

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3921837号
(P3921837)

(45) 発行日 平成19年5月30日(2007.5.30)

(24) 登録日 平成19年3月2日(2007.3.2)

(51) Int. Cl.

G06F 17/30 (2006.01)

F I

G06F 17/30 210D

G06F 17/30 170A

請求項の数 9 (全 21 頁)

(21) 出願番号	特願平10-276313	(73) 特許権者	000005496
(22) 出願日	平成10年9月30日(1998.9.30)		富士ゼロックス株式会社
(65) 公開番号	特開2000-112949(P2000-112949A)		東京都港区赤坂九丁目7番3号
(43) 公開日	平成12年4月21日(2000.4.21)	(74) 代理人	100092152
審査請求日	平成15年10月20日(2003.10.20)		弁理士 服部 毅巖
		(72) 発明者	福島 裕人
			神奈川県足柄上郡中井町境430 グリー
			ンテクなかい 富士ゼロックス株式会社内
		(72) 発明者	桂林 浩
			神奈川県足柄上郡中井町境430 グリー
			ンテクなかい 富士ゼロックス株式会社内
		審査官	辻本 泰隆
最終頁に続く			

(54) 【発明の名称】 情報判別支援装置、情報判別支援プログラムを記録した記録媒体及び情報判別支援方法

(57) 【特許請求の範囲】

【請求項1】

情報の内容によって複数の情報群を分類することを支援する情報判別支援装置において

前記情報群内の各情報を評価単位に分割する情報分割手段と、
各評価単位に対して形態素解析を行うことで語彙を抽出し、抽出された語彙からなるキーワード、連続した固有名詞または普通名詞の語彙を結合することにより得られる結合キーワード、および1文章中の普通名詞または固有名詞の語彙と、その文章中の動詞の直前に存在する普通名詞もしくはサ変名詞の語彙とのペアからなるキーワードペアを含むキーワード群を生成し、すべての評価単位から生成されたキーワード群に含まれるキーワード、結合キーワード、およびキーワードペアを要素とする評価ベクトル空間を生成し、前記評価ベクトル空間の要素に対応するキーワード、結合キーワード、およびキーワードペアが各評価単位のキーワード群に含まれるか否かに基づいて、各評価単位の記述内容の特徴を示す評価ベクトルを生成する評価基準数値化手段と、

評価単位同士の評価ベクトルを比較することで評価単位間の類似度を求め、類似する評価単位双方に対して同じ分類番号を付与する類似情報分類手段と、

同じ分類番号が付与された評価単位同士を視覚的に判別できるように各情報内の評価単位を表示する情報表示手段と、

を有することを特徴とする情報判別支援装置。

【請求項2】

10

20

前記情報分割手段は、前記情報群内の各情報から少なくともタイトルと見出しとを削除し、削除処理後の各情報を分割することを特徴とする請求項 1 記載の情報判別支援装置。

【請求項 3】

前記評価基準数値化手段は、キーワード、結合キーワード、およびキーワードペアに対して、構成する語彙の数に応じた重み付けを行い、キーワード、結合キーワード、およびキーワードペアの重みの値を、各評価単位の前記評価ベクトルにおける対応する要素に付与することを特徴とする請求項 1 記載の情報判別支援装置。

【請求項 4】

前記類似情報分類手段は、各情報内の分類番号の同じ評価単位を集め、分類番号の同じ他の評価単位との間の類似度の合計値に応じた順番で各評価単位を並べることを特徴とする請求項 1 記載の情報判別支援装置。

10

【請求項 5】

前記評価基準数値化手段は、生成したキーワード群からキーワード中の動詞と 1 文字の語彙からなるキーワードとを削除することを特徴とする請求項 1 記載の情報判別支援装置。

【請求項 6】

前記評価基準数値化手段は、各評価単位のキーワード群から、一定の割合以上のキーワード群に存在するキーワードを削除することを特徴とする請求項 1 記載の情報判別支援装置。

【請求項 7】

前記評価基準数値化手段は、各評価単位のキーワード群から、一定の割合以下のキーワード群にしか存在しないキーワードを削除することを特徴とする請求項 1 記載の情報判別支援装置。

20

【請求項 8】

情報の内容によって情報群を分類することを支援する情報判別支援プログラムを記録したコンピュータ読み取り可能な記録媒体において、

前記情報群内の各情報を評価単位に分割する情報分割手段、

各評価単位に対して形態素解析を行うことで語彙を抽出し、抽出された語彙からなるキーワード、連続した固有名詞または普通名詞の語彙を結合することにより得られる結合キーワード、および 1 文章中の普通名詞または固有名詞の語彙と、その文章中の動詞の直前に存在する普通名詞もしくはサ変名詞の語彙とのペアからなるキーワードペアを含むキーワード群を生成し、すべての評価単位から生成されたキーワード群に含まれるキーワード、結合キーワード、およびキーワードペアを要素とする評価ベクトル空間を生成し、前記評価ベクトル空間の要素に対応するキーワード、結合キーワード、およびキーワードペアが各評価単位のキーワード群に含まれるか否かに基づいて、各評価単位の記述内容の特徴を示す評価ベクトルを生成する評価基準数値化手段、

30

評価単位同士の評価ベクトルを比較することで評価単位間の類似度を求め、類似する評価単位双方に対して同じ分類番号を付与する類似情報分類手段、

同じ分類番号が付与された評価単位同士を視覚的に判別できるように各情報内の評価単位を表示する情報表示手段、

40

としてコンピュータを機能させることを特徴とする情報判別支援プログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項 9】

情報の内容によって複数の情報群を分類することをコンピュータで支援する情報判別支援方法において、

情報分割手段が、前記情報群内の各情報を評価単位に分割し、

評価基準数値化手段が、各評価単位に対して形態素解析を行うことで語彙を抽出し、抽出された語彙からなるキーワード、連続した固有名詞または普通名詞の語彙を結合することにより得られる結合キーワード、および 1 文章中の普通名詞または固有名詞の語彙と、その文章中の動詞の直前に存在する普通名詞もしくはサ変名詞の語彙とのペアからなるキ

50

ーワードペアを含むキーワード群を生成し、すべての評価単位から生成されたキーワード群に含まれるキーワード、結合キーワード、およびキーワードペアを要素とする評価ベクトル空間を生成し、前記評価ベクトル空間の要素に対応するキーワード、結合キーワード、およびキーワードペアが各評価単位のキーワード群に含まれるか否かに基づいて、各評価単位の記述内容の特徴を示す評価ベクトルを生成し、

類似情報分類手段が、評価単位同士の評価ベクトルを比較することで評価単位間の類似度を求め、類似する評価単位双方に対して同じ分類番号を付与し、

情報表示手段が、同じ分類番号が付与された評価単位同士を視覚的に判別できるように各情報内の評価単位を表示する、

ことを特徴とする情報判別支援方法。

10

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は収集した情報を分類するための情報判別支援装置、情報判別支援プログラムを記録した記録媒体及び情報判別支援方法に関し、特に収集した情報を取捨選択する行為を支援するための情報判別支援装置、情報判別支援プログラムを記録した記録媒体及び情報判別支援方法に関する。

【0002】

【従来の技術】

近年、インターネットなどのコンピュータネットワークの普及と様々な情報データベースサービスにより、情報検索サービスや電子メール、電子新聞、ネットニュースなどから、多種多様の電子化された情報を得ることが可能になっている。このため、各種情報サービスから類似した情報、もしくは関連のない情報を含めて多量の情報が届けられ、各情報の判読に忙殺されるという事態が発生しつつある。このような状況下において、氾濫する情報から各個人に有用な情報のみを選択、分類、もしくはフィルタリングして、情報の判読にかかる負荷を軽減させる技術が存在する。

20

【0003】

このような情報の取捨選択を支援するための従来技術としては、例えば、特開平5 - 266087号公報の「全文検索装置」がある。この全文検索装置では、検索対象となるテキストを、評価単位と評価単位の集まった抽出単位とに分け、評価単位をユーザが入力したキーワードの出現率で評価する。そして、各評価単位のキーワードの出現率から抽出単位の評価を行う。これにより、例えば、キーワードの出現率の高い文章が集まっている段落を選択してユーザに提示することで、ユーザは文書中の関連ある部分のみを見ることができる。

30

【0004】

また、特開平9 - 6799号公報の「文書分類装置及び文書検索装置」では、辞書に登録された単語の文書中の出現回数をその単語の特徴ベクトルとし、出現する単語の特徴ベクトルの要素を足して文書の特徴ベクトルを生成する。この文書の特徴ベクトルを用いて自動的に文書を分類し、ユーザに提示する。

【0005】

40

また、特開平9 - 44514号公報の「関連情報判定方法及び装置」では、検索結果から部分情報として、要約やアブストラクト、見出しなどを取り出し、これら部分情報間において共通の文字または文字列がどれくらい存在するかを閾値と比較して、検索結果間の関連性を求め、ユーザに提示する。これにより、検索結果の理解に関して、ユーザの負担を軽減する。

【0006】

これらの他にも、文書管理ツールや要約作成ツールにおいて、文書の関連性を評価し、関連ある文書などを集めて表示すると共に、代表となる文書や要約文を提示し、その概要を比較させるものが存在する。

【0007】

50

【発明が解決しようとする課題】

しかし、上記の従来技術には、以下のような問題点がある。

特開平5-266087号公報の「全文検索装置」では、ユーザは、装置が抽出した関連部分のみを見ることができるが、検索結果は分類されておらず、装置が抽出した個々の検索結果を全て見ることになる。すると、同じものでも複数回見なければならず、検索結果間の関係は、関連部分を見ながらユーザが判断することになる。このように、関連部分を持つ情報の取捨選択の指針がない。

【0008】

特開平9-6799号公報の「文書分類装置及び文書検索装置」では、文書全体の文書特徴ベクトルを用いて分類するので、情報の部分的な特徴が全体の中に埋もれてしまい、関連する部分はユーザが探し出さなければならない。

10

【0009】

特開平9-44514号公報の「関連情報判定方法及び装置」では、要約や見出しを用いて部分情報を比較しているが、要約や見出しに含まれる情報によって全て表現されるわけではなく、見出しなどには、注意を引くための誇張などが含まれ、正確に内容を反映していない場合がある。

【0010】

また、要約や重要文を抽出する記述では、情報中の代表的な部分のみを抽出するため、一部分に埋もれた関連情報に関する内容が要約に現れるとは限らない。逆に、抽出された部分が類似していたとしても、その情報のうちどれくらいの部分が関連するものなのか判らない。さらに、その要約や重要文が、文書のどの部分を示しているか判らないため、要約を見て関連ありと判断しても、その関連情報を得るためには、関連情報を含む文書を読まなければならない。文書中から必要な情報を抽出する作業は、人に委ねられる。

20

【0011】

このように、検索及び分類の結果、関連度の高い情報が集められ、もしくは分類されてユーザに提示されたとしても、それはシステムが判断した結果であり、ユーザはその結果をもとに再度情報の中身を判読し、各情報の必要性の有無を判別し取捨選択を行う必要がある。ところが、従来技術では、提示された情報のどの部分を読めば必要性の有無を判断できるのかが判らなかった。そのため、ユーザは不必要な文書の内容を大量に読まされることになり、非効率的であった。

30

【0012】

本発明はこのような点に鑑みなされたものであり、情報群の取捨選択の指針となる情報を提示できる情報判別支援装置及び情報判別支援方法を提供することを目的とする。

【0013】**【課題を解決するための手段】**

本発明では上記課題を解決するために、情報の内容によって複数の情報群を分類することを支援する情報判別支援装置において、前記情報群内の各情報を評価単位に分割する情報分割手段と、各評価単位に対して形態素解析を行うことで語彙を抽出し、抽出された語彙からなるキーワード、連続した固有名詞または普通名詞の語彙を結合することにより得られる結合キーワード、および1文章中の普通名詞または固有名詞の語彙とその文章中の動詞の直前に存在する普通名詞もしくはサ変名詞の語彙とのペアからなるキーワードペアを含むキーワード群を生成し、すべての評価単位から生成されたキーワード群に含まれるキーワード、結合キーワード、およびキーワードペアを要素とする評価ベクトル空間を生成し、前記評価ベクトル空間の要素に対応するキーワード、結合キーワード、およびキーワードペアが各評価単位のキーワード群に含まれるか否かに基づいて、各評価単位の記述内容の特徴を示す評価ベクトルを生成する評価基準数値化手段と、評価単位同士の評価ベクトルを比較することで評価単位間の類似度を求め、類似する評価単位双方に対して同じ分類番号を付与する類似情報分類手段と、同じ分類番号が付与された評価単位同士を視覚的に判別できるように各情報内の評価単位を表示する情報表示手段と、を有することを特徴とする情報判別支援装置が提供される。

40

50

【 0 0 1 4 】

このような情報判別支援装置によれば、情報群が与えられると、情報分割手段により、各情報が評価単位に分割される。すると、評価基準数値化手段により、各評価単位に含まれる語彙に基づいてキーワード群が生成され、評価単位毎のキーワード群を解析することで各評価単位の特徴を示す評価ベクトルが生成される。さらに、類似情報分類手段により、評価単位同士の評価ベクトルが比較され、評価単位間の類似度が求められる。そして、情報表示手段により、同じ分類番号が付与された評価単位同士を視覚的に判別できるように各情報内の評価単位が表示される。

【 0 0 1 5 】

また、上記課題を解決するために、情報の内容によって情報群を分類することを支援する情報判別支援プログラムを記録したコンピュータ読み取り可能な記録媒体において、前記情報群内の各情報を評価単位に分割する情報分割手段、各評価単位に対して形態素解析を行うことで語彙を抽出し、抽出された語彙からなるキーワード、連続した固有名詞または普通名詞の語彙を結合することにより得られる結合キーワード、および1文章中の普通名詞または固有名詞の語彙とその文章中の動詞の直前に存在する普通名詞もしくはサ変名詞の語彙とのペアからなるキーワードペアを含むキーワード群を生成し、すべての評価単位から生成されたキーワード群に含まれるキーワード、結合キーワード、およびキーワードペアを要素とする評価ベクトル空間を生成し、前記評価ベクトル空間の要素に対応するキーワード、結合キーワード、およびキーワードペアが各評価単位のキーワード群に含まれるか否かに基づいて、各評価単位の記述内容の特徴を示す評価ベクトルを生成する評価基準数値化手段、評価単位同士の評価ベクトルを比較することで評価単位間の類似度を求め、類似する評価単位双方に対して同じ分類番号を付与する類似情報分類手段、同じ分類番号が付与された評価単位同士を視覚的に判別できるように各情報内の評価単位を表示する情報表示手段、としてコンピュータを機能させることを特徴とする情報判別支援プログラムを記録したコンピュータ読み取り可能な記録媒体が提供される。

【 0 0 1 6 】

このような記録媒体に記録された情報判別支援プログラムをコンピュータに実行させれば、上記本発明に係る情報判別支援装置に必要な機能がコンピュータ上に実現される。

【 0 0 1 7 】

【 発明の実施の形態 】

以下、本発明の実施の形態を図面を参照して説明する。

図1は、本発明の原理構成図である。本発明の情報判別支援装置は、情報分割手段1、評価基準数値化手段2及び類似情報分類手段3からなる。

【 0 0 1 8 】

情報分割手段1は、情報群内の各情報4に形態素解析を行い、評価単位4aに分割する。評価基準数値化手段2は、各評価単位4aに含まれる語彙に基づいてキーワード群を生成し、評価単位毎のキーワード群を解析することで情報群全体の内容における各評価単位の記述内容の特徴を数値化する。その結果、評価単位毎の評価ベクトル4bが得られる。

【 0 0 1 9 】

類似情報分類手段3は、評価単位同士の評価ベクトル4bを比較することで評価単位間の類似度を求め、さらに各情報中及び情報間の各評価単位の類似度に基づいて複数の情報間の類似度を算出する。そして、互いに類似する情報同士の集合に分類する。これにより、類似情報同士の集まりである類似情報群5が複数生成される。

【 0 0 2 0 】

このようにして類似情報群5に分類された情報を、情報中の類似する評価単位を選択して比較提示することで、ユーザによる情報の類似性・関連性の判断を容易にし、有用な情報の取捨選択の手間を軽減することができる。

【 0 0 2 1 】

次に、本発明の情報判別支援装置により情報を分類し、分類結果をユーザに提示するための実施の形態を説明する。

10

20

30

40

50

図 2 は、本発明の実施の形態に係る情報判別支援装置の構成を示すブロック図である。本実施の形態に係る情報判別支援装置は、以下の要素で構成される。

【 0 0 2 2 】

情報入力部 1 0 は、類似判別を行うべき情報、例えば検索結果として得られた情報群を入力する。情報格納部 2 0 は、入力された情報群や類似度判別を行った結果を格納する。情報表示部 3 0 は、類似判別結果などを画面上に表示する。情報分割部 4 0 は、情報を評価単位に分割する。評価基準数値化部 5 0 は、全情報中の各評価単位の内容の位置づけを数値化して、評価ベクトルを生成する。類似情報分類部 6 0 は、各情報中及び情報間の各評価単位の類似度と、情報中の類似する評価単位の分布と、情報中の類似する評価単位の含有度を用いて複数の情報間の類似度を算出し類似する情報を分類する。

10

【 0 0 2 3 】

このような情報判別支援装置に情報群を与えると、情報入力部 1 0 により、与えられた情報群が情報格納部 2 0 に格納される。格納された各情報は、情報分割部 4 0 により評価単位に分割される。すると、評価基準数値化部 5 0 により、評価単位毎の評価ベクトルが求められる。次に、類似情報分類部 6 0 によって、各情報中及び情報間の各評価単位の類似度と、情報中の類似する評価単位の分布と、情報中の類似する評価単位の含有度を用いて複数の情報間の類似度が算出され、類似する情報が分類される。情報の分類結果は、情報格納部 2 0 に格納されると共に、情報表示部 3 0 によって画面上に表示される。

【 0 0 2 4 】

次に、情報判別支援装置の主な構成要素の詳細について説明する。なお、本装置においては、類似判別のための情報としてテキストを用いる。ここでテキストとは、コード化された文字情報をいい、報告書、特許明細書、議事録などの文書、電子メール、電子会議室に貼り付けられた意見、ホームページなどインターネット上に流れるテキスト情報など、電子化されたテキスト情報一般を含む。

20

【 0 0 2 5 】

まず、情報分割部 4 0 の詳細について説明する。

図 3 は、情報分割部の内部構成を示す図である。情報分割部 4 0 は、レイアウト判別部 4 1、タイトル削除部 4 2、及び評価単位生成部 4 3 で構成される。レイアウト判別部 4 1 は、入力された情報からテキスト部分を認識する。タイトル削除部 4 2 は、情報中のタイトルや見出し、著者名、出典などの部分を削除する。評価単位生成部 4 3 は、抽出された

30

【 0 0 2 6 】

図 4 は、情報判別支援装置の処理の流れを示すフローチャートである。この処理は、対象となる情報群が情報入力部 1 0 によって入力されたときに開始される。この処理をステップ番号に沿って説明する。

[S 1] レイアウト判別部 4 1 が、入力された情報をテキスト部分とそれ以外の部分に分割する。

[S 2] レイアウト判別部 4 1 が、写真や図の部分を取り除き、テキスト部分のみを抽出する。そして、抽出したテキスト部分をタイトル削除部 4 2 に渡す。

[S 3] タイトル削除部 4 2 が、レイアウト情報とテキストのフォントやサイズ、句読点の有無などから、本文であるか、見出しやタイトルであるかを判断し、タイトルや見出しの部分を削除する。このとき、著者名や出典、リファレンスなどの情報も削除する。

40

[S 4] タイトル削除部 4 2 が、テキスト部分のみになった情報を、評価単位生成部 4 3 に送る。

[S 5] 評価単位生成部 4 3 が、テキスト情報を章や段落などの区切りを考慮しながら、評価単位に分割する。例えば、ユーザが指定した評価単位の文字数を 2 0 0 文字程度とすると、2 0 0 文字前後において、空白行などレイアウト上の区切りがあればその部分で、なければ、改行を示すリターンを検出し、それもなければ読点を検出して情報を分割する。なお、図や表に付随する説明文などは、それらを一塊にして処理し、文字数が評価単位文字数より多い場合は分割を行う。

50

【S6】評価単位生成部43が、生成された評価単位を評価基準数値化部50へ送る。

【0027】

ここで、具体例を用いて、情報分割部40の処理内容を説明する。

図5は、本装置において処理の対象となる電子化された文書例を示す図である。図に示す文書100には、タイトル110と著者名120が記載されている。その下には、左半分に章題130と文章140とが記載され、右半分に写真170, 180が添付されている。その下には、さらに章題150と文章160とが記載されている。

【0028】

このような文書が情報分割部40に入力されると、レイアウト判別部41においてテキスト部分とそれ以外の部分に分割される。この例では、タイトル110、著者名120、章題130, 150及び文章140, 160がテキスト部分であり、写真170, 180の部分がテキスト以外の部分である。そこで、レイアウト判別部41により、テキスト部分が抽出される。抽出されたテキスト部分は、タイトル削除部42によって、レイアウト情報やテキストのフォントやサイズ、句読点の有無などから、本文以外の記載内容であるタイトル110、著者名120、章題130, 150が削除される。残った文章140, 160の記載内容であるテキスト情報は、評価対象としてタイトル削除部42によって評価単位生成部43に送られる。

【0029】

なお、この実施の形態では、文章として内容を記述している部分のみを、評価対象の情報としている。これは、タイトルなどは相手の気を引くために、もしくは相手をごまかすために、内容と異なる表現や誇張した表現を用いる場合があり、情報の内容を的確に表現しているとは言えないことが多々あるからである。そのため、タイトルなどに入っている語彙は、あえて評価対象から外している。

【0030】

評価単位生成部43に送られたテキスト情報は、章や段落などの区切りを考慮しながら、評価単位に分割される。ここで、生成された評価単位は、評価基準数値化部50へ送られる。

【0031】

このようにして、文書100中のテキストが評価単位に分割され、評価基準数値化部50に渡される。

ところで、文書などは章や段落を持つ場合が多いが、すべての情報がこのように、意味的な区切りを持っているわけではない。本発明では、章や段落など意味的に区切られていると思われる部分を考慮しつつも、意味的な区切りが認識できず、テキストが続く情報については、文書の認識単位として指定された文字数で情報を区切り、評価単位を生成する。

【0032】

この方法においては、意味的に同じ部分が区切られてしまう可能性が考えられるが、同じ内容が続くならば、それらの情報中に出現する語彙は類似する可能性が高く、区切られたそれぞれの情報が類似する情報として提示されるため問題はない。

【0033】

また、区切られることで、出現語彙が分離され、それぞれの情報の類似度が低下したとすると、それは関連語彙が広範囲に分散しているためであり、情報の密度が低く類似度は低いと考えられる。よって、連続する情報の文字数による分割による影響は少い。

【0034】

次に、評価基準数値化部50の詳細を説明する。

図6は、評価基準数値化部の内部構成を示す図である。図に示す各構成要素の機能を以下に示す。

【0035】

キーワード抽出部51は、テキストを形態素解析し特定品詞の語彙のみをキーワードとして抽出する。結合キーワード生成部52は、抽出されたキーワードから特定品詞が連続している部分を抽出し、キーワードを結合して新たな結合キーワードとする。結合キーワ

10

20

30

40

50

ード重み付け部 53 は、キーワードの結合数に応じて結合キーワードに重みを付ける。キーワードペア生成部 54 は、あるキーワードと同じ文章中に離れて存在する別の特定品詞のキーワードとを組み合わせ、1つのキーワードペアとする。キーワードペア重み付け部 55 は、生成したキーワードペアにキーワード間の距離を考慮した重みを付ける。評価ベクトル空間生成部 56 は、生成した結合キーワード及びキーワードペアを用い、情報間で重複する割合の高いもしくは重複のない結合キーワード及びキーワードペアを削除した上で、類似度評価のための評価ベクトル空間を生成する。評価ベクトル生成部 57 は、各評価単位毎に情報中に存在する結合キーワード及びキーワードペアを検出し評価ベクトルを生成する。

【0036】

このような評価基準数値化部 50 における処理手順を以下に示す。

図 7 は、評価基準数値化部の処理手順を示すフローチャートの前半である。

〔S11〕キーワード抽出部 51 が、以下のステップ S12 ～ステップ S19 の処理を行っていない情報を選択し、その情報から抽出された評価単位群を処理対象とする。

〔S12〕キーワード抽出部 51 が、処理対象とした評価単位群の未処理の評価単位を 1 つ選択し、その評価単位の形態素解析を行い、品詞単位の語彙に分解する。

〔S13〕キーワード抽出部 51 が、各語彙の中で特定の品詞以外の語彙を削除する。本実施の形態では、固有名詞、普通名詞、サ変名詞、動詞を抽出し、他の品詞の語彙を削除する。

〔S14〕キーワード抽出部 51 が、品詞分解された評価単位中の情報を、句点やピリオドで区切られる 1 文章単位で情報格納部 20 に格納する。さらに連続して存在する固有名詞、普通名詞には、次の品詞と連続していたこと示す記号、例えばハイフンを付けて格納する。

〔S15〕結合キーワード生成部 52 が、連続した固有名詞、普通名詞を検出し、これらを結合した新たなキーワードを生成する。例えば「情報・分類・装置」のように連続した 3 つの普通名詞の場合、本実施の形態では、1 つ目と 2 つ目とを結合した「情報分類」と、1 つ目、2 つ目、及び 3 つ目を結合した「情報分類装置」を新たなキーワードとする。

【0037】

なお、本実施の形態では用いないが、これら意外にも、3 つの普通名詞の組み合わせ方は多数存在し、3 つの語彙を 2 つずつ用いて組み合わせ、語彙の前後を入れ替えて 6 つの結合ペアを生成してもよい。

〔S16〕結合キーワード重み付け部 53 が、結合キーワード生成部 52 の生成した結合キーワードに対して重み付けを行う。ここでは、通常の 1 語彙のキーワードを重み「1」とし、2 つのキーワードを結合した結合キーワードには重み「2」を付与し、3 つのキーワードを結合した結合キーワードには重み「3」を付与する。

〔S17〕キーワードペア生成部 54 が、キーワード抽出部 51 で抽出したキーワード群の中で、離れて存在するが特定の関係にある品詞を組み合わせ、キーワードペアを生成する。

【0038】

例えば、「情報を分類する」というフレーズがあった場合、「情報」という名詞と「分類」というサ変名詞を組み合わせ、これらのペアを 1 つのキーワードとして用いる。ただし、この場合「情報」と「分類」という語彙は、それぞれ様々な場面で用いられることが予想され、これらをそれぞれキーワードとして用いると、「情報の分類」とは無関係なものまで、関連情報として抽出される可能性がある。そこで、これら関連の深い語彙を組み合わせ、共に存在した場合のみ関連性を認めることで、語彙の意味を限定してキーワードとして用いることを可能とする。本実施の形態では、1 文章中の普通名詞、固有名詞と、その文章中の動詞の直前に存在する普通名詞もしくはサ変名詞を組み合わせ、キーワードペアとする。

【0039】

なお、本実施の形態では用いないが、形容詞と名詞や形容詞と動詞など、同時に使われる

10

20

30

40

50

ことで意味を限定するような組み合わせは、同様に用いることが可能である。

〔S 1 8〕キーワードペア重み付け部 5 5 が、キーワードペアに対して重み付けを行う。これは、通常の 1 語彙のキーワードを重み「1」とし、1 つの語彙のキーワードと組み合わせを行ったキーワードペアには重み「2」を付与し、2 つの語彙を結合した結合キーワードと組み合わせを行ったキーワードペアには重み「3」を付与する。

〔S 1 9〕キーワードペア生成部 5 4 は、これらキーワード群の生成処理終了後、キーワード中の動詞と 1 文字の語彙からなるキーワードとを削除する。これは、動詞は「する」や「行う」など、汎用性が高く意味を限定し難いためであり、1 文字の語彙も「今」や「何」といったものが多く含まれるためである。

【0 0 4 0】

なお、本実施の形態では、重みの決定に際し語彙の出現回数は考慮しない。これは、汎用性のある語彙であれば、文章中に複数回出現するのが普通であり、かといって、特定の語彙が複数回出現した場合に、その出現語彙に関する内容の情報であるかということ、そうでない場合が多いからである。すなわち、情報の内容は、特定の語彙のみでは、情報の持つ意味の範囲が広過ぎて限定できない。通常は、対象と目的や対象と方法など、複数の語彙により情報の内容を限定し、ユーザの目的に合った情報を取捨選択していると考えられる。よって、本実施の形態では、情報の内容は、特定語彙の出現回数ではなく、語彙の組み合わせの方がその内容を表現するのに適していると考え、結合キーワードもしくはキーワードペアとなる語彙の重みのみを考慮する。

【0 0 4 1】

また、複数の語彙を結合した重みの大きいキーワード含む文書群は、特にキーワードが示す内容に関して一致している可能性が高い。例えば、「情報分類装置」の場合、「情報分類」では、「情報分類機関」や「情報分類の研究者」など装置と直接関連のないものも含まれるが、「情報分類装置」であれば、より関連の高いもののみが選ばれる。また、このような長いキーワードが一致した場合は、そのキーワードを構成する個々のキーワードも一致するため、一致したキーワードの重みだけでなく、一致するキーワード数も多くなり、情報の類似度は飛躍的に高くなる。

【0 0 4 2】

図 8 は、評価基準数値化部の処理手順を示すフローチャートの後半である。

〔S 2 0〕キーワードペア生成部 5 4 は、情報中のすべての評価単位の処理が終了したか否かを判断する。情報中のすべての評価単位の処理が終了したのであればステップ S 2 1 に進み、そうでなければ次の評価単位の処理を行うべき旨の指令をキーワード抽出部 5 1 に送信し、処理をステップ S 1 2 に進める。

〔S 2 1〕キーワードペア生成部 5 4 は、すべての情報に対する処理が終了したか否かを判断する。情報中のすべての評価単位の処理が終了したのであればステップ S 2 2 に進み、そうでなければ次の情報の処理を行うべき旨の指令をキーワード抽出部 5 1 に送信し、処理をステップ S 1 1 に進める。

【0 0 4 3】

以上の処理により、各情報中の評価単位に対応するキーワード群が生成される。生成されたキーワード群は、情報格納部 2 0 により記憶装置などに格納される。全ての情報の評価単位についてキーワード群が生成された後、ステップ S 2 2 以降の処理が行われる。

〔S 2 2〕評価ベクトル空間生成部 5 6 が、各評価単位のキーワード群から、一定の割合以上のキーワード群に存在するキーワードを、各評価単位のキーワード群から削除する。これは、大多数の評価単位に存在するキーワードがあった場合、そのキーワードは評価単位を分類するためには何も寄与しないために行われる処理である。例えば、全評価単位の 7 割以上に含まれるキーワードは削除する。

〔S 2 3〕評価ベクトル空間生成部 5 6 が、各評価単位のキーワード群から、一定の割合以下のキーワード群にしか存在しないキーワードを削除する。これは、全評価単位中の少数の評価単位にしか存在しないキーワードは、評価単位を分類するためには何も寄与しないために行われる処理である。例えば、全評価単位の 1 割以下にしか含まれないキーワ

10

20

30

40

50

ドは削除する。

【 0 0 4 4 】

なお、ステップ S 2 2 , S 2 3 で行った削除処理の基準となるキーワードが含まれる割合は、情報の分類の状況やユーザの好みで変更することが可能である。

[S 2 4] 評価ベクトル空間生成部 5 6 が、不必要と思われるキーワードを削除した全評価単位のキーワード群（すべての情報の各評価単位のキーワード群すべて）を用いて、評価ベクトル空間を生成する。評価ベクトル空間は、全ての評価単位のキーワード群からキーワード、結合キーワード、キーワードペアを抽出し、全種類のキーワード、結合キーワード、キーワードペアを 1 つずつそろえたものである。評価ベクトル空間生成時には、キーワード群中に重複して存在するものがあっても、その出現回数や個数は考慮されない。また、その出現回数や個数に関係なく、存在するキーワードは評価ベクトル空間に 1 つだけ含まれる。

10

[S 2 5] 評価ベクトル生成部 5 7 が、評価ベクトル空間を用いて、各評価単位の評価ベクトルを生成する。評価ベクトルを生成するには、まず、その評価単位のキーワード群を用い、評価ベクトル空間の要素中にキーワード群中のキーワード、結合キーワード、キーワードペアがあれば、一致した要素にキーワードの持つ重みを与える。ここで、キーワード群中に存在しない評価ベクトル空間の要素は 0 となる。このように重みの値が付与された評価ベクトル空間の要素が、各評価単位の評価ベクトルとなる。

【 0 0 4 5 】

このようにして、評価単位に分割された情報に基づいて、部分類似度算出のための、評価ベクトルが生成される。

20

ここで、評価基準数値化部 5 0 における処理を、具体例を用いて説明する。

【 0 0 4 6 】

図 9 は、評価単位となる情報の例を示す図である。これは、処理対象の情報に記載された文章に含まれる評価単位 7 1 である。評価基準数値化部 5 0 がこのような評価単位 7 1 を取得すると、キーワード抽出部 5 1 が評価単位 7 1 の記載内容を品詞単位の語彙に分解する。

【 0 0 4 7 】

図 1 0 は、品詞分解をした結果を示す図である。品詞分解が行われると、図のような語彙のリスト 7 2 が生成される。例えば、「マルチメディア時代にネットワークに期待されるサービスは何か。」という文は、「マルチメディア」（普通名詞）、「メディア」（普通名詞）、「時代」（普通名詞）、「に」（名詞接続助詞）、「ネットワーク」（普通名詞）、「に」（名詞接続助詞）、「期待」（サ変名詞）、「さ」（動詞）、「れる」（動詞性接尾詞）、「サービス」（サ変名詞）、「は」（副助詞）、「何」（普通名詞）、「か」（終助詞）、「。」（句点）という語彙に分解される。

30

【 0 0 4 8 】

次に、キーワード抽出部 5 1 により、評価単位の文から得られた複数の語彙の中から、特定の品詞以外の語彙が削除され、1 文章毎に情報格納部 2 0 に格納される。

【 0 0 4 9 】

図 1 1 は、不要な語彙を削除し 1 文章毎に格納したキーワード群を示す図である。この例に示した評価単位 7 1 の文章は、3 つの文で構成されるため、3 つのキーワード群 7 3 a ~ 7 3 c が生成されている。例えば、キーワード群 7 3 a は、「マルチメディア時代にネットワークに期待されるサービスは何か。」という文から生成されたものである。この文の中で、「に」、「れる」、「は」、「か」、「。」という語彙は、固有名詞、普通名詞、サ変名詞、動詞のいずれでもないため削除されている。また、「マルチ」や「メディア」は、次の固有名詞もしくは普通名詞と連続していたため「 - 」の記号が付加されている。

40

【 0 0 5 0 】

このようにして生成されたキーワード群を用いて、結合キーワード及びキーワードペアが生成され、さらに重みが付けられる。

50

図 1 2 は、1 文章毎の結合キーワードとキーワードペアを含んだキーワード群の例を示す図である。キーワード群 7 4 a ~ 7 4 c 中のハイフンで繋がれた語彙のペアは、キーワードペアであることを示す。また、各キーワードに付けられている数字は、1 語彙のキーワードの重みを 1 とした場合の、各キーワードに付けられた重みである。例えば、キーワード群 7 4 a では、「マルチ」や「メディア」は単一の語彙であるため、重みの値は 1 である。また、「マルチメディア時代」は、「マルチ」、「メディア」、「時代」の 3 つの語彙で構成されるため、重みの値は 3 である。

【 0 0 5 1 】

この文章毎のキーワード群から、重複するキーワードを削除し、各種類のキーワードを 1 つずつ含んだものが、評価単位のキーワード群となっている。このように文章毎に生成されたキーワード群が統合され、評価単位に対するキーワード群となる。

10

【 0 0 5 2 】

図 1 3 は、評価単位のキーワード群を示す図である。このキーワード群 7 5 と同様のキーワード群が、すべての情報の評価単位毎に生成され、それらのキーワード群から入力された情報群における評価ベクトル空間が生成される。

【 0 0 5 3 】

図 1 4 は、評価ベクトル空間を示す図である。この評価ベクトル空間 7 6 は、簡略化のため、前述の評価単位とは関連しない評価ベクトル空間を示している。

ここで、図 1 5 のようなキーワード群が存在した場合を考える。

【 0 0 5 4 】

20

図 1 5 は、評価単位のキーワード群の例を示す図である。このキーワード群 7 7 について、図 1 4 に示した評価ベクトル空間 7 6 での評価を行うと、図 1 6 のようになる。

【 0 0 5 5 】

図 1 6 は、評価単位の評価ベクトルの生成状況を示す図である。このように、評価ベクトル 7 8 は、評価ベクトル空間 7 6 中の各要素に対応する数値の列で表される。評価ベクトル 7 8 中の値を持つ要素数と要素の位置が、評価ベクトル空間 7 6 中での評価単位の方角を表わし、評価ベクトル要素の値の大きさがベクトルの大きさを表す。この方向と大きさがその評価単位の情報の特徴を表し、評価ベクトルが類似した方向を向き、類似した大きさならば、その評価単位の内容は類似していると考えられる。

【 0 0 5 6 】

30

このように、情報の分類において、単語や語彙単位の出現頻度ではなく、関連性のある単語や語彙を組み合わせると結合キーワードもしくはキーワードペアとして用いることにより、個々の単語や語彙の意味を限定して類似度を判定することができ、従来より高い精度での分類を実現することを可能とする。

【 0 0 5 7 】

次に、類似情報分類部の詳細を説明する。類似情報分類部は、評価基準数値化部で得られた評価ベクトル空間と評価単位の評価ベクトルを用いて、情報を分類すると共に、情報間の部分（評価単位）の類似性を示し、情報の構成の類似性や類似部分の分布、必要な情報の位置を一瞥できるようにするものである。

【 0 0 5 8 】

40

図 1 7 は、類似情報分類部の内部構成を示す図である。図に示す各構成要素の機能を以下に示す。

類似部分判別部 6 1 は、情報の評価単位毎の評価ベクトルからベクトル間の内積もしくは相対角度とベクトル間距離を求め、評価単位毎の類似度判定を行いグループ分けを行う。情報類似度判定部 6 2 は、情報間の類似する評価単位の含有度及び分布状態から情報間の類似度を判別する。インデックス情報検出部 6 3 は、類似評価単位中からそのグループの代表となる評価単位及び評価ベクトルを検出する。類似部分提示部 6 4 は、複数の類似する評価単位を類似度順などで順次比較表示し、内容の類似度を確認する。類似状況提示部 6 5 は、情報間の類似度を確認するために、情報中及び情報間の評価単位の類似部分とその分布が判るように一覧表示する。

50

【 0 0 5 9 】

図 1 8 は、類似情報分類部の処理手順を示すフローチャートの前半である。

[S 3 1] 類似部分判別部 6 1 が、評価ベクトルを用いて評価単位間の類似度を算出する。類似度の判別は、2つの評価単位の評価ベクトルからその内積を求め、内積の値を類似判定値として、その大きさによって類似性を判別する。類似判定のための閾値は、情報の分類状況やユーザの好みによって変更することが可能である。この他にも、ベクトル間の角度とベクトルの大きさから、類似度を判別してもよい。

[S 3 2] 類似部分判別部 6 1 が、類似度判別の結果、互いに類似している評価単位の双方に対して、類似する相手の評価単位の識別番号、その類似判定値、及び同じ分類であることを示す分類番号を付与する。

10

[S 3 3] 類似部分判別部 6 1 が、すべての評価単位間の類似度判別が終了したか否かを判断する。類似度判別が終了したのであればステップ S 3 4 に進み、そうでなければステップ S 3 1 に進む。なお、評価単位間の類似度判別は、同一情報内の評価単位同士であるか、異なる情報内の評価単位同士であるかに関わらず、すべての評価単位間で行う。

[S 3 4] 情報類似度判定部 6 2 が、比較対象となる2つの情報を構成する評価単位の中で、互いに類似する評価単位のそれぞれの情報内における含有度を求める。

[S 3 5] 情報類似度判定部 6 2 が、類似する評価単位の含有度と、互いに類似する評価単位の類似判定値とから、情報間の類似性を求める。そして、類似性の値が一定の閾値を超えた情報間は類似性があると判断し、同じカテゴリーに分類する。類似性の算出方法としては、例えば、互いに類似する評価単位の含有度に、類似する評価単位間の類似判定値を乗算する。

20

【 0 0 6 0 】

なお、この例では、情報間の類似判定は、大まかな範囲での区分けにとどめる。これは、詳細な判定値を用いて文書を順位付けしても必ずしもユーザの意図と一致しないため、まずは大局的な判断での分類を行い、以後の細かな類似判別はユーザに委ねるためである。

[S 3 6] 情報類似度判定部 6 2 が、同じカテゴリーに分類された情報に対して、同一のカテゴリーラベルを付与する。なお、情報間の類似度には特に順位付けは行わないが、結果の表示においては、他の評価単位と類似する評価単位の含有度を指針として、含有度の大きいものを優先して表示する。

[S 3 7] 情報類似度判定部 6 2 は、すべての情報間の類似度判別が終了したか否かを判断する。判別が終了していればステップ S 3 8 に進み、そうでなければステップ S 3 4 に進む。

30

【 0 0 6 1 】

図 1 9 は、類似情報分類部の処理手順を示すフローチャートの後半である。

[S 3 8] インデックス情報検出部 6 3 が、類似判別において情報間毎に行ったカテゴリー分けが適切かどうかを分類された情報間で見直し、複数のカテゴリーに分類された情報については、含有度の高い方を優先し、他方との関連に関しては、補足情報として情報格納部 2 0 に格納する。

[S 3 9] インデックス情報検出部 6 3 が、ステップ S 3 8 で行った見直しの処理において、分類したカテゴリー内の情報群の中で最も多く含まれている種類の分類番号を特定し、その分類番号を持つ評価単位を全て抽出する。そして、この同じ分類番号を持つ評価単位の評価ベクトルの中で、中心となる評価ベクトルを選び出し、これをこのカテゴリーのインデックス情報として登録する。この中心となるベクトルの抽出は、例えば、各評価ベクトルのうち最大角度を持つ組み合わせを抽出し、それらの中間にあるもの、それぞれの評価ベクトルと同じ角度を持つものを探し出すことで、カテゴリーの中心となる評価ベクトルを選出できる。

40

【 0 0 6 2 】

インデックス情報は、そのインデックス情報の評価単位が、カテゴリーの情報を表わす代表情報として示される。また、その評価単位から抽出されたキーワード群は、そのカテゴリーのキーワードとして用いられる。さらに、インデックス情報は、新たに入力された情

50

報を分類する場合における、類似性判定の指針としても利用する。

【0063】

これら分類された情報は、類似部分提示部64と類似状況提示部65により、その情報間の関係や内容がユーザに示され、ユーザ自身の手で、必要な情報を取り出すことが可能となる。

【S40】類似部分提示部64が、分類したカテゴリー毎に、その情報内の評価単位の分類番号の同じ評価単位を集め、他の評価単位との間の類似順に各評価単位を並べ、表を作成する。類似順は、分類番号の同じ他の評価単位との間の類似度の合計値を比較することで求める。作成した表は、情報格納部20に格納する。格納された表の内容は、情報表示部30によって表示される。

10

【0064】

なお、類似順に並べられた評価単位には、他のカテゴリーにある類似評価単位へのリンクが張られている。これによりユーザは、必要とする情報に類似した評価単位を選択することで、収集し分類した情報全体の中で、必要な部分のみを見ることができ、またその評価単位を含む情報がどのカテゴリーに含まれるかを知ること、新たな関連情報を見つけることができる。

【S41】類似状況提示部65が、分類したカテゴリー内の情報を、類似する評価単位の含有度の大きい順に並べた表を作成し、情報格納部20に格納する。情報格納部20に格納された表は、情報表示部30によって表示される。

【S42】類似状況提示部65が、各情報毎にその評価単位の類似する部分がどこか、どれくらい存在するかを一瞥でき、且つ他の情報との比較が容易なように、一覧にして提示するための表を作成し、情報格納部20に格納する。情報格納部20に格納された表は、情報表示部30によって表示される。

20

【0065】

これによりユーザは、任意の評価単位の内容を理解することで、その評価単位を含む情報の内容を推定でき、評価単位の分布や情報中に含まれる評価単位の種類などから、情報が有用であるかどうかの判断を行うことができる。また、情報間の含まれる評価単位の分布を比較することで、内容の類似性を確認でき、装置の分類結果の確認を行うこともできる。

【0066】

図20は、類似状況提示部による情報間の類似状況表示の例を示す図である。この図において、同じ模様の部分は同じ分類番号の評価単位であることを示す。これから文書などの情報間で、類似部分がどれくらいある文書なのかが容易に判る。

30

【0067】

従来は必要な内容が含まれると思われる情報、例えば文書を収集し、その中から必要な部分情報を探し出していた。本発明によれば、情報の構成の類似性や類似部分の分布、必要な情報の位置を一瞥でき、必要な部分情報を探し出してから、その部分情報が含まれる文書を取り出すことが可能となる。

【0068】

また本発明によれば、評価ベクトル空間生成部を用いて生成した評価ベクトル空間を用いて、入力文書や送られてくる情報をフィルタリングし、指定の類似度の評価単位を持つ文書のみを選択し、ユーザに提示するような情報フィルタリング機能を提供することも可能である。

40

【0069】

また本発明によれば、評価ベクトル空間生成部を用いて生成した評価ベクトル空間を用いて、データベースなどを検索し、規定の類似度の評価単位を持つ情報のみを選択し、ユーザに提示する情報検索装置として用いることも可能である。

【0070】

また、ユーザが求めるものは、必要な情報を含む文書ではなく、文書中に存在する情報である。これまでの検索や分類は、欲しい情報を含んでいると思われる文書を集めるもので

50

あり、欲しい情報そのものを提示するものではなかった。本発明は、欲しい情報であると思われる部分が情報全体にどのように分布しているかを提示することができ、ユーザは、情報間の類似や差異を確認した上で、情報の有無を判断できる。

【0071】

さらに、本発明は情報を含む媒体の有用性の判断を促し、また直接的に有用な情報を提示するものであり、これまでの情報の検索や分類で行われていたような、必要な情報の密度を上げるものではなく、情報の取捨選択を効率化するものである。

【0072】

なお、上記の実施の形態では、情報の分類数は不定であったが、インデックス情報となる情報数を指定することで、分類数を指定するようにしてもよい。

10

また、上記の実施の形態では、インデックス情報をシステムが求め出していたが、ユーザが任意にインデックス情報を指定できるようにしてもよい。この場合、類似情報分類部はユーザの指定した情報をインデックス情報として情報を分類する。

【0073】

また、生成された評価ベクトル空間を用いて入力文書群を検索し、規定の類似度の評価単位を持つ文書のみを選択し、ユーザに提示するような情報検索機能を備えさせることもできる。

【0074】

また、上記の処理機能は、コンピュータによって実現することができる。その場合、情報判別支援装置が有すべき機能の処理内容は、コンピュータで読み取り可能な記録媒体に記録されたプログラムに記述しておく。そして、このプログラムをコンピュータで実行することにより、上記処理がコンピュータで実現される。コンピュータで読み取り可能な記録媒体としては、磁気記録装置や半導体メモリ等がある。市場に流通させる場合には、CD-ROM (Compact Disk Read Only Memory) やフロッピーディスク等の可搬型記録媒体にプログラムを格納して流通させたり、ネットワークを介して接続されたコンピュータの記憶装置に格納しておき、ネットワークを通じて他のコンピュータに転送することもできる。コンピュータで実行する際には、コンピュータ内のハードディスク装置等にプログラムを格納しておき、メインメモリにロードして実行する。

20

【0075】

【発明の効果】

30

以上説明したように本発明の情報判別支援装置では、各情報内の評価単位毎の類似度を求めるようにしたため、これらの情報の分類及び評価単位間の類似情報をユーザに提示すれば、欲しい情報であると思われる部分が情報全体にどのように分布しているかを提示することができ、ユーザによる情報の類似性・関連性の判断が容易となり、有用な情報の取捨選択の手間が軽減される。

【0076】

また、本発明の情報判別支援プログラムを記録したコンピュータ読み取り可能な記録媒体では、記録媒体に記録された情報判別支援プログラムをコンピュータに実行させることにより、情報群を互いに類似する情報に分類し、且つ各情報内の評価単位毎の類似度も求めるような処理をコンピュータに実行させることができる。

40

【図面の簡単な説明】

【図1】 本発明の原理構成図である。

【図2】 本発明の実施の形態に係る情報判別支援装置の構成を示すブロック図である。

【図3】 情報分割部の内部構成を示す図である。

【図4】 情報判別支援装置の処理の流れを示すフローチャートである。

【図5】 本装置において処理の対象となる電子化された文書例を示す図である。

【図6】 評価基準数値化部の内部構成を示す図である。

【図7】 評価基準数値化部の処理手順を示すフローチャートの前半である。

【図8】 評価基準数値化部の処理手順を示すフローチャートの後半である。

【図9】 評価単位となる情報の例を示す図である。

50

- 【図10】 品詞分解をした結果を示す図である。
 【図11】 不要な語彙を削除し1文章毎に格納したキーワード群を示す図である。
 【図12】 1文章毎の結合キーワードとキーワードペアを含んだキーワード群の例を示す図である。
 【図13】 評価単位のキーワード群を示す図である。
 【図14】 評価ベクトル空間を示す図である。
 【図15】 評価単位のキーワード群の例を示す図である。
 【図16】 評価単位の評価ベクトルの生成状況を示す図である。
 【図17】 類似情報分類部の内部構成を示す図である。
 【図18】 類似情報分類部の処理手順を示すフローチャートの前半である。
 【図19】 類似情報分類部の処理手順を示すフローチャートの後半である。
 【図20】 類似状況提示部による情報間の類似状況表示の例を示す図である。

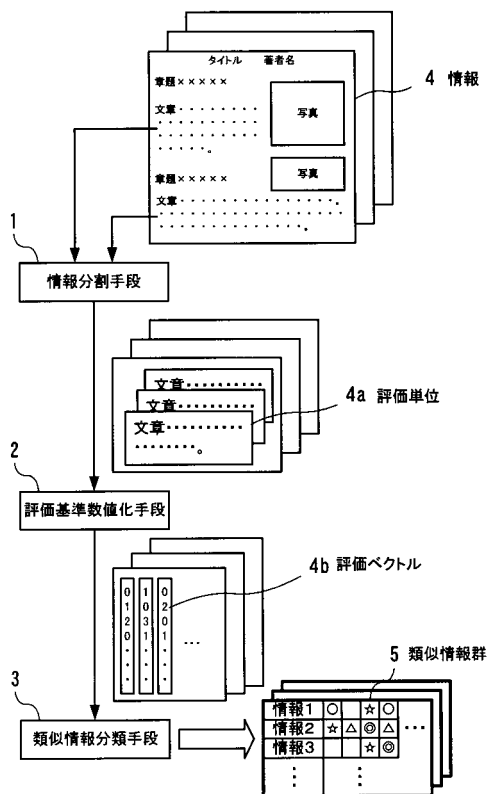
【符号の説明】

- 1 情報分割手段
- 2 評価基準数値化手段
- 3 類似情報分類手段
- 4 情報
- 4 a 評価単位
- 4 b 評価ベクトル
- 5 類似情報群

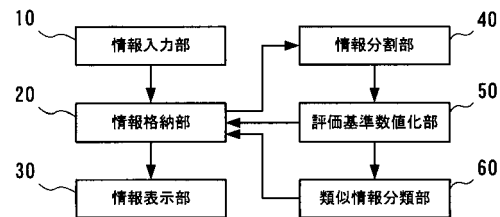
10

20

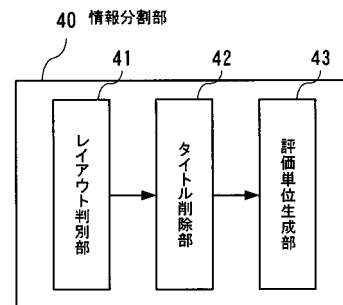
【図1】



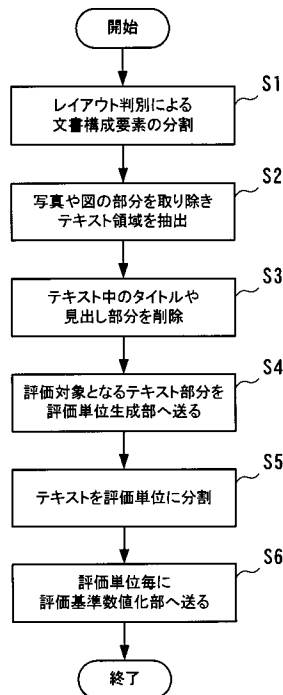
【図2】



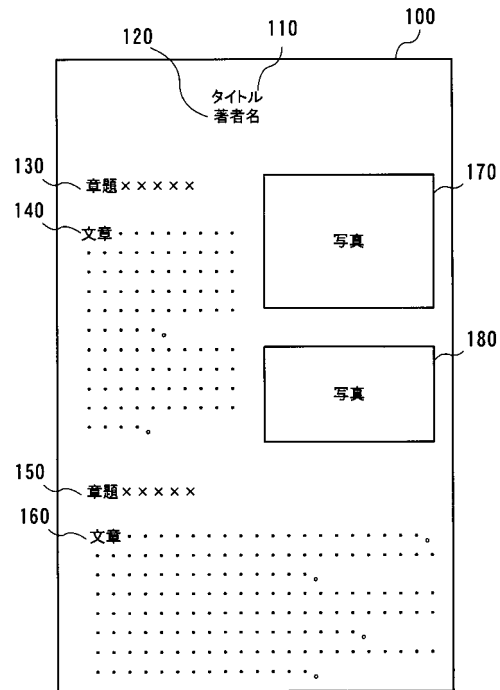
【図3】



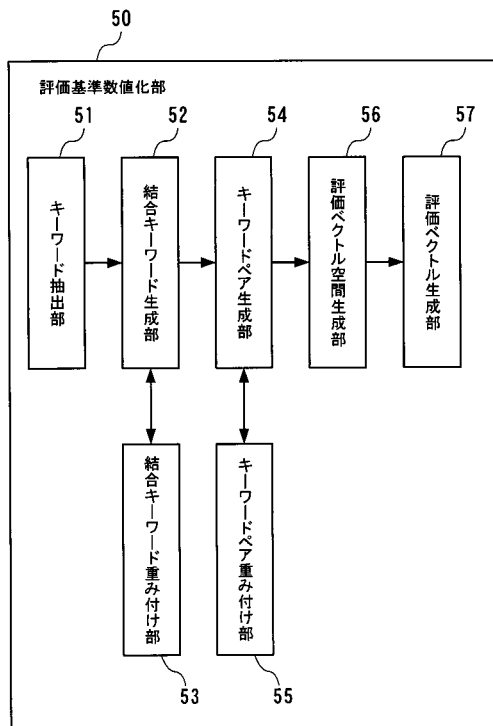
【図 4】



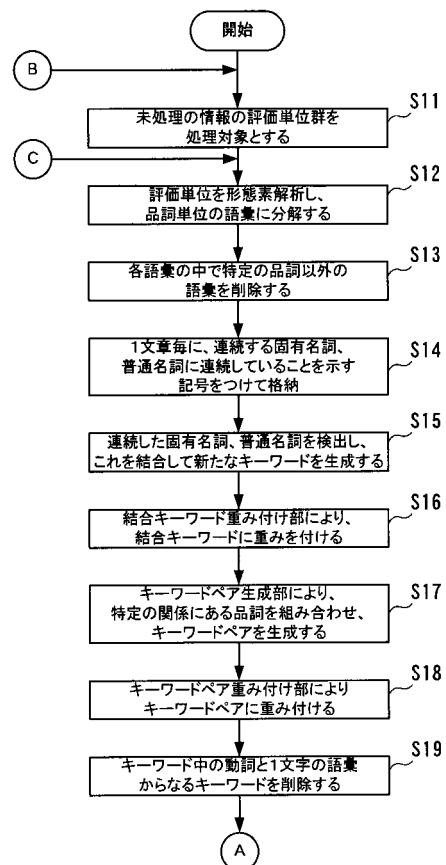
【図 5】



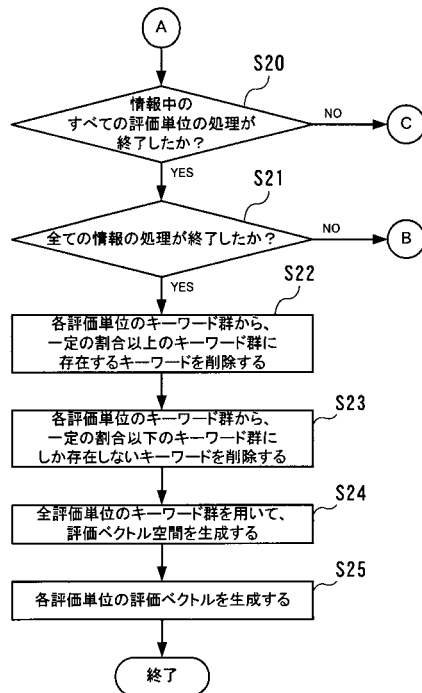
【図 6】



【図 7】



【図 8】



【図 9】

71

マルチメディア時代にネットワークに期待されるサービスは何か。今、マルチメディアを取り巻く種々の動きがある。大規模な社会インフラとして、電子取引、電子図書館や電子自治体行政サービスなどの実験が行われている。

【図 10】

72

マルチメディア	普通名詞
メディア	普通名詞
時代に	普通名詞
ネットワーク	名詞接続助詞
(普通名詞
)	括弧
に	名詞接続助詞
期待	サ変名詞
される	動詞
サービス	動詞性接尾辞
は	サ変名詞
何か	副助詞
。	普通名詞
今	普通名詞
、	読点
マルチ	普通名詞
メディア	普通名詞
を取り巻く	格助詞
種々の	動詞
動き	普通名詞
がある	名詞接続助詞
。	普通名詞
大規模な	格助詞
社会	普通名詞
インフラ	普通名詞
として	名詞接続助詞
、	動詞
電子	読点
取引	普通名詞
、	普通名詞
電子	読点
図書館	普通名詞
や	名詞
電子化	普通名詞
自治体	名詞接続助詞
行政	サ変名詞
サービス	普通名詞
などの	サ変名詞
実験	副助詞
が	名詞接続助詞
行わ	サ変名詞
れて	格助詞
いる	動詞
。	動詞性接尾辞
	動詞性接尾辞
	句点

【図 11】

73a

マルチ	普通名詞
メディア	普通名詞
時代に	普通名詞
ネットワーク	普通名詞
期待	サ変名詞
さ	動詞
サービス	サ変名詞
何	普通名詞

73b

今	普通名詞
マルチ	普通名詞
メディア	普通名詞
取り巻く	動詞
種々	普通名詞
動き	普通名詞
ある	動詞

73c

社会	普通名詞
して	動詞
電子	普通名詞
取引	普通名詞
電子	普通名詞
図書館	普通名詞
電子化	サ変名詞
自治体	普通名詞
行政	普通名詞
サービス	サ変名詞
実験	サ変名詞
行わ	動詞

【図 1 2】

マルチ	1	74a
メディア	1	
時代	1	
マルチメディア	2	
マルチメディア時代	3	
ネットワーク	1	
期待	1	
マルチー期待	2	
メディアー期待	2	
時代ー期待	2	
マルチメディアー期待	3	74b
マルチメディア時代ー期待	4	
ネットワークー期待	2	
サービス	1	
マルチ	1	
メディア	1	74c
マルチメディア	2	
種々	1	
動き	1	
マルチー動き	2	
メディアー動き	2	
マルチメディアー動き	3	
種々ー動き	2	
社会	1	
電子	1	
取引	1	
電子取引	2	
電子	1	
図書館	1	
電子図書館	2	
電子化	1	
自治体	1	
行政	1	
自治体行政	2	
サービス	1	
実験	1	
社会ー実験	2	
電子ー実験	2	
取引ー実験	2	
電子取引ー実験	3	
電子ー実験	2	
図書館ー実験	2	
電子図書館ー実験	3	
電子化ー実験	2	
自治体ー実験	2	
行政ー実験	2	
自治体行政ー実験	3	
サービスー実験	2	

【図 1 3】

マルチ	1	75
メディア	1	
時代	1	
マルチメディア	2	
マルチメディア時代	3	
ネットワーク	1	
期待	1	
マルチー期待	2	
メディアー期待	2	
時代ー期待	2	
マルチメディアー期待	3	
マルチメディア時代ー期待	4	
ネットワークー期待	2	
サービス	1	
マルチ	1	
メディア	1	
マルチメディア	2	
種々	1	
動き	1	
マルチー動き	2	
メディアー動き	2	
マルチメディアー動き	3	
種々ー動き	2	
社会	1	
電子	1	
取引	1	
電子取引	2	
電子	1	
図書館	1	
電子図書館	2	
電子化	1	
自治体	1	
行政	1	
自治体行政	2	
サービス	1	
実験	1	
社会ー実験	2	
電子ー実験	2	
取引ー実験	2	
電子取引ー実験	3	
電子ー実験	2	
図書館ー実験	2	
電子図書館ー実験	3	
電子化ー実験	2	
自治体ー実験	2	
行政ー実験	2	
自治体行政ー実験	3	
サービスー実験	2	

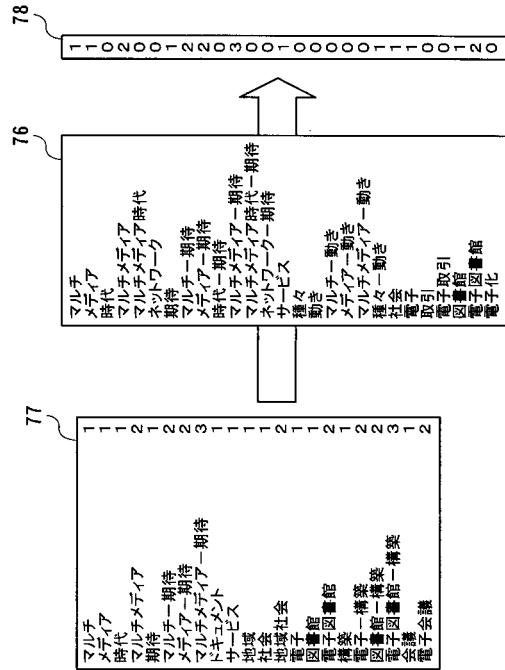
【図 1 4】

マルチ	76
メディア	
時代	
マルチメディア	
マルチメディア時代	
ネットワーク	
期待	
マルチー期待	
メディアー期待	
時代ー期待	
マルチメディアー期待	
マルチメディア時代ー期待	
ネットワークー期待	
サービス	
種々	
動き	
マルチー動き	
メディアー動き	
マルチメディアー動き	
種々ー動き	
社会	
電子	
取引	
電子取引	
図書館	
電子図書館	
電子化	

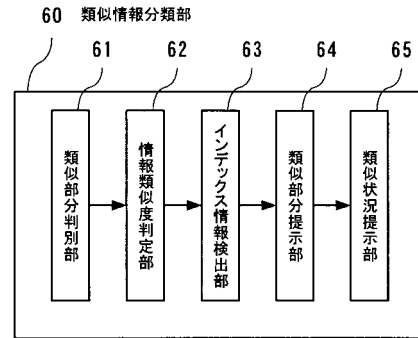
【図 1 5】

マルチ	1	77
メディア	1	
時代	1	
マルチメディア	2	
期待	1	
マルチー期待	2	
メディアー期待	2	
マルチメディアー期待	3	
ドキュメント	1	
サービス	1	
地域	1	
社会	1	
地域社会	2	
電子	1	
図書館	1	
電子図書館	2	
構築	1	
電子ー構築	2	
図書館ー構築	2	
電子図書館ー構築	3	
会議	1	
電子会議	2	

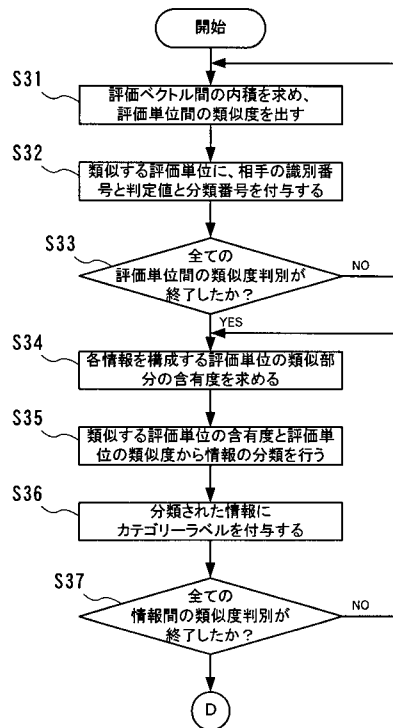
【図 16】



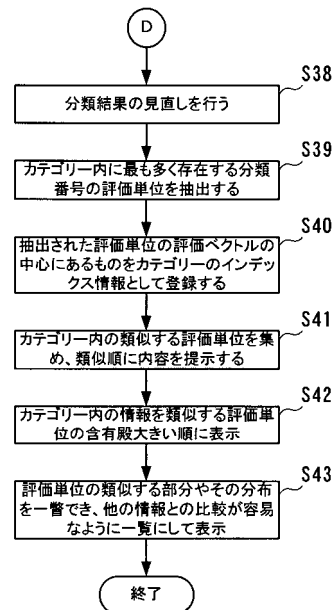
【図 17】



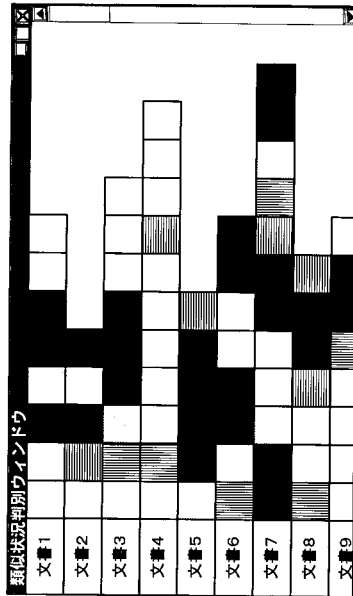
【図 18】



【図 19】



【図 20】



フロントページの続き

- (56)参考文献 特開平10-207911(JP,A)
特開平10-254894(JP,A)
特開平08-137895(JP,A)
特開平9-218878(JP,A)

- (58)調査した分野(Int.Cl., DB名)
G06F 17/30