築され、構成の構築や保守管理などが容易となる。

【0011】

また、本発明では、請求項1に記載の類義語検索装置であって、前記単語関連情報取得手段は、単語に関する単語情報に、その単語に関する説明文に出現する単語に関する関連単語情報を複数有する関連単語集合情報を構成する関連単語テーブルが1つのデータ構造として関連付けられた単語関連情報が単語毎に複数格納されるテーブル構造の関連単語テーブルに基づいて、前記単語関連情報を取得する構成とすることが好ましい。この構成では、単語に関する説明文から抽出された単語は、単語を特徴付ける関連する単語であることから、説明文から抽出した単語の集合を関連単語集合情報を構成して単語関連情報を構成して複数格納するテーブル構造の関連単語テーブルを利用することで、所定の単語に対して特徴付く単語が検索されることとなる。このため、説明文から抽出した単語の集合に関連付く単語を類義語としているため、単語にその単語に関連し特徴付ける単語を関連付けた簡単なテーブル構造を利用して容易に適切な類義語の検索が可能となる。

【0012】

さらに、本発明では、請求項1に記載の類義語検索装置であって、前記単語関連情報取得手段は、単語に関する単語情報に、その単語のシソーラス項目となる単語群を構成する各単語に関する関連単語情報を複数有する関連単語集合情報を構成する関連単語テーブルが1つのデータ構造として関連付けられ、単語毎に複数格納されるテーブル構造の関連単語テーブルに基づいて、前記単語関連情報を取得する構成とすることが好ましい。この構成では、単語のシソーラス項

10

20

30

40

50

目となる単語群を構成する各単語は、単語を特徴付ける関連した単語であることから、シソーラス項目の単語群を関連単語集合情報として単語関連情報を構成してテーブル構造に構築した関連単語テーブルを利用することで、所定の単語に対して特徴付く単語が検索されることとなる。このため、シソーラス項目における単語群に関連付く単語を類義語とすることで、単語にその単語に関連し特徴付ける単語を関連付けた簡単なテーブル構造を利用して容易に適切な類義語の検索が可能となる。

【0013】

そして、本発明では、前記単語情報取得手段にて前記所定の単語に関する単語情報を取得すると、前記関連単語テーブルを作成する関連単語テーブル生成手段を具備した構成とすることが好ましい。このことにより、単語情報取得手段にて所定の単語に関する単語情報を取得すると、関連単語テーブル生成手段が関連単語テーブルを作成する。このことにより、別途関連単語テーブルを記憶しておく必要がなく、構成の簡略化が図れる。10

【0014】

また、本発明では、前記関連単語テーブルを記憶する関連単語記憶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、所定の単語に対する類似語の検索の際、あらかじめ関連単語記憶手段に記憶した関連単語テーブルを利用すればよく、処理の高速化が図れる。

【0017】

また、本発明では、前記関連単語テーブルの単語関連情報は、前記単語情報の単語の分類に関する単語分類情報を有し、前記関連単語検索手段は、前記所定の単語に対応する単語関連情報の前記単語分類情報と同一の単語分類情報を有した前記関連単語テーブルにおける単語関連情報から検索する構成とすることが好ましい。この構成では、関連単語テーブルの単語関連情報における単語情報の単語の分類に関する単語分類情報と同一の単語分類情報を有した関連単語テーブルにおける単語関連情報から、所定の単語に対応する単語関連情報の関連単語集合情報に類似する関連単語集合情報を備えた単語関連情報を検索するので、簡単なテーブル構造で、所定の単語に対する類義語が所定の関連単語テーブルの一部の検索のみで容易に検索される。20

【0020】

また、本発明では、前記類義語選出手段にて類義語として選出した単語情報を、前記所定の単語に対応する単語情報に関連付けて1つのデータ構造とした類義語情報を生成し、各種情報を記憶する記憶手段に複数記憶させる類義語情報生成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、類義語選出手段で類義語として選出した単語情報を、類義語情報生成手段により、所定の単語に対応する単語情報を関連付けて1つのデータ構造とした類義語情報を生成して記憶手段に記憶させるので、1度演算した所定の単語に対する類義語が蓄積され、所定の単語に対する類義語の検出が容易となる。30

【0021】

さらに、本発明では、前記文章記憶手段は、前記文章を複数含む文献毎の集合とした前記文章情報を複数有する文献情報を複数記憶するテーブル構造に構築された構成とすることが好ましい。この構成では、文書を複数含む文献毎の集合とした文章情報を複数有する文献情報を複数記憶するテーブル構造に構築した文章記憶手段を利用して文献を複数蓄積する構成を利用して文献で利用される適切な類義語が検出され、文献毎に文献情報として複数記憶する簡単な文章記憶手段の構成を利用して文章記憶手段の構築や保守管理が容易となるとともに、例えば所定の単語および類義語に基づく文献の検索も容易となる。40

【0022】

そして、本発明では、前記関連単語検索手段は、前記所定の単語が含まれる前記文献の文献情報における前記文章情報の文章に含まれ前記所定の単語の近傍に位置する前記他の単語に対応する単語関連情報から検索する構成とすることが好ましい。この構成では、関連単語検索手段により、所定の単語が含まれる文献の文献情報における文章情報の文章に含まれる他の単語に対応する単語関連情報から検索するので、所定の単語を含む文献は所50

定の単語に対する類義語の存在頻度が比較的に高いことから、一部の文献から効率よく類義語が検索される。

【0023】

また、本発明では、前記文献情報は、前記文献の分類に関する文献分類情報を有し、前記関連単語検索手段は、前記所定の単語が含まれる前記文献の文献情報の文献分類情報と同一の文献分類情報を有した前記文献情報における前記文章情報の文章に含まれ前記所定の単語の近傍に位置する前記他の単語に対応する単語関連情報から検索する構成とすることが好ましい。この構成では、関連単語検索手段により、所定の単語が含まれる文献の文献情報に設けられ文献の分類に関する文献分類情報と同一の文献分類情報を有した文献情報における文章情報の文章に含まれ所定の単語の近傍に位置する他の単語に対応する単語関連情報から検索するので、所定の単語を含む文献と同一の分類の文献は所定の単語に対する類義語の存在頻度が比較的に高いことから、一部の文献から効率よく類義語が検索される。

10

【0024】

さらに、本発明では、前記文献情報は、前記文献に関する書誌情報を有し、前記関連単語検索手段は、前記所定の単語が含まれる前記文献の文献情報に関連付けられた書誌情報と同一の書誌情報を有した前記文献情報における前記文章情報の文章に含まれ前記所定の単語の近傍に位置する前記他の単語に対応する単語関連情報から検索する構成とすることが好ましい。この構成では、関連単語検索手段により、所定の単語が含まれる文献の文献情報に設けられ文献に関する書誌情報と同一の書誌情報を有した文献情報における文章情報の文章に含まれ所定の単語の近傍に位置する他の単語に対応する単語関連情報から検索するので、所定の単語を含む文献と同一の書誌情報の文献は所定の文語に対する類義語の存在頻度が比較的に高いことから、一部の文献から効率よく類義語が検索される。

20

【0025】

また、本発明の情報検索装置は、本発明に記載の類義語検索装置と、この類義語検索装置にて類義語として選出した前記単語情報および前記単語情報取得手段で取得した単語情報のうちの少なくともいずれか1つを含む文章に関する文章情報の検索を実行する検索手段と、を具備したことを特徴とする。

【0026】

この発明では、本発明に記載の類義語検索装置にて類義語として選出した単語情報および単語情報取得手段で取得した単語情報のうちの少なくともいずれか1つを含む文章に関する文章情報の検索を検索手段にて実行する。このことにより、文章情報の検索のための単語に対して適切な類義語での文章検索が可能となり、文章の検索漏れが減少し、適切に文章が検索される。

30

【0027】

さらに、本発明の情報検索装置は、本発明に記載の類義語検索装置と、この類義語検索装置にて類義語として選出した前記単語情報および前記単語情報取得手段で取得した単語情報のうちの少なくともいずれか1つを含む文献情報の検索を実行する検索手段と、を具備したことを特徴とする。

【0028】

40

この発明では、本発明に記載の類義語検索装置にて類義語として選出した単語情報および単語情報取得手段で取得した単語情報のうちの少なくともいずれか1つを含む文献情報の検索を検索手段にて実行する。このことにより、文献情報の検索のための単語に対して適切な類義語での文献検索が可能となり、文献の検索漏れが減少し、適切に文献が検索される。

【0029】

そして、本発明では、本発明に記載の情報検索装置であって、前記検索手段は、前記類義語検索装置にて選出した前記単語情報および前記単語情報取得手段で取得した単語情報の単語を、入力操作可能な操作手段における入力操作に応じて選択および変更のうちの少なくともいずれか一方を実施可能に前記表示手段に表示させ、前記入力操作に応じて前記

50

単語の選択および変更のうちの少なくともいずれか一方が実施されて選択された単語および変更された単語のうちの少なくともいずれか一方の単語に関する単語情報に基づいて検索を実行する構成とすることが好ましい。この構成では、検索手段により、類義語検索装置にて選出した単語情報および類義語検索装置の単語情報取得手段で取得した単語情報の単語を、入力操作可能な操作手段における入力操作に応じて選択および変更のうちの少なくともいずれか一方を実施可能に表示手段に表示させる。そして、検索手段は、入力操作に応じて選択や変更が実施されて選択された単語および変更された単語のうちの少なくともいずれか一方の単語に関する単語情報に基づいて、文章や文献の検索をする。このことにより、所定の単語や所望の類義語による所望の文章や文献が得られる。

【0030】

また、本発明の類義語検索方法は、コンピュータにより、複数の単語にて構成された文章に関する文章情報を複数記憶する記憶手段を利用して所定の単語に対する類義語を検索する類義語検索方法であって、前記コンピュータは、所定の単語に関する単語情報を取得すると、単語に関する単語情報に、その単語の予め定められた範囲内に出現する前記文章中の他の単語を前記単語情報の単語に関連する関連単語として、該関連単語に関する関連単語情報を複数有する関連単語集合情報を、1つのデータ構造として関連付けられた単語関連情報を複数記憶するテーブル構造に構築された関連単語記憶領域に基づき、前記取得した所定の単語に対応する前記単語関連情報を取得し、この取得した単語関連情報の関連単語集合情報に類似する関連単語集合情報を備えた単語関連情報を検索し、この検索した単語関連情報の単語情報を前記所定の単語に対する類義語として選出する際、前記所定の単語に対応する単語関連情報の関連単語集合情報を前記関連単語テーブルにおける単語関連情報の関連単語集合情報の類似状態に関する類似度情報を演算する時、前記所定の単語に対応する単語関連情報の関連単語集合情報を構成する各関連単語情報について、該関連単語情報と文字列が一致する関連単語情報を関連単語集合情報に有する、前記関連単語テーブルにおける前記単語関連情報の単語情報の数を一致単語数として、前記関連単語テーブルに複数記憶された単語関連情報の単語情報についての総単語数を前記一致単語数で除算することで重み度情報を演算し、この重み度情報に基づいて前記類似度情報を演算し、この演算した類似度情報を基づいて前記単語関連情報を検出することを特徴とする。

【0031】

この発明では、請求項1に記載の類義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。

【0032】

そして、本発明の検索プログラムは、コンピュータを、本発明に記載の類義語検索装置、または、本発明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。

【0033】

この発明では、例えば汎用のコンピュータなどを利用しインストールすることにより本発明に記載の類義語検索装置、または、本発明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できとなり、本発明の利用促進が大幅に図れる。

【0034】

また、本発明の検索プログラムは、本発明に記載の類義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。

【0035】

この発明では、例えば汎用のコンピュータなどを利用しインストールすることにより本発明に記載の類義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。

【0038】

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。

10

20

30

40

50

【発明を実施するための最良の形態】**【0039】****〔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、本発明の類義語検索装置を備え、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ます。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において類義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、関連単語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7は、類義語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。

10

【0040】

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。

20

【0041】

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。

30

【0042】

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。

【0043】

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。

40

【0044】

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2

50

20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、さらには光学的テキスト読取装置(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可能ないずれの構成が適用できる。

【0045】

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。

10

【0046】

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も記憶する。なお、端末記憶手段240としては、HD(Hard Disc)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、各種情報を記憶可能な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としてもよい。

20

【0047】

端末処理手段250は、例えばCPU(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210が接続される通信制御ポート、端末入力手段220が接続されるキー入力ポート、端末表示手段230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240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250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251と、検索要求手段252と、検索結果取得手段253と、表示制御手段としても機能する出力制御手段254と、などを備えている。

30

【0048】

キーワード設定手段251は、端末入力手段220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220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:IME)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。

40

【0049】

検索要求手段252は、キーワード設定手段251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300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200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252は、生成した検索要求情報を端末通信手段210によりネットワーク110を介してサーバ装置300へ送信する。

【0050】

50

検索結果取得手段 253 は、ネットワーク 110 を介して端末通信手段 210 によりサーバ装置 300 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 240 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。

【0051】

出力制御手段 254 は、端末入力手段 220 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 253 にて取得した検索結果情報を適宜処理し、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する。また、出力制御手段 254 は、サーバ装置 300 から出力される単語候補を、端末入力手段 220 による入力操作にて選択可能に端末表示手段 230 で複数表示させる制御をする。

10

【0052】

サーバ装置 300 は、端末装置 200 からのキーワードを含む文献情報を検索して端末装置 200 へ提供する装置で、ネットワーク 110 を介して端末装置 200 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 300 は、例えば図 3 に示すように、サーバ通信手段 310 と、サーバ入力手段 320 と、サーバ表示手段 330 と、記憶手段として機能するサーバ記憶手段 340 と、演算手段としての類義語検索装置であるサーバ処理手段 350 と、などを備えている。

【0053】

20

サーバ通信手段 310 は、ネットワーク 110 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 350 へ出力する。また、サーバ通信手段 310 は、サーバ処理手段 350 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。

【0054】

30

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、端末入力手段 220 と同様に、操作ボタンや操作つまみなどを備えた構成に限られない。

【0055】

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。

40

【0056】

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、関連単語記憶手段としての関連単語記憶領域 343 と、類義語情報記憶領域 344 と、などを備えている。

【0057】

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

50

れている。文献情報341Aは、例えば図4に示すように、本文情報341A1と、識別情報341A2と、などを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。本文情報341A1は、複数の単語にて構成された文章や文節などの例えばテキスト形式の複数の文章情報を有している。なお、文献情報341A1には、画像データ形式の図面情報などが含まれていてもよい。識別情報341A2は、文献情報341Aを特定する文献固有情報である例えばID(identification)番号などである。なお、文献情報341Aには、出願番号や出願人、著者などの書誌情報である付帯情報などをも適宜関連付けられている。

【0058】

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。

【0059】

関連単語記憶領域343は、例えば図6に示すように、関連単語テーブルを記憶、すなわち単語関連情報343Aを複数記憶するテーブル構造に構築されている。この単語関連情報343Aは、単語情報342A1と、関連単語集合情報343A2と、などを1つのデータ構造に関連付けられて構成されている。さらに、関連単語集合情報343A2は、単語情報342A1の形態素解析などによる文献からの抽出の際に、その単語に近接する位置、例えばその単語の前後それぞれ3～5単語分の範囲内に位置する他の単語に関する関連単語情報343A21を複数備えた構成で、単語情報342A1にそれぞれ関連付けられて単語関連情報343Aを構成する。ここで、図6は、説明の都合上、単語情報342A1として、「鉛筆」、「ボールペン」、「絵具」を例示し、「鉛筆」であれば、「紙」、「記入」、「筆記」、「記載」、「削る」、「訂正」、「芯」、「H B」、「消しゴム」がそれぞれ近接する位置の関連する単語である関連単語情報343A21として例示した図である。なお、関連単語情報343A21は、文献における1つの文章や文節に限らず、ある単語に対して前後の文章や文節に亘って抽出してもよく、また前後それぞれ3～5単語分に限られない。また、単語の前後に位置する他の単語の全てを関連単語情報343A21とする場合に限らず、他の単語の出現頻度を演算し、頻度が高い順から所定数を選出したものを関連単語情報343A21としたり、抽出した他の単語のうち、その単語の分類に関連する分類の単語を関連単語情報343A21としたり、キーワードが含まれる文献に対応する分類に関連する単語を関連単語情報343A21とするなどしてもよい。また、文献から抽出する場合に限らず、単語に関する説明が記載された用語辞典に関するデータベースを利用して抽出した単語を関連単語情報343A21としたり、その単語のシソーラス項目となる単語群の各単語を関連単語情報343A21としたりしてもよい。

【0060】

類義語情報記憶領域344は、例えば図7に示すように、類義語関連情報344Aを複数記憶するテーブル構造に構成されている。この類義語関連情報344Aは、単語情報342A1と、類義語集合情報344A2と、アイコン情報344A3と、などを1つのデータ構造に関連付けられて構成されている。さらに、類義語集合情報344A2は、単語情報342A1の単語に対する類義語としてサーバ処理手段350で設定された単語情報342A1を類義語に関する類義語情報344A21として設けられている。ここで、図7は、説明の都合上、単語情報342A1として「鉛筆」を例示し、この「鉛筆」に対する類義語として「万年筆」、「ボールペン」、「筆記具」、「絵具」が検出されてスコア順にリストアップした図である。なお、類義語集合情報344A2は、サーバ処理手段3

10

20

30

40

50

50にて複数設定されている場合には複数の類義語情報344A21を有したデータ構造に構成される。そして、本実施の形態では、例えば図7に示すように、類義語集合情報344A2として、類義語として設定された各類義語情報344A21に、サーバ処理手段350で演算した後述するスコアに関する情報が関連付けられ、例えばスコア順に並べられたデータ構造に構成して説明するが、単に類義語情報344A21の集合としたデータ構造とするなどしてもよい。また、アイコン情報344A3は、別途サーバ記憶手段340の記憶された各種アイコン情報を特定する識別情報である。なお、アイコン情報344A3は、識別情報に限らず、アイコンの画像データなどとしてもよい。このアイコンとしては、例えばキーワードが表示されるなど、何の類義語を一覧表示させるものであるかを認識できる表示形態で表示される。

10

【0061】

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。

【0062】

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、単語関連情報取得手段352と、関連単語検索手段353と、類義語選出手段354と、検索手段355と、報知手段356と、課金演算手段357と、決済手段358と、関連単語テーブル生成手段359と、などを備えている。なお、検索要求情報取得手段351と、単語関連情報取得手段352と、関連単語検索手段353と、類義語選出手段354と、により本発明の類義語検索装置が構成され、この類義語検索装置と検索手段355とにより本発明の情報検索装置が構成される。

20

【0063】

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所定の単語であるキーワードを取得し、単語関連情報取得手段352へ出力する。

30

【0064】

単語関連情報取得手段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認識し、サーバ記憶手段340の関連単語記憶領域343に基づいて、キーワードに対応する単語情報342A1を有する単語関連情報343Aを取得する。この取得した単語関連情報343Aは、関連単語検索手段353へ出力される。

【0065】

関連単語検索手段353は、単語関連情報取得手段352から出力されるキーワードに対応する単語関連情報343Aの関連単語集合情報343A2に類似する関連単語集合情報343A2を備えた単語関連情報343Aを検索する。この関連単語集合情報343A2の類似判断としては、各種方法が利用できる。例えば、関連単語集合情報343A2の関連単語情報343A21の各単語の文字列を一連とした文字列の編集距離やリーベンシユタイン距離などの距離を演算する方法や、各関連単語情報343A21の単語の重み付けに基づいて類似度を演算する方法などが例示できる。本実施の形態では、重み付けに基づいて類似度を演算する構成について例示する。すなわち、関連単語検索手段353は、例えば重み度情報演算手段353Aと、類似度演算手段353Bと、類義語候補選出手段353Cと、などを備えている。

40

【0066】

50

重み度情報演算手段 3 5 3 A は、単語関連情報取得手段 3 5 2 から取得したキーワードに対応する単語関連情報 3 4 3 A の関連単語集合情報 3 4 3 A 2 における各関連単語情報 3 4 3 A 2 1 と、関連単語記憶領域 3 4 3 に記憶された各単語関連情報 3 4 3 A の関連単語集合情報 3 4 3 A 2 における各関連単語情報 3 4 3 A 2 1 との同一性に関する重み度情報を演算する。なお、本実施の形態では、あらかじめ構築した関連単語記憶領域 3 4 3 に基づいて演算して説明するが、要求に基づいて関連単語記憶領域 3 4 3 を構築して演算するなどしてもよい。なお、処理の高速化や処理負荷の低減などの点では、あらかじめ関連単語記憶領域 3 4 3 を構築しておくことが好ましいことから、本実施の形態では、あらかじめ構築した構成で説明する。

【0067】

10

そして、関連単語集合情報 3 4 3 A 2 の同一性に関する重み度情報の演算として、例えばキーワードが「鉛筆」の場合、図 6 に示す「紙」、「記入」、「筆記」、「記載」、「削る」、「訂正」、「芯」、「H B」、「消しゴム」と一致する関連単語情報 3 4 3 A 2 1 を有した関連単語集合情報 3 4 3 A 2 における重み度を演算する。すなわち、キーワードの各関連単語情報 3 4 3 A 2 1 の単語の文字列を q_1, \dots, q_n (上記「鉛筆」の場合、 n は 9)、関連単語記憶領域 3 4 3 の各単語関連情報 3 4 3 A における各関連単語情報 3 4 3 A 2 1 の単語の文字列を 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。

【0068】

20

〔数1〕

$$W(q_i, d_j, q, d) = \log (\text{総単語数} / (\text{q}_i = \text{d}_j \text{ の出現する単語数})) \\ \times (\text{q}_i \text{ のキーワードの関連単語集合情報内の出現数}) \\ \times (\text{d}_j \text{ の単語の関連単語集合情報内の出現数})$$

【0069】

30

類似度演算手段 3 5 3 B は、演算した重み度に基づいて、関連単語記憶領域 3 4 3 の各単語の関連単語集合情報 3 4 3 A 2 のキーワードの関連単語集合情報 3 4 3 A 2 に対する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、キーワードの各関連単語情報 3 4 3 A 2 1 と同一の各単語の各関連単語情報 3 4 3 A 2 1 における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。

【0070】

〔数2〕

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

【0071】

なお、類似度演算手段 3 5 3 B 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_{nN} (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} の各関連単語情報 3 4 3 A 2 1 の文字列 d_1, \dots, d_t のうち、キーワードの各関連単語情報 3 4 3 A 2 1 の文字列 q_1 と一致する文字列 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関連単語情報 3 4 3 A 2 1 の文字列 $q_2 \sim q_n$ 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。

【0072】

40

そして、類義語候補選出手段 3 5 3 C は、類似度演算手段 3 5 3 B にて演算した類似度 S に基づいて、キーワードに対する類義語の候補を選出する。具体的には、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

50

単語を類義語候補単語として複数選出する。なお、選出される類似候補単語の数は、端末装置200における入力設定により検索条件として設定してもよい。これら選出された類義語候補単語の単語関連情報343Aは、類義語選出手段354へ出力される。

【0073】

類義語選出手段354は、関連単語検索手段353で選出されて出力される単語関連情報343Aから単語情報342A1を読み取り、適宜絞り込み処理にて選出した単語情報342A1をキーワードに対する類義語として選出する。また、類義語選出手段354は、選出した類義語の単語情報342A1をキーワードの単語情報342A1に関連付けて類義語関連情報344Aを生成し、類義語情報記憶領域344に記憶させる。すなわち、類義語選出手段354は、頻度認識手段354Aと、検索単語設定手段354Bと、類義語情報生成手段354Cと、を備えている。なお、類義語選出手段354は、絞り込み処理をせず、関連単語検索手段353で選出されて出力される単語関連情報343Aから単語情報342A1を読み取り、それら単語情報342A1の単語をキーワードに対する類義語として設定したり、関連単語検索手段353で選出された単語関連情報343Aの単語情報342A1を端末装置200で選択あるいは修正された単語を類義語として設定したりするなどしてもよい。10

【0074】

頻度認識手段354Aは、類義語候補単語として選出された各単語関連情報343Aから単語情報342A1をそれぞれ読み取り、さらに絞り込み処理を実施する。この絞り込み処理としては、いずれの方法が利用できるが、本実施の形態では、例えば類義語候補単語の文献に出現する出現頻度および類義語候補単語が出現する文献頻度のうちの少なくともいざれか一方を認識する処理を例示する。具体的には、頻度認識手段354Aは、単語出現一覧記憶領域342に基づいて、類義語候補単語の単語情報342A1に関連付けられた文献の識別情報341A2の数を計数する。そして、計数した識別情報341A2の数が多い順に、例えばあらかじめサーバ記憶手段340に記憶しておいたスコアを読み取って、識別情報341A2の数が多い順に、「100点」、「70点」、「50点」、...などのスコアを付与し、スコアを文献頻度としたり、全体の文献数に対する計数した文献数の割合を演算し、この演算値を文献頻度としたり、計数した数をそのまま文献頻度としたりする。20

【0075】

検索単語設定手段354Bは、頻度認識手段354Aにて文献頻度が演算されてスコア付けされた単語情報342A1のうち、文献頻度が所定の頻度より低い単語情報342A1以外を選出する。具体的には、検索単語設定手段354Bは、類義語候補単語に対応する文献頻度があらかじめ設定された閾値より極めて低いものは、関連単語検索手段353で類似度が比較的に高い場合でも、その単語を排除して他の類義語候補単語をキーワードに対する類義語として、その単語情報342A1を選出する。すなわち、文献頻度が極めて低い単語は、文献で利用されない不適切な単語あるいは誤入力された単語として判断できるので、文献検索のキーワードとして成り立たないとして除外する選別処理をする。さらに、検索単語設定手段354Bは、選出した類義語に対応する単語情報342A1を端末装置200で出力可能に送信させ、端末装置200から確認の入力操作に基づいて単語情報342A1を設定する処理をする。すなわち、検索単語設定手段354Bは、表示制御手段としての一覧表示制御手段354B1と、単語情報設定手段354B2と、を備えている。40

【0076】

一覧表示制御手段354B1は、選出した単語情報342A1に基づいて、端末表示手段230で類義語として一覧表示されるとともに、端末入力手段220における入力操作により一覧表示される類義語が選択さらには修正可能な図示しない一覧表示出力フォームを作成する。そして、一覧表示制御手段354B1は、一覧表示出力フォームをサーバ通信手段310にてネットワーク110を介して端末装置200へ送信させ、端末表示手段230に表示させる処理をする。50

【0077】

単語情報設定手段354B2は、端末装置200の端末表示手段230に表示される一覧表示出力フォームに従って、端末入力手段220の入力操作により所定の類義語が選択あるいは修正された設定入力を、ネットワーク110を介して認識する。そして、単語情報設定手段354B2は、認識した設定入力に基づいて、端末装置200で入力操作による選択あるいは修正された類義語を文献検索のためのキーワードとして設定する。

【0078】

また、類義語選出手段354の類義語情報生成手段354Cは、類義語として選出した単語情報342A1を、キーワードの単語情報342A1に関連付けて1つのデータ構造とした類義語関連情報344Aを生成し、類義語情報記憶領域344に記憶させる。この類義語関連情報344Aの生成の際、類義語情報生成手段354Cは、端末表示手段230でキーワードの類義語を一覧表示させるための端末入力手段220での入力操作可能なアイコン表示を端末表示手段230に表示させるアイコン情報344A3を関連付ける。そして、類義語情報生成手段354Cは、アイコン情報344A3をサーバ通信手段310にてネットワーク110を介して端末装置200へ送信させ、端末表示手段230にアイコンを表示させる処理をする。

10

【0079】

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、キーワードおよび類義語選出手段354で選出・設定された類義語を含む文献情報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基づき、キーワードおよび類義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を読み取って文献情報341Aを検出する。そして、検索手段355は、検出した文献情報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、文献情報341Aを検出できなかった場合、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356へ適宜出力される。

20

【0080】

報知手段356は、検索手段355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。このことにより、端末装置200は、検索結果取得手段253にて取得した検索結果情報を適宜処理し、文献情報341Aの本文情報341A1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し、検索結果が端末装置200で報知される。

30

【0081】

課金演算手段357は、検索要求情報や検索手段355にて検索し報知手段356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356により端末装置200で確認可能に送信される。

40

【0082】

決済手段358は、課金演算手段357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を介して各金融機関が管理する付加価値通信網(Value-Added Network: V

50

A N)であるいわゆる金融VAN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する。

【0083】

関連単語テーブル生成手段359は、単語関連情報343Aを生成して関連単語記憶領域343に複数記憶させて関連単語テーブルを構築する。例えば、サーバ入力手段320としてOCRにて文献が読み込まれたりキー入力にて入力されたりして設定入力された文献情報341Aや、ネットワーク110を介して外部のサーバ装置から送信された文献情報341Aなどに基づいて、文献の文章や文節を認識し、形態素解析などにより単語を抽出して単語情報342A1を生成する。これら生成した単語情報342A1のうち、既に関連単語記憶領域343に記憶されている単語情報342A1以外の各単語情報342A1について、その単語に近接する位置の他の単語をそれぞれ認識する。これら認識した他の単語の単語情報342A1を関連単語情報343A21として単語情報342A1に関連付けて単語関連情報343Aを生成し、関連単語記憶領域343に記憶させ、関連単語テーブルを構築する。

【0084】

〔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100における所定の文献情報341A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。

10

【0085】

文献情報検索システム100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200の端末入力手段220を適宜入力操作することで、サーバ装置300とネットワーク110を介して各種情報を送受信可能に端末装置200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300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230に表示させる画面表示にしたがって端末入力手段220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250のキーワード設定手段251が、端末入力手段220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S101)。

20

【0086】

この後、利用者が画面表示に従って端末入力手段220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250の検索要求手段252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300に要求する旨の検索要求情報を生成し(ステップS102)、端末通信手段210によりネットワーク110を介してサーバ装置300へ送信させる処理をする(ステップS103)。

30

【0087】

そして、ステップS103で端末装置200から送信された検索要求情報を、サーバ通信手段310でサーバ装置300が受信すると(ステップS104)、サーバ処理手段350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、サーバ処理手段350は、キーワードに対する類義語を検索する類義語検索工程を実施する。

40

【0088】

すなわち、サーバ処理手段350の単語関連情報取得手段352は、取得したキーワードを認識し、キーワードに対応する単語情報342A1を有する単語関連情報343Aを関連単語記憶領域343から検索して取得する(ステップS106)。このステップS1

50

06で取得した単語関連情報343Aに基づいて、サーバ処理手段350の関連単語検索手段353により、キーワードの単語関連情報343Aの関連単語集合情報343A2を構成する各関連単語情報343A21と、関連単語記憶領域343に記憶された各単語関連情報343Aの関連単語集合情報343A2における各関連単語情報343A21との同一性に関する重み度情報を演算する重み度演算工程を実施する(ステップS107)。具体的には、例えばキーワードが「鉛筆」の場合、図6に示す「紙」、「記入」、「筆記」、「記載」、「削る」、「訂正」、「芯」、「H B」、「消しゴム」と一致する関連単語情報343A21を有した関連単語集合情報343A2における重み度を、上述した数1に基づいて演算する。

【0089】

10

このステップS107における重み度の演算結果に基づいて、サーバ処理手段350の類似度演算手段353Bが、関連単語記憶領域343の各単語の関連単語集合情報343A2のキーワードの関連単語集合情報343A2に対する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類似度演算工程を実施する(ステップS108)。具体的には、キーワードの各関連単語情報343A21と同一の各単語の各関連単語情報343A21における重み度の総和を類似度Sとして、上述した数2に基づいて演算する。このステップS108の後、サーバ処理手段350の類義語候補選出手段353Cが、演算された類似度Sに基づいて、例えば類似度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となる単語をキーワードに対する類義語の候補である類義語候補単語として複数選出する(ステップS109)。

【0090】

20

このステップS109の後、サーバ処理手段350は、類義語選出手段354の頻度認識手段354Aにより、類義語候補単語として選出した単語関連情報343Aの単語情報342A1の文献頻度を単語出現一覧記憶領域342に基づいて演算する絞り込み工程を実施する(ステップS110)。そして、このステップS110にて演算した文献頻度に基づいて、例えば文献頻度があらかじめ設定された閾値より低いものを排除して他の類義語候補単語をキーワードに対する類義語として、その単語情報342A1を選出する(ステップS111)。

【0091】

30

そして、サーバ処理手段350は、ステップS106～S111の類義語検索工程で選出した類義語の単語情報342A1を、一覧表示させるべく端末装置200へ送信する処理をする(ステップS112)。この選出した類義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。

【0092】

このステップS112により、サーバ装置300で選出された類義語に関する一覧表示フォームを端末通信手段210で端末装置200が受信すると(ステップS113)、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を適宜制御して端末入力手段220での入力操作により類義語を選択さらには修正などが可能に画面表示させる。そして、一覧表示フォームの画面表示にしたがって所定の類義語が選択あるいは修正されると(ステップS114)、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義語をサーバ装置300へ送信させる(ステップS115)。この類義語の選択あるいは修正された単語情報342A1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S116)、サーバ処理手段350の検索単語設定手段354Bが選択あるいは修正された類義語の単語情報342A1を検索要求のキーワードとして確定する(ステップS117)。なお、例えば、検索要求時のキーワードに代えて、類義語をキーワードとして文献検索せらるるようにしたり、類義語が選択されずにキーワードのみで文献検索せたり、キーワードとともに類義語をキーワードとして文献検索せたりするなど、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、文献検索のためのキーワードが設定される。

【0093】

40

50

このステップ S 117 の後、設定された類義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、サーバ処理手段 350 の検索手段 355 は、ステップ S 104 で受信した検索要求情報に基づいて、ステップ S 117 で確定された検索要求のキーワードおよび類義語を含む文献情報 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 S 118）。すなわち、キーワードおよび類義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、検索手段 355 により、検出した文献情報 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 357 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 S 119）。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 356 により、検索手段 355 で生成した検索結果情報を、課金演算手段 357 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 S 120）。

【0094】

このステップ S 120 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 S 121）、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 S 122）、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。

【0095】

一方、ステップ S 120 の後、サーバ装置 300 は、決済手段 358 により、課金演算手段 357 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 S 123）、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。

【0096】

ここで、キーワードとして例えば「鉛筆」である場合、図 6 に示すように、「鉛筆」に関連付けられた単語群である関連単語集合情報 343A2 と、他の単語である例えば「ボールペン」および「絵具」の関連単語集合情報 343A2 との類似性を比較すると、「鉛筆」に対して「絵具」より「ボールペン」の方が類似していることがわかる。すなわち、「鉛筆」には、「絵具」より「ボールペン」の方が意味的に近い単語として判断できる。このため、類似状態に関する類似度に基づいて、図 7 に示すように、「鉛筆」に対して、意味的に近い順に類義語集合情報 344A2 が構成される。

【0097】

さらに、仮に「鉛筆」の関連単語集合情報 343A2 に対して「ボールペン」および「ボールペン」の関連単語集合情報 343A2 の類似度が同じであった場合でも、頻度認識手段 354A により単語出現一覧記憶領域 342 に基づいて類義語の候補である「ボールペン」と「ボールペン」との文献頻度を演算している。そして、「ボールペン」は意味をなすが、「ボールペン」は意味をなさず偶然に文献情報記憶領域 341 の文献中に誤入力された単語が抽出された単語である。したがって、図 5 に示すように、「ボールペン」の文献頻度は高いが、意味をなさない「ボールペン」の文献頻度は極めて低くなる。そして

10

20

30

40

50

、検索単語設定手段 354B により文献頻度の低い類義語候補単語を排除することで、より適切な類義語が設定される。すなわち、意味をなさない単語「ポールペン」を文献の検索キーワードとして設定することは目的合理性から反するため、意味をなす「ポールペン」を類義語として設定、すなわち検索キーワードとしてより適切であると優先度をおくことが認知的判断に照らしてポピュラリズム的な定義によりかなっている。このため、文献頻度をも加味することで、より適切な類義語が検出される。

【0098】

〔実施の形態の作用効果〕

上述したように、上記実施の形態では、単語に関する単語情報 342A1 にその単語に関連しその単語を特徴付ける単語に関する関連単語情報 343A21 を複数有する関連単語集合情報 343A2 を 1 つのデータ構造として関連付けた単語関連情報 343A を複数記憶するテーブル構造に関連単語記憶領域 343 を構築する。そして、端末装置 200 からネットワーク 110 を介して検索要求情報を検索要求情報取得手段 351 が取得したことを認識すると、単語関連情報取得手段 352 により、関連単語記憶領域 343 に基づいて検索要求情報に含まれるキーワードに対応する単語情報 342A1 を有した単語関連情報 343A を取得する。この取得した単語関連情報 343A の関連単語集合情報 343A2 に類似する関連単語集合情報 343A2 を備えた単語関連情報 343A を、関連単語記憶領域 343 に基づいて関連単語検索手段 353 にて検索する。検出した単語関連情報 343A の単語情報 342A1 を、類義語選出手段 354 により所定の単語に対する類義語として選出させる。この単語に関連し特徴付ける単語の集合における類似性を判断する構成により、キーワードとの特徴の類似性が判断できるので、関連単語記憶領域 343 を利用して関連単語集合情報 343A2 の類似性の判断により、単語にその単語を特徴付ける単語群である関連単語集合情報 343A2 を関連付けた簡単なテーブル構造を利用して容易に適切な類義語を検索できる。

【0099】

そして、関連単語記憶領域 343 として、複数の単語にて構成された文章に出現する単語情報 342A1 に、その単語の文章における近傍に位置する他の単語に関する関連単語情報 343A21 を複数有する関連単語集合情報 343A2 が 1 つのデータ構造として関連付けられ、文章に出現する各単語についてそれぞれ構成された複数の単語関連情報 343A を格納するテーブル構造に構築している。この文章における単語の近傍に位置する他の単語の集合は、その単語の機能や効能または作用あるいは効果、これら機能などを奏す、あるいは奏させるものなどの比較的に関連する単語であるので、他の単語の集合をその単語に関連付けた単語関連情報 343A を複数格納するテーブル構造に構築される関連単語記憶領域 343 の関連単語テーブルを利用することで、キーワードに対して関連する単語が検索されることとなる。このため、その他の単語の集合に関連付く単語を類義語として、単語にその単語に関連する単語を関連付けた簡単なテーブル構造の関連単語記憶領域 343 を利用して容易に適切な類義語の検索ができるとともに、例えば用語辞典などの言語資源を利用する必要がなく、構成の簡略化や適切な類義語検索のためのデータベースにおける保守管理の容易性などが得られる。さらに、あらかじめ単語関連情報 343A を複数記憶するテーブル構造の関連単語テーブルをサーバ記憶手段 340 に関連単語記憶領域 343 として構築している。このため、あらかじめ構築した関連単語テーブルを利用して単語に関連する単語群の類似性を判断すればよく、処理の高速化が容易に得られる。

【0100】

また、文章に関する文章情報を複数記憶する文献情報記憶領域 341 に基づいて単語関連情報 343A を複数格納するテーブル構造の関連単語テーブルを構築している。このため、検索対象の文章の集合となる文献を蓄積するデータベースである文献情報記憶領域 341 を利用するので、例えば用語辞典などの言語資源を利用する必要がなく、また例えば関連単語記憶領域 343 としてあらかじめ構築せずに類義語の検索の際に関連単語テーブルを構築することで関連単語テーブルを記憶しておく領域を削減でき、さらなる構成の簡

10

20

30

40

50

略化が図れる。また、文章を複数含む文献毎の集合として文章情報を複数有する文献情報 3 4 1 A を複数記憶するテーブル構造に構築された文献情報記憶領域 3 4 1 を利用している。このため、文献毎に文章情報がまとまった文献情報を複数記憶するテーブル構造の文献情報記憶領域 3 4 1 を利用して、文献で利用される積雪な類義語が検索され、文章情報毎に複数格納するテーブル構造に比してデータベースの構築や保守管理などが容易にできるとともに、取得したキーワードや選出した類義語に基づく文献の検索も容易にできる。さらに、文献情報の文章情報の文章における近傍に位置する他の単語を関連単語情報 3 4 3 A 2 1 として関連付けている。このため、キーワードを含む文献はそのキーワードに対する類義語の存在頻度が比較的に高いことから、全ての文章情報から検索する必要がなく、一部の文献から効率よく異義語が検索される。

10

【0101】

さらに、単語関連情報 3 4 3 A の単語情報 3 4 2 A 1 として、形態素解析などにより文章から抽出した単語の情報としている。このため、既存の形態素解析などにより簡単に単語関連情報 3 4 3 A を生成でき、簡単な処理で関連単語テーブルが構築され、類義語を検索するための構成や新たな単語関連情報 3 4 3 A の追加や修正などの保守管理なども容易にできる。

【0102】

そして、関連単語検索手段 3 5 3 により、キーワードに対応する単語関連情報 3 4 3 A の関連単語集合情報 3 4 3 A 2 に対して、関連単語記憶領域 3 4 3 の関連単語テーブルにおける各単語関連情報 3 4 3 A の関連単語集合情報 3 4 3 A 2 の類似状態に関する類似度情報を演算する。そして、関連単語検索手段 3 5 3 により、類似度情報に基づいて単語関連情報 3 4 3 A を検出している。このため、簡単な構造の関連単語テーブルを利用した簡単な演算により、関連する単語群の類似状態に応じた類義語の検索が得られる。

20

【0103】

さらに、関連単語検索手段 3 5 3 における類似度情報の演算として、キーワードに対応する単語関連情報 3 4 3 A の関連単語集合情報 3 4 3 A 2 を構成する複数の関連単語情報 3 4 3 A 2 1 と、関連単語記憶領域 3 4 3 の関連単語テーブルにおける各単語関連情報 3 4 3 A の関連単語集合情報 3 4 3 A 2 を構成する複数の関連単語情報 3 4 3 A 2 1 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し、この重み度情報に基づいて関連単語検索手段 3 5 3 における類似度情報の演算を実施している。このため、関連単語テーブルを利用した類似状態に応じた類義語の検索のための類似度情報の演算が容易にでき、処理負荷を低減でき、処理の高速化が容易に図れる。

30

【0104】

そして、類義語選出手段 3 5 4 で類義語として選出した単語情報 3 4 2 A 1 を、類義語情報生成手段 3 5 4 C により、キーワードに対応する単語情報 3 4 2 A 1 に関連付けて 1 つのデータ構造とした類義語情報を生成してサーバ記憶手段 3 4 0 に記憶させ、類義語情報を複数記憶するテーブル構造の類義語情報記憶領域 3 4 4 を構築している。このため、キーワードに対し 1 度演算して設定された類義語が関連付けられて蓄積されるので、類義語情報記憶領域 3 4 4 を参照することにより、キーワードに対する類義語を演算することなく検出可能となり、類義語の検出が迅速かつ容易に得られる。

40

【0105】

さらに、類義語選出手段 3 5 4 では、頻度認識手段 3 5 4 A により、単語出現一覧記憶領域 3 4 2 を利用して類似候補単語の文献頻度を認識させている。このため、構成が簡単な単語出現一覧記憶領域 3 4 2 で類義語候補単語の文献頻度を一義的に認識することができ、簡単な構成で処理の高速化が容易に得られ、キーワードに対する類義語の選出が迅速に得られる。さらに、類義語候補単語のうち、誤入力や不適切な単語などは関連単語検索手段 3 5 3 にて仮に類似度が高い単語でも排除しているため、より適切な類義語の検出、さらには適切な文献情報 3 4 1 A の検索が得られる。

【0106】

そして、類義語選出手段 3 5 4 により、選出した類義語を端末入力手段 2 2 0 の入力操

50

作により選択および修正のうちの少なくともいずれか一方を可能に端末表示手段 230 に表示させ、入力操作に応じて選択あるいは修正された単語情報 342A1 に基づいて検索手段 355 にて文献検索を実行している。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るとともに、利用者の所望とする適切な文献検索が得られる。

【0107】

さらに、キーワードにより文献検索する構成におけるキーワードに対する類義語を検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。

【0108】

また、単語情報 342A1 にその単語が出現する文献情報 341A の識別情報 341A 2 を関連付けた単語出現一覧記憶領域 342 を構築している。このため、この単語出現一覧記憶領域 342 を類義語候補単語の絞り込み処理のための文献頻度の演算と、文献検索との双方に利用でき、単語出現一覧テーブルの共用が得られ、より構成の簡略化が得られるとともにデータの保守管理が容易となる。

10

【0109】

そして、ネットワーク 110 を介して接続する端末装置 200 からの検索要求によりサーバ装置 300 で類義語さらには文献を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 200 の構成でも、漏れのない適切な類義語や文献の検索が容易にでき、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 300 にて統括的に容量の大きな文献情報 341A や各種テーブルなどを管理でき、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小さく記憶容量も比較的に少ないものを用いても容易に類義語さらには文献の検索ができ、利用の拡大が容易に図れる。

20

【0110】

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。

【0111】

〔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。

30

【0112】

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 341 を有する記録媒体から文献情報 341A を適宜検索する装置構成としたり、文献情報記憶領域 341 をネットワーク 110 を介して読み取る構成としたり、端末装置 200 に類義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、類義語を選出する手段を設け、サーバ装置 300 では検索処理するのみの構成としたり、類義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 300 で実施するシステム構成としたり、サーバ記憶手段 340 の構成をサーバ装置 300 と切り離してネットワーク 110 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。

40

【0113】

また、文献を検索するためのキーワードに対する類義語を検索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に適用、さらには単に類義語を検索する専用装置など、独立した構成とするなどしてもよい。

【0114】

そして、類義語の検索としては、上述したように、関連単語集合情報 343A2 を構成する関連単語情報 343A21 の重み度を演算する構成に限らず、編集距離などの距離を演算する構成やルールテーブルを利用する構成など、いずれの方法で関連単語集合情報 343A2 の類似性を判断してもよい。

【0115】

50

また、関連単語テーブルを構成する単語関連情報343Aの単語情報342A1や関連単語情報343A21を文献から抽出して説明したが、上述したように、文献に限らず、例えば所定の単語に対して用語辞典の説明文から関連単語を抽出したり、シソーラス項目となる単語群を構成する単語を関連単語として抽出したりしてもよい。さらには、文献に限らず、文章から適宜抽出した単語の近傍に位置する他の単語を関連単語として関連付けるなどしてもよい。また、文献としては、産業財産権における公報などに限らず、新聞や雑誌、各種書籍など、各種刊行物をも対象とすることができます。さらには、形態素解析により抽出する構成に限らず、あらかじめキー操作により設定入力にて関連単語テーブルを構築してもよく、あらかじめ記憶領域に構築しておく構成に限らず、類義語の検索処理時に関連単語を抽出して関連単語集合情報343A2を生成して類似性をその都度判断するなどしてもよい。

10

【0116】

そして、キーワードに対応する単語関連情報343Aの関連単語集合情報343A2と類似判断する単語関連情報343Aとして、関連単語記憶領域343の他の全ての単語関連情報343Aとしたが、例えば単語関連情報343Aに分類などの情報を関連付け、同一の分類内の単語関連情報343Aとの類似判断により検索するなどしてもよい。この分類などの情報としては、例えば、「鉛筆」であれば「筆記用具」などの単語における分類に関する単語分類情報、単語関連情報343Aの単語情報342A1を抽出した文献情報341Aに関連付けられた付帯情報である「特許」や「実用新案」、あるいは「公開」や「登録」などの種別、文献の分類に関する国際特許分類などの文献分類情報、発明者や出願人、著者名、出願番号などの書誌情報などが例示できる。これらの構成により、類義語の存在が比較的に高い分類などの情報が同一の範囲内でのみ類似性を判断するので、処理負荷の低減や処理の高速化が得られ、より簡単な構成で迅速な類義語の検索ができる。

20

【0117】

また、類似性の判断として関連単語情報343A21の重み度に基づいて類似度を演算してスコア付けして説明したが、この方法に限られない。また、スコア付けしなくてもよい。

【0118】

さらに、関連単語検索手段353にて検出した後にさらに文献頻度を認識して絞り込み処理して説明したが、上述したように、絞り込み処理を実施せず、検出した類義語候補単語を類義語として設定したり、端末装置200に類義語候補を選択や修正可能に表示させて報知し、選択や修正に応じて類義語候補単語の絞り込み処理を実施して類義語を設定したりしてもよい。また、絞り込み処理として文献頻度に限らず、その単語の文献における出現頻度に基づいて絞り込み処理をしてよい。さらには、関連単語検索手段353により検索した単語関連情報343Aの関連単語集合情報343A2や、類義語選出手段354により選出した単語情報342A1に対応する単語関連情報343Aの関連単語集合情報343A2に類似する関連単語集合情報343A2を備えた単語関連情報343Aをさらに検索させてもよい。そしてさらには、検出あるいは選出した複数の単語関連情報343Aの関連単語集合情報343A2をそれぞれ構成する関連単語情報343A21を備えた関連単語集合情報343A2を生成し、この生成した単語関連集合情報343A2に類似する関連単語集合情報343A2を有した単語関連情報343Aを選出するなどしてもよい。

30

【0119】

また、類義語情報を生成して複数記憶するテーブル構造の類義語情報記憶領域344を構築する構成について説明したが、類義語情報を生成しなくてもよい。また、類義語情報を生成する場合でも、アイコン表示をしなくてもよい。なお、このアイコン表示の際、例えばその類義語の利用頻度が高くなるに従って表示形態を変更してもよい。さらには、アイコンの操作により表示される類義語の集合として、類義語の削除や修正、追加などが可能としてもよく、これらの削除や修正、追加の内容を類義語情報記憶領域344に反映させたり、端末装置200のみに記憶させておいたりしてもよい。

40

50

【0120】

そして、端末装置200に選出した類義語を選択や修正可能に表示させて報知して説明したが、報知することなく選出した類義語で文献検索を実施してもよい。

【0121】

さらに、キーワードおよび類義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に報知する際に類義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、類義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義語を検索して文献検索を実施し、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、類義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

【0122】

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みに限ることなく、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】**【0123】**

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ます。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における類義語の検索や単に類義語を検索する専用装置としても利用できる。

20

【図面の簡単な説明】**【0124】**

【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。

30

【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。

【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として示す模式図である。

【図6】前記一実施の形態における関連単語記憶領域のテーブル構造を概念的な関連単語テーブルとして示す模式図である。

【図7】前記一実施の形態における類義語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。

40

【符号の説明】**【0125】**

100...情報検索装置としての文献情報検索システム

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...文章記憶手段としての文献情報記憶領域

341A...文献情報

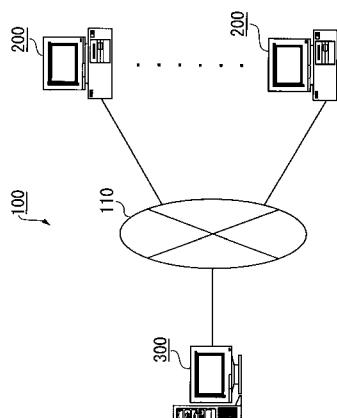
342A1...単語情報

50

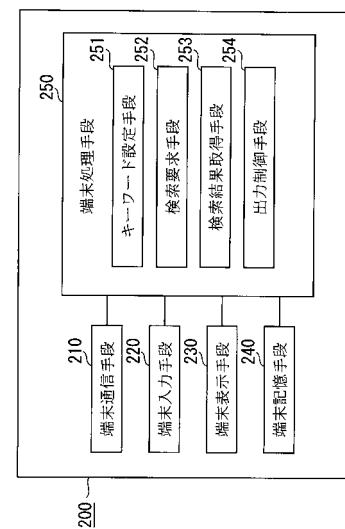
3 4 3 ... 関連単語テーブルを構築する関連単語記憶領域
 3 4 3 A ... 単語関連情報
 3 4 3 A 2 ... 関連単語集合情報
 3 4 3 A 2 1 ... 関連単語情報
 3 5 0 ... 演算手段としての類義語検索装置を構成するサーバ処理手段
 3 5 1 ... 単語情報取得手段としても機能する検索要求情報取得手段
 3 5 2 ... 単語関連情報取得手段
 3 5 3 ... 関連単語検索手段
 3 5 3 A ... 重み度演算手段
 3 5 3 B ... 類似度演算手段
 3 5 4 ... 類義語選出手段
 3 5 4 B 1 ... 表示制御手段としての一覧表示制御手段
 3 5 4 B 2 ... 単語関連情報設定手段としての単語情報設定手段
 3 5 4 C ... 類義語情報生成手段
 3 5 5 ... 検索手段

10

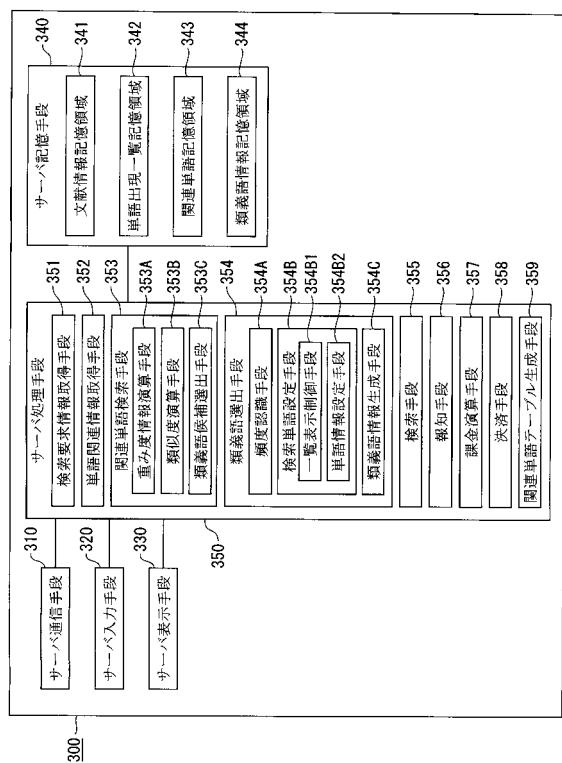
【図1】



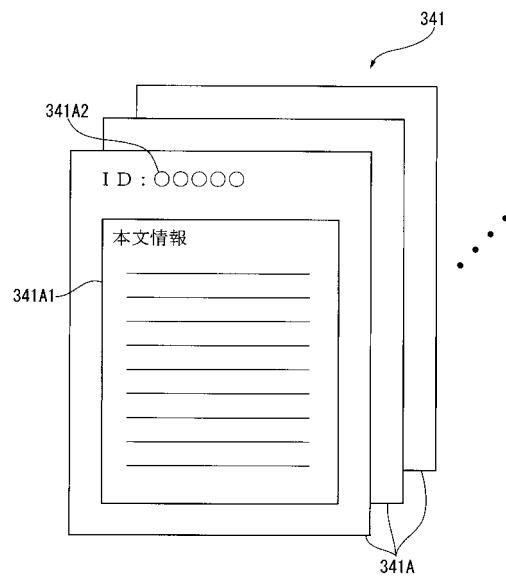
【図2】



【 义 3 】



【 図 4 】



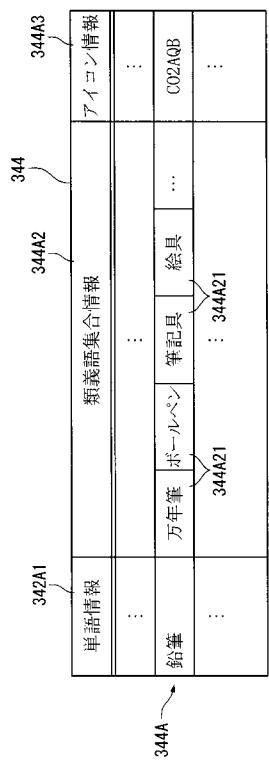
【 义 5 】

		342A1	341A2	341A2	
単語一覧		出現文献の識別番号			
342A	A A B A	○○○△○	○○△×○	
	：	
	ボールペン	○○△△×	△○××○	
	：	
	ボールペン	×○△○○			
	：	

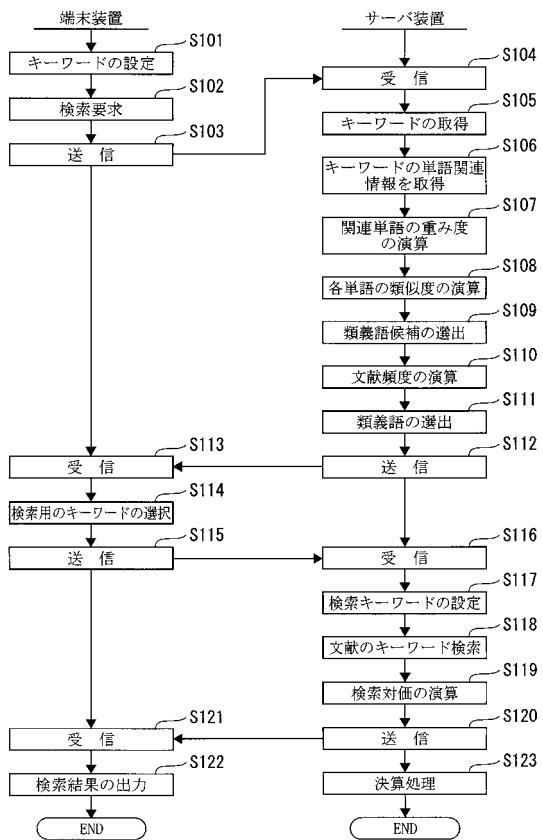
【図6】

単語情報		関連単語集合情報	
:	343A21	343A21 :	343A21
鉛筆	紙、記入、筆記、記載、削る、訂正、芯、H B、消しゴム	343A21	
ボールペン	紙、記入、筆記、記載、ペン、インク、ブルー、万年筆	343A21	
絵具	紙、キャンバス、混ぜる、塗る、描写、油、赤、青、原色、筆	343A21	343A21

【図7】



【図8】



フロントページの続き

審査官 波内 みさ

(56)参考文献 特開2000-222427(JP, A)
特開平05-028199(JP, A)
特開2003-256447(JP, A)

(58)調査した分野(Int.Cl., DB名)

G 06 F 17 / 30

G 06 F 17 / 28