

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3864235号  
(P3864235)

(45) 発行日 平成18年12月27日(2006.12.27)

(24) 登録日 平成18年10月13日(2006.10.13)

(51) Int. Cl.

G06F 17/30 (2006.01)

F I

G06F 17/30 350C

G06F 17/30 210A

G06F 17/30 360Z

請求項の数 14 (全 23 頁)

(21) 出願番号 特願2002-151537 (P2002-151537)  
 (22) 出願日 平成14年5月24日(2002.5.24)  
 (65) 公開番号 特開2003-345829 (P2003-345829A)  
 (43) 公開日 平成15年12月5日(2003.12.5)  
 審査請求日 平成14年6月10日(2002.6.10)

(73) 特許権者 000233538  
 株式会社 日立東日本ソリューションズ  
 宮城県仙台市青葉区本町2-16-10  
 (74) 代理人 100064414  
 弁理士 磯野 道造  
 (74) 代理人 100111545  
 弁理士 多田 悦夫  
 (72) 発明者 塚原 朋哉  
 宮城県仙台市青葉区本町二丁目16番10  
 号 日立東北ソフトウェア株式会社内  
 (72) 発明者 橋山 俊和  
 宮城県仙台市青葉区本町二丁目16番10  
 号 日立東北ソフトウェア株式会社内

最終頁に続く

(54) 【発明の名称】 情報検索システムおよび情報検索プログラム

(57) 【特許請求の範囲】

【請求項1】

情報の検索を行うための検索情報を含む各種の情報が入力され、検索の結果を表示する操作装置と、検索対象である収集情報、前記収集情報を特徴付ける単語である複数のラベル及び前記複数のラベルと前記収集情報との関連度であるスコアが格納されたデータベースと、前記操作装置及び前記データベースと接続され、前記検索情報を基に情報の検索を行い、検索結果を出力する検索処理装置を含んでなる情報検索システムであって、

前記操作装置は、

前記検索情報を入力するための第1のエリアを有した表示画面を表示し、

前記検索処理装置は、

前記第1のエリアを介して、入力された前記検索情報を分析することによって、この検索情報を基に、前記複数のラベル及び前記スコアを前記データベースから抽出し、

前記操作装置は、

前記抽出された複数のラベルの表示位置を、前記スコアに基づく関係の度合いに従って、前記収集情報を基準として定め、

前記定められた表示位置に従って、前記抽出された複数のラベルを前記操作装置の表示画面上の第2のエリアに、前記関係の度合いを示す図として表示し、

前記検索処理装置は、

前記第2のエリアを介して、前記表示された複数のラベルの中からラベルが指定されると、前記データベースに格納された収集情報から、前記指定されたラベルに関係付けられ

10

20

た検索結果を選択し、

前記操作装置は、

前記指定されたラベルに関係付けられて選択された検索結果を前記操作装置の表示画面  
上の第3のエリアに表示し、

前記第2のエリアを介して、前記表示された複数のラベルの中からさらにラベルが指定  
されると、前記さらに指定されたラベルを視覚的に区別できるように表示し、

前記検索処理装置は、

前記検索結果から、前記さらに指定されたラベルに関係付けられた検索結果をさらに絞  
り込み、

前記操作装置は、

前記さらに絞り込まれた検索結果を、前記第3のエリアに表示することを特徴とする情  
報検索システム。

【請求項2】

前記検索処理装置は、

前記第1のエリアを介して、入力された前記検索情報を基に、

前記収集情報と前記検索情報との該当の度合いである活性値を伝播させることによって  
、前記収集情報から複数の検索結果を選択することを特徴とする請求項1に記載の情報検  
索システム。

【請求項3】

前記操作装置は、

前記選択された検索結果と、検索情報をすべて含んでいないが関連している文書である  
検索結果との両方を前記第3のエリアに表示することを特徴とする請求項1または請求項  
2に記載の情報検索システム。

【請求項4】

前記操作装置は、

前記抽出されたラベルに関係する情報あるいは当該ラベルの重要度を、前記第2のエリ  
アに、当該ラベルに付随して表示することを特徴とする請求項1から請求項3のいずれか  
一項に記載の情報検索システム。

【請求項5】

前記操作装置は、

前記第2のエリアを介して、前記抽出されたラベルの表示の指定が行われると、前記指  
定に基づき、前記第2のエリアにラベルを表示することを特徴とする請求項1から請求項  
4のいずれか一項に記載の情報検索システム。

【請求項6】

前記操作装置は、

前記第2のエリアを介して、前記抽出されたラベルの非表示の指定が行われると、前記  
指定に基づき、前記第2のエリアにラベルを非表示にすることを特徴とする請求項1から  
請求項4のいずれか一項に記載の情報検索システム。

【請求項7】

前記操作装置は、

前記第2のエリアを介して、前記抽出されたラベルの表示数が指定されると、前記指  
定された数に基づいてラベルを、前記第2のエリアに表示することを特徴とする請求項1  
から請求項6のいずれか一項に記載の情報検索システム。

【請求項8】

前記操作装置は、

前記選択された検索結果の表示順を入力するエリアである第4のエリアを表示し、

前記第4のエリアを介して、前記選択された検索結果の表示順として、スコア順が指定  
されると、前記指定の順序に基づいて前記選択された検索結果を前記第3のエリアに表  
示することを特徴とする請求項1から請求項7のいずれか一項に記載の情報検索システム。

【請求項9】

10

20

30

40

50

前記操作装置は、

前記選択された検索結果の表示順を入力するエリアである第4のエリアを表示し、

前記第4のエリアを介して、前記選択された検索結果の表示順として、前記指定されたラベルを多く含む順が指定されると、前記指定の順序に基づいて前記選択された検索結果を前記第3のエリアに表示することを特徴とする請求項1から請求項7のいずれか一項に記載の情報検索システム。

【請求項10】

前記操作装置は、

前記選択された検索結果の表示順を入力するエリアである第4のエリアを表示し、

前記第4のエリアを介して、前記選択された検索結果の表示順として、前記収集情報の属性である日付順が指定されると、前記指定の順序に基づいて前記選択された検索結果を前記第3のエリアに表示することを特徴とする請求項1から請求項7のいずれか一項に記載の情報検索システム。

【請求項11】

前記操作装置は、

前記第2のエリアを介して、前記第2のエリアから特定のエリアが選択されたとき、前記選択されたエリアを拡大して表示することを特徴とする請求項1から請求項10のいずれか一項に記載の情報検索システム。

【請求項12】

前記操作装置は、

検索対象を切り替えるエリアである第5のエリアを表示し、

前記第5のエリアを介して、検索対象の切り替え操作が為されると、前記検索対象を切り替えて検索することを特徴とする請求項1から請求項11のいずれか一項に記載の情報検索システム。

【請求項13】

前記操作装置は、

ラベルの属性を選択するエリアである第6のエリアを表示し、

前記検索処理装置は、

前記第6のエリアを介して、選択されたラベルの属性を基に検索を行うことを特徴とする請求項1から請求項12のいずれか一項に記載の情報検索システム。

【請求項14】

情報の検索を行うための検索情報を含む各種の情報が入力され、検索の結果を表示する操作装置と、検索対象である収集情報、前記収集情報を特徴付ける単語である複数のラベル及び前記複数のラベルと前記収集情報との関連度であるスコアが格納されたデータベースと、前記操作装置及び前記データベースと接続され、前記検索情報を基に情報の検索を行い、検索結果を出力する検索処理装置を含んでなる情報検索システムにおける情報検索プログラムであって、

前記操作装置に、

前記検索情報を入力するための第1のエリアを有した表示画面を表示させ、

前記検索処理装置に、

前記第1のエリアを介して、入力された前記検索情報を分析することによって、この検索情報を基に、前記複数のラベル及び前記スコアを前記データベースから抽出させ、

前記操作装置に、

前記抽出された複数のラベルの表示位置を、前記スコアに基づく関係の度合いに従って、前記収集情報を基準として定め、

前記定められた表示位置に従って、前記抽出された複数のラベルを前記操作装置の表示画面上の第2のエリアに、前記関係の度合いを示す図として表示させ、

前記検索処理装置に、

前記第2のエリアを介して、前記表示された複数のラベルの中からラベルが指定されると、前記データベースに格納された収集情報から前記指定されたラベルに関係付けられた

10

20

30

40

50

検索結果を選択させ、  
前記操作装置に、  
前記指定されたラベルに関係付けられて選択された検索結果を前記操作装置の表示画面  
上の第3のエリアに表示させ、  
前記第2のエリアを介して、前記表示された複数のラベルの中からさらにラベルが指定  
されると、前記さらに指定されたラベルを視覚的に区別できるように表示させ、  
前記検索処理装置に、  
前記検索結果から、前記さらに指定されたラベルに関係付けられた検索結果をさらに絞  
り込み、  
前記操作装置に、  
前記さらに絞り込まれた検索結果を、前記第3のエリアに表示させることを特徴とする  
情報検索プログラム。

【発明の詳細な説明】

【発明の属する技術分野】

【0001】

本発明はデータベースの検索方法および検索システムあるいは検索のためのコンピュ  
タプログラムに関する。

【従来の技術】

【0002】

大量に蓄えられている蓄積情報の中から目的とする情報を検索する例として、検索キー 20  
を含む、あるいは検索キーと類似する文書の検索を行う方法が知られている。

文書検索では、全文検索のように指定された検索キーが含まれる文書を検索する方法が  
ある。また、文書中で使われている単語の頻度等を文書の特徴ベクトルとし、検索キーの  
特徴ベクトルと類似する文書を見つける概念検索の方法がある。このような技術は例えば  
公開特許公報、特開平7-114572号に記載されている。

【0003】

また公開特許公報 特開2002-123551には、検索語が多数存在する場合に発  
生する検索処理速度の低下や操作性の低下を防止する内容が記載されている。

【発明が解決しようとする課題】

【0004】

検索処理においてはより利用し易いシステム等が望まれている。

本発明は検索のための操作がより便利なシステム及びコンピュータプログラムを提供す  
ることを目的とする。

【0005】

本発明では、データベースなどの検索対象を検索しようとする検索情報に基づき、以下  
の実施の形態で述べている如きラベルを複数抽出し、これらのラベルから特定のラベルを  
指定することにより、指定されたラベルに基づき選別して上記データベースの最終的な検  
索結果を出力するようにしたことである。

本発明によれば出力結果を見てラベル指定をさらに変更し、検索結果の最適化を図るこ  
とが可能である。

【発明の実施の形態】

【0006】

本発明に係る実施の形態について説明する。図1に情報検索システムの構成を示す。本  
システムは、複数の操作装置102とこれらの操作装置102からの検索要求に基づき検  
索を実行する検索処理装置101とを備えている。操作装置102は、利用者が検索の操  
作を行うための装置で、各操作装置102は入力装置としてのキーボードやマウスと出力  
装置としてのディスプレイを有している。また検索処理装置101は、情報特定手段（図  
示せず）や類似度計算手段（図示せず）および情報蓄積手段（図示せず）を備えており、  
さらにインデクス情報データベース103や蓄積情報データベース104、ファイル情報  
データベース105に接続されている。インデクス情報データベース103や蓄積情報デ

10

20

30

40

50

ータベース１０４、ファイル情報データベース１０５は検索処理を実行する上で利用するデータベースである。

【０００７】

図２は上記情報検索システムが検索処理を行う上で、例えば類似情報検索の処理を行う場合に必要な構成、および上記構成の動作とデータベースの利用の関係を示す動作説明図である。検索キー入力手段２０９とディスプレイ等の出力装置２１５は操作装置１０２が有する構成である。また検索処理装置１０１は、情報入力手段２０１と特徴抽出手段２０２、情報特定手段２０３、類似度計算手段２０４、情報蓄積手段２０５とを備え、さらに検索キー該当情報検索手段２１０と活性値設定手段２１１、活性値伝播手段２１２、選択手段２１３および出力手段２１４とを備えている。ここで、情報入力手段２０１を操作装置１０２が有し、操作装置１０２から検索処理装置１０１に情報を入力する構成としてもよい。また、検索処理装置１０１の出力手段２１４により表示するための出力情報のデータを生成し、生成したデータを操作装置１０２に送信し表示手段２１６により表示してもよい。

10

【０００８】

上記手段である特徴抽出手段２０２や情報特定手段２０３、類似度計算手段２０４、情報蓄積手段２０５、検索キー該当情報検索手段２１０、活性値設定手段２１１、活性値伝播手段２１２、選択手段２１３および出力手段２１４は、検索処理装置１０１を構成する計算機に、上記機能を実現するためのコンピュータプログラムをロードして実行することで実現される。インデクス情報データベース１０３や蓄積情報データベース１０４、ファイル情報データベース１０５は検索処理装置１０１の内部メモリに設けてもよいが、構成および動作を理解し易くするために、検索処理装置１０１の外にあるように記載した。

20

情報検索システムは検索キー入力手段２０９から入力された入力情報に基づき検索処理を行い、その結果を出力する機能を備えている。例えば情報検索システムは、検索キー２０９から入力された入力情報と類似する内容を記載する文書等を蓄積されている情報から検索したり、あるいは上記入力情報に関係する知識を蓄積されている文書から収集したり、またはパターン検索などを行う機能を備えている。

【０００９】

情報検索システムの処理内容を大きく分けると、蓄積時の処理と検索時の処理と出力時の処理とに分けることができる。蓄積時や検索時の処理は、主に検索処理装置１０１とインデクス情報データベース１０３や蓄積情報データベース１０４、ファイル情報データベース１０５で行われる。上記検索処理装置１０１とインデクス情報データベース１０３や蓄積情報データベース１０４、ファイル情報データベース１０５は、サーバマシンとして動作する。

30

【００１０】

蓄積時の処理に係る検索処理装置１０１の構成は、情報入力手段２０１と特徴抽出手段２０２、情報特定手段２０３、類似度計算手段２０４、情報蓄積手段２０５とである。また検索時の処理に係る構成は、検索キー入力手段２０９と検索キー該当情報検索手段２１０、活性値設定手段２１１、活性値伝播手段２１２、選択手段２１３とである。出力時の処理は主にクライアントマシンである操作装置１０２で行う。しかし、サーバ側のマシンである検索処理装置１０１で出力処理まで行うようにすることも可能である。

40

【００１１】

以下では、先ず蓄積処理、検索処理、出力処理の順に説明する。この説明では、“情報”として文書を例にとり、検索キー即ち検索のキーとなる、文書や単語などの情報（検索情報）に関連して、類似している文書群をまとめて検索する場合の例について説明する。

図３と図４に蓄積時の処理の流れを示す。図３のステップ３０１で、図２に示す情報入力手段２０１により文書、またはＵＲＬなどの文書やディレクトリへのポイントが情報検索システムに入力される。ステップ３０２で、特徴抽出手段２０２により入力された情報である文書中から特徴が抽出される。この特徴抽出動作では、文書を対象とする場合には形態素解析などの単語抽出手段を利用することができる。ステップ３０３で、抽出された

50

特徴を利用して、情報特定手段 2 0 3 により必要に応じて入力された情報がより細かな単位で捉えられる。例えば入力された情報が文章の場合には、情報特定手段は文書全体、または章や節、形式段落や語彙的連鎖などを単位とした意味的なまとまりを単位として特定する。このように情報特定手段 2 0 3 により、文章より細かな単位を特定することにより、これらの単位は意味的なまとまりを検索の単位とすることが可能となる。このことにより、検索対象をよりの確に選別できる。すなわち検索処理部分以外の冗長な部分が少なくなるため、検索キーを適切に与えれば、冗長な部分からの干渉の影響を受けにくい状態で検索対象を文書中の部分からよりの確に選別でき、精度の高い検索が行える。また文書全体が検索結果として得られるのではなく、文章のさらに細かい単位で検索できるので、検索結果を読む場合文書全体を読む必要がなく、文書中の関係部分を特定できる。従って検索者の検索後の負担を少なくすることができる。

10

#### 【 0 0 1 2 】

ステップ 3 0 4 と 3 0 5 で、情報特定手段 2 0 3 により細かな単位となった文書中の部分（以下単に文書と呼ぶ）のそれぞれについて、類似度計算手段 2 0 4 により類似度計算を行う。類似度計算手段 2 0 4 では、情報特定手段 2 0 3 から渡された文書と、それまでに蓄積されている文書との間で、内容の類似度合いである類似度を計算する。類似度を計算する際、文書の内容は複数の単語（特に重要な単語）で代表され、同様の単語の組を持つ文書は同様の内容を表す性質を利用する。この性質を利用すると、類似度は、それぞれの文書内の単語の重要度をベクトルの要素としたときのベクトルの内積や、係り受けなどの単語の相互の関係で表される。ここで、単語の重要度は、単語が出現する文書数で全文書数を割った値の関数、例えば  $\log$ （全文書数 / 注目語の出現文書数）といった関数の値と、単語がその文書内に出現する回数の積等で表される値である。

20

#### 【 0 0 1 3 】

次に計算量の削減に効果のある処理方法を説明する。類似度計算手段 2 0 4 により新しく情報が蓄えられる場合、それまでに蓄積されている情報との間で類似度を計算する必要がある。また、単語の重要度の変更などによる蓄積情報間の類似度の再計算の必要が生じたとき、蓄積情報間の類似度を計算するために、任意の蓄積情報 2 つの間の類似度を計算する必要がある。蓄積情報数を  $n$  とすると、計算量は  $n \times n$  となるため、蓄積情報数が多くなると多くの計算時間が必要となる。そこで、全ての情報間で類似度を計算するのではなく、あらかじめ情報をいくつかに分類し、分類された情報の集まりの間のみで類似度を計算することにより、計算時間の減少をはかることができる。さらに、処理を分散できる環境があれば、あらかじめ分類された情報間の類似度を並列に計算できるため、計算時間の短縮が可能となる。

30

#### 【 0 0 1 4 】

あらかじめ情報をいくつかに分類する際、文書内の重要語や共起語の偏りにより分類する方法を用いることができる。また、特許情報のように既に分類されているものはその分類情報を用いることができる。あらかじめ文書をいくつかのグループに分割する場合、複数のグループに属することが可能な文書は複数のグループに属すると考え、それぞれのグループ内で類似度を計算する。以上の処理により、類似度の計算時間を短縮することができる。

40

#### 【 0 0 1 5 】

次にインデクス情報データベース 1 0 3 へのデータの蓄積を説明する。情報蓄積手段 2 0 5 により、計算された閾値以上の類似度とともに文書を蓄積情報データベース 1 0 4 に登録し蓄積する。また抽出された特徴はインデクス情報データベース 1 0 3 に蓄積し、文書のファイル名の情報はファイル情報データベース 1 0 5 に蓄積する。

#### 【 0 0 1 6 】

蓄積情報データベース 1 0 4 に格納されるデータの形式を図 5 に示す。情報  $id$  は情報特定手段 2 0 3 により特定された情報に付けられる番号である。本説明のように検索対象として文書を用いた場合には、文書中の小部分に一意に付けられた番号が情報  $id$  となる。どの文書のどの位置から切り出された情報であるのかを、文書に一意に付けられた文書

50

i d と、先頭からの文字数によって情報の開始位置と終了位置で表す。

【0017】

文書内の単語とその重要度も保持され、類似度計算に用いられる。ここで、データ中の s t p はリストの終わりを表すための情報であり、例えば特定の値（マイナス1）である。他の図6と図7中の s t p も同様の情報であり、リストの終わりを表す。類似度計算手段204により計算された類似度は、図5に示すように、計算対象となった情報 i d と共にリンクの重み、例えば“重み値”として保持される。計算対象となった情報 i d は、互いに類似する蓄積情報としてリンクを有する。リンクの重みは、各情報 i d の類似度に応じた値で設定される。

【0018】

インデクス情報データベース103に格納されるデータの形式を図6に示す。単語 i d は特徴抽出手段202により抽出された特徴に一意に付けられる番号である。実際の特徴を意味する単語は文字列として保持する。検索用に使われる単語の出現位置は、出現する文書 i d と、その文書中での出現位置のリストによって表される。

【0019】

ファイル情報データベース105に格納されるデータの形式を図7に示す。文書 i d は蓄積される文書に一意に付けられる番号である。文書のファイル名は文字列で保持する。

【0020】

図8は検索処理の流れを示す。検索処理では、操作装置である操作部の情報入力手段、例えば検索キー入力手段209により検索を行うための検索情報が入力される。ステップ801で、検索情報に基づき、検索部である検索キー該当情報検索手段210により検索情報に該当する文書が検索される。検索部210は、特徴抽出手段202により検索情報から単語を抽出し、インデクス情報データベース103と蓄積情報データベース104とを参照し検索情報に該当する文書を検索する。検索情報に該当する文書とは、検索情報を含む文書や、検索情報の共起語や同義語を含む文書である。

【0021】

ステップ802で、活性値付与手段211により、検索情報に該当する文書に該当の度合いによって初期活性値が付与される。初期活性値は、文書中での検索情報にあたる単語の出現頻度の和や、文書中での検索情報にあたる単語の重要度の和、検索情報と文書とのベクトルの内積値などである。

【0022】

ステップ803で、検索情報に該当する文書全てに初期活性値が付与されると、活性値伝播手段212で付与された初期活性値にリンクの重みをかけた値をリンク先に伝播し、活性値を加算する処理を行う。伝播される活性値は、活性値付与手段211により付与された初期活性値である。活性値伝播処理は複数回繰り返してもよい。複数回繰り返す場合は、前回の活性値伝播処理によって伝播した活性値を初期活性値として同様の処理を繰り返す。

【0023】

ステップ804で、活性値伝播処理が終わったとき、活性状態判定手段213により閾値以上となる活性値を持つ情報が収集され、収集された情報はその活性値とともに出力処理に渡される。

【0024】

図9と図10を用いて、出力の処理の流れおよび動作を説明する。収集された文書は、文書へのポイントであるノード1001の形で出力手段214により、操作部102のディスプレイ等の出力装置215に出力される。ノードは他のノードと反発するが、類似している文書同士を表すノード同士はリンクの重みで表される類似度に比例した力で引き合うようにノードを二次元の画面上配置する。この配置により、類似した情報群がまとまるため、検索者は求める情報を視覚的に知ることができる。表示されるノードは、活性値に応じてノードを強調した形で表示する。ノードが選択されるとノード先の文書が別画面などに表示される。

10

20

30

40

50

## 【0025】

また、ノードを表示するとき、ノードを特徴付ける文書中の重要な単語であるラベル1002もノードとともに表示される。表示される単語（ラベル）は、例えば、単語（ラベル）である文書中の単語の重要度と、文書の活性値の積を、単語を含んでいる文書全てについて求めた和が大きな順に選択することで求める。ラベルは重要度の大きさを引力としてノードの近傍に配置される。複数の文書に共通に含まれるラベルは、それらの文書を表すノードの近傍に表示されるため、ノードの集まりが表す内容を知ることができる。表示されたラベルのうち、目的とする文書に関係のある単語の選択していくことで目的とする文書を絞り込むことができる。即ち検索処理に基づき重要な単語（ラベル）がその相互の関係に基づいて二次元的に表示され、この単語を選択する操作により検索結果から更に絞込みが可能となる。

10

## 【0026】

また利用者が検索の専門家でない場合でも抵抗感無く利用できるように、検索された文書をノード（節）の形で画面に表示するのではなく、リスト（一覧）形式で文書のタイトルや内容を別画面に表示しておき、関係する複数の単語（ラベル）が視覚的に目に付くようにを表示しておき、これら複数の単語（ラベル）から利用者がカットアンドトライの形式で、適切と思われる単語（ラベル）の選択し、これに基づき文書が選択されるようにすると、たいへん利用しやすくなる。

## 【0027】

この場合、ラベルの表示位置は上記のようにノード（表示はされない）の位置を決めた後に近傍に配置することで決定できる。また、ラベルとノードとの関連度合いをベクトルの要素としたラベルのベクトルを生成し、ベクトルの類似度合いを引力としてラベル同士を反発するようにラベルを配置してもよい。なお、出力された情報の表示や絞り込みなどの操作に関しては後述する。

20

## 【0028】

具体例を用いて類似情報検索の処理について説明する。例えば利用者が、次のような検索の要求を行った場合について述べる。検索の要求は、「コレステロールが低いとき原因は何か、どのような病気になるのか、医学的な情報が欲しい」である。この検索の前に予めデータベースに次のデータが保持されるとする。即ち次の3つの文書が蓄積情報データベース104に蓄積されている場合を示す。

30

文書A「コレステロールが低いと脳出血が起こる危険性がある。」

文書B「低脂血症のうえ高血圧も合併していると血管が弱まり脳出血の危険性が増すため、甲状腺や栄養状態をチェックする。」

文書C「低コレステロール食品は、豆腐・うどん・かまぼこなどである。」

ここで文書Aから文章Cがいずれも一文で記載されているが、これは説明を簡潔に理解し易くするために短くしているのであり、特に一文である必要はない。文書蓄積時に、蓄積した文書から特徴が抽出される。特徴抽出手段202は、形態素解析を用い名詞・形容詞の単語を抽出すると、各蓄積文書から抽出される特徴となる単語は以下の特徴AからCとして示す単語となる。

## 【0029】

40

特徴Aは「コレステロール、低い、脳出血、危険性」である。特徴Bは「低脂血症、高血圧、合併、血管、脳出血、危険性、甲状腺、栄養、状態、チェック」である。特徴Cは「コレステロール、食品、豆腐、うどん、かまぼこ」である。以上の情報は図7に示す形式で、インデクス情報データベース103に蓄積される。

## 【0030】

情報特定手段203により、形式段落や、同様の語が続く範囲である語彙的連鎖等の意味的なまとまりや、文書全体が情報の単位となる。例としてあげた文書AからCの場合、この一文が例えば単位となる。情報特定手段203により特定した単位について、類似度計算手段204により類似度計算を行う。類似度計算手段204での情報間の類似度の計算方法としては、文書中の単語の重要度（出現頻度やtf-idf法等により求まる値）

50



を要素としたベクトル同士の内積などを用いることができる。

類似度を文書中の単語の出現頻度を要素としたベクトルの内積として、文書 A から C についてそれぞれの文書間の類似度を計算すると次のようになる。

類似度  $A B = 2$  (脳出血 + 危険性)

類似度  $A C = 1$  (コレステロール)

類似度  $B C = 0$

となる。情報蓄積手段 205 で、この類似度を文書間のリンクの重み(重み値)として蓄積情報データベース 104 に保持しておく。計算対象となった文書は、互いに類似する蓄積情報としてリンクを有する。

#### 【0031】

次に具体的な検索を説明する。文書 A から C に対して、「コレステロール」という検索情報により検索が行われる場合を例として説明する。検索キー該当情報検索手段 210 により、「コレステロール」を含む文書 A と文書 C が検索される。活性値付与手段 211 により、該当の度合いに応じて活性値を付与する。この例では、文書中の検索情報にあたる単語の出現頻度によって活性値を付与する方法をとる。この方法により、文書 A、文書 C ともに活性値 1.0 が付与される。付与された活性値は、活性値伝播手段 212 によりリンク先の文書に伝播される。伝播される値は、リンクもとの活性値とリンクの重みとの積で表される値である。文書 A からは文書 B に 2.0 の値が、文書 C に 1.0 の値が伝播、加算される。文書 C からは文書 A に 1.0 の値が伝播、加算される。なお、活性値の伝播は並列に行われる。

#### 【0032】

活性値の伝播が終わった後、活性状態判定手段 213 により最終的な活性状態が判定される。この例では、(文書 A の活性値)は 2.0、(文書 B の活性値)は 2.0、(文書 C の活性値)は 2.0、となり、活性値の大きな文書 A、B、C が検索結果として収集対象となる。これにより、従来の検索では検索できない文書 B も検索結果として収集できる。

次に出力手段 214 の処理を説明する。出力手段 214 は、文書 A から C に関する情報 id と活性値を受け取り、ディスプレイ等の出力装置 215 に出力を行う。このとき、情報である文書はノードとして表され、このノードがクリックやドラッグなどにより選択されると、別画面にあるいは画面上の異なる表示エリアに検索された情報の内容が表示される。

#### 【0033】

ノードは、各ノードが反発するように、そしてリンクの重みにしたがって類似したノード同士が引き合うように配置される。例としてあげた文書 A から C では、文書 A - B 間に「脳出血・危険性」に関して、文書 A - C 間に「コレステロール」に関してリンクが張られているため、図 10 に示すように、文書 A と文書 B の近傍に単語「脳出血・危険性」が表示される。また文書 A と文書 C の近傍に「コレステロール」配置される。図では、ノード間のリンクを点線で表している。

#### 【0034】

また、各文書で重要な単語も文書の近傍に表示され、文書がどのような内容を持っているかを表す。図 10 では、特徴抽出処理によって抽出された単語をいくつか表示している。これらの単語のうち、医学的な内容を述べている「脳出血」や「高血圧」といった、検索者が目的とする情報に関連する単語を選択していくことで、コレステロールが低いときの医学的な情報が文書 A から、そして関連する情報が文書 B から得られることがわかり、ユーザは目的とする情報の集まりを容易に理解できる。なお、この画面ではノードとラベルの関係がわかりやすいようにノード 1001 とラベル 1002 の両方を表示した例を示したが、ノードを別画面にあるいは画面上の他の表示エリアに表示することが可能である。この方が操作がやりやすくなる場合がある。またノードの出力を、文書のタイトルを出力する形式で表示画面の他のエリアに表示することが可能であり、この場合例えばリスト形式で表示し、一方ず 10 の表示エリアにはノードを表示しない形式でラベルを表示する

10

20

30

40

50

ことができる。この方が理解し得やすく、またその後の操作が簡単になる場合がある。このような出力画面の詳細な説明については次に説明する。

#### 【0035】

上記図1から図10を用いて説明した検索処理や蓄積処理において、文章を例として説明しているが文献を対象としても基本動作は同じである。文章のほうがよりきめ細かい検索結果が得られ、利用者が利用し易いので文章を例とした。

#### 【0036】

次に複数の利用者が操作装置102である操作部から行う検索処理について説明する。但し基本的な動作は上記動作と同じである。まずステップ1101で、クライアントが行う類似情報検索の操作を示すフローを図11に示す。クライアント側のマシンである操作装置102において情報検索のプログラムを立ち上げると、すなわち検索のためのアイコンを選択するなどにより検索のためのプログラムを実行すると、図12のような画面がディスプレイに表示される。この画面には以下の3つの表示エリアおよび操作ボタンが設けられており、利用者が利用しやすくなっている。表示エリア1201は、情報を検索するために検索情報を入力するフィールド即ち表示エリアである。表示エリア1201に検索のための情報が利用者の入力に応じて順に表示される。この表示内容を見て訂正や削除、追加が可能である。検索開始の指示を送る操作ボタン1202が表示される。この操作ボタンは他の印刷やその他のメニュー共にエリア1201の更に上側の位置に設けても良い。この実施の形態のように表示した場合、他のメニューより大きく表示でき、不慣れな人にも利用し易い効果がある。検索を実施するための検索ボタン1202を選択すると、これに基づき検索実行の指示が送られ、検索が始まる。

#### 【0037】

表示エリア1203は、検索結果の情報を表示するためのエリアであり、また表示エリア1204はラベルを表示するためのエリアである。検索結果とラベルは対比した形で表示でき、ラベルを選択しながら更に絞り込む動作がたいへん行いやすくなる。上述の説明例を図12を用いて説明すると、エリア1201に、検索情報である検索要求「コレステロールが低いとき原因は何か、どのような病気になるのか、医学的な情報が欲しい」を入力する。入力に基づくエリア1201の検索情報の表示内容を確認して、検索ボタン1202を選択する。この検索指示に基づきエリア1203には図10のノードである検索結果1001がリストの形式で表示される。即ち文章AとBとCとが表示される。それと共にエリア1204に、検索情報に関連する単語であるラベルが表示される。即ち単語「豆腐」、「かまぼこ」、「コレステロール」、「危険性」、「脳出血」、「血管」、「甲状腺」、「高血圧」が表示される。

#### 【0038】

このラベル表示の中から単語「豆腐」を選択すると文章A、B、Cの中から文章Cが選択される。また単語「コレステロール」を選択すると文章AとCが選択される。

更に他の具体例を図13を用いて説明する。図のエリア1201に、検索情報として「品質 計画」を入力する。図13は上記入力に基づいて検索情報が表示されている例である。図11のステップ1102の説明の如く、表示されている検索ボタン1202を選択する。

#### 【0039】

ここで、検索情報として入力された言葉は、「品質」と「計画」が両方含まれる文書、または、どちらかが含まれる文書を期待して入力したものとする。もちろん前述のように「品質」、「計画」に関連する類似文書群をまとめて検索する検索手法を用いてもよい。検索ボタンが押されるとステップ1102、検索情報がサーバに送信されステップ1103検索処理が実行される。ステップ1104で、検索結果のデータがサーバからクライアント側の操作部102に送られる。

#### 【0040】

ステップ1105で、サーバ側の装置101から受信した検索結果のデータを基に図14の画面が表示される。図14の表示エリア1203には、検索情報「品質 計画」に関

10

20

30

40

50

する検索結果が表示される。検索結果表示エリア 1 2 0 1 に表示されている「絞り込み情報」1 4 0 2 の画面下方に検索された文書のタイトル一覧 1 4 0 3 が示されている。タイトルのみではなく、文書の要約や検索情報と文書との適合度などを追加して表示してもよい。利用者はこの絞り込まれた文書一覧から目的とする文書を選び、ファイル名をクリックするなどの選択する操作に基づき、選択されたファイルの表示や取得等が行える。

#### 【0 0 4 1】

図 1 4 に表示の「関連情報」1 4 0 4 には検索されたが絞り込みの対象にならなかった検索結果、本例では文書（選択されたラベル全てを含まない文書）が表示される。この図では、検索結果として収集された文書は検索情報を全て含む文書としたため、全ての文書が「絞り込み情報」となっている。検索方法として、上述の類似情報検索方法などを用いた場合は、検索直後に、検索情報を全て含んでいないが関連している文書が「関連情報」の下に表示されることとなる。

10

#### 【0 0 4 2】

「絞り込み情報」と「関連情報」の下の文書一覧は「絞り込み情報」、「関連情報」を選択すると、例えばクリックするなどの操作で表示や非表示の切り替えが可能となる。

ラベル表示画面 1 2 0 4 には、ノードである文書を特徴付ける単語、例えばラベル 1 4 0 5 が表示される。表示されるラベルは、検索されたノード（文書）の 1 つ、または複数の文書に対して、それらの特徴付けるものとして自動的に選ばれた単語である。単語はそれを特長付けるアイコンであってもかまわない。表示したくないラベルはあらかじめデータベースに登録しておくことにより表示対象からはずすことができる。

20

#### 【0 0 4 3】

ラベルはクリックやドラッグにより選択することが可能である。また、選択されたラベルを再選択することで選択を解除することも可能である。図に示す検索直後の状態では、検索情報として入力された単語がラベルとして選択されている。選択されたラベルは選択状態にあることを示すように視覚的に区別できる表示、即ち強調表示される。この例では太線で囲んである。

ラベルの選択を行うことで、選択されたラベルを含む文書が絞り込まれ、絞り込まれた文書とそれ以外の文書が分類され、その結果が画面左側の情報表示エリア 1 2 0 3 に表示される。

#### 【0 0 4 4】

30

ラベルの表示位置は、関連する（同様の文書を特徴付ける）ラベル同士は近傍に、関連が薄くなるに従い距離が大きくなるように配置される。このため、検索者は目的とする文書を特徴付ける複数のラベルを一度に見つけることができるため、カルタ取のように画面の隅々までラベルを探す労力を減らすことができる。ラベルは、表示画面上で、二次元的に相対的な位置関係を保ちながら、ラベル表示画面内に収まる範囲で最大に拡大されて表示される。また、ラベルを手動で検索者の好むように動かすことも可能である。

ラベル表示エリア 1 2 0 4 に表示されたラベルの中から適切なラベルを選択することで、文書を絞り込むことができる（図 1 1 のステップ 1 1 0 6）。図 1 5 はその様子を示したものである。例えば、図 1 4 の「品質 計画」の検索結果のうち、作りこみ不良の低減と摘出不良の増大のために品質計画について知りたい場合、ラベル表示エリアのラベルからラベル「不良」と「摘出」とを順に選択することで目的とする「品質計画書」を抽出できる様子を示している。これにより検索者は、絞り込み前の検索された文書の内容を順に見ていき目的とする情報を探す手間や、新たな検索情報を入力し絞り込み検索を行う労力を減らすことができる。

40

#### 【0 0 4 5】

図 1 5 は、図 1 4 のラベル「品質」1 5 0 1 とラベル「計画」1 5 0 2 が選択された状態からラベル「不良」1 5 0 3 を選択した様子を示している。「絞り込み情報」1 5 0 4 には選択したラベル「品質や計画や不良」を全て含む文書 1 5 0 5 が表示されている。「関連情報」1 5 0 6 には検索された文書のうち「絞り込み情報」の文書以外の文書 1 5 0 7 が表示されている。絞り込みからは外れているが、表示したいあるいはこれらの情報を

50

取得したい場合は、このエリアから操作が可能である。

#### 【 0 0 4 6 】

図 1 6 ではさらにラベル「抽出」1 6 0 1 を選択した状態を示す。なお、絞り込み情報として、選択したラベル全てを含む文書の代わりに、ラベルを 1 つでも含む文書としてもよい。指定方法は一般的な検索方法と同様に、検索情報中に「+」や「|」などの記号で検索情報を区切る方法等で指定可能である。

同様に、品質計画のうち、ISO 9001 に備えて品質システムについて知りたい場合を図 1 7 に示す。ラベル表示画面から「品質」1 7 0 1、「計画」1 7 0 2、「内部」1 7 0 3、「監査」1 7 0 4 とを選択することで目的とする「品質システムの維持管理」1 7 0 5 等を抽出することができる。

10

初期状態では選択されている検索情報に相当するラベルの選択を解除し、ラベル表示画面に表示されているラベルのうちからより適切なラベルを選択することで、絞り込み情報により適切なノードを集めることも可能である。

#### 【 0 0 4 7 】

図 1 5 から図 1 7 まで示した絞り込みによって、最初から品質計画書を得る目的で「品質 計画 不良 抽出」といった目的文書を特定する検索情報が思いつかない場合でも、思いついた検索情報、例えば「品質 計画」で検索を行った後、目的とする文書に関連するラベルを選択していくことで目的とするノード、例えば文書を得ることができる。この操作により、検索者が検索時に必要な文書について分類を進めることができるため、あらかじめクラスタリング等の手法で文書を分類しておく場合に比べ、検索者の目的とする文書のみの必要十分な文書群が得られやすい。

20

#### 【 0 0 4 8 】

図 1 3 において、検索情報を入力せずに検索を行うなどの手段により、検索対象の文書の傾向を知ることにも可能である。検索対象の文書すべてを対象にした代表となるラベルを提示することで、検索者はどのような文書が蓄積されているのかを知ることができる。

分類結果をワープロソフトや表形式ソフト、html や xml などの各種ファイル形式で出力することで、文書の傾向をまとめた資料、例えば社内文書の分類資料、などの作成も容易になる。

#### 【 0 0 4 9 】

図 1 4 から図 1 7 において、ラベルを選択する場合、選択したラベルを検索情報入力フィールドに反映するなどの手段により、選択したラベルで再検索を行うことも可能である。この操作により検索者は再び検索情報を入力する手間がなく、選択したラベルを中心とする文書の集まりと、検索された文書の特徴付けるラベル、つまり選択されたラベルに関連するラベル、の集まりを得ることができるため、検索者の興味ある文書と文書の傾向を知ることができる。

30

逆に検索された文書を 1 つ、または、複数選択することで、関連しているラベルを選択状態にすることもできる。これにより選択された文書に含まれる情報がラベルで代表されるため、文書の内容を読む前に内容を予想することが可能となり、目的とする文書の発見が容易になるとともに、不必要な文書を読むことが抑制できる。

#### 【 0 0 5 0 】

40

図 1 8 は、図 1 2 を用いて説明した実施の形態にさらにいくつかの機能を加えたときの立ち上げ画面を示している。検索情報を含まない文書も検索することができる類似検索機能を加え、類似検索を行うボタン 1 8 0 1 を選択すると、知りたい文書を特定する検索情報が思い浮かばないときでも、ある程度関連する検索情報により類似する文書を広く多く検索することが可能となる。なお、多くの文書が検索されてもラベルを選択することで文書を絞り込むことができるため、検索結果に目的とする文書があるのかわからない、というのを抑制することができる。

#### 【 0 0 5 1 】

また、検索対象となる文書のソースを選択する検索対象情報選択フィールド 1 8 0 2 の選択により検索対象を切り替えることで、その検索対象についてのみ文書で、不要な検

50

索結果を得ることを抑制することができる。また、検索対象となる文書のソースを切り替えの検索対象の傾向分析を行えば、検索対象ごとにどのような文書が蓄積されているのを知ることができる。

また、表示するラベルの属性を選択するラベル属性選択フィールド 1 8 0 3 により表示するラベルの属性（例えば地名や人名など）を切り替えることで、表示するラベルの属性を指定できるため、検索者がラベルを選択することが容易になる。例えば、牛肉の種類について文書を求めている場合はラベルの属性として地名を選択し、牛肉を検索情報として検索を行えば、図 1 9 に示すように牛肉に関連する地名がラベル 1 9 0 1 として表示されるため、検索者が目的とする文書を選択することが容易になる。

#### 【 0 0 5 2 】

10

図 1 8 の情報表示画面の「絞り込み情報」、「関連情報」それぞれに表示する文書の順序を設定する操作エリア 1 8 0 4 を設け、この設定により、検索結果のスコア順、選択されたラベルを多く含む順（図ではラベルランク順）、日付の新しい順に表示する日付順などが選択できる。

#### 【 0 0 5 3 】

図 2 0 は、情報表示画面の変形例である。複数の文書をまとめて選択したい場合には、文書の前のチェックボックス 2 0 0 1 をチェックすることで絞り込み対象ラベルを選択すれば、キーボードでの操作や図 2 3 の右クリックメニュー 2 3 0 1 の中に選択したノードである文書の表示や取得などの項目を追加することで実現することができる。

#### 【 0 0 5 4 】

20

図 2 1 は、情報表示画面にラベル 2 1 0 1 や 2 1 0 2 も組み込んだ例を示している。ラベル表示画面が表示できないような簡易な検索画面などでも、ラベルの前のチェックボックス 2 1 0 3 をチェックしていくことで、ラベル表示画面でラベルを選択することと同様の機能を使用できる。

また、この画面のラベル一覧 2 1 0 1 は「ラベル」2 1 0 2 をクリックするなどの手段で表示や非表示の切り替えが可能である。

#### 【 0 0 5 5 】

図 2 2 は、図 2 1 のようにラベル 2 2 0 1 が表示された状態で、ラベルが選択されるとラベルに関連するノードである文書 2 2 0 2 が表示される様子を示したものである。これにより検索者はラベルに関連する文書を容易に知ることができる。この機能はラベル表示画面でもシフトキーを押しながらラベルをクリックするなどの手段によりポップアップウィンドウなどで表示することにより実現することもできる。ラベルに関連する文書が得られると、ディレクトリ構造のファイル（文書）の分類を行うことも容易となる。例えば社内文書の管理者などが、文書の分類やディレクトリ検索のための文書分類のメンテナンスに役立てることができる。

30

#### 【 0 0 5 6 】

図 2 3 は、情報表示画面において第 1 の操作、例えば右クリックを押したときの画面である。右クリックの事例の如き第 1 の操作によりメニュー 2 3 0 1 が表示され、選択した文書の操作に関する機能が表示されている。例では「品質計画書」の文書 2 3 0 2 に対して、以下の機能が選択可能である。1.「この文書を表示する」：選択された文書の内容を画面に表示する。2.「詳細情報」：文書のタイトルや作成者、文書の先頭部分や選択されたラベルに関する要約などの詳細情報表示機能。3.「フォントサイズ変更」：フォントの大きさを変える機能。4.「色変更」：表示している文書のタイトルの背景色やフォントカラーを変換する機能。5.「類似情報検索」：選択された文書を検索情報として類似する文書を検索する機能。なお、これらの機能はキーボードでの操作も可能である。

40

#### 【 0 0 5 7 】

図 2 4 はラベル表示画面の変形例である。ラベル 2 4 0 1 や 2 4 0 2 のうち、検索情報に相当するラベル 2 4 0 1 を円 2 4 0 3 で囲むことで目立たせたものである。これにより検索者は検索情報のラベルを容易に見つけることができ、検索情報と他のラベルとの関連も理解することができる。

50

## 【0058】

図25はラベル表示画面の変形例である。各ラベル2501が示している文書の数やラベルの後ろにカッコ付きで表したものである。これにより検索者はラベルが示す文書の分布の様子を知ることができる。また、文書の数やラベルの重要度（スコア）を表示してもよい。また図26はラベル表示画面の変形例である。重要なラベル2601ほど大きく表したものである。図27はラベル表示画面の変形例である。ラベルの重要度やラベルの属性、例えば地名や人名など、あるいはラベル同士2701, 2702, 2703の関連度により色分けしたものである。

## 【0059】

図28はラベル表示画面の変形例である。ラベル2801の表示を二次元ではなく三次元により、ラベル間の相互の位置関係がよりわかりやすくなり、検索者は関連するラベル同士を選びやすくなる。この画面は画面内でドラッグすることにより任意の方向に回転することができるため、奥行き方向の距離も視覚的に理解できる。

## 【0060】

図29はラベル表示画面で右クリックを押したときの画面である。右クリックによりメニュー2901が表示され、選択したラベルの操作に関する機能が上側に、全体の操作に関する機能が下側に表示されている。例ではラベル「内部」2902のラベルに対して、以下の機能が選択可能である。1.「このラベルを表示しない」：選択ラベルを非表示にする機能。そのかわりに別のラベルを出すことも可能であり、検索者は自分の知りたい文書に関するラベルの選択を続けていくことができる。2.「ラベルの詳細情報」：ラベルが表す文書数や重要度などの詳細情報表示機能。3.「フォントサイズ変更」：フォントの大きさを変える機能。4.「色変更」：ラベルの背景色やフォントカラーを変換する機能。5. また全体の操作として、以下の機能が選択可能である。6.「再検索」：選択されたラベルで再検索を行う機能。7.「ラベルの表示数変更」：表示するラベルの上限を変更する機能。ラベル表示画面近傍にラベル数を変更するためのラジオボタンやスライダー等を用意することでも実現可能。8.「表示ラベルの追加」：画面に表示されていないが表示したいラベルを追加することができる機能。非表示にしたラベルの再表示や、もともと表示されていないラベルを追加して表示する機能。9.「UNDO」：ラベルの選択状態を1つ前の状態に戻す機能。10.「前回の検索結果」：前回検索したときの状態に戻す機能。なお、これらの機能はキーボードでの操作も可能である。

## 【0061】

ウィンドウの拡大縮小を行ったとき、表示されるラベル同士の相対的な位置関係はそのままにウィンドウの大きさにともなってラベル間の距離を拡大縮小することで検索者がラベルを見失うことを防ぐことができる。また、図30のように表示されるラベル同士の相対的な位置関係はそのままにウィンドウを拡大縮小することで、ウィンドウを拡大したときにできる新たな領域に新たなラベル3001を追加して提示することで、検索者のラベル選択の幅を広げることができる。

さらに、ラベル表示画面内をドラッグで選択した領域を拡大して表示することもできる。これにより、検索者は興味あるラベルの集まりのみを拡大して見ることができる。

## 【0062】

図31は検索された文書もノード3101としてラベル表示画面に表示した例である。この画面によりラベルと文書との関連を視覚的に理解することができる。ラベル表示画面に表示された文書を選択、例えばシフトキーを押しながらクリックやドラッグなど、することで、選択された文書の取得や表示、選択された文書を対象とした表示領域の拡大、それらの文書についてのみのラベルの再表示などが行える。

表示する文書数は、ラベル表示画面で右クリックを押されたときのメニュー（図29）に追加するなどの手段により調整可能である。また、選択されたラベルに関する文書のみを表示するようにもできる。

## 【0063】

図32は文書とラベルを表3201の形式であらわしたものである。文書、ラベルを選

10

20

30

40

50

押し、右クリックなどの操作で、この画面に切り替えることで文書とラベルの関係を知ることができるため、検索者が目的とする文書を選びやすくなり、また、文書の分析も容易になる。

#### 【 0 0 6 4 】

図 3 3 から 3 7 を用いて操作についてさらに説明する。ここで同じ参照番号は先に説明したものと同じ内容を示す。図 3 3 では品質計画で検索した結果、抽出された文書をエリア 1 2 0 3 に表示している。上記品質計画の検索結果に基づいて抽出されたラベルをエリア 1 2 0 4 に表示している。ここで選択するラベルを変えるとエリア 1 2 0 3 に表示される文書が変更される。この状況を図 3 4 で説明する。図 3 4 では選択するラベルを増やした例で、この結果文書が絞り込まれている。逆にラベルを減らすとラベルによって絞り込まれた文書が増加する。

10

#### 【 0 0 6 5 】

図 3 5 と 3 6 はラベルの表示を見易くするための部分拡大表示の例である。図 3 5 で拡大したい領域をマウスのドラッグなどにより選択すると、図 3 6 に示すように、その領域が拡大表示される。なお元に戻す場合は、ラベルを含まない領域を選択する。その他の方法でも良い。

#### 【 0 0 6 6 】

また図 3 3 の表示状態で、ラベルの表示数を増減したい場合は、ラベル数選択エリア 3 3 0 3 でラベル数を指定することで可能になる。図 3 7 は図 3 3 のラベル数選択エリア 3 3 0 3 でラベル数 2 5 を指定した場合である。より詳しい、より適切な絞り込みが可能になる。

20

#### 【 0 0 6 7 】

上記説明では、ラベルとして単語を例示して説明した。単語は内容を端的に示すことで優れているが、ラベルとして単語以外にもアイコンやマークを使用できる。

以上の実施の形態によれば、検索情報の準備に苦勞せずとも類似情報を容易に一度に収集することが可能である。また収集された類似情報群から目的とする情報を選び出すのも容易である。

#### 【 発明の効果 】

#### 【 0 0 6 8 】

本発明によれば検索処理の操作性が向上する。

30

#### 【 図面の簡単な説明 】

#### 【 0 0 6 9 】

【 図 1 】 類似情報検索装置のシステム構成図である。

【 図 2 】 システム全体の処理フローチャートである。

【 図 3 】 蓄積処理のフローチャートである。

【 図 4 】 類似度計算処理のフローチャートである。

【 図 5 】 蓄積情報データベースのデータの説明図である。

【 図 6 】 インデクス情報データベースのデータの説明図である。

【 図 7 】 ファイル情報データベースのデータの説明図である。

【 図 8 】 検索処理のフローチャートである。

40

【 図 9 】 出力処理のフローチャートである。

【 図 1 0 】 出力画面の説明図である。

【 図 1 1 】 クライアント側の処理フローチャートである。

【 図 1 2 】 検索画面の説明図である。

【 図 1 3 】 検索の説明図である。

【 図 1 4 】 出力画面の説明図である。

【 図 1 5 】 出力画面の説明図である。

【 図 1 6 】 出力画面の説明図である。

【 図 1 7 】 出力画面の説明図である。

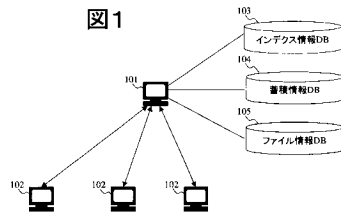
【 図 1 8 】 出力画面の説明図である。

50

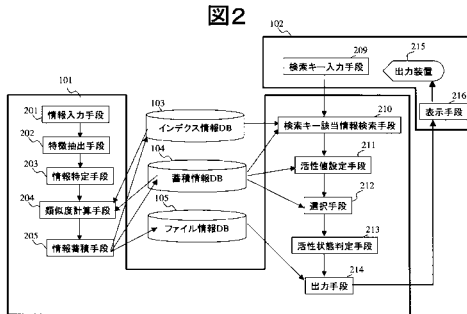
【図 19】	出力画面の説明図である。	
【図 20】	情報表示画面の説明図である。	
【図 21】	情報表示画面の説明図である。	
【図 22】	情報表示画面の説明図である。	
【図 23】	情報表示画面の説明図である。	
【図 24】	ラベル表示画面の説明図である。	
【図 25】	ラベル表示画面の説明図である。	
【図 26】	ラベル表示画面の説明図である。	
【図 27】	ラベル表示画面の説明図である。	
【図 28】	ラベル表示画面の説明図である。	10
【図 29】	ラベル表示画面の説明図である。	
【図 30】	出力画面の説明図である。	
【図 31】	出力画面の説明図である。	
【図 32】	出力画面の説明図である。	
【図 33】	情報表示画面の説明図である。	
【図 34】	情報表示画面の説明図である。	
【図 35】	情報表示画面の説明図である。	
【図 36】	情報表示画面の説明図である。	
【図 37】	情報表示画面の説明図である。	
【符号の説明】		20
【0070】		
101	サーバ	
102	クライアント	
102	インデクス情報データベース	
103	蓄積情報データベース	
104	ファイル情報データベース	
201	情報入力手段	
202	特徴抽出手段	
203	情報特定手段	
204	類似度計算手段	30
205	情報蓄積手段	
209	検索キー入力手段	
210	検索キー該当情報検索手段	
211	活性値付与手段	
212	活性値伝播手段	
213	活性状態判定手段	
214	出力手段	
215	ディスプレイ等の出力装置	
216	表示手段	
1403	絞り込まれた文書のタイトル	40
1801	類似検索開始ボタン	
2403	円状のマーカー	



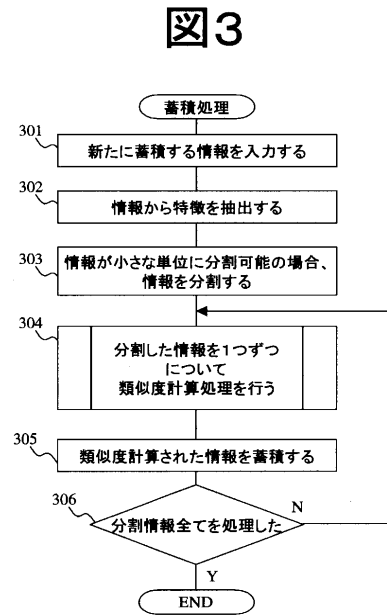
【図 1】



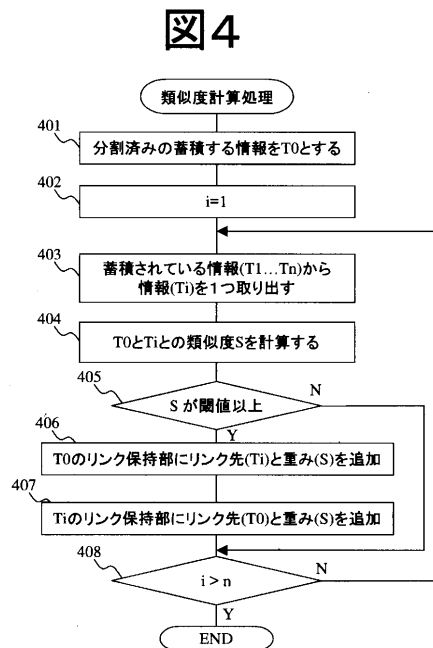
【図 2】



【図 3】



【図 4】



【図 5】

図 5

情報id	情報	類似度計算のための情報	リンクの重み
0001	(文書id, 情報開始位置, 情報終了位置)	(単語id, 重要度), (単語id, 重要度), ... sp	(情報id, 重み), (情報id, 重み), ... sp
...	...	...	...

【図 6】

図 6

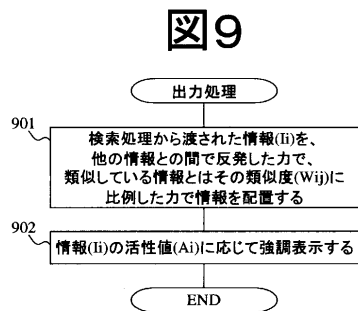
単語id	単語	単語の出現位置
0001	文字列	(文書id, (位置, 位置, ..., sp)), (文書id, (位置, 位置, ..., sp)), ... sp
...	...	...

【図 7】

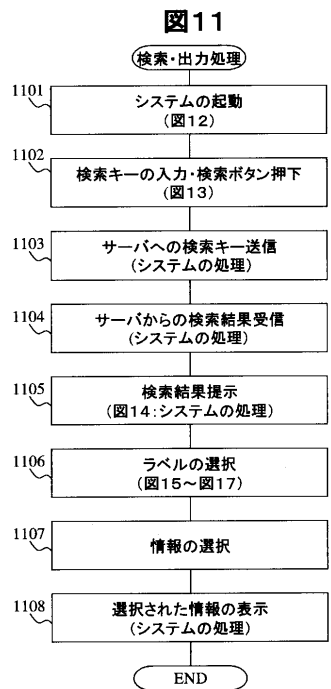
図 7

文書id	ファイル名
0001	文字列
...	...

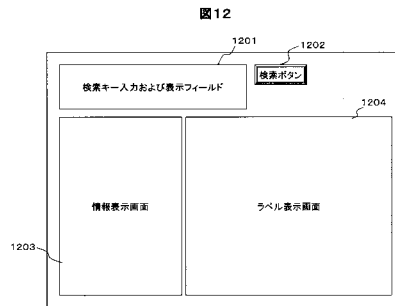
【 図 9 】



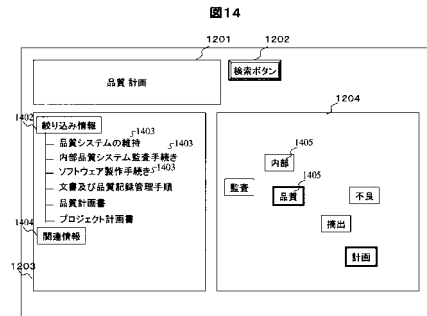
【 図 1 1 】



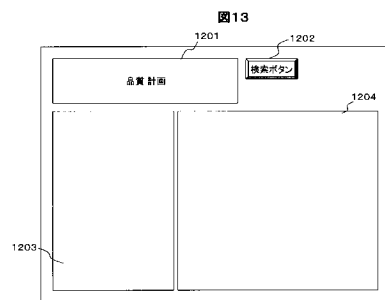
【図 12】



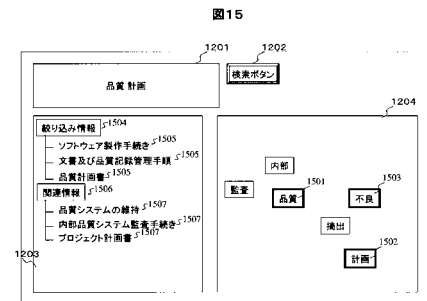
【図 14】



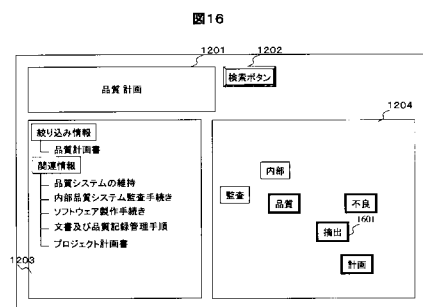
【図 13】



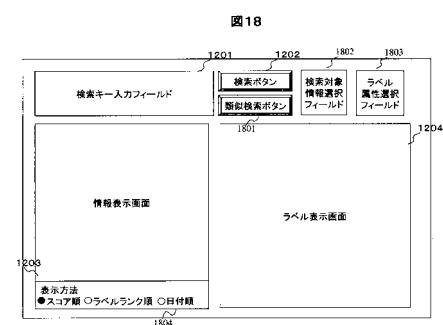
【図 15】



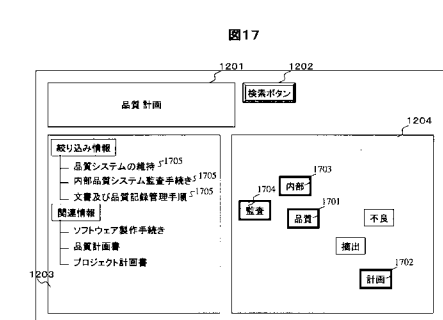
【図 16】



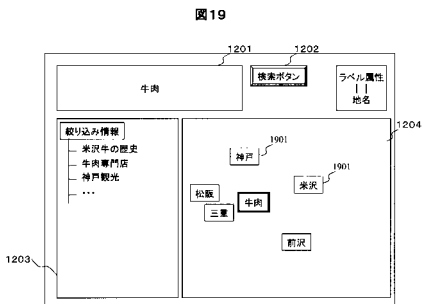
【図 18】



【図 17】

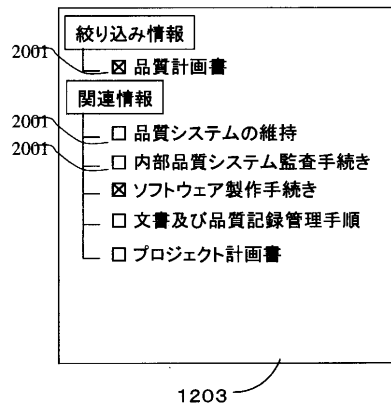


【図 19】



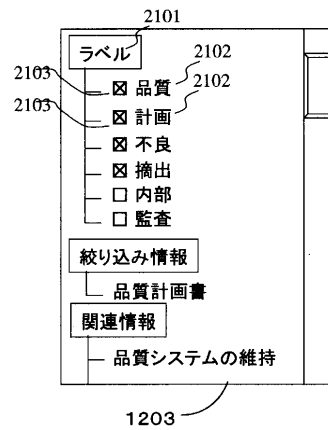
【図 20】

図20



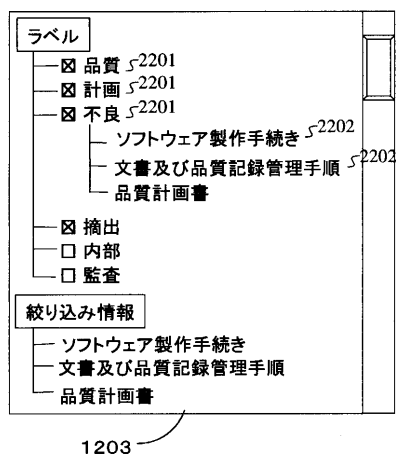
【図 21】

図21



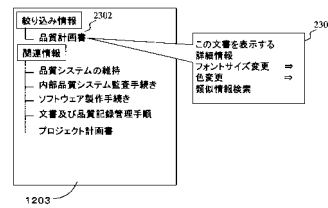
【図 22】

図22



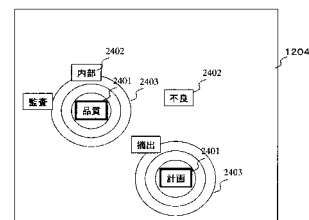
【図 23】

図23

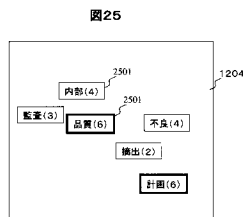


【図 24】

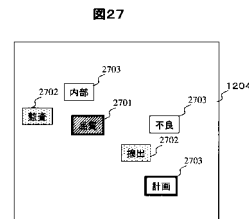
図24



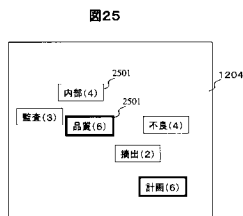
【図 25】



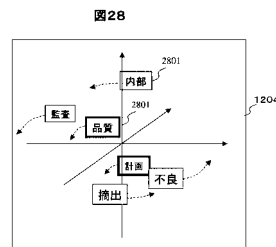
【図 27】



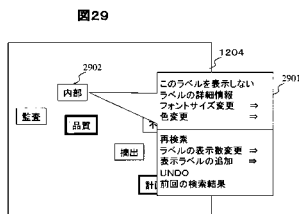
【図 26】



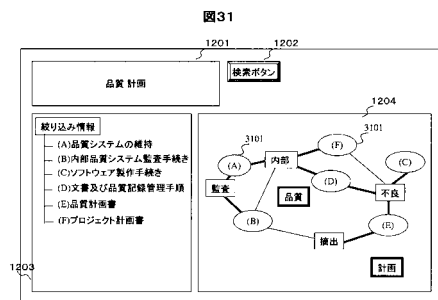
【図 28】



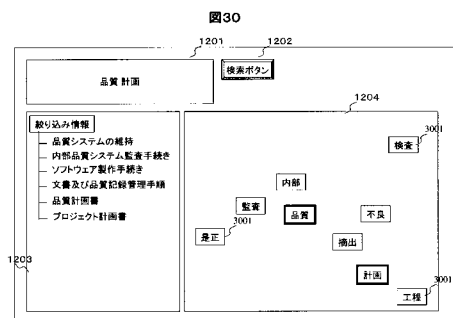
【図 29】



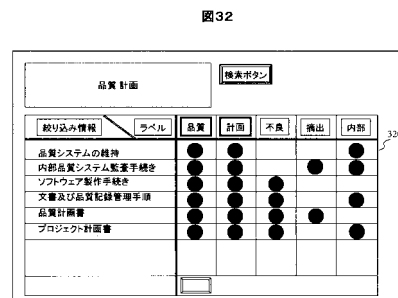
【図 31】



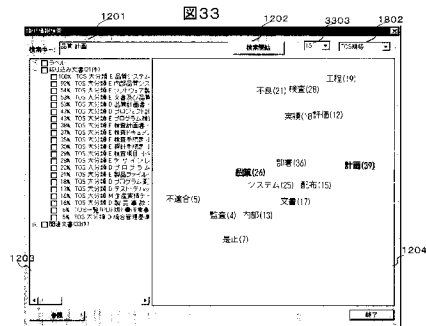
【図 30】



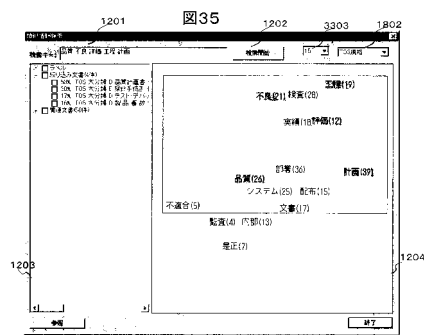
【図 32】



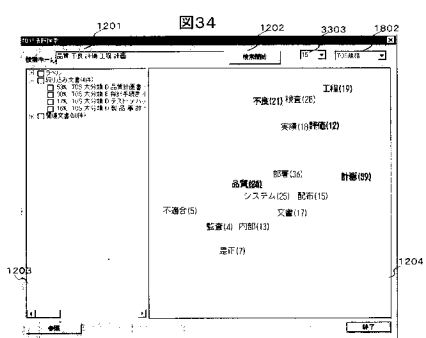
【図 3 3】



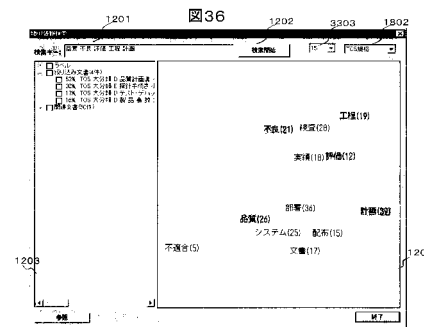
【図 3 5】



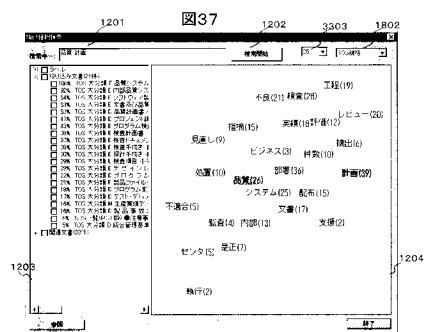
【図 3 4】



【図 3 6】



【図 3 7】



---

フロントページの続き

(72)発明者 佐藤 俊也

宮城県仙台市青葉区本町二丁目16番10号 日立東北ソフトウェア株式会社内

審査官 鈴木 和樹

(56)参考文献 特開平10-334105(JP,A)

特開2000-331032(JP,A)

丹羽芳樹, 動的な共起解析を用いた対話的文書検索支援, 情報処理学会研究報告96-FI-43, 1996年9月13日, 第96巻, 第88号, p.41-48

舘村純一, インタラクティブ視覚化による文献集合からの情報獲得支援, 日本ソフトウェア科学会第13回大会論文集, 1996年9月27日, p.365-368

舘村純一, DocSpace: 文献空間のインタラクティブ視覚化, インタラクティブシステムとソフトウェアIV: 日本ソフトウェア科学会 WISS'96, 近代科学社, 1996年12月4日, p.11-20

野口喜洋、外4名, 検索型ナビゲーションを実現したホームページ知的検索システムの開発, 利用者指向の情報システムシンポジウム論文集, 社団法人情報処理学会, 1996年12月5日, 第96巻, 第11号, p.91-98

村田剛志, 対象の参照度に基づく視覚化手法, 第43回知識ベースシステム研究会資料SIG-KBS-9803, 社団法人人工知能学会, 1999年3月23日, p.99-104

Junichi Tatemura外1名, DocSpace: an Interactive Document Space Visualizer that Combines Querying and Navigation, Proc. of International Symposium on Research, Development and Practice in Digital Libraries 1997 (ISDL'97), 1997年12月18日, p.107-114

塚原朋哉、外2名, 情報の体系化と視覚的検索方法, 日立TO技報, 日立東北ソフトウェア株式会社, 2002年12月12日, 第8号, p.5-11

(58)調査した分野(Int.Cl., DB名)

G06F 17/30

JSTPlus(JDream2)

NR1サイバーパテント