

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2006-31577

(P2006-31577A)

(43) 公開日 平成18年2月2日(2006.2.2)

(51) Int. Cl. F I テーマコード (参考)
G06F 17/30 (2006.01) G06F 17/30 220Z 5B075
 G06F 17/30 170A
 G06F 17/30 360Z

審査請求 未請求 請求項の数 11 O L (全 16 頁)

(21) 出願番号	特願2004-212475 (P2004-212475)	(71) 出願人	504278835 美馬 秀樹 東京都港区六本木七丁目10番18号
(22) 出願日	平成16年7月21日(2004.7.21)	(71) 出願人	504278857 トレンド株式会社 東京都港区六本木七丁目10番18号
		(71) 出願人	593036165 西華産業株式会社 東京都千代田区丸の内三丁目3番1号
		(74) 代理人	100080838 弁理士 三浦 光康
		(72) 発明者	美馬 秀樹 東京都港区六本木七丁目10番18号
		Fターム(参考)	5B075 ND03 NS10 PQ02 PQ13 UU06

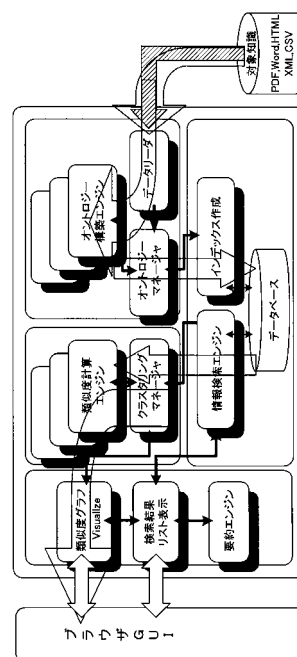
(54) 【発明の名称】 情報の検索俯瞰方法および装置

(57) 【要約】

【課題】 本発明は検索対象に対して、あるキーワードとの関連性のみならず、その検索された情報間の関連性を動的かつリアルタイムに計算し、情報を整理した上で新たなカテゴリを自動的に生成しユーザに提示することができ、質問文とのマッチングの度合いに関係なく、検索された対象間の関連性により内容の確認を行うことが支援されるため、より効率的に意図する対象に到達する可能性が高まる情報の検索俯瞰方法および装置を得るにある。

【解決手段】 検索対象の情報を計算機上において整理し、俯瞰する情報の検索俯瞰方法であって、検索対象である情報からテキスト情報を抽出し、索引付けを行うステップと、ユーザからの検索質問の入力に対し与えられた検索質問にマッチする情報を検索し、前記検索質問とのマッチングの度合いを計算し、情報間の関連性やマッチングの度合いを反映させて所定の位置に配置するとともに視覚化し、俯瞰表示処理するステップとで情報の検索俯瞰方法を構成している。

【選択図】 図2



【特許請求の範囲】**【請求項 1】**

検索対象の情報を計算機上において整理し、俯瞰する情報の検索俯瞰方法であって、検索対象である情報からテキスト情報を抽出し、索引付けを行うステップと、ユーザからの検索質問の入力に対し与えられた検索質問にマッチする情報を検索し、前記検索質問とのマッチングの度合いを計算し、情報間の関連性やマッチングの度合いを反映させて所定の位置に配置するとともに視覚化し、俯瞰表示処理するステップとを備えることを特徴とする情報検索俯瞰方法。

【請求項 2】

検索対象の情報を計算機上において整理し、俯瞰する情報の検索俯瞰方法であって、検索対象である情報からテキスト情報を抽出し、索引付けを行うステップと、ユーザからの検索質問の入力に対し与えられた検索質問にマッチする情報を検索し、前記検索質問とのマッチングの度合いを計算し、情報間の関連性やマッチングの度合いを反映させて所定の位置に配置するとともに視覚化し、俯瞰表示処理するステップとを備え、PDF、Word、HTML、XML、CSV等の検索対象のテキスト情報から用語抽出エンジンを介してそこに含まれる用語を自動抽出するステップと、この抽出ステップから抽出された用語の相互の関連性を用語分類エンジンを介して計算して分類するステップと、この分類ステップで分類された用語の情報を扱い、関連性抽出エンジンを介して対象とする情報の特徴量を検出し、その特徴量を基に情報の関連性を計算することで分類するステップと、ユーザからの検索質問の入力に対し、情報分類エンジンを介して与えられた検索質問にマッチする情報を検索し、検索質問とのマッチングの度合いを整理するステップと、このステップで出力された検索情報と、前記分類ステップにより計算された情報間の関連性やマッチングの度合いを反映させて、視覚化エンジンを介して視覚化し、俯瞰表示の処理をするステップとを備えることを特徴とする情報検索俯瞰方法。

【請求項 3】

用語や検索キーワード等の言葉の関連性の情報をその関連性の度合いを反映させる位置に視覚化し、その視覚化により表現された関連する言葉を直接ポインティングデバイスにより指定するか、自動的に計算された質問拡張候補をその一覧より選択することで、検索に指定する質問文を自動的に作成し、検索の絞り込みを支援するステップをさらに備えることを特徴とする請求項 1 あるいは 2 のいずれかに記載の情報検索俯瞰方法。

【請求項 4】

前記視覚化ステップは、整理された情報に対して、リアルタイムで関連性の再計算を行うことと計算結果を視覚化に反映することで、just-in-timeに情報を俯瞰することができるステップであることを特徴とする請求項 1、2、3 のいずれかに記載の情報検索俯瞰方式。

【請求項 5】

前記視覚化ステップでは、情報間の関連性やマッチングの度合いを反映させる位置にそれぞれの情報が配置され、情報間の関連性やマッチングの度合いにより、情報間に関連付けマークを設ける処理をすることを特徴とする請求項 1、2、3、4 のいずれかに記載の情報検索俯瞰方法。

【請求項 6】

用語分類エンジンを介して計算して分類するステップでは、テキスト情報に対してインデクシング（索引を付与する）を行うステップを含むことを特徴とする請求項 1、2、3、4、5 のいずれかに記載の情報の検索俯瞰方式。

【請求項 7】

検索対象の情報のテキスト情報を複数の言語に翻訳して分類する多言語翻訳ステップをさらに備えることを特徴とする請求項 1、2、3、4、5、6 のいずれかに記載の情報検索俯瞰方法。

【請求項 8】

検索対象の情報を計算機上において整理し、俯瞰する情報の検索俯瞰させる処理を実行さ

10

20

30

40

50

せるためのプログラムであって、PDF、Word、HTML、XML、CSV等の検索対象のテキスト情報からそこに含まれる用語を自動抽出する用語抽出手段と、抽出された用語の相互の関連性を計算して分類する用語分類手段と、分類された用語の情報をを用い、対象とする情報の特徴量を検出し、その特徴量を基に情報の関連性を計算することで分類する関連性抽出手段と、ユーザからの検索質問の入力に対し、入力装置を介して与えられた検索質問にマッチする情報を検索し、検索質問とのマッチングの度合いを整理する情報分類手段と、出力された検索情報と、前記分類手段により計算された情報間の関連性やマッチングの度合いを反映させて視覚化し、俯瞰表示の処理をする視覚化手段とを備えることを特徴とする情報検索俯瞰装置。

【請求項 9】

用語や検索キーワード等の言葉の関連性の情報をその関連性の度合いを反映させる位置に視覚化し、その視覚化により表現された関連する言葉を直接ポインティングデバイスにより指定するか、自動的に計算された質問拡張候補をその一覧より選択することで、検索に指定する質問文を自動的に作成し、検索の絞り込みを支援する手段をさらに備えることを特徴とする請求項 8 に記載の情報検索俯瞰装置。

10

【請求項 10】

前記視覚化手段では、情報間の関連性やマッチングの度合いを反映させる位置にそれぞれの情報が配置され、情報間の関連性やマッチングの度合いにより、情報間に関連付けマークを設けることを特徴とする請求項 8 あるいは 9 のいずれかに記載の情報検索俯瞰装置。

【請求項 11】

検索対象の情報を計算機上において整理し、俯瞰する情報の検索俯瞰させる処理を実行させるためのプログラムであって、PDF、Word、HTML、XML、CSV等の検索対象のテキスト情報からそこに含まれる用語を自動抽出する用語抽出エンジンと、抽出された用語の相互の関連性を計算して分類する用語分類エンジンと、分類された用語の情報をを用い、対象とする情報の特徴量を検出し、その特徴量を基に情報の関連性を計算することで分類する関連性抽出エンジンと、ユーザからの検索質問の入力に対し、入力装置を介して与えられた検索質問にマッチする情報を検索し、検索質問とのマッチングの度合いを整理する情報分類エンジンと、出力された検索情報と、前記分類エンジンにより計算された情報間の関連性やマッチングの度合いを反映させて視覚化し、俯瞰表示の処理をする視覚化エンジンとを備えるプログラムを記憶したコンピュータ読み取り可能な記憶媒体。

20

30

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ある蓄積された情報に対して、ユーザにより指定するキーワードに関連する情報の検索と検索された情報の相互の関連性により整理を行い、その結果をユーザに提示する情報の検索俯瞰方式および装置に関する。

【背景技術】

【0002】

ある情報を計算機により検索する場合、計算機上に蓄積された情報に対し、指定された質問文にマッチする情報を探しだし、質問文とのマッチングの度合い（スコア）をVSM（Vector Space Model）方式や、k 近傍等の統計確率を基にした方式により計算し、計算されたスコア順に並べてリスト化して表示するのが一般的であった。

40

また、例えば、検索サイトやフーで提供されているような、カテゴリ検索におけるカテゴリ内の情報を利用し、検索された情報をその属するカテゴリ内の情報に基づき表示（リスト表示）あるいはグラフィカルに表示する検索方式は提案されているが、基本的には、あらかじめ決められた静的なカテゴリ内の情報を利用するため、ユーザの要求や視点に対して十分な対応ができないという問題があった。

【特許文献 1】特になし

【発明の開示】

50

【発明が解決しようとする課題】

【0003】

近年のインターネットやIT技術の発展に伴い、文献やデータベース等のアクセス可能な知識資源の生成サイクルが短縮化されており、蓄積される情報の量も爆発的に増加しつつある。しかし、それら情報の膨大さと、新たな情報が日々追加されるリアルタイム性により、意図する情報を見つけだすことが非常に困難となっている状況において、有用な情報を容易に獲得するための技術の一つとして情報の自動分類技術が重要視されている。

例えば、ハイパーリンクによる情報の関連性の定義と、その関連に基づいた分類や、あらかじめ分類されたカテゴリ情報により情報を分類し、提示する方式がある。しかし、上述の情報のリアルタイム性や、ユーザの視点の違いにより、必ずしも静的な分類が情報の獲得に有効だと言えないのが現状である。

これに対し、新たな情報に対しても動的に適切な自動分類を行う方式、つまり情報を自動的に整理し、ユーザの要求に応じてJust In Timeに提示できる方式が望ましいと考えられる。

【0004】

本発明は以上のような従来欠点に鑑み、検索対象に対して、あるキーワードとの関連性のみならず、その検索された情報間の関連性を動的かつリアルタイムに計算し、情報を整理した上で新たなカテゴリを自動的に生成しユーザに提示する情報の検索俯瞰方法および装置を提供することを目的とする。

さらに、従来検索システムでは、質問文とマッチングの度合いにより計算されたスコアを基に、一番スコアのよいものより内容を確認することが一般的であるため、例えば意図するものがリストの最後にあったような場合などはそこにたどり着くのは非常に労力を要するものであった。それに対し、本発明では、質問文とのマッチングの度合いに関係なく、検索された対象間の関連性により内容の確認を行うことが支援されるため、より効率的に意図する対象に到達する可能性が高まる情報の検索俯瞰方法および装置を提供することを目的とする。

【0005】

本発明の前記ならびにそのほかの目的と新規な特徴は次の説明を添付図面と照らし合わせて読むと、より完全に明らかになるであろう。

ただし、図面はもっぱら解説のためのものであって、本発明の技術的範囲を限定するものではない。

【課題を解決するための手段】

【0006】

上記目的を達成するために、本発明は、検索対象の情報を計算機上において整理し、俯瞰する情報の検索俯瞰方法であって、検索対象である情報からテキスト情報を抽出し、索引付けを行うステップと、ユーザからの検索質問の入力に対し与えられた検索質問にマッチする情報を検索し、前記検索質問とのマッチングの度合いを計算し、情報間の関連性やマッチングの度合いを反映させて所定の位置に配置するとともに視覚化し、俯瞰表示処理するステップとで情報の検索俯瞰方法を構成している。

【0007】

また、本発明は、検索対象の情報を計算機上において整理し、俯瞰する情報の検索俯瞰させる処理を実行させるためのプログラムであって、PDF、Word、HTML、XML、CSV等の検索対象のテキスト情報からそこに含まれる用語を自動抽出する用語抽出手段と、抽出された用語の相互の関連性を計算して分類する用語分類手段と、分類された用語の情報をを用い、対象とする情報の特徴量を検出し、その特徴量を基に情報の関連性を計算することで分類する関連性抽出手段と、ユーザからの検索質問の入力に対し、入力装置を介して与えられた検索質問にマッチする情報を検索し、検索質問とのマッチングの度合いを整理する情報分類手段と、出力された検索情報と、前記分類エンジンにより計算された情報間の関連性やマッチングの度合いを反映させて視覚化し、俯瞰表示の処理をする視覚化手段とで情報検索俯瞰装置を構成している。

10

20

30

40

50

【発明の効果】

【0008】

以下の説明から明らかなように、本発明にあつては次に列挙する効果が得られる。

【0009】

(1) 検索対象の情報を計算機上において整理し、俯瞰する情報の検索俯瞰方法であつて、検索対象である情報からテキスト情報を抽出し、索引付けを行うステップと、ユーザからの検索質問の入力に対し与えられた検索質問にマッチする情報を検索し、前記検索質問とのマッチングの度合いを計算し、情報間の関連性やマッチングの度合いを反映させて所定の位置に配置するとともに視覚化し、俯瞰表示処理するステップとを備えているので、情報検索の結果を、リストのみによらず、整理された形で俯瞰することができるので、膨大な検索対象や検索結果に対しても効率的に意図する情報を獲得することができる。

10

【0010】

(2) 前記(1)によって、複数の種類の情報を検索対象とし、複数の種類の情報間の関連性を縦断的に抽出し、俯瞰化を行うことで、情報の時間軸上の繋がりや、情報の階層的繋がりをもより明確に表現することができるので、情報の前後や上下の繋がりを知ることによって、効率的に情報を理解することができる。

【0011】

(3) 請求項2～7も前記(1)および(2)と同様の効果が得られる。

【0012】

(4) 請求項8も検索対象の情報を計算機上において整理し、俯瞰する情報の検索俯瞰させる処理を実行させるためのプログラムであつて、PDF、Word、HTML、XML、CSV等の検索対象のテキスト情報からそこに含まれる用語を自動抽出する用語抽出手段と、抽出された用語の相互の関連性を計算して分類する用語分類手段と、分類された用語の情報を用い、対象とする情報の特徴量を検出し、その特徴量を基に情報の関連性を計算することで分類する関連性抽出手段と、ユーザからの検索質問の入力に対し、入力装置を介して与えられた検索質問にマッチする情報を検索し、検索質問とのマッチングの度合いを整理する情報分類手段と、出力された検索情報と、前記分類手段により計算された情報間の関連性やマッチングの度合いを反映させて視覚化し、俯瞰表示の処理をする視覚化手段とからなるので、前記(1)および(2)と同様の効果が得られる。

20

【0013】

(5) 請求項9～11も前記(1)および(2)と同様の効果が得られる。

30

【発明を実施するための最良の形態】

【0014】

以下、図面に示す実施するための最良の形態により、本発明を詳細に説明する。

【実施例1】

【0015】

図1ないし図18に示す本発明の第1の実施するための最良の第1の形態において、図1は、本発明の情報の検索俯瞰方法を実施するための情報端末機1の一構成例を示すブロック図で、この情報端末機1は、データを入力するための入力装置2と、入力されたデータを処理する中央処理装置(CPU)3と、入力されたデータがそのまま、あるいは中央処理装置3により処理された後に格納されるメモリ4と、前記中央処理装置3が処理したデータが出力される出力装置5とを備えている。前記情報端末機1は、従来と同様の構成のものを使用しているが、さらに効率のよい入力装置2、CPU3、メモリ4、出力装置5等を任意に使用することができる。

40

前記中央処理装置3は、前記入力装置2を介して入力された初期データをそのままメモリに記憶させる。

次に、前記メモリ4に記憶されたデータを検索するような指令が前記入力装置2を介して入力された場合、前記中央処理装置3は、その指令に従ってデータを検索した後、検索されたデータを処理し、出力装置5で出力する。

【0016】

50

前記情報端末機 1 を用いて処理される、本発明の情報の検索俯瞰方法の一例を説明する。図 1 は本システムをアクティベートする、あるいは処理する情報端末機の基本構成を示す図で、図 2 は本システムの構成の概念図である。本実施例の情報の検索俯瞰方法では、一般的に使用されている P D F、W o r d、H T M L、X M L、C S V 等の形式を含むテキスト情報を対象として、意図する知識の検索と、図 3 に示すような知識の構造化を行うことを目的とし、情報の検索、及び整理を行うための事前の情報抽出とインデクシング処理機能、および情報検索と情報の整理（情報の関連性計算と視覚化）処理機能の 2 系統の処理により構成される。図 4 および図 5 は、それぞれ情報抽出とインデクシング処理機能および情報の整理処理機能の概略的な流れを示すフロー図である。

【 0 0 1 7 】

10

ここで、本発明の情報の検索俯瞰方法の処理ステップを図 6、図 8 および図 1 0 を参照して説明する。

図 6 に示すように、検索対象の情報を前記情報端末機 1 上において整理し、視覚化して俯瞰状態で表示する情報の検索俯瞰方法であって、P D F、W o r d、H T M L、X M L、C S V 等の検索対象のテキスト情報からそこに含まれる用語を用語抽出エンジン 6 を介して自動抽出する用語抽出ステップ 7 と、この抽出ステップ 7 で抽出された用語の相互の関連性を用語分類エンジン 8 を介して計算し、分類する用語分類ステップ 9 と、この分類ステップ 9 で分類された用語の情報を対象とする情報の特徴量を検出し、その特徴量を基に関連性抽出エンジン 1 0 を介して情報の関連性を計算する分類する関連性抽出ステップ 1 1 と、ユーザ（情報の検索者）に検索質問の入力を促し、入力された検索質問にマ 20 ッチする情報を検索し、検索質問とのマッチングの度合いを情報分類エンジン 1 2 を介して分類、整理する情報分類ステップ 1 3 と、このステップ 1 3 で整理された検索情報と、前記関連性抽出ステップ 1 1 により処理された情報間の関連性やマッチングの度合いを反映させて視覚化エンジン 1 4 を介して、前記情報間の関連性やマッチングの度合いを反映させる位置にそれぞれの情報が配置され、情報間の関連性やマッチングの度合いにより、情報間に関連付けマーク 1 5、本実施例ではラインを設けて整理し、視覚化して俯瞰表示の処理をする視覚化ステップ 1 6 とを備えている。

20

なお、本実施例では、検索対象の情報は、一般的なデータベースに蓄積される情報を指し、特に限定はしていないが、本発明の情報の検索俯瞰方法を使用する場所や環境に応じて、例えば大学や教育現場等で利用する場合には、前記情報は、特許、論文等の内容とそ 30 れらの意味的類似性による関連性が認識される情報が含まれる。

30

【 0 0 1 8 】

前記用語抽出ステップ 7、用語分類ステップ 9、関連性抽出ステップ 1 1、情報分類ステップ 1 3 および視覚化ステップ 1 6 は、基本的に情報の管理上、二系統の処理に分割することができる。

まず、図 7 および図 8 に示す処理においては、まず情報ソースよりテキスト情報を抽出し、抽出されたテキストからオントロジー情報を抽出する処理と、テキスト情報に対してインデクシング（索引を付与する）を行う処理に分かれる。すなわち前記用語抽出ステップ 7、用語分類ステップ 9、関連性抽出ステップ 1 1 である。そして、それぞれの処理結果を、それぞれオントロジー情報はオントロジーデータベース 1 7 へ、およびインデクシ 40 ング情報は索引データベース 1 8 に格納する。

40

また、図 9 および図 1 0 に示す処理においては、ユーザ（情報の検索者）から情報端末機を介して入力されるキーワードや文章等の検索質問を、外部の検索エンジンを介する情報検索処理により、前記索引データベース 1 8 に格納された索引情報を利用し関連する情報を検索する。さらに、検索された情報に対し、前記オントロジーデータベース 1 7 に格納されたベースオントロジー情報を参照することで、各情報に含まれるオントロジー情報と用語関連の関連度を示すオントロジー情報を利用し、情報間の関連性の計算を行い、その結果を基に画面上に視覚化を行う。すなわち、前記情報分類ステップ 1 3 および視覚化ステップ 1 6 である。

ここで、「オントロジー」（O n t o l o g y）とは、語彙と語彙の関係等の概念の体

50

系を表し、例えば、「歯科医」と「歯医者」は同義語であるとか、「歯科医」は「医者」の下位概念である等の論理的関係、および「歯科医」は「歯」を治療する等の要素的關係等を定義する。

また、「俯瞰化」とは、情報及びその構造を上位の視点から観察し、理解を促進することを指す。

【0019】

以下、各処理エンジンでの処理作業の概要を説明する。

前記用語抽出ステップ7、用語分類ステップ9、関連性抽出ステップ11で使用するエンジンは、図8に示すような前記用語抽出エンジン6、用語分類エンジン8、関連性抽出エンジン10より構成され、これらをオントロジー抽出エンジンと呼ぶことができる。

10

【0020】

前記用語抽出エンジン6は、まずPDF、Word、HTML、XML、CSV等の検索対象のテキスト情報から抽出されるテキストを、例えばC/NC-value手法により、テキストに含まれる用語を自動抽出する。

なお、前記用語抽出エンジン6は、現状の膨大な情報と、それらが日々更新されるリアルタイム性により、用語に関する命名規則を100%規制することが不可能であろう現状では、ターミノロジーの知見を利用し、自動処理を介した知識獲得の効率化を行うことは必然である。さらには、用語自体の定義があいまいな(用語と非用語の区別が難しい)状況では、一定の指標による自動用語認識を介した処理を行うことは、尚更、重要な意味を持つ。

20

本発明を構成するシステムにおいても、用語抽出エンジン6として「C/NC-value」の用語認識技術を基にした用語の自動認識を利用する。「C-value」とは、用語構成に関する基本語彙の組み合わせパターンと用語の対象ドメインにおける出現頻度、さらには、用語のネ스팅に関する性質に注目し、スコア付けを行うことで用語の高精度な自動認識を行う。また、NC-valueでは、候補となる用語の実際の文書上でのコンテキスト中にある語彙とのコロケーションの情報を用いて、用語としての確からしさ(termhood)の指標を求め、求めた指標を基に候補となる用語の再順序付けを行う。我々の行った実験では、本方法により、英語および日本語に関しても、ドメインによらず、上位の候補では90%以上の正解率を得られることが示されている。本エンジンのこのような対象ドメインや対象言語への非依存性は、本システムを複数の言語や複

30

数の分野の情報を対象とした情報の整理、俯瞰を行うための、重要な特徴となる。なお、用語候補として、形態素解析の結果に対し、その頻度の高いものから最良の数、もしくは最良の頻度以上の形態素を用いることで同様の情報の関連性の計算を行うことも可能である。

【0021】

前記用語分類エンジン8においては、用語抽出の結果と文脈抽出処理によりテキストから抽出した文脈の情報を利用し、平均相互情報量を含む用語間の類似性計算処理を用いることで用語の自動分類を行う。

図11に示すように、用語自体の命名規則が不明瞭である状況においては、語彙結合のためのハイフンの使用法や、“leukemia”と“leukaemia”のような綴り方(orthographic)のゆれ、“NF-kappa B”、“NF-kB”等の省略法(acronym)のゆれ、さらには“human clones”と“clones of humans”のような統語的(syntactic)なゆれ、等の用語のバリエーションに関する様々な問題が生じる。このような問題に対し、C-valueのような用語の対象ドメインにおける相対的頻度を基本的スコアとして利用した手法では、本質的な解決を行うことは難しい。さらには、知識統合のためには、“eye surgery”と“ophthalmological surgery”のように、意味的には同義である関係に対しても、同一クラス

40

の用語として自動で認識できることが望ましい。つまり、用語の認識と統合には、図11に示すように、複数のプロセスによる用語クラスの自動認識技術が不可欠となる。本発明では、orthographic、morphological、syntacticさらにacronymに関する用語バリエーションの認識手法を構築し、用語抽出処理に組み込む。さらに、意味的バリエーシ

50

ョンに関しては、コンテキスト語彙との共起確率を基に得られる統計的類似度として、平均相互情報量を用いた用語分類手法を利用する。

なお、用語の分類手法として、分類対象である用語のテキストにおける文脈の情報のみではなく、用語と動詞との関係や、共起する用語との関係等を利用し、ベクトルスペースモデル等を利用した他の類似性の計算手法や、サポートベクターマシンや、決定木等の機械学習による分類手法を利用することでも同様の用語分類を行うことは可能である。

【0022】

前記用語抽出エンジン6では、図9に示すように、テキスト情報を対象に、語彙辞書、およびオントロジー情報を利用することで形態素解析を行い、その結果に対して語構成ルールを基に用語候補を抽出する。さらに得られた用語候補に対して頻度分析を含む用語スコアの計算処理を行い、指定のスコア以上の用語候補を用語リストとして出力する。 10

【0023】

図12は一抽出処理により抽出されたオントロジー情報を基に、情報間の関連性を抽出する処理のフローを示したものである。情報の関連性抽出エンジン10は、前記用語抽出エンジン6と用語分類エンジン8により抽出された用語とその関連性の情報を用いて、用語間の意味的関連性を定量的に計算するためのエンジンで、情報を特徴付ける情報としての用語間の意味的関連性を計算することで、情報間の意味的関連性の計算を行う。

【0024】

前記視覚化エンジン14は、前記関連性抽出エンジン10による情報間の意味的関連性を、対象とする情報間の組み合わせのすべてに対して計算することで、情報間の意味的関連を計算する。さらに、画面へ情報の関連性を描画に際し、各情報をノードNに割り当て、意味的関連性が大きいほど、ノード間の物理的距離を短くし、またノード間のリンクの太さ、すなわち前記関連付けマーク15を太くし、すべての対象とする情報を画面上の任意個所に最適配置することで、情報間の関連性を視覚化することができる。 20

ここで、「ノード」とは、一般的には、パソコンなどの装置を接続するネットワーク(LAN)上の接合点や、インターネット上の中継点、また、そこに設置されるコンピュータのことであり、いくつもの構成要素が相互に接続されている状態を構成する個々の要素、すなわち意味的関連性を有する各情報をノードと呼ぶ。

【0025】

以下、本実施例の具体例として、本実施例に係る装置を論文や特許等の情報の検索俯瞰システムに応用した例を説明する。本システムは本発明を用いて実現されるものである。図13は、本発明の情報の検索俯瞰システムの処理フローの概略図を示している。 30

上述の実施例に係る情報端末機1を構成する中央処理装置3の動作は、コンピュータが読み取り可能な言語で記述されたコンピュータプログラムによっても実行可能である。

コンピュータプログラムにより中央処理装置3を動作させる場合には、例えば、中央処理装置3にプログラム記憶用のメモリを設け、そのメモリにコンピュータプログラムを格納する。中央処理装置はメモリからそのコンピュータプログラムを読み出すことにより、そのコンピュータプログラムに従って、上述のような動作を実行する。

さらには、そのようなコンピュータプログラムを格納した記憶媒体を情報端末機にセットすることにより、中央処理装置がその記憶媒体からそのコンピュータプログラムを読み出し、そのコンピュータプログラムに従って、上述のような動作を実行するようにすることも可能である。 40

【0026】

次に、前記情報分類ステップ13で整理された検索情報と、前記関連性抽出ステップ11により処理された情報間の関連性やマッチングの度合いを反映させて前記視覚化エンジン14を用いて前記情報を整理して視覚化し、俯瞰表示の処理をする視覚化ステップ16においては、図14に示すような、情報の関連性を視覚化されたマッピングを表示することができる。

例えば、検索情報として「カーボンナノチューブ」を検索すると、関連性のあるテキスト情報が、まるで地図のようにウインド内に表示され、情報間の関連性が強ければ強いほど 50

、太く、濃く関連付けマーク15で結ばれる。

また、図15に示すような処理をすることにより、図16に示すような情報の関連性を、本実施例では「カテゴリA」、「カテゴリB」、「カテゴリC」のカテゴリ別に表示することもできる。

【0027】

なお、本実施例では、論文や特許等の情報検索俯瞰システムに応用した例を説明したが、本発明はこれに限らず、例えば図17および図18に示すように、知識ソースとして既存のデータベースを取り込む形で支援システムを組織することにより、例えば小学校や中学校等のクローズドネットワーク内で活用することもできる。

{発明を実施するための異なる形態}

10

【0028】

次に、図19ないし図23に示す本発明を実施するための異なる形態につき説明する。なお、これらの本発明を実施するための異なる形態の説明に当って、前記本発明を実施するための最良の第1の形態と同一構成部分には同一符号を付して重複する説明を省略する。

【実施例2】

【0029】

図19および図20に示す本発明を実施するための第2の形態において、前記本発明を実施するための最良の第1の形態と主に異なる点は、整理された情報に対して、リアルタイムで関連性の再計算を行うとともに、計算結果を視覚化に反映することで、just-in-timeに情報を俯瞰することができる視覚化エンジン14Aを用いた点で、このように構成された視覚化ステップ16Aを用いることにより、前記本発明を実施するための最良の第1の形態と同様な作用効果が得られる。

20

【実施例3】

【0030】

図21および図22に示す本発明を実施するための第3の形態において、前記本発明を実施するための最良の第1の形態と主に異なる点は、前記用語抽出エンジン6を介して、PDF、Word、HTML、XML、CSV等の検索対象のテキスト情報から抽出されるテキストに含まれる用語を自動抽出する用語抽出ステップ7と同時、あるいはその処理後に、抽出された用語を多言語翻訳エンジン19を介して、複数の言語に翻訳し、多言語データベース20に蓄積する多言語翻訳ステップ21を用いた点で、このように構成された情報の検索俯瞰方法にすることにより、前記本発明を実施するための最良の第1の形態と同様な作用効果が得られるとともに、異なる複数の言語テキストの選択に躊躇することなく本発明の俯瞰方法を使用することができるとともに、情報間の関連性も、より緊密なものにすることができる。

30

【実施例4】

【0031】

図23に示す本発明を実施するための第4の形態において、前記本発明を実施するための最良の第1の形態と主に異なる点は、検索対象の情報を情報端末機1上において整理し、俯瞰する情報の検索俯瞰させる処理を実行させるためのプログラムであって、前記用語抽出エンジン6、用語分類エンジン8、関連性抽出エンジン10、情報分類エンジン12、視覚化エンジン14とを備えるプログラムを記憶したコンピュータ読み取り可能な記憶媒体22を用いた点で、このように構成しても、前記本発明を実施するための最良の第1の形態と同様な作用効果が得られるとともに、前記情報端末機1において使用しても、各種のコマンドを含むプログラムとして実現することができる。

40

本実施例において、「記憶媒体」の語は、データを記録することができるあらゆる媒体を含み、例えば前記記憶媒体としては、CD-ROMやPDなどのディスク型の記憶媒体、磁気テープ、MO、DVD-ROM、DVD-RAM、フレキシブルディスク、RAMやROM等のメモリーチップ、EPROM(Erasable Programmable Read Only Memory)、EEPROM(Electrically

50

Erasable Programmable Read Only Memory)、スマートメディア(登録商標)、フラッシュメモリ、コンパクトフラッシュ(登録商標)などの書き換え可能なカード型ROM、ハードディスクがあり、その他プログラムの格納に適していれば、いかなる手段も用いることができる。

前記記憶媒体22は、コンピュータが読み取り可能なプログラム用言語を用いて上述のマイクロコンピュータの各機能をプログラミングし、そのプログラムをプログラムの記録が可能な上記の記憶媒体に記録することにより、作成することができる。また、記憶媒体として、サーバに備え付けられたハードディスクを用いることも可能である。

また、本実施例の前記記憶媒体22は、ネットワークを介して、格納されるコンピュータプログラムを他のコンピュータにより読み取ることによっても、作成や使用することができる。

10

なお、情報端末機としてのコンピュータは、パーソナルコンピュータ、デスクトップ型コンピュータ、ノート式コンピュータ、モバイルコンピュータ、ラップトップ式コンピュータ、ポケットコンピュータ、サーバーコンピュータ、クライアントコンピュータ、ワークステーション、ホストコンピュータ等を用いることができる。

【0032】

なお、前記本発明の異なる実施の形態では主に第1の実施の形態を基にして説明したが、本発明はこれに限らず、各実施の形態に使用された構成を組み合わせて使用しても同様な作用効果が得られる。

【産業上の利用可能性】

20

【0033】

本発明は、情報を検索俯瞰方法および装置を利用、開発、製造等する産業、業界等で利用される。

【図面の簡単な説明】

【0034】

【図1】本発明を実施するための最良の第1の形態の情報端末機の一例の説明図。

【図2】本発明を実施するための最良の第1の形態の処理フロー図。

【図3】本発明を実施するための最良の第1の形態の知識の構造化の概念図。

【図4】本発明を実施するための最良の第1の形態のオントロジー抽出処理フロー図。

【図5】本発明を実施するための最良の第1の形態の検索俯瞰化処理フロー図。

30

【図6】本発明を実施するための最良の第1の形態の処理工程図。

【図7】本発明を実施するための最良の第1の形態のオントロジー抽出処理の概念図。

【図8】本発明を実施するための最良の第1の形態のオントロジー抽出処理の概略図。

【図9】本発明を実施するための最良の第1の形態の用語抽出処理の概念図。

【図10】本発明を実施するための最良の第1の形態の検索俯瞰化処理の概略図。

【図11】本発明を実施するための最良の第1の形態の用語分類処理フロー図。

【図12】本発明を実施するための最良の第1の形態の情報の関連性の計算処理フロー図。

【図13】本発明を実施するための最良の第1の形態の検索俯瞰システムの処理フローの概略図。

40

【図14】本発明を実施するための最良の第1の形態の情報の視覚化の参考図。

【図15】本発明を実施するための最良の第1の形態のカテゴリ別処理の概念図。

【図16】本発明を実施するための最良の第1の形態の情報の視覚化の他の参考図。

【図17】本発明を実施するための最良の第1の形態の検索俯瞰システムの参考図。

【図18】本発明を実施するための最良の第1の形態の情報の視覚化の更なる参考図。

【図19】本発明を実施するための第2の形態の処理工程図。

【図20】本発明を実施するための第2の形態の概略説明図。

【図21】本発明を実施するための第3の形態の処理工程図。

【図22】本発明を実施するための第3の形態の概略説明図。

【図23】本発明を実施するための第4の形態の概略説明図。

50

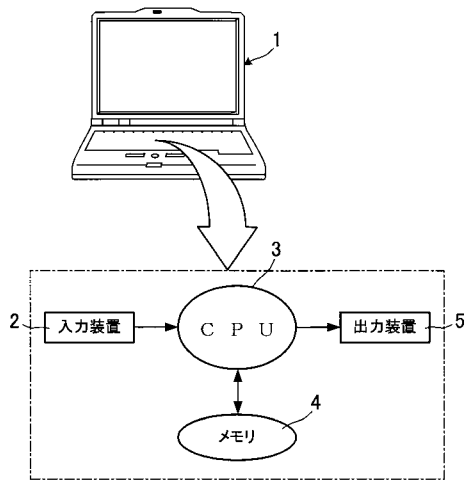
【符号の説明】

【0035】

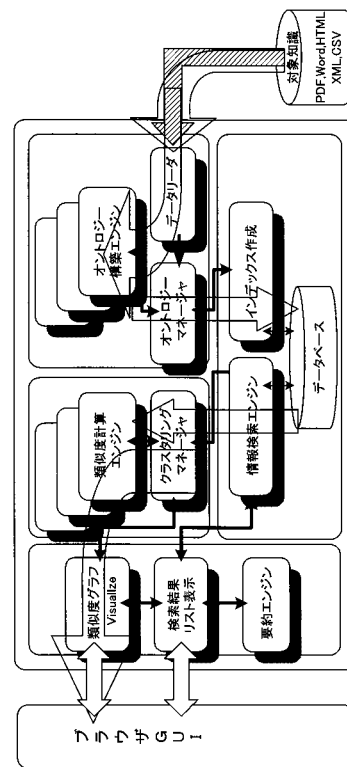
- 1 : 情報端末機、
- 3 : 中央処理装置、
- 5 : 出力装置、
- 7 : 用語抽出ステップ、
- 9 : 用語分類ステップ、
- 11 : 関連性抽出ステップ、
- 13 : 情報分類ステップ、
- 15 : 関連付けマーク、
- 17 : オントロジーデータベース、
- 19 : 多言語翻訳エンジン、
- 21 : 多言語翻訳ステップ、
- N : ノード、

- 2 : 入力装置、
- 4 : メモリ、
- 6 : 用語抽出エンジン、
- 8 : 用語分類エンジン、
- 10 : 関連性抽出エンジン、
- 12 : 情報分類エンジン、
- 14、14A : 視覚化エンジン、
- 16、16A : 視覚化ステップ、
- 18 : 索引データベース、
- 20 : 多言語データベース、
- 22 : 記憶媒体、
- W : ウインド。

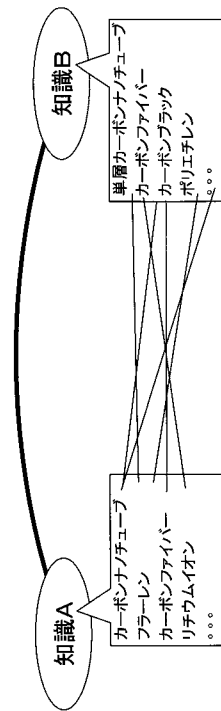
【図1】



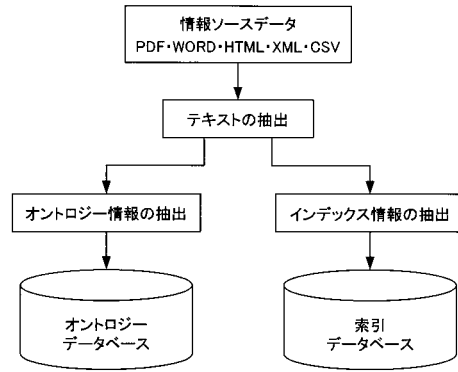
【図2】



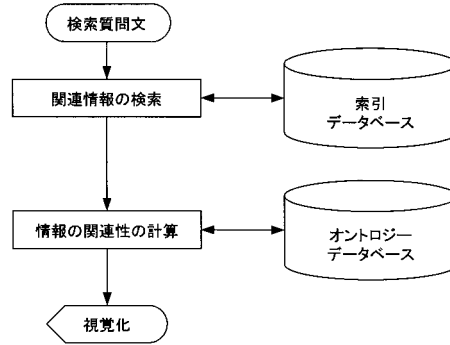
【 図 3 】



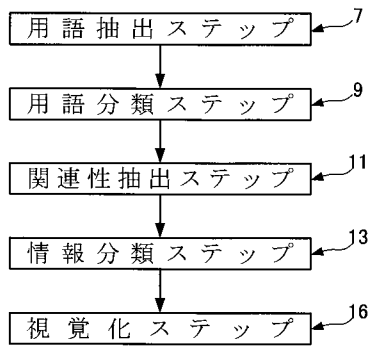
【 図 4 】



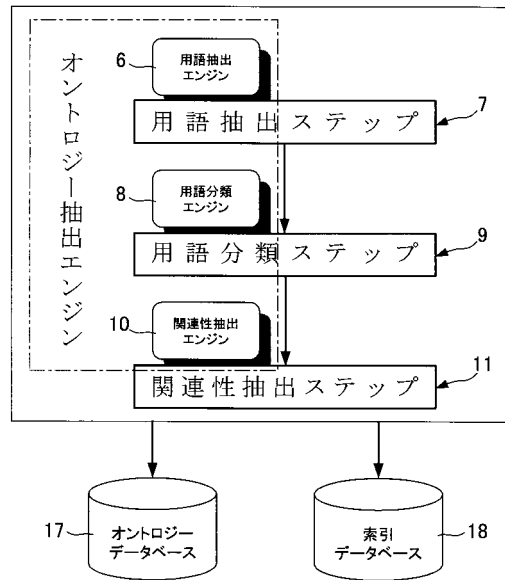
【 図 5 】



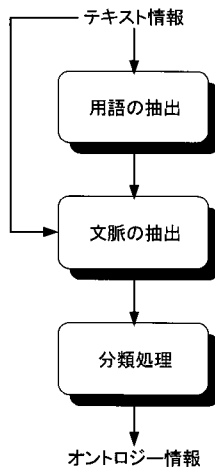
【 図 6 】



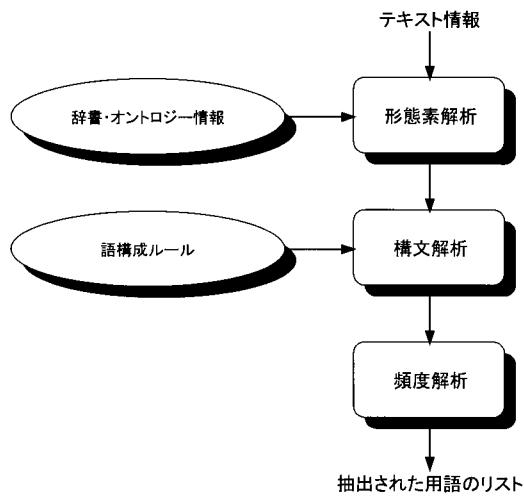
【 図 8 】



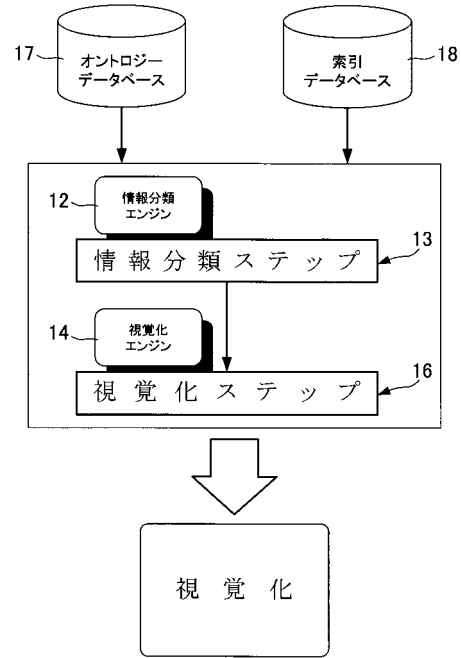
【 図 7 】



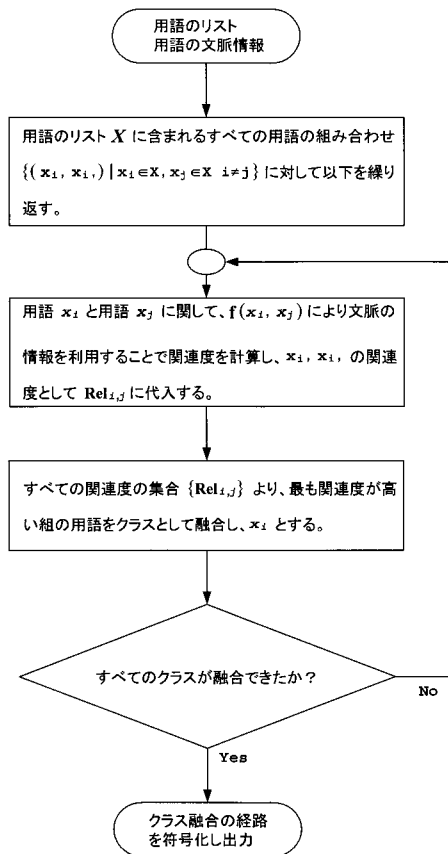
【 図 9 】



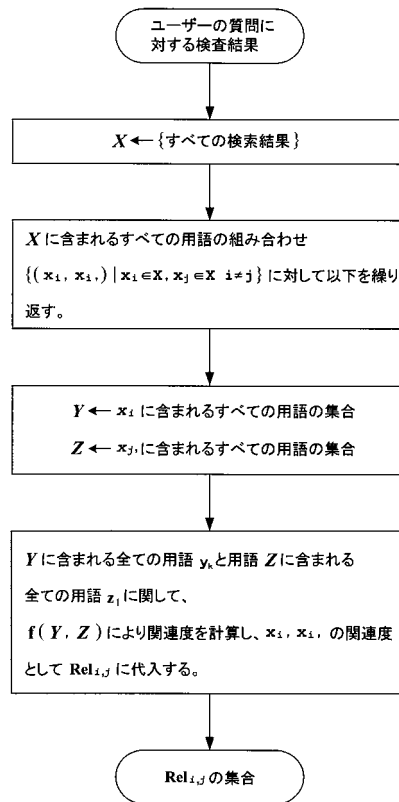
【 図 10 】



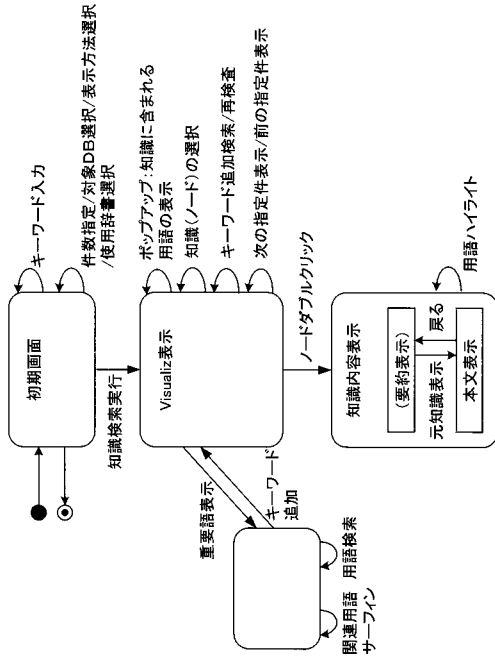
【 図 11 】



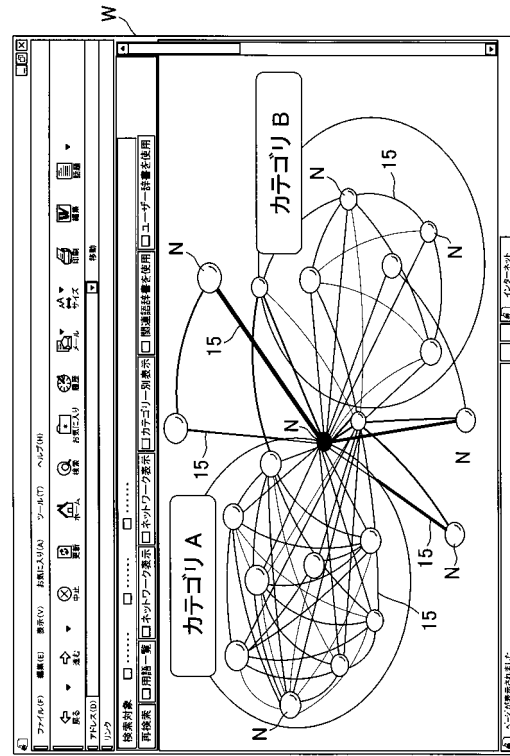
【 図 12 】



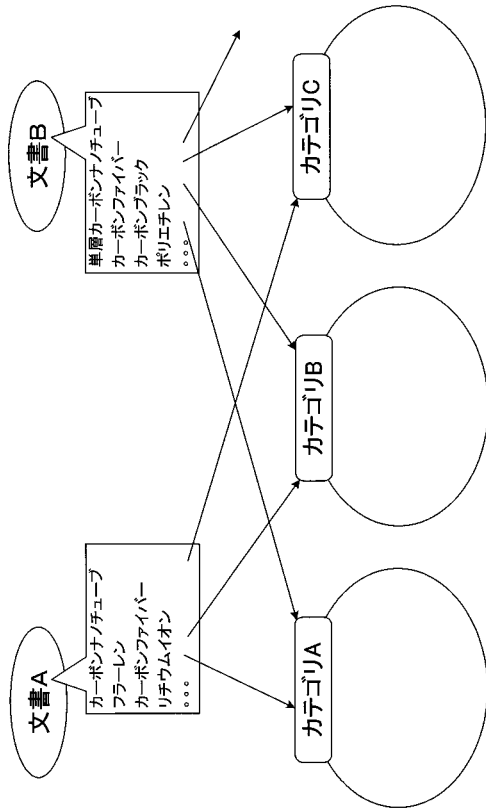
【 図 1 3 】



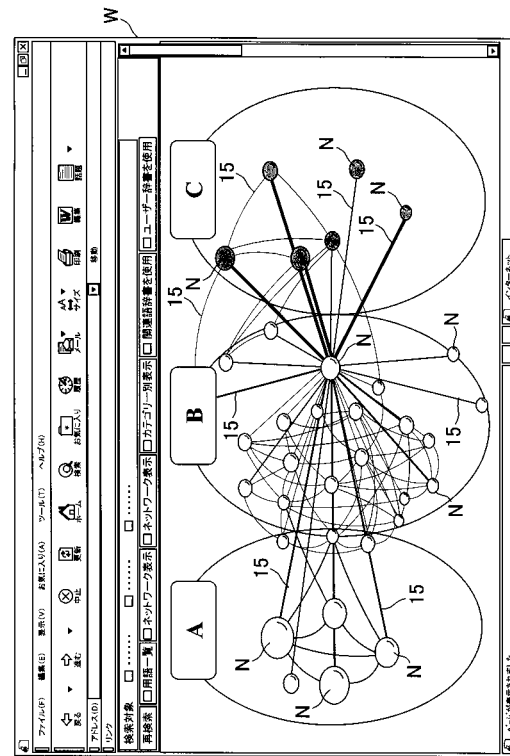
【 図 1 4 】



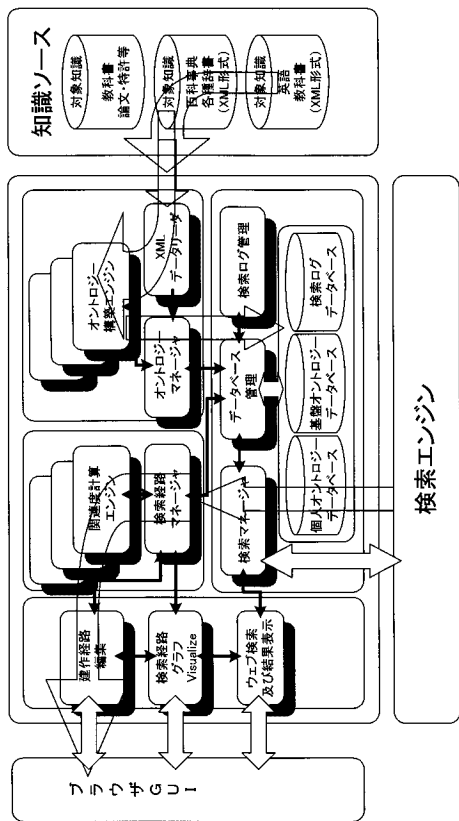
【 図 1 5 】



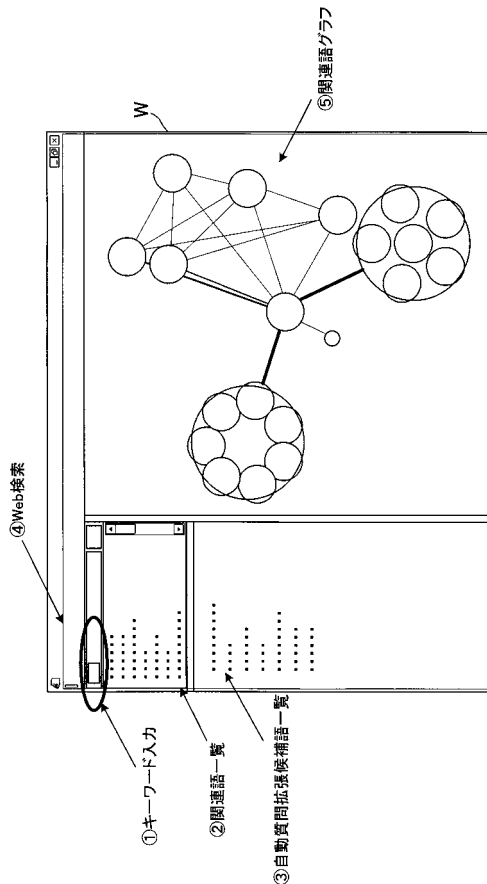
【 図 1 6 】



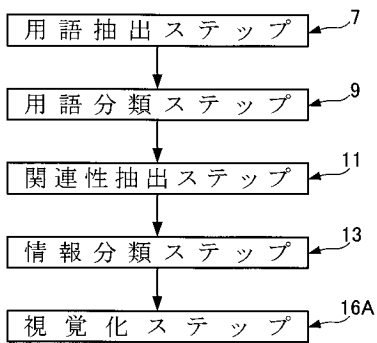
【 図 1 7 】



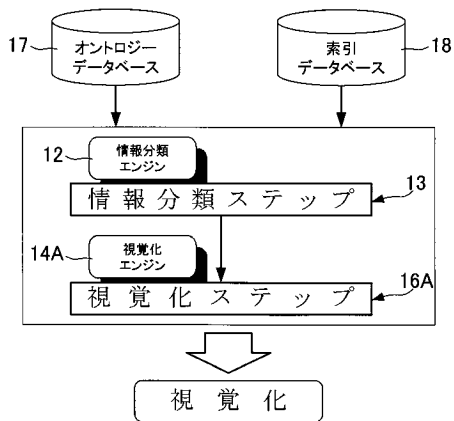
【 図 1 8 】



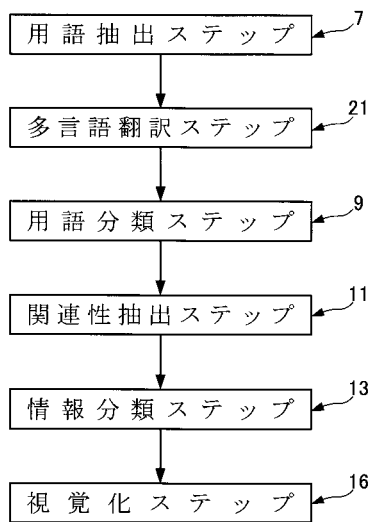
【 図 1 9 】



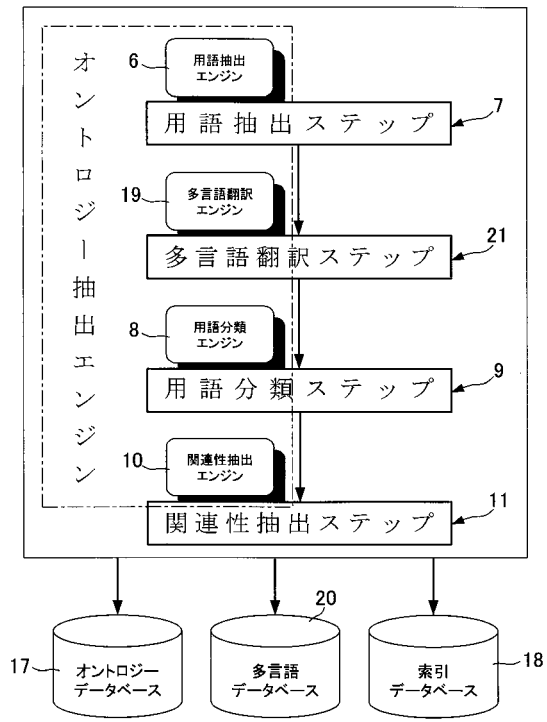
【 図 2 0 】



【 図 2 1 】



【図 2 2】



【図 2 3】

