(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5023176号
(P5023176)

(45)発行日　平成24年9月12日(2012.9.12)　(24)登録日　平成24年6月22日(2012.6.22)

(51)Int.Cl.　F I
G06F 17/30 (2006.01)　G06F 17/30 ２１０Ｄ
G06Q 10/00 (2012.01)　G06F 17/30 １７０Ａ
G06F 17/30 ３６０Ｚ
G06F 17/30 ２２０Ｚ
G06F 17/60 １５０

請求項の数 5　(全 34 頁)

(21)出願番号　特願2010-64821(P2010-64821)
(22)出願日　平成22年3月19日(2010.3.19)
(65)公開番号　特開2011-198111(P2011-198111A)
(43)公開日　平成23年10月6日(2011.10.6)
審査請求日　平成22年3月19日(2010.3.19)

(73)特許権者　000003078
株式会社東芝
東京都港区芝浦一丁目１番１号
(73)特許権者　301063496
東芝ソリューション株式会社
東京都港区芝浦一丁目１番１号
(74)代理人　100108855
弁理士　蔵田　昌俊
(74)代理人　100091351
弁理士　河野　哲
(74)代理人　100088683
弁理士　中村　誠
(74)代理人　100109830
弁理士　福原　淑弘

最終頁に続く

(54)【発明の名称】特徴語抽出装置及びプログラム

(57)【特許請求の範囲】
【請求項１】
文書ＩＤ及び内容テキスト情報を有する複数の文書を記憶する文書記憶手段と、
カテゴリＩＤ毎に１つ以上の文書ＩＤを関連付けて記憶するカテゴリ記憶手段と、
前記文書記憶手段内の文書毎に、当該文書の文書ＩＤと、当該文書の内容テキスト情報から抽出された文書特徴語とを関連付けて記憶する文書特徴語記憶手段と、
前記カテゴリ記憶手段内で関連したカテゴリＩＤ及び１つ以上の文書ＩＤと、当該カテゴリＩＤに関連したカテゴリ特徴語とを関連付けて記憶するカテゴリ特徴語記憶手段と、
前記文書記憶手段内の文書毎に内容テキスト情報を形態素解析し、当該形態素解析の結果から文書特徴語を抽出し、当該抽出した文書特徴語と、当該文書特徴語に対応する文書の文書ＩＤとを関連付けて前記文書特徴語記憶手段に書き込む文書特徴語抽出手段と、
前記複数の文書により構成される文書集合を入力とし、当該文書集合に含まれる文書ＩＤに関連した文書特徴語が、当該文書ＩＤの文書中で出現する文書数を算出する出現文書数算出手段と、
前記出現文書数算出手段により算出された文書数に基づいて、全文書中におけるカテゴリＩＤに関連した文書に対する当該文書特徴語の特徴度を算出する特徴度算出手段と、
この文書特徴語に当該特徴度を付加したカテゴリ特徴語を作成し、当該作成したカテゴリ特徴語と当該カテゴリ特徴語に関連したカテゴリＩＤ及び１つ以上の文書ＩＤとを関連付けて前記カテゴリ特徴語記憶手段に書き込むカテゴリ特徴語作成手段と、
前記カテゴリ記憶手段内のカテゴリＩＤ毎に、当該カテゴリＩＤに関連した文書ＩＤの

個数を含むカテゴリ個数データを提示するカテゴリ個数提示手段と、
前記カテゴリ個数提示手段によるカテゴリ個数データの提示中、いずれかのカテゴリ個数データの選択を受け付けるカテゴリ個数データ選択受付手段と、
前記カテゴリ個数データ選択受付手段による選択を受け付けたカテゴリ個数データのカテゴリＩＤに関連したカテゴリ特徴語のうち、特徴度が上位のカテゴリ特徴語における文書特徴語をカテゴリ特徴語として提示するカテゴリ特徴語提示手段と、
前記カテゴリ個数提示手段によるカテゴリ個数データの提示中、複数個のカテゴリ個数データの各カテゴリＩＤの和集合である複数の比較対象からなる比較対象集合の選択を受け付ける比較対象集合選択受付手段と、
前記比較対象集合選択受付手段による選択を受け付けた比較対象集合内の各カテゴリＩＤに関連付けられた各文書ＩＤに関連した文書特徴語毎に、前記出現文書数算出手段により算出された文書数に基づいて、各比較対象における当該文書特徴語の相違性を表す度合いの相違特徴度を算出し、相違特徴度が上位の文書特徴語をカテゴリ相違特徴語として送出するカテゴリ相違特徴語送出手段と、
前記カテゴリ相違特徴語送出手段により送出されたカテゴリ相違特徴語を提示するカテゴリ相違特徴語提示手段と、
を備えたことを特徴とする特徴語抽出装置。
【請求項２】
請求項１に記載の特徴語抽出装置において、
前記出現文書数算出手段は、
前記文書記憶手段内の文書特徴語毎に、前記文書記憶手段の全ての文書中で当該文書特徴語が出現する文書数を算出する全文書中出現文書数算出手段と、
前記カテゴリ記憶手段内のカテゴリＩＤ毎に、当該カテゴリＩＤに関連付けられた文書ＩＤに関連した文書特徴語が当該文書ＩＤの文書中で出現する文書数を算出するカテゴリ文書中出現文書数算出手段と、
前記比較対象集合選択受付手段による選択を受け付けた比較対象集合内の各カテゴリＩＤに関連付けられた各文書ＩＤに関連した文書特徴語毎に、当該各文書ＩＤに関連した全ての文書中で当該文書特徴語が出現する文書数を算出する第１文書数算出手段と、
前記比較対象集合選択受付手段による選択を受け付けた比較対象集合内の比較対象毎に、当該比較対象内の各カテゴリＩＤに関連付けられた各文書ＩＤに関連した文書特徴語が当該各文書ＩＤの文書中で出現する文書数を算出する第２文書数算出手段と、
を備えたことを特徴とする特徴語抽出装置。
【請求項３】
請求項１または２に記載の特徴語抽出装置において、
前記選択を受け付けた比較対象集合内の各カテゴリＩＤに関連付けられた各文書ＩＤに関連した文書特徴語毎に、前記第１文書数算出手段により算出された文書数と、前記第２文書数算出手段により算出された文書数とに基づいて、前記比較対象集合における当該文書特徴語の共通性を表す度合いの共通特徴度を算出し、共通特徴度が上位の文書特徴語をカテゴリ共通特徴語として送出するカテゴリ共通特徴語送出手段と、
前記カテゴリ共通特徴語送出手段により送出されたカテゴリ共通特徴語を提示するカテゴリ共通特徴語提示手段と、
を更に備えたことを特徴とする特徴語抽出装置。
【請求項４】
請求項１または２に記載の特徴語抽出装置において、
前記カテゴリ個数提示手段によるカテゴリ相違特徴語の提示中、複数のカテゴリ相違特徴語からなる着目語集合の選択を受け付ける着目語集合選択受付手段と、
前記着目語集合選択受付手段による選択を受け付けた着目語集合と、前記カテゴリ特徴語記憶手段内のカテゴリ特徴語との関連度を算出し、関連度の高いカテゴリ特徴語に関連付けられたカテゴリＩＤに関連したカテゴリ個数データを強調表示する関連カテゴリ提示手段と、

を更に備えたことを特徴とする特徴語抽出装置。
【請求項５】
文書記憶手段、カテゴリ記憶手段、文書特徴語記憶手段及びカテゴリ特徴語記憶手段を備えた特徴語抽出装置のプログラムであって、
前記特徴語抽出装置を、
文書ＩＤ及び内容テキスト情報を有する複数の文書を前記文書記憶手段に書き込む文書書込手段、
カテゴリＩＤ毎に１つ以上の文書ＩＤを関連付けて前記カテゴリ記憶手段に書き込むカテゴリ書込手段、
前記文書記憶手段内の文書毎に内容テキスト情報を形態素解析し、当該形態素解析の結果から文書特徴語を抽出し、当該抽出した文書特徴語と、当該文書特徴語に対応する文書の文書ＩＤとを関連付けて前記文書特徴語記憶手段に書き込む文書特徴語抽出手段、
前記文書記憶手段内の文書特徴語毎に、前記文書記憶手段の全ての文書中で当該文書特徴語が出現する文書数を算出する全文書中出現文書数算出手段、
前記カテゴリ記憶手段内のカテゴリＩＤ毎に、当該カテゴリＩＤに関連付けられた文書ＩＤに関連した文書特徴語が当該文書ＩＤの文書中で出現する文書数を算出するカテゴリ文書中出現文書数算出手段、
前記全文書中出現文書数算出手段により算出された文書数と、前記カテゴリ文書中出現文書数算出手段により算出された文書数とに基づいて、全文書中におけるカテゴリＩＤに関連した文書に対する当該文書特徴語の特徴度を算出する特徴度算出手段、
この文書特徴語に当該特徴度を付加したカテゴリ特徴語を作成し、当該作成したカテゴリ特徴語と当該カテゴリ特徴語に関連したカテゴリＩＤ及び１つ以上の文書ＩＤとを関連付けて前記カテゴリ特徴語記憶手段に書き込むカテゴリ特徴語作成手段、
前記カテゴリ記憶手段内のカテゴリＩＤ毎に、当該カテゴリＩＤに関連付けられた文書ＩＤの個数を含むカテゴリ個数データを提示するカテゴリ個数提示手段、
前記カテゴリ個数データの提示中、いずれかのカテゴリ個数データの選択を受け付けるカテゴリ個数データ選択受付手段、
前記カテゴリ個数データ選択受付手段による選択を受け付けたカテゴリ個数データのカテゴリＩＤに関連したカテゴリ特徴語のうち、特徴度が上位のカテゴリ特徴語における文書特徴語をカテゴリ特徴語として提示するカテゴリ特徴語提示手段、
前記カテゴリ個数提示手段によるカテゴリ個数データの提示中、複数個のカテゴリ個数データの各カテゴリＩＤの和集合である複数の比較対象からなる比較対象集合の選択を受け付ける比較対象集合選択受付手段、
前記比較対象集合選択受付手段による選択を受け付けた比較対象集合内の各カテゴリＩＤに関連付けられた各文書ＩＤに関連した文書特徴語毎に、当該各文書ＩＤに関連した全ての文書中で当該文書特徴語が出現する文書数を算出する第１文書数算出手段と、
前記比較対象集合選択受付手段による選択を受け付けた比較対象集合内の比較対象毎に、当該比較対象内の各カテゴリＩＤに関連付けられた各文書ＩＤに関連した文書特徴語が当該各文書ＩＤの文書中で出現する文書数を算出する第２文書数算出手段と、
前記比較対象集合選択受付手段による選択を受け付けた比較対象集合内の各カテゴリＩＤに関連付けられた各文書ＩＤに関連した文書特徴語毎に、前記第１文書数算出手段により算出された文書数と、前記第２文書数算出手段により算出された文書数とに基づいて、各比較対象における当該文書特徴語の相違性を表す度合いの相違特徴度を算出し、相違特徴度が上位の文書特徴語をカテゴリ相違特徴語として送出するカテゴリ相違特徴語送出手段と、
前記カテゴリ相違特徴語送出手段により送出されたカテゴリ相違特徴語を提示するカテゴリ相違特徴語提示手段、
として機能させるためのプログラム。
【発明の詳細な説明】
【技術分野】

【０００１】
本発明は、電子化された大量の文書に対し、各文書の内容を特徴づける単語である特徴語を抽出するための特徴語抽出装置及びプログラムに関する。
【背景技術】
【０００２】
特許調査やアンケート分析などにおいては、特徴語を抽出し、複数の文書集合の内容や傾向を比較したいというニーズがある。例えば、特許調査においては、自社と競合他社との各年代の特許出願の傾向を比較したいニーズがある。この種の調査の質は、選定する比較範囲や特徴語に影響される。
【０００３】
しかしながら、適切な比較範囲や特徴語は、調査の目的や文書集合の内容によって異なる。このため、比較範囲や特徴語の選定には、文書集合の内容に関連した知識や、目的を踏まえた調査自体に関するスキルが必要である。
【０００４】
これに対し、特許文献１では適切な分析軸を提示する技術を提案している。特許文献１に記載の技術は、データに含まれる属性毎に単語を抽出し、単語の属性値毎の出現頻度を集計手段で集計し、集計した出現頻度を視認し易いようにユーザに表示するとともに、単語の出現傾向から分析に適した属性を抽出しユーザに提示する。これにより、ユーザに適切な分類軸の選択を支援する。
【０００５】
また、特許文献２に記載の技術では、分析に使用する適切な特徴語の選定のために、比較する属性（例えば、作成日）について、属性値毎（例えば、月毎）に抽出される特徴語の共起関係に基づき、各属性において相違点を提示する。これにより、文書集合の内容をより好適に分析可能としている。なお、特許文献２に記載の技術で用いる「共起」については、例えば、特許文献３にまとめられている。
【先行技術文献】
【特許文献】
【０００６】
【特許文献１】特開２００６－１７１９３１号公報
【特許文献２】特開２００２－２４５０７０号公報
【非特許文献】
【０００７】
【非特許文献１】内山将夫，中條清美，山本英子，井佐原均．「英語教育のための分野特徴単語の選定尺度の比較」，自然言語処理，11 (3)，165-197，2004．
【非特許文献２】岸田和明．「検索実験における評価指標としての平均精度の性質」，情報処理学会論文誌：データベース，第４３巻，第ＳＩＧ２（ＴＯＤ１３）号（２００２）．
【非特許文献３】相澤彰子．「共起に基づく類似性尺度」，オペレーションズ・リサーチ，２００７年１１月号，pp.706(20)－712(26)．
【発明の概要】
【発明が解決しようとする課題】
【０００８】
しかしながら、以上のような特許文献１，２に記載の技術では、通常は特に問題ないが、本発明者の検討によれば、以下に述べる点で改良の余地がある。
【０００９】
例えば、特許文献１に記載の技術では、ユーザに提示される分析軸が事前に文書データの属性として定義されている必要がある。そのため、提示される分析軸が事前に定義された属性に限られるため、意図する分析を行えない点で改良の余地がある。
【００１０】
特許文献２に記載の技術は、各属性値に対する相違点を表すことにより、文書集合の内

容の明確化を図っている。このため、分析の対象が属性値に縛られ、ユーザが任意の範囲で文書集合を比較できない点で改良の余地がある。

【００１１】

また、特許文献２に記載の技術は、膨大な文書集合を比較分析する場合、ユーザが文書集合の中で何に着目すべきかを把握できない場合がある。例えば、着目する「画像認識」の技術を先行調査するために、数千・数万の特許文献を出願人と出願年月（１９９０年～２００８年の各月）でクロス分析する場合を考える。出願人として数１０～１００社程度の各企業を各行に配置し、出願年月として１００個程度の各月を各列に配置した場合、クロス分析のマトリックス全体として１万前後のセルが構成される。

【００１２】

これらのセルは、「画像認識」に関連する多数の特許文献が含まれるセルや、「画像認識」に無関係の多数の特許文献が含まれるセルなどがあり、着目する「画像認識」との関連度にはムラがある。企業毎や出願年毎でも同様のことが言える。

【００１３】

これに対し、ユーザは、着目する技術に関連が強い企業や出願年に関するセルの文書集合に比較範囲を絞り込むことで、より精緻な調査を行いたいというニーズがある。

【００１４】

しかしながら、特許文献１，２に記載の技術では、着目すべき比較範囲の絞り込みを支援できず、比較範囲を柔軟に変更することもできない。また、着目する技術に関連の強いセルの特徴語を参照すればユーザは意識しなかった関連技術を発見できるが、特許文献１，２に記載の技術では、文書集合の内容理解までに留まり、新たに着目すべき特徴語の参照を支援することはできない。

【００１５】

本発明は上記実情を考慮してなされたもので、事前に定義された属性に限らずに分析軸の候補として特徴語を提示できると共に、着目すべき比較範囲の絞り込みや、着目すべき特徴語の参照を支援し得る特徴語抽出装置及びプログラムを提供することを目的とする。

【課題を解決するための手段】

【００１６】

本発明の一つの局面は、特徴語抽出装置であって、文書ＩＤ及び内容テキスト情報を有する複数の文書を記憶する文書記憶手段と、カテゴリＩＤ毎に１つ以上の文書ＩＤを関連付けて記憶するカテゴリ記憶手段と、前記文書記憶手段内の文書毎に、当該文書の文書ＩＤと、当該文書の内容テキスト情報から抽出された文書特徴語とを関連付けて記憶する文書特徴語記憶手段と、前記カテゴリ記憶手段内で関連したカテゴリＩＤ及び１つ以上の文書ＩＤと、当該カテゴリＩＤに関連したカテゴリ特徴語とを関連付けて記憶するカテゴリ特徴語記憶手段と、前記文書記憶手段内の文書毎に内容テキスト情報を形態素解析し、当該形態素解析の結果から文書特徴語を抽出し、当該抽出した文書特徴語と、当該文書特徴語に対応する文書の文書ＩＤとを関連付けて前記文書特徴語記憶手段に書き込む文書特徴語抽出手段と、前記文書記憶手段内の文書特徴語毎に、前記文書記憶手段の全ての文書中で当該文書特徴語が出現する文書数を算出する全文書中出現文書数算出手段と、前記カテゴリ記憶手段内のカテゴリＩＤ毎に、当該カテゴリＩＤに関連付けられた文書ＩＤに関連した文書特徴語が当該文書ＩＤの文書中で出現する文書数を算出するカテゴリ文書中出現文書数算出手段と、前記全文書中出現文書数算出手段により算出された文書数と、前記カテゴリ文書中出現文書数算出手段により算出された文書数とに基づいて、全文書中におけるカテゴリＩＤに関連した文書に対する当該文書特徴語の特徴度を算出する特徴度算出手段と、この文書特徴語に当該特徴度を付加したカテゴリ特徴語を作成し、当該作成したカテゴリ特徴語と当該カテゴリ特徴語に関連したカテゴリＩＤ及び１つ以上の文書ＩＤとを関連付けて前記カテゴリ特徴語記憶手段に書き込むカテゴリ特徴語作成手段と、前記カテゴリ記憶手段内のカテゴリＩＤ毎に、当該カテゴリＩＤに関連付けられた文書ＩＤの個数を含むカテゴリ個数データを提示するカテゴリ個数提示手段と、前記カテゴリ個数データの提示中、いずれかのカテゴリ個数データの選択を受け付けるカテゴリ個数データ選択受

付手段と、前記カテゴリ個数データ選択受付手段による選択を受け付けたカテゴリ個数データのカテゴリＩＤに関連したカテゴリ特徴語のうち、特徴度が上位のカテゴリ特徴語における文書特徴語をカテゴリ特徴語として提示するカテゴリ特徴語提示手段と、前記カテゴリ特徴語提示手段によるカテゴリ個数データの提示中、複数個のカテゴリ個数データの各カテゴリＩＤの和集合である複数の比較対象からなる比較対象集合の選択を受け付ける比較対象集合選択受付手段と、前記比較対象集合選択受付手段による選択を受け付けた比較対象集合内の各カテゴリＩＤに関連付けられた各文書ＩＤに関連した文書特徴語毎に、当該各文書ＩＤに関連した全ての文書中で当該文書特徴語が出現する文書数を算出する第１文書数算出手段と、前記比較対象集合選択受付手段による選択を受け付けた比較対象集合内の比較対象毎に、当該比較対象内の各カテゴリＩＤに関連付けられた各文書ＩＤに関連した文書特徴語が当該各文書ＩＤの文書中で出現する文書数を算出する第２文書数算出手段と、前記比較対象集合選択受付手段による選択を受け付けた比較対象集合内の各カテゴリＩＤに関連付けられた各文書ＩＤに関連した文書特徴語毎に、前記第１文書数算出手段により算出された文書数と、前記第２文書数算出手段により算出された文書数とに基づいて、各比較対象における当該文書特徴語の相違性を表す度合いの相違特徴度を算出し、相違特徴度が上位の文書特徴語をカテゴリ相違特徴語として送出するカテゴリ相違特徴語送出手段と、前記カテゴリ相違特徴語送出手段により送出されたカテゴリ相違特徴語を提示するカテゴリ相違特徴語提示手段と、を備えた特徴語抽出装置である。

【００１７】

なお、本発明の一つの局面は、装置として表現したが、これに限らず、方法、プログラム又はプログラムを記憶したコンピュータ読取り可能な記憶媒体として表現してもよい。

【００１８】

（作用）

このような本発明の一つの局面においては、カテゴリＩＤに関連した文書ＩＤの個数を含むカテゴリ個数データの提示中に、選択を受け付けたカテゴリ個数データのカテゴリＩＤに関連したカテゴリ特徴語のうち、特徴度が上位のカテゴリ特徴語における文書特徴語をカテゴリ特徴語として提示する。

【００１９】

また、本発明の一つの局面においては、カテゴリ個数データの提示中、複数個のカテゴリ個数データの各カテゴリＩＤの和集合である複数の比較対象からなる比較対象集合の選択を受け付けると、選択を受け付けた比較対象集合内の各カテゴリＩＤに関連付けられた各文書ＩＤに関連した文書特徴語毎に、各比較対象における当該文書特徴語の相違性を表す度合いの相違特徴度を算出し、相違特徴度が上位の文書特徴語をカテゴリ相違特徴語として提示する。

【００２０】

このように、選択したカテゴリ個数データに応じて、カテゴリ特徴語やカテゴリ相違特徴語といった特徴語を提示する構成により、事前に定義された属性に限らずに分析軸の候補として特徴語を提示できると共に、着目すべき比較範囲の絞り込みや、着目すべき特徴語の参照を支援することができる。

【発明の効果】

【００２１】

以上説明したように本発明によれば、事前に定義された属性に限らずに分析軸の候補として特徴語を提示できると共に、着目すべき比較範囲の絞り込みや、着目すべき特徴語の参照を支援できる。

【図面の簡単な説明】

【００２２】

【図１】本発明の一実施形態に係る特徴語抽出装置の構成を示すブロック図である。

【図２】同実施形態における文書記憶部を説明するための模式図である。

【図３】同実施形態におけるカテゴリ記憶部を説明するための模式図である。

【図４】同実施形態における特徴語記憶部を説明するための模式図である。

【図５】同実施形態における特徴語抽出部の動作を説明するためのフローチャートである。
【図６】同実施形態におけるカテゴリ特徴語抽出部の動作を説明するためのフローチャートである。
【図７】同実施形態におけるカテゴリ共通特徴語抽出部の動作を説明するためのフローチャートである。
【図８】同実施形態におけるカテゴリ個数データをセルに提示した画面例を示す模式図である。
【図９】同実施形態における比較対象集合を選択した画面例を示す模式図である。
【図１０】同実施形態における共通特徴語を表示した画面例を示す模式図である。
【図１１】同実施形態におけるカテゴリ相違特徴語抽出部の動作を説明するためのフローチャートである。
【図１２】同実施形態における共通特徴語と相違特徴語を表示した画面例を示す模式図である。
【図１３】同実施形態における関連カテゴリ提示部の動作を説明するためのフローチャートである。
【図１４】同実施形態におけるユーザ操作・提示部の動作を説明するためのフローチャートである。
【図１５】同実施形態におけるカテゴリ特徴語を表示した画面例を示す模式図である。
【図１６】同実施形態における関連カテゴリのセルを強調表示した画面例を示す模式図である。
【図１７】同実施形態におけるカテゴリ特徴語から着目語を選択したときの画面例を示す模式図である。
【図１８】同実施形態における比較対象の絞り込みと特徴語の表示例を示す模式図である。
【図１９】同実施形態における着目語の変更と関連カテゴリの表示例を示す模式図である。
【図２０】同実施形態におけるクロス分析の画面例を示す模式図である。
【図２１】同実施形態における他のクロス分析の画面例を示す模式図である。
【図２２】同実施形態における更に他のクロス分析の画面例を示す模式図である。
【図２３】同実施形態におけるグラフ表示の画面例を示す模式図である。
【発明を実施するための形態】
【００２３】
以下、本発明の一実施形態について図面を用いて説明する。なお、以下の装置は、装置毎に、ハードウェア構成、又はハードウェア資源とソフトウェアとの組合せ構成のいずれでも実施可能となっている。組合せ構成のソフトウェアとしては、予めネットワーク又は記憶媒体から対応する装置のコンピュータにインストールされ、対応する装置の機能を実現させるためのプログラムが用いられる。また、以下の説明で用いられる用語と記号の定義は、次の表１及び表２に示す通りである。
【００２４】

【表１】

［表１］

用語と記号	定義
全文書集合 docAll	分析対象の全文書を含む文書集合
カテゴリ cat	複数の所属文書から構成される文書集合。 カテゴリ特徴語の抽出単位。
比較対象 cmp	複数のカテゴリ(cat1, cat2, ...)によって指定される文書集合。 指定されたカテゴリの和集合。 cmp = cat1 ∪ cat2 ∪ ... (cat i はカテゴリ)
比較範囲 tgtDocs	複数の比較対象(比較対象集合)によって指定される文書集合。 複数の比較対象の和集合。 tgtDocs = cmp1 ∪ cmp2 ∪ ... ∪ cmpM (cmp i は比較対象)
差集合cmpDocs_i	比較範囲tgtDocsから比較対象cmp_iを除いた差集合 cmpDocs_i = tgtDocs - cmp_i
比較対象集合 tgtSet	複数の比較対象を要素とする集合。 tgtSet = {cmp1, cmp2, ...} (cmp i は比較対象)
df(t, docSet)	文書集合docSet中の単語tの出現頻度。 docSetとしては、以下の文書集合を引数とする。 - 全文書集合docAll - カテゴリcat - 比較対象cmp - 比較範囲tgtDocs
|docSet|	文書集合docSet中に含まれる文書数 docSetとしては、以下の文書集合を引数とする。 - 全文書集合docAll - カテゴリcat - 比較対象cmp - 比較範囲tgtDocs

【００２５】

【表２】

［表２］

用語と記号	定義
特徴度score(t, cat)	全文書集合docAllにおける、カテゴリcatに対する単語tの単語特徴度
共通特徴度 com(t, tgtSet)	比較対象集合tgtSetに対する、単語tの共通特徴度。 比較対象cmp_iの評価値eval(t,cmp_i)の総和 $\mathrm{com(t,\ tgtSet)} = \sum_{i} \mathrm{eval(t, cmp\_i)}$
相違特徴度 diff(t, cmp)	比較対象cmpに対する、単語tの相違特徴度
関連度 rel(cat, tgtTerms)	カテゴリcatと、着目語集合tgtTermsとの関連度。 平均精度、 又は着目語集合tgtTerms内の単語ｔの特徴度score(t, cat)の総和 $\mathrm{rel(cat, tgtTerms)} = \sum_{t} \mathrm{score(t, cat)}$
共通特徴語 comTerms	比較対象集合cmpに対して共通点を表す単語の集合。 comTerms = {term1, term2, ...} (term i は単語)
相違特徴語 iffTerms(cmp)	比較対象集合cmpに対する相違点を表す単語の集合。 diffTerms(cmp) = {term1, term2, ...} (term i は単語) 引数cmpにより、指定される比較対象の相違特徴語を返す。 また、diffTerms と表記した場合は、複数のカテゴリcmpについて、それぞれの相違特徴語 diffTerms(cmp)を、要素とする集合とする。 diffTerms = {diffTerms(cmp1), diffTerms(cmp2), ...} (cmp i は比較対象))
着目語集合 tgtTerms	複数の単語から構成される集合。 tgtTerms = {term1, term2, ...} (term i は単語)

【００２６】

なお、以下で説明する実施形態においては、複数の文書からなる文書集合を入力とし、その文書集合に含まれる文書ＩＤに関連した文書特徴語が、当該文書ＩＤの文書中で出現する文書数（文書の中で出現する特徴後の数を含む概念）を算出する処理（出現文書数算出機能）に、特に特徴があるといえる。

【００２７】

従って、全文書中出現文書数算出機能と、カテゴリ文書中出現文書数算出機能と、第１文書数算出機能と、第２文書数算出機能とを例にして、出現文書数算出機能を説明していく。なぜならば、上記の各表で定義されている全文書集合、カテゴリ、比較対象集合（比較対象の集合）、比較対象（カテゴリの集合）などについて、各入力の種類は異なっていても、これらはいずれも文書集合（文書ＩＤの集合）と換言することができるからである。

【００２８】

図１は本発明の一実施形態に係る特徴語抽出装置の構成を示すブロック図であり、図２乃至図４は同装置内の各記憶部１０，２０，３０を説明するための模式図である。この特徴語抽出装置は、文書記憶部１０、カテゴリ記憶部２０、特徴語記憶部３０、特徴語抽出

部４０及びユーザ操作・提示部５０を備えている。

【００２９】

文書記憶部１０は、各部４０，５０から読出／書込可能な記憶装置であり、図２に示すように、文書データ１０ｄを記憶している。文書データ１０ｄは、各文書を識別する文書ＩＤとしての文書１１ｄと、内容テキスト情報（文字列情報）１２ｄとしての文書名１２ｄ及び／又は本文１５ｄとを有する複数の文書を電子化したデータであり、ここでは特許文献の例が図示されている。なお、文書データ１０ｄは、文書ＩＤ１１ｄと内容テキスト情報に加え、出願日１３ｄや出願人１４ｄなどの属性値を有していてもよい。

【００３０】

カテゴリ記憶部２０は、各部４０，５０から読出／書込可能な記憶装置であり、図３に示すように、カテゴリＩＤ２１ｃ毎に１つ以上の文書ＩＤからなる所属文書情報２２ｃを関連付けて記憶している。ここで、カテゴリＩＤ及び所属文書情報２２ｃの集合をカテゴリデータ２０ｃと呼ぶ。１つのカテゴリＩＤ２１ｃに関連付けられた所属文書情報２２ｃ内の文書ＩＤの集合は特徴語抽出の最小単位となる文書集合を示しており、この文書集合がカテゴリとも呼ばれる。例えば、カテゴリＩＤ＝Ｃ０１で識別されるカテゴリは、文書ＩＤ＝Ｄ１７、Ｄ２３、Ｄ４１で識別される文書が所属している。このカテゴリデータ２０ｃは、予め与えられるものである。例えば、文書クラスタリングなどの文書分類技術による分類結果をカテゴリデータ２０ｃとしてもよく、文書の作成年や作成者といった属性値によって分割される文書の集合をカテゴリデータ２０ｃとしてもよい。さらに、一つの文書がただ一つのカテゴリに所属するようなカテゴリ構造でもよく、１つの文書が複数のカテゴリに所属するようなカテゴリ構造でもよい。カテゴリデータ２０ｃは、カテゴリＩＤ２１ｃと所属文書情報２２ｃ以外にも、カテゴリ名やラベルなどの属性情報を有していてもよい。

【００３１】

特徴語記憶部３０は、各部４０，５０からから読出／書込可能な記憶装置であり、図４に示すように、文書特徴語データ３０ｄｔ及びカテゴリ特徴語データ３０ｃｔを記憶する。

【００３２】

文書特徴語データ３０ｄｔは、文書記憶部１０内の文書毎に、当該文書の文書ＩＤ３１ｄｔと、当該文書の内容テキスト情報から抽出された文書特徴語３２ｄｔとを関連付けたデータである。この文書特徴語３２ｄｔは、文書特徴語抽出部４１において、文書記憶部１０に記憶された文書データの内容テキスト情報を形態素解析して得られた単語の集合から、不要語を除去して抽出された単語の集合である。不要語の除去では、名詞や未知語といった品詞で、特徴語として利用する単語の条件に合致しない単語や、”こと”や”もの”という一般性が高く特徴語として不適切な単語を排除する。反対に、文書中に１回しか出現しないような出現頻度が極端に少ない単語も不要語として排除してもよい。特許文献やメール文書といった特徴語抽出の対象となる文書の種類や、調査や分析といった特徴語抽出の目的などに応じて、保持する品詞の種類を変更することができる。この例では、文書特徴語データ３０ｄｔとして文書特徴語３２ｄｔを単語のみで保持しているが、文書中での単語の出現回数ＴＦを各文書特徴語３２ｄｔの当該単語に関連付けて保持してもよい。ＴＦは、特徴語抽出において、単語の特徴語を求める際の１つの指標として利用することができる。

【００３３】

カテゴリ特徴語データ３０ｃｔは、カテゴリ記憶部２０内のカテゴリＩＤ２１ｃ及び文書所属情報２２ｃと同一のカテゴリＩＤ３１ｃｔ及び所属文書情報３２ｃｔと、当該カテゴリＩＤ３１ｃｔに関連したカテゴリ特徴語３３ｃｔとを関連付けたデータである。カテゴリ特徴語３３ｃｔは、所属文書情報３２ｃｔ内の文書ＩＤに関連した文書特徴語３２ｄｔである各単語と、当該各単語に付加された特徴度とからなる。

【００３４】

特徴語抽出部４０は、文書特徴語抽出部４１、カテゴリ特徴語抽出部４２、カテゴリ共

通特徴語抽出部４３及びカテゴリ相違特徴語抽出部４４を備えている。なお、カテゴリ共通特徴語抽出部４３及びカテゴリ相違特徴語抽出部４４は、いずれか一方があれば文書集合の分析が可能なため、いずれか一方を残し、他方を省略することも可能である。

【００３５】

文書特徴語抽出部４１は、文書記憶部１０内の文書毎に内容テキスト情報を形態素解析し、形態素解析の結果から文書特徴語を抽出し、当該抽出した文書特徴語と、当該文書特徴語に対応する文書の文書ＩＤとを関連付けた文書特徴語データ３０ｄｔを文書特徴語記憶部３０に書き込む機能をもっている。ここで、文書特徴語の抽出は、例えば形態素解析の結果から、文書中に１回しか出現してないなど、特徴語抽出において不要な単語（不要語）を排除する処理により実行すればよい。

【００３６】

カテゴリ特徴語抽出部４２は、以下の各機能(f42-1)～(f42-5)をもっている。

(f42-1)　文書記憶部１０内の文書特徴語毎に、文書記憶部１０の全ての文書中で当該文書特徴語が出現する文書数を算出する全文書中出現文書数算出機能。

【００３７】

(f42-2)　カテゴリ記憶部２０内のカテゴリＩＤ２１ｃ毎に、当該カテゴリＩＤ２１ｃに関連付けられた文書ＩＤに関連した文書特徴語が当該文書ＩＤの文書中で出現する文書数を算出するカテゴリ文書中出現文書数算出機能。

【００３８】

(f42-3)　全文書中出現文書数算出機能により算出された文書数と、カテゴリ文書中出現文書数算出機能により算出された文書数とに基づいて、全文書中におけるカテゴリＩＤ２１ｃに関連した文書に対する当該文書特徴語の特徴度を算出する特徴度算出機能。文書特徴語の特徴度は、カテゴリに属する文書の文書特徴語の統計情報に基づいて算出される。

【００３９】

(f42-4)　この文書特徴語に当該特徴度を付加したカテゴリ特徴語３３ｃｔを作成する機能。

【００４０】

(f42-5)　当該作成したカテゴリ特徴語３３ｃｔと当該カテゴリ特徴語３３ｃｔに関連したカテゴリＩＤ３１ｃｔ及び所属文書情報３２ｃｔとを関連付けたカテゴリ特徴語データ３０ｃｔを特徴語記憶部３０に書き込む機能。

【００４１】

カテゴリ共通特徴語抽出部４３は、以下の各機能(f43-1)～(f43-3)をもっている。

【００４２】

(f43-1)　ユーザ操作により共通・相違特徴語提示部５３が選択を受け付けた比較対象集合内の各カテゴリＩＤ２１ｃに関連付けられた所属文書情報２２ｃの各文書ＩＤに関連した文書特徴語毎に、当該各文書ＩＤに関連した全ての文書中で当該文書特徴語が出現する文書数を算出する第１文書数算出機能。

【００４３】

(f43-2)　選択を受け付けた比較対象集合内の比較対象毎に、当該比較対象内の各カテゴリＩＤ２１ｃに関連付けられた所属文書情報２２ｃの各文書ＩＤに関連した文書特徴語が当該各文書ＩＤの文書中で出現する文書数を算出する第２文書数算出機能。

【００４４】

(f43-3)　選択を受け付けた比較対象集合内の各カテゴリＩＤ２１ｃに関連付けられた所属文書情報２２ｃの各文書ＩＤに関連した文書特徴語毎に、第１文書数算出機能により算出された文書数と、第２文書数算出機能により算出された文書数とに基づいて、比較対象集合における当該文書特徴語の共通性を表す度合いの共通特徴度を算出し、共通特徴度が上位の文書特徴語をカテゴリ共通特徴語として共通・相違特徴語提示部５３に送出するカテゴリ共通特徴語送出機能。ここで、共通特徴度は、各比較対象集合における各特徴語の共通性を表す度合いであり、比較対象集合に属する文書の文書集合の統計情報に基づい

て算出される。

【００４５】

カテゴリ相違特徴語抽出部４４は、以下の各機能(f44-1)～(f44-3)をもっている。

【００４６】

(f44-1) ユーザ操作により共通・相違特徴語提示部５３が選択を受け付けた比較対象集合内の各カテゴリＩＤ２１ｃに関連付けられた所属文書情報２２ｃの各文書ＩＤに関連した文書特徴語毎に、当該各文書ＩＤに関連した全ての文書中で当該文書特徴語が出現する文書数を算出する第１文書数算出機能。

【００４７】

(f44-2) 選択を受け付けた比較対象集合内の比較対象毎に、当該比較対象内の各カテゴリＩＤ２１ｃに関連付けられた所属文書情報２２ｃの各文書ＩＤに関連した文書特徴語が当該各文書ＩＤの文書中で出現する文書数を算出する第２文書数算出機能。

【００４８】

(f44-3) 選択を受け付けた比較対象集合内の各カテゴリＩＤ２１ｃに関連付けられた所属文書情報２２ｃの各文書ＩＤに関連した文書特徴語毎に、第１文書数算出機能により算出された文書数と、第２文書数算出機能により算出された文書数とに基づいて、各比較対象における当該文書特徴語の相違性を表す度合いの相違特徴度を算出し、相違特徴度が上位の文書特徴語をカテゴリ相違特徴語として共通・相違特徴語提示部５３に送出するカテゴリ相違特徴語送出機能。ここで、相違特徴度は、各比較対象集合における各特徴語の相違性を表す度合いであり、それぞれの比較対象に属する文書の文書集合の統計情報に基づいて算出される。

【００４９】

ユーザ操作・提示部５０は、画面提示部５１、カテゴリ特徴語提示部５２、相違・共通特徴語提示部５３及び関連カテゴリ提示部５４を備えている。なお、関連カテゴリ提示部５４は、文書集合の分析に必須ではなく、省略してもよい。

【００５０】

画面提示部５１は、ユーザの操作に応じて、各記憶部１０，２０，３０を参照して画面データを作成する機能と、当該画面データに基づいて画面を提示する機能とをもっている。ここで、画面データとしては、例えば、カテゴリ記憶部２０内のカテゴリＩＤ２１ｃ毎に、当該カテゴリＩＤ２１ｃに関連付けられた文書所属情報２２ｃ内の文書ＩＤの個数を含むカテゴリ個数データを各セルに提示したクロス分析画面の画面データ、選択された特徴語を分析軸にするようにカテゴリ個数データを修正して各セルに提示したクロス分析画面の画面データ、提示中のクロス分析画面の画面データに基づくグラフ表示画面の画面データ、あるいは、提示中のグラフ表示画面の画面データに基づくクロス分析画面の画面データ、などがある。表示形式は、クロス表示やグラフ表示以外にも、文書集合を平面上に楕円などで表現したマップ表示や、コンピュータのファイルシステムで使われるようなフォルダ表示でもよい。

【００５１】

カテゴリ特徴語提示部５２は、ユーザによるカテゴリの選択を受け付け、カテゴリ特徴語記憶部３０から選択されたカテゴリに対応するカテゴリ特徴語データを取得する。取得したカテゴリ特徴語データに基づき、該カテゴリにおいて特徴度が上位の特徴語をカテゴリ特徴語として、ユーザに提示する。

【００５２】

相違・共通特徴語提示部５３は、ユーザによる比較対象集合の選択を受け付け、選択された比較対象集合を特徴語抽出部４０に送出する機能と、特徴語抽出部４０から受けた共通特徴語を提示する機能と、特徴語抽出部４０から各々の比較対象に対するそれぞれの相違特徴語を受けると、これらの相違特徴語を各々の比較対象に対応づけて提示する機能とをもっている。

【００５３】

関連カテゴリ提示部５４は、ユーザによる着目語集合の選択を受け付け、特徴語記憶部

３０に記憶されたカテゴリ特徴語データに基づき、その着目語集合と各カテゴリとの関連度を算出し、関連度が大きいカテゴリを関連カテゴリとして、該当するカテゴリ個数データを強調表示する機能をもっている。ここで、関連度が大きいカテゴリとしては、関連度がしきい値以上のカテゴリとしたが、これに限らず、関連度が上位ｓ個以内のカテゴリとしてもよい。

【００５４】

次に、以上のように構成された特徴語抽出装置の動作を図５乃至図２３のフローチャートや模式図を参照しながら説明する。

【００５５】

（特徴語抽出部４１の動作：図５）

特徴語抽出部４１は、概略的には、文書記憶部１０内の文書毎に内容テキスト情報を形態素解析し、当該形態素解析の結果から文書特徴語を抽出し、当該抽出した文書特徴語と、当該文書特徴語に対応する文書の文書ＩＤとを関連付けて特徴語記憶部３０に書き込む処理を実行する（Ｓ１～Ｓ４）。

【００５６】

具体的には、特徴語抽出部４１は、文書記憶部１０から、分析対象の全ての文書データの集合である全文書集合ｄｏｃＡｌｌを取得する（Ｓ１）。

【００５７】

次に、特徴語抽出部４１は、この全文書集合ｄｏｃＡｌｌに含まれる文書データｄｏｃ毎に、ステップＳ３とステップＳ４の処理を繰り返す（Ｓ２）。

【００５８】

すなわち、特徴語抽出部４１は、文書データｄｏｃ毎に内容テキスト情報を形態素解析する（Ｓ３）。また、特徴語抽出部４１は、この形態素解析の結果から、特徴語抽出の対象とする品詞以外の単語や、”こと”、”もの”などの不要語を排除して抽出した単語群を文書特徴語とする。しかる後、特徴語抽出部４１は、抽出した文書特徴語と文書ＩＤとを関連付けた文書特徴語データを特徴語記憶部３０に書き込む（Ｓ４）。

【００５９】

例えば、図２に示す文書データについて、本文２５を分析対象の内容テキスト情報とした場合、ステップＳ３とステップＳ４の手順により、図４に示すように、文書特徴語データ３０ｄｔが特徴語記憶部３０に書き込まれる。

【００６０】

（カテゴリ特徴語抽出部４２の動作：図６）

カテゴリ特徴語抽出部４２は、概略的には、文書記憶部１０内の文書特徴語毎に、文書記憶部１０の全ての文書中で当該文書特徴語が出現する文書数ｄｆ（ｔ、ｄｏｃＡｌｌ）を算出する全文書中出現文書数算出処理（Ｓ１１～Ｓ１３）と、カテゴリ記憶部２０内のカテゴリＩＤ２１ｃ毎に、当該カテゴリＩＤ２１ｃに関連付けられた所属文書情報２２ｃの文書ＩＤに関連した文書特徴語３２ｄｔが当該文書ＩＤの文書中で出現する文書数ｄｆ（ｔ、ｃａｔ）を算出するカテゴリ文書中出現文書数算出処理（Ｓ１４～Ｓ１８）と、全文書中出現文書数算出処理により算出された文書数ｄｆ（ｔ、ｄｏｃＡｌｌ）と、カテゴリ文書中出現文書数算出処理により算出された文書数ｄｆ（ｔ、ｃａｔ）とに基づいて、全文書中におけるカテゴリＩＤ２１ｃに関連した文書に対する当該文書特徴語３２ｄｔの特徴度ｓｃｏｒｅ（ｔ、ｃａｔ）を算出する特徴度算出処理（Ｓ１９）と、この文書特徴語３２ｄｔに当該特徴度ｓｃｏｒｅ（ｔ、ｃａｔ）を付加したカテゴリ特徴語３３ｃｔを作成し、当該作成したカテゴリ特徴語３３ｃｔと当該カテゴリ特徴語に関連したカテゴリＩＤ３１ｃｔ（カテゴリＩＤ２１ｃと同一値）及び１つ以上の文書ＩＤを含む所属文書情報３２ｃｔ（所属文書情報２２ｃと同一値）とを関連付けて特徴語記憶部３０に書き込む処理（Ｓ２０）とを実行する。

【００６１】

具体的には、カテゴリ特徴語抽出部４２は、特徴語記憶部３０から全ての文書ｄｏｃＡｌｌの文書特徴語データを取得する（Ｓ１１）。

【００６２】
次に、カテゴリ特徴語抽出部４２は、ステップＳ１１によって得られた文書特徴語データに含まれる文書特徴語ｔ毎に、ステップＳ１３の処理を繰り返す（Ｓ１２）。
【００６３】
すなわち、カテゴリ特徴語抽出部４２は、文書特徴語ｔ毎に、全ての文書ｄｏｃＡｌｌの文書特徴語データを参照しながら、全文書集合ｄｏｃＡｌｌ中で当該文書特徴語ｔが出現する文書数ｄｆ（ｔ、ｄｏｃＡｌｌ）を求める処理（Ｓ１３）を繰り返す。
【００６４】
しかる後、カテゴリ特徴語抽出部４２は、カテゴリ記憶部２０から全てのカテゴリデータ２０ｃを取得する（Ｓ１４）。
【００６５】
また、カテゴリ特徴語抽出部４２は、全てのカテゴリｃａｔについて、カテゴリＩＤ２１ｃ毎に、ステップＳ１６～Ｓ２０の処理を繰り返す（Ｓ１５）。
【００６６】
さらに、カテゴリ特徴語抽出部４２は、当該カテゴリＩＤ２１ｃに関連付けられた所属文書情報２２ｃの文書ＩＤに関連した文書特徴語を特徴語記憶部３０から読み出すことにより、特徴語記憶部３０から、カテゴリｃａｔに所属する複数の文書について、それぞれの文書の文書特徴語データを取得する（Ｓ１６）。
【００６７】
続いて、カテゴリ特徴語抽出部４２は、取得した文書特徴語データに含まれる文書特徴語ｔ毎に、ステップＳ１８の処理を繰り返す（Ｓ１７）。
【００６８】
カテゴリ特徴語抽出部４２は、文書特徴語ｔ毎に、ステップＳ１６で取得した文書特徴語データを参照しながら、カテゴリｃａｔに所属する複数の文書中で、文書特徴語ｔが出現する文書数ｄｆ（ｔ、ｃａｔ）を求める（Ｓ１８）。
【００６９】
カテゴリ特徴語抽出部４２は、ステップＳ１３で求めた文書数ｄｆ（ｔ、ｄｏｃＡｌｌ）と、ステップＳ１８で求めた文書数ｄｆ（ｔ、ｃａｔ）に基づき、全文書集合ｄｏｃＡｌｌにおけるカテゴリｃａｔに対する特徴語ｔの特徴度ｓｃｏｒｅ（ｔ、ｃａｔ）を算出する（Ｓ１９）。
【００７０】
具体的には、特徴度ｓｃｏｒｅ（ｔ、ｃａｔ）は、各文書数ｄｆ（ｔ、ｄｏｃＡｌｌ），ｄｆ（ｔ、ｃａｔ）に基づいて、表３に示すように、共通パラメータａ，ｂ，ｃ，ｄ，ｎを算出した後、表４に示す如き、いずれかの統計指標として算出される。
【００７１】
【表３】

［表３］

特徴度(t,cat)算出用の共通パラメータ	特徴度(t,cat)算出用の共通パラメータの算出式
a	a = df(t, cat)
b	b = df(t, docAll) − df(t, cat)
c	c = |cat| − df(t, cat)
d	d = |docAll| − df(t, docAll) − |cat| + df(t, cat)
n	n = a + b + c + d = |docAll|

【００７２】

【表４】

［表４］

統計指標の例	特徴度(t, cat)の算出式の例
対数尤度比 LLR	LLR' = a log(a n/((a+b)(a+c)))＋b log(b n/((a+b)(b+d))) ＋c log(c n/((c+d)(a+c)))＋d log(d n/((c+d)(b+d))) if((ad−bd) ＜ 0) LLR = −LLR' else LLR = LLR' score(t, cat) = LLR
ダイス係数 Dice	Dice = 2a/((a+b)＋(a+c)) score(t, cat) = Dice
イエーツ補正 χ２乗値 Yates'	$\mathrm{Yates}' = \dfrac{n\left(\lvert ad-bc\rvert - \dfrac{n}{2}\right)^2}{(a+b)(c+d)(a+c)(b+d)}$ if((ad-bd) < 0) Yates = −Yates' else Yates = Yates' score(t, cat) = Yates
自己相互情報量 MI	MI = log(a n/((a+b)(a+c))) score(t, cat) = MI

【００７３】

ここでは、対数尤度比ＬＬＲという統計指標として、各単語ｔの特徴度ｓｃｏｒｅ（ｔ，ｃａｔ）を算出している。但し、統計指標は、対数尤度比ＬＬＲに限らず、例えば、ダイス係数Ｄｉｃｅ、イエーツ補正χ２乗値Ｙａｔｅｓ’又は自己相互情報量ＭＩ等としてもよい。なお、各統計指標にはそれぞれ特徴があるため、各統計指標の特徴に応じて、得られる特徴語の傾向が異なる。

【００７４】

例えば、ダイス係数Ｄｉｃｅは、カテゴリｃａｔ内で単語ｔが出現する文書数ｄｆ（ｔ、ｃａｔ）の大きい単語ｔ（カテゴリｃａｔに多く含まれる単語（高頻度））を高く評価する。

【００７５】

イエーツ補正χ２乗値Ｙａｔｅｓ’は、全文書集合ｄｏｃＡｌｌ中での出現確率に対し、カテゴリｃａｔ中での出現確率が高い単語を高く評価する。結果として、イエーツ補正χ２乗値Ｙａｔｅｓ’は、対数尤度比ＬＬＲやダイス係数Ｄｉｃｅを利用した場合よりも、比較的低頻度の単語が特徴語として抽出されやすい。

【００７６】

自己相互情報量ＭＩは、全文書集合ｄｏｃＡｌｌ中での出現確率と、カテゴリｃａｔ中での出現確率とで偏りの大きい単語を高く評価する。但し、自己相互情報量ＭＩは、低頻度語を過大評価する傾向があるため、利用する場合、ｄｆ（ｔ、ｃａｔ）が極端に小さい単語を特徴語から排除するなどの処理が必要となる。以上の各統計量の詳細については、非特許文献１に記載されている。

【００７７】

カテゴリ特徴語抽出部４２は、ステップＳ１９で算出した各特徴語の特徴度ｓｃｏｒｅ（ｔ，ｃａｔ）を、その特徴語に付加したカテゴリ特徴語情報３３ｃｔとして、カテゴリｃａｔのカテゴリデータ２０ｃに付加したカテゴリ特徴語データ３０ｃｔを特徴語記憶部３０に格納する（Ｓ２０）。

【００７８】

（カテゴリ共通特徴語抽出部４３の動作：図７）

画面提示部５１は、図８に示すように、カテゴリ記憶部２０内のカテゴリＩＤ２１ｃ毎

に、当該カテゴリＩＤ２１ｃに関連付けられた文書ＩＤの個数を含むカテゴリ個数データを提示した画面Ｇ１０を表示する。例えば、画面Ｇ１０内のセルｃ１，ｃ２は、図示しないカテゴリＩＤ２１毎に表示されており、各セルｃ１，ｃ２内の値“７５”，“５０”がカテゴリ個数データに相当している。

【００７９】

相違・共通特徴語提示部５３は、画面提示部５１によるカテゴリ個数データの提示中、ユーザの操作により、複数個のカテゴリ個数データの各カテゴリＩＤの和集合である複数の比較対象ｃｍｐ_ｉからなる比較対象集合ｔｇｔＳｅｔの選択を受け付ける。例えば図９に示す場合、第１の比較対象ｃｍｐ１は、実線枠ｆ１で囲まれた５つのカテゴリ個数データ“６５”，“５０”，“６９”，“７５”，“７２”の各カテゴリＩＤの和集合であり、第２の比較対象ｃｍｐ２は、点線枠ｆ２で囲まれた５つのカテゴリ個数データ“１０”，“２１”，“４５”，“５３”，“３５”の各カテゴリＩＤの和集合である。

【００８０】

カテゴリ共通特徴語抽出部４３は、概略的には、相違・共通特徴語提示部５３により選択を受け付けた比較対象集合ｔｇｔＳｅｔ内の各カテゴリＩＤ２１ｃに関連付けられた所属文書情報２２ｃの各文書ＩＤに関連した文書特徴語毎に、当該各文書ＩＤに関連した全ての文書（ｔｇｔＤｏｃｓ）中で当該文書特徴語が出現する文書数ｄｆ（ｔ、ｔｇｔＤｏｃｓ）を算出する第１文書数算出処理（Ｓ２１～Ｓ２５）と、選択を受け付けた比較対象集合ｔｇｔＳｅｔ内の比較対象ｃｍｐ_ｉ毎に、当該比較対象ｃｍｐ_ｉ内の各カテゴリＩＤ２１ｃに関連付けられた所属文書情報２２ｃの各文書ＩＤに関連した文書特徴語が当該各文書ＩＤの文書中で出現する文書数ｄｆ（ｔ、ｃｍｐ）を算出する第２文書数算出処理（Ｓ２６～Ｓ２９）と、選択を受け付けた比較対象集合ｔｇｔＳｅｔ内の各カテゴリＩＤ２１ｃに関連付けられた所属文書情報２２ｃの各文書ＩＤに関連した文書特徴語毎に、第１文書数算出処理により算出された文書数ｄｆ（ｔ、ｔｇｔＤｏｃｓ）と、第２文書数算出処理により算出された文書数ｄｆ（ｔ、ｃｍｐ）とに基づいて、比較対象集合ｔｇｔＳｅｔにおける当該文書特徴語の共通性を表す度合いの共通特徴度ｃｏｍ（ｔ、ｔｇｔＳｅｔ）を算出し、共通特徴度ｃｏｍ（ｔ、ｔｇｔＳｅｔ）が上位の文書特徴語をカテゴリ共通特徴語として相違・共通特徴語提示部５３に送出する処理を実行する（Ｓ３０～Ｓ３１）。

【００８１】

具体的には、カテゴリ共通特徴語抽出部４３は、ユーザ操作・提示部５０から、ユーザによって選択された各カテゴリ個数データに対応する複数の比較対象ｃｍｐ（各文書ＩＤ）からなる比較対象集合ｔｇｔＳｅｔを取得する（Ｓ２１）。

【００８２】

カテゴリ共通特徴語抽出部４３は、比較対象集合ｔｇｔＳｅｔに含まれる全ての比較対象ｃｍｐの和集合をとり、比較範囲ｔｇｔＤｏｃｓを求める（Ｓ２２）。

【００８３】

カテゴリ共通特徴語抽出部４３は、比較範囲ｔｇｔＤｏｃｓに含まれる全ての文書ＩＤに関連した文書特徴語データを、特徴語記憶部３０から取得する（Ｓ２３）。

【００８４】

カテゴリ共通特徴語抽出部４３は、ステップＳ２３で取得した文書特徴語データに含まれる全ての特徴語ｔについて、ステップＳ２５を繰り返す（Ｓ２４）。

【００８５】

カテゴリ共通特徴語抽出部４３は、ステップＳ２３で取得した文書特徴語データを参照しながら、比較範囲ｔｇｔＤｏｃｓに含まれる文書ＩＤに関連した文書の中で、特徴語ｔが出現する文書数ｄｆ（ｔ、ｔｇｔＤｏｃｓ）を求める（Ｓ２５）。

【００８６】

カテゴリ共通特徴語抽出部４３は、比較対象集合ｔｇｔＳｅｔに含まれる比較対象ｃｍｐ毎に、ステップＳ２７～Ｓ２９の処理を繰り返す（Ｓ２６）。

【００８７】

カテゴリ共通特徴語抽出部４３は、比較対象ｃｍｐの文書ＩＤに関連する文書特徴語データを、特徴語記憶部３０から取得する（Ｓ２７）。

【００８８】

カテゴリ共通特徴語抽出部４３は、ステップＳ２７で取得した文書特徴語データに含まれる全ての特徴語ｔについて、ステップＳ２９の処理を繰り返す（Ｓ２８）。

【００８９】

カテゴリ共通特徴語抽出部４３は、ステップＳ２７で取得した文書特徴語データを参照しながら、比較対象ｃｍｐの文書ＩＤに関連した文書の中で、単語ｔが出現する文書数ｄｆ（ｔ、ｃｍｐ）を求める（Ｓ２９）。

【００９０】

カテゴリ共通特徴語抽出部４３は、ステップＳ２５で算出した比較範囲ｔｇｔＤｏｃｓ内で単語が出現する文書数ｄｆ（ｔ、ｔｇｔＤｏｃｓ）と、ステップＳ２９で算出した各比較対象ｃｍｐ内で単語が出現する文書数ｄｆ（ｔ、ｃｍｐ）に基づき、比較範囲ｔｇｔＤｏｃｓ内の文書に含まれる全ての単語ｔについて、比較対象集合ｔｇｔＳｅｔにおける共通特徴度ｃｏｍ（ｔ、ｔｇｔＳｅｔ）を算出する（Ｓ３０）。

【００９１】

具体的には、共通特徴度ｃｏｍ（ｔ、ｔｇｔＳｅｔ）を算出する場合、始めに、各文書数ｄｆ（ｔ、ｔｇｔＤｏｃｓ），ｄｆ（ｔ、ｃｍｐ）に基づいて、表５に示すように、共通パラメータａ’，ｂ’，ｃ’，ｄ’，ｎ’を算出した後、表６に示す如き、いずれかの統計指標として評価値ｅｖａｌ（ｔ，ｃｍｐ_ｉ）を算出する。

【００９２】

【表５】

［表５］

評価値 eval(t,cmp_i) 算出用の 共通パラメータ	評価値eval(t,cmp_i)算出用の共通パラメータの算出式
a’	a’ = df(t, tgtDocs)
b’	b’ = df(t, tgtDocs) − df(t, cmp_i)
c’	c’ = |cmp| − df(t, cmp_i)
d’	d’ =|tgtDocs|−df(t, tgtDocs)−|cmp_i|+df(t, cmp_i)
n’	n’ = a’ + b’ + c’ + d’ = |tgtDocs|

【００９３】

【表６】

［表６］

統計指標の例	評価値eval(t,cmp_i)の算出式の例
対数尤度比 LLR	LLR' = a' log(a' n' /((a' +b')(a' +c'))) +b' log(b' n' /((a' +b')(b' +d'))) +c' log(c' n' /((c' +d')(a' +c'))) +d' log(d' n' /((c' +d')(b' +d'))) if((a' d' −b' d') < 0) LLR = −LLR' else LLR = LLR' eval(t,cmp_i) = LLR
イエーツ補正 χ２乗値 Yates'	$Yates' = \frac{n'\left(\left\|a'd'-b'c'\right|-\frac{n'}{2}\right)^2}{(a'+b')(c'+d')(a'+c')(b'+d')}$ if((a' d' −b' d') < 0) Yates = −Yates' else Yates = Yates' eval(t,cmp_i) = Yates
自己相互情報量 MI	MI = log(a' n' /((a' +b')(a' +c'))) eval(t,cmp_i) = MI

【００９４】

続いて、比較対象ｃｍｐ_ｉの評価値ｅｖａｌ（ｔ，ｃｍｐ_ｉ）の総和を算出し、得られた総和の値を、比較範囲ｔｇｔＤｏｃｓにおける単語ｔの共通特徴度ｃｏｍ（ｔ，ｔｇｔＳｅｔ）とする。

【００９５】

この指標では、より多くの比較対象ｃｍｐ_ｉに特徴語として含まれ、かつそれぞれの比較対象ｃｍｐ_ｉで、より高い評価値ｅｖａｌ（ｔ，ｃｍｐ_ｉ）を持つ単語ほど、共通特徴語として高く評価される。

【００９６】

ここでは、例えば対数尤度比ＬＬＲという統計指標を用いて、単語の共通特徴語ｃｏｍ（ｔ、ｔｇｔＳｅｔ）を求めている。なお、対数尤度比ＬＬＲに代えて、前述したイエーツχ２乗値や自己相互情報量ＭＩなどの統計指標を用いてもよい。

【００９７】

このような統計指標において、全文書集合ｄｏｃＡｌｌにおける各特徴語ｔの出現頻度ｄｆ（ｔ、ｄｏｃＡｌｌ）や、各カテゴリｃａｔにおける各単語の出現頻度ｄｆ（ｔ、ｃａｔ）も利用してもよい。

【００９８】

しかる後、カテゴリ共通特徴語抽出部４３は、ステップＳ３０で算出した各特徴語の共通特徴度ｃｏｍ（ｔ、ｔｇｔＳｅｔ）について、上位ｒ個の単語をｔｇｔＳｅｔの共通特徴語ｃｏｍＴｅｒｍｓとして、ユーザ操作・提示部５０に送出する（Ｓ３１）。

【００９９】

ここで、ｒとは共通特徴語、相違特徴語及びカテゴリ特徴語の提示において、提示する特徴語の個数の設定値であり、事前に設定されてもよく、特徴語抽出を行う都度設定されてもよい。また、共通特徴度ｃｏｍ（ｔ、ｔｇｔＳｅｔ）が上位ｒ個以内の特徴語を共通特徴語としたが、これに限らず、共通特徴度ｃｏｍ（ｔ、ｔｇｔＳｅｔ）がしきい値以上の特徴語を共通特徴語としてもよい。

【０１００】

相違・共通特徴語提示部５３は、図１０に示すように、ステップＳ３１で送出されたｒ個のカテゴリ共通特徴語をリストＬｃｏｍに提示する。

【０１０１】
（カテゴリ相違特徴語抽出部４４の動作：図１１）
画面提示部５１は、図８に示したように、カテゴリ記憶部２０内のカテゴリＩＤ２１ｃ毎に、当該カテゴリＩＤ２１ｃに関連付けられた文書ＩＤの個数を含むカテゴリ個数データを提示した画面Ｇ１０を表示する。
【０１０２】
相違・共通特徴語提示部５３は、画面提示部５１によるカテゴリ個数データの提示中、ユーザの操作により、複数個のカテゴリ個数データの各カテゴリＩＤの和集合である複数の比較対象ｃｍｐ_ｉからなる比較対象集合ｔｇｔＳｅｔの選択を受け付ける。
【０１０３】
カテゴリ相違特徴語抽出部４４は、概略的には、図９に示したように相違・共通特徴語提示部５３により選択を受け付けた比較対象集合ｔｇｔＳｅｔ内の各カテゴリＩＤ２１ｃに関連付けられた所属文書情報２２ｃの各文書ＩＤに関連した文書特徴語毎に、当該各文書ＩＤに関連した全ての文書（ｔｇｔＤｏｃｓ）中で当該文書特徴語が出現する文書数ｄｆ（ｔ、ｔｇｔＤｏｃｓ）を算出する第１文書数算出処理（Ｓ４１～Ｓ４５）と、選択を受け付けた比較対象集合ｔｇｔＳｅｔ内の比較対象ｃｍｐ_ｉ毎に、当該比較対象ｃｍｐ_ｉ内の各カテゴリＩＤ２１ｃに関連付けられた所属文書情報２２ｃの各文書ＩＤに関連した文書特徴語が当該各文書ＩＤの文書中で出現する文書数ｄｆ（ｔ、ｃｍｐ）を算出する第２文書数算出処理（Ｓ４６～Ｓ４９）と、選択を受け付けた比較対象集合ｔｇｔＳｅｔ内の各カテゴリＩＤ２１ｃに関連付けられた所属文書情報２２ｃの各文書ＩＤに関連した文書特徴語毎に、第１文書数算出処理により算出された文書数ｄｆ（ｔ、ｔｇｔＤｏｃｓ）と、第２文書数算出処理により算出された文書数ｄｆ（ｔ、ｃｍｐ）とに基づいて、各比較対象ｃｍｐ_ｉにおける当該文書特徴語の相違性を表す度合いの相違特徴度ｄｉｆｆ（ｔ、ｃｍｐ）を算出し、相違特徴度ｄｉｆｆ（ｔ、ｃｍｐ）が上位の文書特徴語をカテゴリ相違特徴語として相違・共通特徴語提示部５３に送出する処理を実行する（Ｓ５０～Ｓ５１）。
【０１０４】
具体的には、カテゴリ相違特徴語抽出部４４は、前述したステップＳ２１～Ｓ２９と同様に、ステップＳ４１～Ｓ４９を実行する。なお、ステップＳ４１～Ｓ４９に代えて、ステップＳ２１～Ｓ２９の結果をステップＳ５０で用いるようにカテゴリ相違特徴語抽出部４４を変形してもよい。逆に、ステップＳ２１～Ｓ２９に代えて、ステップＳ４１～Ｓ４９の結果をステップＳ３０で用いるようにカテゴリ共通特徴語抽出部４３を変形してもよい。
【０１０５】
ステップＳ４１～Ｓ４９の実行後、カテゴリ相違特徴語抽出部４４は、ステップＳ２５と同様のステップＳ４５で算出した比較範囲ｔｇｔＤｏｃｓ内で単語が出現する文書数ｄｆ（ｔ、ｔｇｔＤｏｃｓ）と、ステップＳ２９と同様のステップＳ４９で算出した各比較対象ｃｍｐ内で単語が出現する文書数ｄｆ（ｔ、ｃｍｐ）に基づき、比較範囲ｔｇｔＤｏｃｓ内の文書に含まれる全ての特徴語ｔについて、各比較対象ｃｍｐに対する相違特徴度ｄｉｆｆ（ｔ、ｃｍｐ）を算出する（Ｓ５０）。
【０１０６】
具体的には、相違特徴度ｄｉｆｆ（ｔ、ｃｍｐ）としては、各文書数ｄｆ（ｔ、ｔｇｔＤｏｃｓ），ｄｆ（ｔ、ｃｍｐ）に基づいて、表７に示すように、共通パラメータａ”，ｂ”，ｃ”，ｄ”，ｎ”を算出した後、表８に示す如き、いずれかの統計指標として相違特徴度ｄｉｆｆ（ｔ、ｃｍｐ）とする。ここでは、例えば、Ｔ統計量を相違特徴度ｄｉｆｆ（ｔ、ｃｍｐ）とする場合について述べる。
【０１０７】

【表７】

［表７］

相違特徴度diff(t,cmp)算出用の共通パラメータ	相違特徴度diff(t,cmp)算出用の共通パラメータの算出式
a”	a” = df(t, cmp_i)
b”	b” = df(t, cmpDocs_i)
c”	c” = |cmp_i| - df(t, cmp_i)
d”	d” = |cmpDocs_i| − df(t, cmpDocs_i)
n”	n” = a” + b” + c” + d” = |tgtDocs|

【０１０８】

【表８】

［表８］

統計指標の例	相違特徴度diff(t,cmp)の算出式の例
T統計量 tscore	tscore = (a” −((a” + b”)(a” + c”)/n”))/$\sqrt{a”}$ diff(t, cmp) = tscore
対数尤度比LLR	LLR’ = a” log(a” n” /((a” +b”)(a” +c”))) +b” log(b” n” /((a” +b”)(b” +d”))) +c” log(c” n” /((c” +d”)(a” +c”))) +d” log(d” n” /((c” +d”)(b” +d”))) if((a” d” −b” d”) < 0) LLR = −LLR’ else LLR = LLR’ diff(t, cmp) = LLR
イエーツ補正χ２乗値 Yates’	$Yates' = \frac{n”\left(\left\|a”d” - b”c”\right\| - \frac{n”}{2}\right)^2}{(a”+b”)(c”+d”)(a”+c”)(b”+d”)}$ if((a” d” -b” d”) < 0) Yates = −Yates’ else Yates = Yates’ diff(t, cmp) = Yates
自己相互情報量 MI	MI = log(a” n” / ((a” +b”)(a” +c”))) diff(t, cmp) = MI

【０１０９】

相違特徴度ｄｉｆｆ（ｔ、ｃｍｐ）は、Ｔ統計量を利用して、単語ｔについて、比較対象ｃｍｐ_ｉと、比較範囲ｔｇｔＤｏｃｓから比較対象ｃｍｐ_ｉを除いた差集合ｃｍｐＤｏｃｓ_ｉとの間の出現頻度の平均の差に基づき、有意性を求める指標である。これにより、比較対象ｃｍｐ_ｉにおいて、比較対象以外の比較範囲（ｃｍｐＤｏｃｓ_ｉ）に比べ、有意に出現頻度の多い単語を相違特徴語として抽出することができる。なお、Ｔ統計量（Ｔスコアともいう）については、例えば非特許文献３に記載されている。また、Ｔ統計量に代えて、特徴度の算出の説明で述べたような対数尤度比ＬＬＲやχ２乗値、自己相互情報量ＭＩなどの統計指標を使ってもよい。

【０１１０】
このような統計指標において、全文書集合ｄｏｃＡｌｌにおける各特徴語ｔの出現頻度ｄｆ（ｔ、ｄｏｃＡｌｌ）や、各カテゴリｃａｔにおける各特徴語の出現頻度ｄｆ（ｔ、ｃａｔ）も利用してもよい。
【０１１１】
カテゴリ相違特徴語抽出部４４は、比較対象集合ｔｇｔＳｅｔに含まれるそれぞれの対象集合ｃｍｐについて、ステップＳ５０で算出した各特徴語の相違特徴度ｄｉｆｆ（ｔ、ｃｍｐ）が上位ｒ個の特徴語を、相違特徴語ｄｉｆｆＴｅｒｍｓ（ｃｍｐ）として、ユーザ操作・提示部５０に送出する（Ｓ５１）。ここで、上位個数ｒは前述した設定値である。また、相違特徴度ｄｉｆｆＴｅｒｍｓ（ｃｍｐ）が上位ｒ個以内の特徴語を相違特徴語としたが、これに限らず、相違特徴度ｄｉｆｆＴｅｒｍｓ（ｃｍｐ）がしきい値以上の特徴語を相違特徴語としてもよい。
【０１１２】
相違・共通特徴語提示部５３は、図１２に示すように、ステップＳ５１で送出された各ｒ個のカテゴリ相違特徴語をリストＬｄｉｆ１，Ｌｄｉｆ２に提示する。
【０１１３】
（関連カテゴリ提示部５４の動作：図１３）
関連カテゴリ提示部５４は、概略的には、例えば相違・共通特徴語提示部５３によるカテゴリ相違特徴語の提示中、ユーザの操作により、複数のカテゴリ相違特徴語からなる着目語集合ｔｇｔＴｅｒｍｓの選択を受けると、当該選択を受け付けた着目語集合ｔｇｔＴｅｒｍｓと、カテゴリ特徴語記憶部３０内のカテゴリ特徴語との関連度ｒｅｌ（ｃａｔ、ｔｇｔＴｅｒｍｓ）を算出し、関連度ｒｅｌ（ｃａｔ、ｔｇｔＴｅｒｍｓ）の高いカテゴリ特徴語に関連付けられたカテゴリＩＤに関連したカテゴリ個数データを強調表示する（Ｓ６１～Ｓ６６）。
【０１１４】
具体的には、関連カテゴリ提示部５４は、ユーザによって選択された複数の単語から構成される着目語集合ｔｇｔＴｅｒｍｓを取得する（Ｓ６１）。なお、着目語集合に含まれる単語としては、前述したカテゴリ相違特徴語に限らず、カテゴリ特徴語やカテゴリ共通特徴語などが適宜、選択可能となっている。
【０１１５】
関連カテゴリ提示部５４は、特徴語記憶部３０から全てのカテゴリ特徴語３３ｃｔを取得する（Ｓ６２）。
【０１１６】
関連カテゴリ提示部５４は、全てのカテゴリデータｃａｔについて、ステップＳ６４とステップＳ６５の処理を繰り返す（Ｓ６３）。
【０１１７】
関連カテゴリ提示部５４は、カテゴリデータｃａｔのカテゴリ特徴語３３ｃｔに含まれる特徴語を特徴度でソートし、特徴語ランキングｔｅｒｍＲｎｋを求める（Ｓ６４）。
【０１１８】
関連カテゴリ提示部５４は、着目語集合ｔｇｔＴｅｒｍｓと、特徴語ランキングｔｅｒｍＲｎｋに基づいて、カテゴリｃａｔと着目語集合ｔｇｔＴｅｒｍｓとの関連度ｒｅｌ（ｃａｔ、ｔｇｔＴｅｒｍｓ）を求める（Ｓ６５）。
【０１１９】
関連度ｒｅｌ（ｃａｔ、ｔｇｔＴｅｒｍｓ）としては、平均精度と呼ばれる統計指標を利用することができる。この統計指標は、特徴語ランキングｔｅｒｍＲｎｋにおいて、着目語集合ｔｇｔＴｅｒｍｓに含まれる単語が、より上位に多く出現する程、高い値をとる指標である。平均精度の詳細については、非特許文献２に記載されている。関連度ｒｅｌ（ｃａｔ、ｔｇｔＴｅｒｍｓ）としては、平均精度以外にも、カテゴリｃａｔのカテゴリ特徴語において、着目語集合ｔｇｔＴｅｒｍｓに存在する単語ｔの特徴度ｓｃｏｒｅ（ｔ、ｃａｔ）を足し合わせた値としてもよい。

【０１２０】
関連カテゴリ提示部５４は、ステップＳ６５により算出された各カテゴリの関連度ｒｅｌ（ｃａｔ、ｔｇｔＴｅｒｍｓ）に基づき、当該関連度ｒｅｌ（ｃａｔ、ｔｇｔＴｅｒｍｓ）がしきい値ｓ以上のカテゴリ特徴語に関連付けられたカテゴリＩＤを、着目語集合ｔｇｔＴｅｒｍｓの関連カテゴリｒｅｌＣａｔｓとして、関連カテゴリｒｅｌＣａｔｓに含まれるカテゴリＩＤに関連したカテゴリ個数データのセルを強調表示する（Ｓ６６）。
【０１２１】
なお、関連カテゴリとしては、関連度ｒｅｌ（ｃａｔ、ｔｇｔＴｅｒｍｓ）がしきい値以上のカテゴリに限らず、関連度ｒｅｌ（ｃａｔ、ｔｇｔＴｅｒｍｓ）が上位ｔ個以内のカテゴリとしてもよい。しきい値ｓや上位個数ｔは、前述した上位個数ｒと同様に、予め設定されていてもよく、関連カテゴリの提示を行う都度設定されてもよい。
【０１２２】
（ユーザ操作・提示部５０の動作：図１４）
次に、以上のような特徴語抽出部４０や関連カテゴリ提示部５４等の処理をユーザ操作に応じて用いるユーザ操作・提示部５０の動作について説明する。なお、文書特徴語抽出部４１及びカテゴリ特徴語抽出部４２の動作（ステップＳ１～Ｓ４，Ｓ１１～Ｓ２０）は予め完了している状態であるとする。
【０１２３】
ユーザ操作・提示部５０は、概略的には、カテゴリＩＤ毎にカテゴリ個数データをセル表示し、ユーザによる選択操作に応じて、カテゴリ特徴語、カテゴリ共通特徴語及びカテゴリ相違特徴語を提示し、また、関連カテゴリを強調して提示する処理を実行する（Ｓ１００～Ｓ１３１）。
【０１２４】
具体的には、ユーザ操作・提示部５０においては、画面提示部５１が、特徴語記憶部３０に記憶された全てのカテゴリデータについて、それぞれのカテゴリを１つのセルとして表示する（Ｓ１００）。
【０１２５】
この表示例としては、図８の画面Ｇ１０に示すようなクロス表示が挙げられる。この例では、文書データは図２に示すような特許文献とし、カテゴリとしては、特許文献の出願人１４ｄの属性値と、出願日１３ｄの上位４桁である出願年の属性値との２つの属性値で予め分類された文書集合を想定する。ユーザは特許文献から競合他社の技術動向を調査する作業中であるものとする。画面Ｇ１０のクロス表示において、一つのセルが１つのカテゴリに相当する。例えば、セルｃ１は、Ｆ社が２００４年に出願した特許文献を含むカテゴリに相当する。なお、表示形式は、クロス表示に限らず、グラフ表示、マップ表示又はフォルダ表示といった任意の表示形式が使用可能となっている。
【０１２６】
ステップＳ１１０～Ｓ１１２は、ユーザによるカテゴリの選択を受け付け、該カテゴリにおけるカテゴリ特徴語を提示するカテゴリ特徴語提示部５２の処理を示している。
【０１２７】
すなわち、カテゴリ特徴語提示部５２は、ステップＳ１００によりセルとして表示されたカテゴリに対して、ユーザがカテゴリｃａｔを選択した場合、ステップＳ１１１とＳ１１２の処理を行う（Ｓ１１０）。
【０１２８】
カテゴリ特徴語提示部５２は、ユーザが選択したカテゴリｃａｔのカテゴリＩＤに関連するカテゴリ特徴語データを、特徴語記憶部３０から取得する（Ｓ１１１）。
【０１２９】
カテゴリ特徴語提示部５２は、取得したカテゴリ特徴語データに含まれる特徴度に基づき、特徴度ｓｃｏｒｅ（ｔ、ｃａｔ）が上位ｒ個の特徴語をカテゴリ特徴語として、ユーザに提示する（Ｓ１１２）。
【０１３０】

例えば、図１５に示すように、ユーザがセル（カテゴリ）ｃ２をマウスのクリックなどにより選択した場合、該カテゴリに対するカテゴリ特徴語をリストＬ２に表示する。これにより、ユーザは、選択したセルｃ２に含まれる文献の内容の特徴を把握することができる。すなわち、選択したセルｃ２に対応するＦ社の２００５年の出願特許におけるカテゴリ特徴語のリストＬ２に“検索”や“Ｗｅｂ”という技術用語が有意に出現していることにより、ユーザは、Ｆ社の２００５年における注力技術としては、検索やＷｅｂなどがあることを把握できる。ユーザは、他に選択したセルｃ３があれば、同様にリストＬ３から、出願年及び企業名の分析軸におけるカテゴリ特徴語を把握することができる。

【０１３１】

ステップＳ１２０～Ｓ１２６は、ユーザによる比較対象集合ｔｇｔＳｅｔの選択を受け付け、選択された比較対象集合ｔｇｔＳｅｔを特徴語抽出部４０に送り、特徴語抽出部４０によって抽出される共通特徴語と相違特徴語を受け取り、ユーザに提示する相違・共通特徴語提示部５３の処理を示している。

【０１３２】

すなわち、相違・共通特徴語提示部５３は、ステップＳ１００によって表示されたカテゴリに対して、比較対象集合ｔｇｔＳｅｔとして複数の比較対象を選択した場合、ステップＳ１２１～Ｓ１２６の処理を行う（Ｓ１２０）。

【０１３３】

相違・共通特徴語提示部５３は、比較対象集合ｔｇｔＳｅｔを特徴語抽出部４０に送る（Ｓ１２１）。特徴語抽出部４０では、カテゴリ共通特徴語抽出部４３が、前述したステップＳ２１～Ｓ３１の処理を実行し、得られた共通特徴語ｃｏｍＴｅｒｍｓを相違・特徴語提示部５３に送出する。

【０１３４】

相違・共通特徴語提示部５３は、特徴語抽出部４０から共通特徴語ｃｏｍＴｅｒｍｓを受け取り、ユーザに提示する（Ｓ１２２）。

【０１３５】

相違・共通特徴語提示部５３は、比較対象集合ｔｇｔＳｅｔを特徴語抽出部４０に送る（Ｓ１２３）。特徴語抽出部４０では、カテゴリ相違特徴語抽出部４４が、前述したステップＳ４１～Ｓ５１の処理を実行し、得られた相違特徴語ｄｉｆｆＴｅｒｍｓを相違・特徴語提示部５３に送出する。

【０１３６】

相違・共通特徴語提示部５３は、特徴語抽出部４０から相違特徴語ｄｉｆｆＴｅｒｍｓを取得する（Ｓ１２４）。

【０１３７】

相違・共通特徴語提示部５３は、比較対象集合ｔｇｔＳｅｔに含まれる全ての比較対象ｃｍｐについて、ステップＳ１２６の処理を繰り返す（Ｓ１２５）。

【０１３８】

相違・共通特徴語提示部５３は、比較対象ｃｍｐに対する相違特徴語ｄｉｆｆＴｅｒｍｓ（ｃｍｐ）をユーザに提示する（Ｓ１２６）。

【０１３９】

ステップＳ１２０～Ｓ１２６における表示例は、図１２に示す通りである。ユーザは、例えば、ユーザがＡ社とＢ社の技術動向を比較したい場合、画面Ｇ１０においてＡ社に関するカテゴリを示す複数のセルを実線枠ｆ１のように選択することにより１つの比較対象を選択する。

【０１４０】

また、もう一つの比較対象として、Ｂ社に関するカテゴリを示す複数のセルを点線枠ｆ２のように選択した場合、実線枠ｆ１と点線枠ｆ２で示される２つの比較対象から構成される比較範囲における共通特徴語リストＬｃｏｍを表示する。

【０１４１】

このように、両社の出願特許における共通特徴語リストＬｃｏｍに有意に出現する技術

用語として、”分類”や”クラスタリング”が提示され、これらの技術がＡ社とＢ社で共通する技術分野であることを把握できる。

【０１４２】

また、実線枠ｆ１で示される比較対象に対する相違特徴語リストＬｄｉｆ１を表示し、点線枠ｆ２で示される比較対象に対する相違特徴語リストＬｄｉｆ２を表示する。このような相違特徴語リストＬｄｉｆ１，Ｌｄｉｆ２の表示により、Ａ社とＢ社の独自性を表す技術を把握することができる。

【０１４３】

また、複数の比較対象に対し、共通特徴語と相違特徴語を表示することにより、単に文書集合に対する特徴語を提示するよりも、比較対象間の特徴を、より明確にユーザに示すことができる。

【０１４４】

ステップＳ１３０、Ｓ１３１は、ステップＳ１１２や、Ｓ１２２，Ｓ１２６による各特徴語の提示中に、これら各特徴語から選択された着目語からなる着目語集合ｔｇｔＴｅｒｍｓを受け付け、その着目語集合と各カテゴリとの関連度を算出し、関連度の高いカテゴリを関連カテゴリとしてユーザに提示する関連カテゴリ提示部５４の処理を示している。

【０１４５】

関連カテゴリ提示部５４は、カテゴリ特徴語提示部５２や、相違・共通特徴語提示部５３によって提示された、カテゴリ特徴語または共通特徴語または相違特徴語から、ユーザが着目語を選択した場合、ステップＳ１３１の処理を行う（Ｓ１３０）。ここで、ユーザは複数の単語を着目語として選択できるものとし、選択された複数の着目語を着目語集合ｔｇｔＴｅｒｍｓとする。また、本実施形態では、提示された特徴語から着目語を選択する場合について説明したが、これに限らず、Ｗｅｂの検索のようにユーザが任意のキーワードを着目語として入力してもよい。

【０１４６】

関連カテゴリ提示部５４は、前述したステップＳ６１～Ｓ６６の処理を実行することにより、着目語集合ｔｇｔＴｅｒｍｓと各カテゴリとの関連度を算出して関連度の高いカテゴリを関連カテゴリとしてユーザに提示する（Ｓ１３１）。

【０１４７】

例えば、図１６に示すように、ユーザは、共通特徴語のリストＬｃｏｍや相違特徴語のリストＬｄｉｆ１，Ｌｄｉｆ２の中から、着目したい単語Ｔｃｏｍ１とＴｄｉｆ２を選択する。ここでは、ユーザは“分類”と“ＸＭＬ”との単語が気になった場合、単語“分類”を示す共通語Ｔｃｏｍ１と、単語“ＸＭＬ”を示す相違特徴語Ｔｄｉｆ２とを着目語として選択する。関連カテゴリ提示部５４は、ユーザの着目語の選択を受けて、着目語との関連度が高い関連カテゴリのセルｃ４を、背景色を変更する等して、強調表示する。

【０１４８】

これによって、ユーザは着目する技術について、調査すべき範囲の糸口をつかむことができる。図１６に示した例では、ユーザが着目した”分類”と”ＸＭＬ”について、企業の観点から見ると、Ｃ社もこれらの技術に関連していることがわかる。さらに、出願年の観点からみると２００６～２００８年の間で、これらの技術に関連する特許が有意に出現していることがわかる。これによって、ユーザは着目している技術について、詳細に調査すべき範囲を明確化でき、効率的に先行技術調査を行うことができる。

【０１４９】

また、着目語の選択は、共通特徴語や相違特徴語だけでなく、カテゴリ特徴語からも選択することができる。例えば図１７に示すように、着目語の選択に加え、セルｃ２におけるカテゴリ特徴語リストＬ２内のカテゴリ特徴語からも着目語を選択した場合には、この選択に応じて、関連カテゴリの表示が変化する。

【０１５０】

画面提示部５１は、ユーザがシステムの終了を選択した場合、処理を終了し、それ以外はステップＳ１１０に処理を戻す（Ｓ１４０）。

【０１５１】
例えば、ステップＳ１１０に処理を戻し、調査を継続する場合の例について説明する。図１８は比較対象の絞り込みと共通特徴語及び相違特徴語の表示例を表す図である。ユーザは、共通特徴語や相違特徴語、カテゴリ特徴語の提示や、着目語指定に対する関連カテゴリの提示を受けて、比較対象の縮小（絞り込み）や拡大といった変更を行うことができる。

【０１５２】
例えば、図１６に示す如き、特徴語や関連語カテゴリの提示中に、ユーザは、図１８に示すように、比較範囲の各枠ｆ１，ｆ２を出願年について２００６～２００８年に絞り込み、新たな点線枠ｆ３によりＣ社を比較企業に選択する。これら各枠ｆ１～ｆ３に基づく比較対象集合ｔｇｔＳｅｔに基づいて、特徴語抽出装置は、提示する共通特徴語や相違特徴語を変化させる。これにより、ユーザはそれまで思いつかなかったが着目すべき技術用語を発見する手がかりとなる。

【０１５３】
図１９は、ユーザによる着目語の変更（追加／削除）と、関連カテゴリの表示例を表す図である。ユーザは、共通特徴語や相違特徴語やカテゴリ特徴語を見ながら、着目語を追加したり、削除したりすることができる。

【０１５４】
例えば、図１８による特徴語の変化や、Ｃ社の相違特徴語を受けて、新たに単語“マイニング”を示す相違特徴語Ｔｄｉｆ３を着目語に追加する。これを受けて、特徴語抽出装置は、提示する関連カテゴリを変化させる。

【０１５５】
これにより、ユーザは、着目語を切り替えながら関連するカテゴリを概観することで、それまで気づいていなかった着目語とカテゴリの関連を発見することができる。先行技術調査であれば、着目している技術を扱っている意外な企業や、ある企業はユーザが認識しているよりも早い年代から着目している技術に関する特許を出願しているといったことを発見する手がかりとなる。

【０１５６】
また、図１５～図１９を用いて述べたように、特徴語抽出装置による「特徴語の提示」と「関連カテゴリ提示」、ユーザによる「比較対象の指定」と「着目語の選択」、というプロセスを繰り返すことで、分析する対象や特徴語を明確化するとともに、それまでユーザが意識していなかったキーワードや、分析対象を発見することができる。特許調査においては、新たに着目すべき技術や、注意すべき競合他社を発見する糸口となる。また、比較対象や着目語を利用することで、適切な比較範囲に対する、適切な単語による、先行技術調査を実現することができる。

【０１５７】
例えば図２０に示す如き、「時系列×企業」の分析軸によるクロス分析の画面Ｇ１０は、図２１及び図２２に示すように、ある企業に対する「時系列×特徴語」の分析軸によるクロス分析の画面Ｇ２０、ある特徴語に対する「時系列×企業」の分析軸によるクロス分析の画面Ｇ３０などのように、任意の分析軸の画面に適用して適切な比較範囲と適切な単語による分析・調査を実現することができる。

【０１５８】
また例えば、ある特徴語に対する「時系列×企業」のクロス分析の画面Ｇ３０は、図２３に示す如き、ある特徴語に対する「時系列×企業」のグラフ表示の画面Ｇ３１に表示形式を変更することができる。なお、表示形式を変更できることは、他のクロス分析の画面Ｇ１０，Ｇ２０でも同様である。

【０１５９】
上述したように本実施形態によれば、カテゴリＩＤに関連付けられた文書ＩＤの個数を含むカテゴリ個数データの提示中に、選択を受け付けたカテゴリ個数データのカテゴリＩＤに関連したカテゴリ特徴語のうち、特徴度が上位のカテゴリ特徴語における文書特徴語

をカテゴリ特徴語として提示する。

【０１６０】

また、カテゴリ個数データの提示中、複数個のカテゴリ個数データの各カテゴリＩＤの和集合である複数の比較対象からなる比較対象集合の選択を受け付けると、選択を受け付けた比較対象集合内の各カテゴリＩＤに関連付けられた各文書ＩＤに関連した文書特徴語毎に、各比較対象における当該文書特徴語の相違性を表す度合いの相違特徴度を算出し、相違特徴度が上位の文書特徴語をカテゴリ相違特徴語として提示する。

【０１６１】

このように、選択したカテゴリ個数データに応じて、カテゴリ特徴語やカテゴリ相違特徴語といった特徴語を提示する構成により、事前に定義された属性に限らずに分析軸の候補として特徴語を提示できると共に、着目すべき比較範囲の絞り込みや、着目すべき特徴語の参照を支援できる。

【０１６２】

補足すると、カテゴリ特徴語を提示することにより、ユーザが各カテゴリに対するカテゴリ特徴語を確認して、文書集合の全体像や、個々のカテゴリの内容を効率よく把握できる。

【０１６３】

また、複数の比較対象間の相違特徴語を提示する構成により、ユーザは着目している任意の範囲における比較対象の相違点を把握することができる。さらに、比較対象集合を絞り込めば、各比較対象における相違点をより詳細に把握できる。一方、比較対象集合を拡大すれば、マクロな視点で相違点を把握でき、全体的な内容の理解を深めることができる。以上により、ユーザは各特徴語を参照しながら、文書集合について内容の理解を進め、分析すべき範囲や、着目すべきキーワードを明確化することができる。

【０１６４】

また、本実施形態によれば、カテゴリ個数データの提示中、複数個のカテゴリ個数データの各カテゴリＩＤの和集合である複数の比較対象からなる比較対象集合の選択を受け付けると、選択を受け付けた比較対象集合内の各カテゴリＩＤに関連付けられた各文書ＩＤに関連した文書特徴語毎に、比較対象集合における当該文書特徴語の共通性を表す度合いの共通特徴度を算出し、共通特徴度が上位の文書特徴語をカテゴリ共通特徴語として提示する構成により、ユーザは自身が任意の着目している範囲における文書の共通点を把握でき、文書集合に対する理解がさらに深められ、分析すべき範囲やキーワードをより明確に捉えることができる。

【０１６５】

さらに、本実施形態によれば、例えば、カテゴリ相違特徴語の提示中、複数のカテゴリ相違特徴語からなる着目語集合の選択を受け付けると、選択を受け付けた着目語集合と、カテゴリ特徴語記憶部３０内のカテゴリ特徴語との関連度を算出し、関連度の高いカテゴリ特徴語に関連付けられたカテゴリＩＤに関連したカテゴリ個数データを強調表示する構成により、ユーザは、着目語の選択に対して提示されるカテゴリを概観することで、自身が着目しているキーワードに関連しているカテゴリを把握でき、それまで気づいていなかった分析対象を発見できる。従って、ユーザは、分析したい事項について、適切な分析対象を把握でき、より精度の高い分析が可能となる。

【０１６６】

このように、ユーザは、相違特徴語・共通特徴語と関連カテゴリを確認しながら、比較対象の選択（絞り込みや拡大）と着目語の選択を繰り返すことで、分析する範囲や着目する特徴語を明確化することができる。これによって、複数の文書集合に対して、ユーザは漏れなく、無駄なく、目的にあった、内容把握や比較調査を効率的に行うことができる。

【０１６７】

なお、上記実施形態に記載した手法は、コンピュータに実行させることのできるプログラムとして、磁気ディスク（フロッピー（登録商標）ディスク、ハードディスクなど）、光ディスク（ＣＤ－ＲＯＭ、ＤＶＤなど）、光磁気ディスク（ＭＯ）、半導体メモリなど

の記憶媒体に格納して頒布することもできる。

【０１６８】

また、この記憶媒体としては、プログラムを記憶でき、かつコンピュータが読み取り可能な記憶媒体であれば、その記憶形式は何れの形態であってもよい。

【０１６９】

また、記憶媒体からコンピュータにインストールされたプログラムの指示に基づきコンピュータ上で稼働しているＯＳ（オペレーティングシステム）や、データベース管理ソフト、ネットワークソフト等のＭＷ（ミドルウェア）等が上記実施形態を実現するための各処理の一部を実行してもよい。

【０１７０】

さらに、本発明における記憶媒体は、コンピュータと独立した媒体に限らず、ＬＡＮやインターネット等により伝送されたプログラムをダウンロードして記憶または一時記憶した記憶媒体も含まれる。

【０１７１】

また、記憶媒体は１つに限らず、複数の媒体から上記実施形態における処理が実行される場合も本発明における記憶媒体に含まれ、媒体構成は何れの構成であってもよい。

【０１７２】

尚、本発明におけるコンピュータは、記憶媒体に記憶されたプログラムに基づき、上記実施形態における各処理を実行するものであって、パソコン等の１つからなる装置、複数の装置がネットワーク接続されたシステム等の何れの構成であってもよい。

【０１７３】

また、本発明におけるコンピュータとは、パソコンに限らず、情報処理機器に含まれる演算処理装置、マイコン等も含み、プログラムによって本発明の機能を実現することが可能な機器、装置を総称している。

【０１７４】

なお、本願発明は、上記実施形態そのままに限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で構成要素を変形して具体化できる。また、上記実施形態に開示されている複数の構成要素の適宜な組合せにより種々の発明を形成できる。例えば、実施形態に示される全構成要素から幾つかの構成要素を削除してもよい。更に、異なる実施形態に亘る構成要素を適宜組合せてもよい。

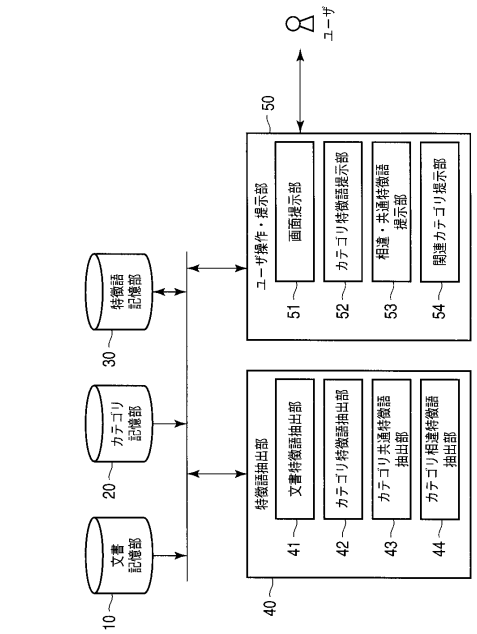
【符号の説明】

【０１７５】

１０…文書記憶部、２０…カテゴリ記憶部、３０…特徴語記憶部、４０…特徴語抽出部、４１…文書特徴語抽出部、４２…カテゴリ特徴語抽出部、４３…カテゴリ共通特徴語抽出部、４４…カテゴリ相違特徴語抽出部、５０…ユーザ操作・提示部、５１…画面提示部、５２…カテゴリ特徴語提示部、５３…相違・共通特徴語提示部、５４…関連カテゴリ提示部。

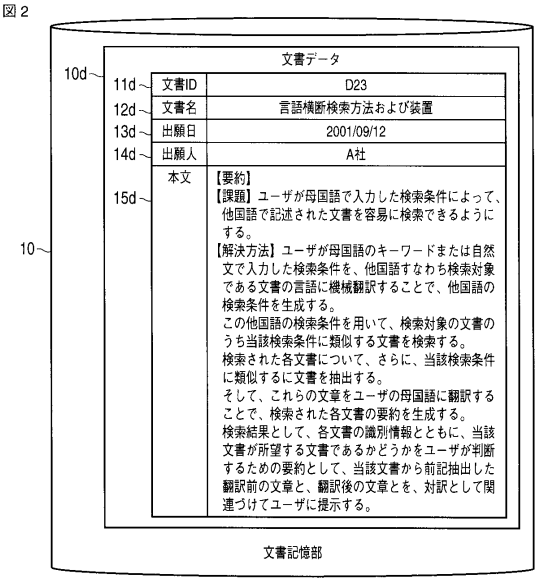
【図１】

図１



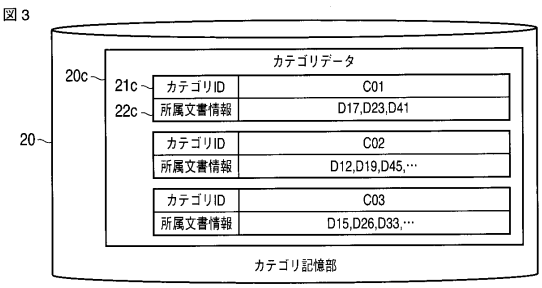
【図２】

図２



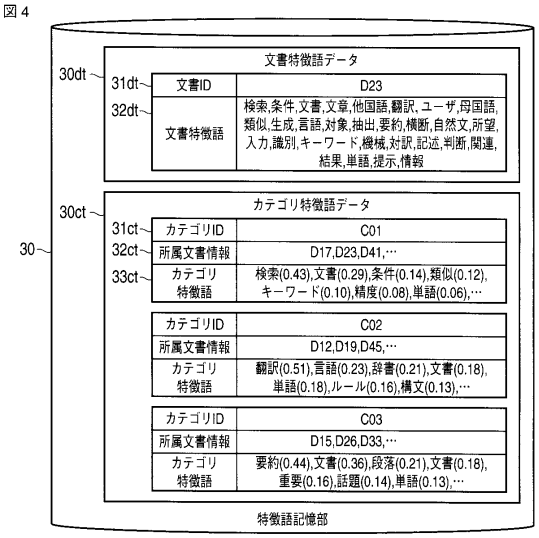
【図３】

図３



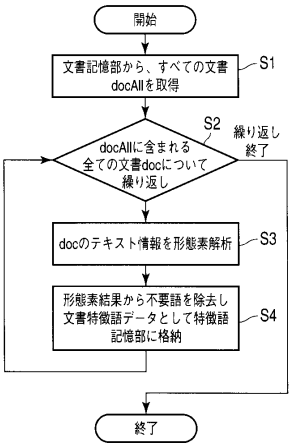
【図４】

図４



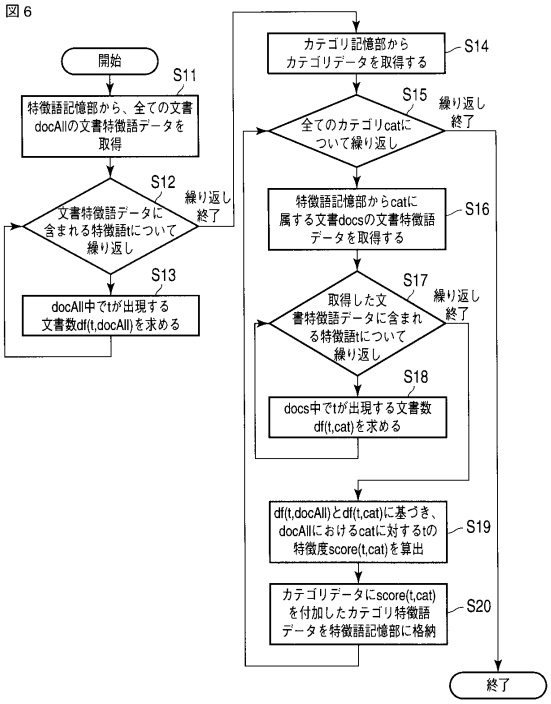
【図５】

図５



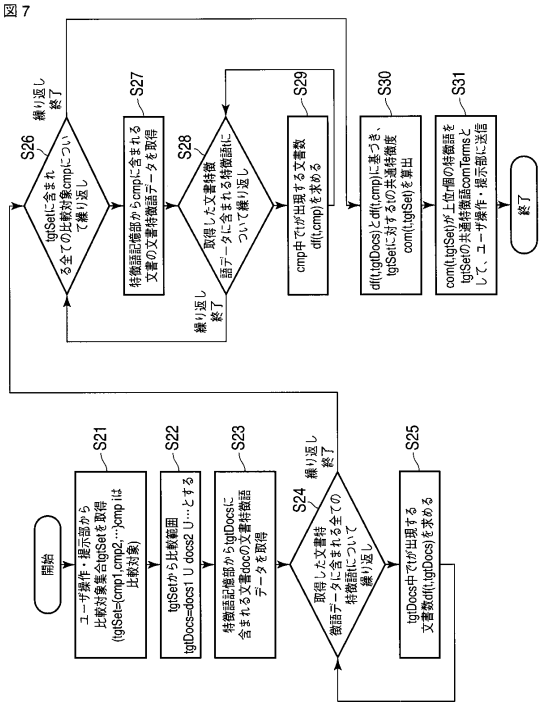
【図６】

図６



【図７】

図７



【図８】

図８

G10

業界		出願年 2004/	2005/	2006/	2007/	2008/
	A社	65	50	69	75	72
	B社	10	21	45	53	35
	C社	25	10	24	36	20
	D社	16	6	12	5	10
	E社	52	50	32	54	43
	F社	75	63	80	69	78

c1　c2

【図９】

図９

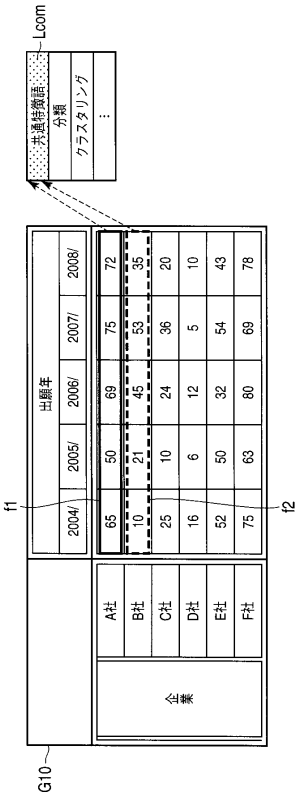
G10　f1

業界		出願年 2004/	2005/	2006/	2007/	2008/
	A社	65	50	69	75	72
	B社	10	21	45	53	35
	C社	25	10	24	36	20
	D社	16	6	12	5	10
	E社	52	50	32	54	43
	F社	75	63	80	69	78

f2

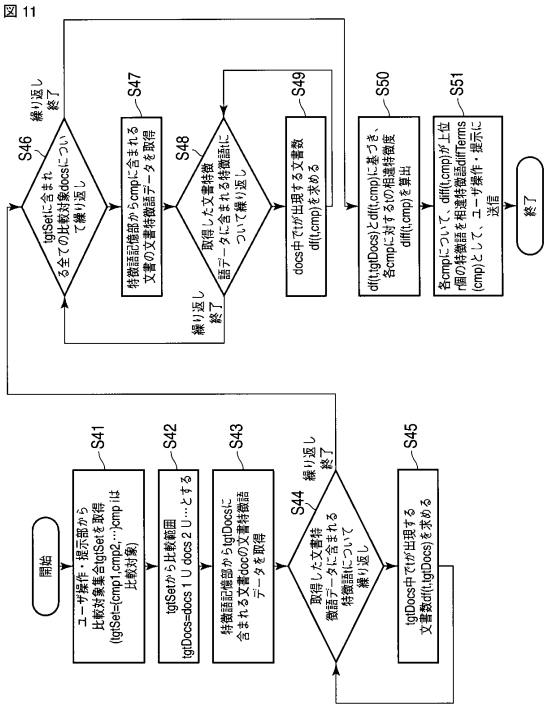
【図１０】

図10



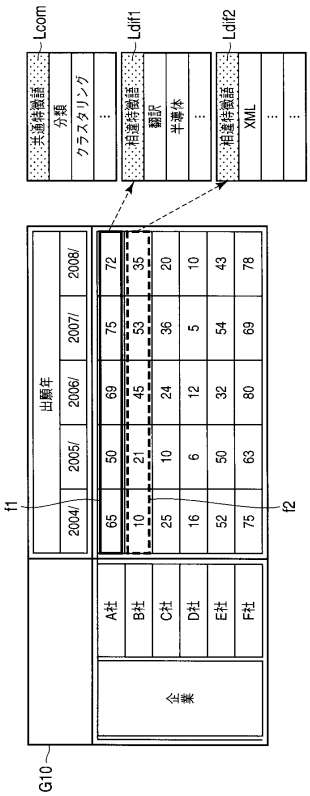
【図１１】

図11



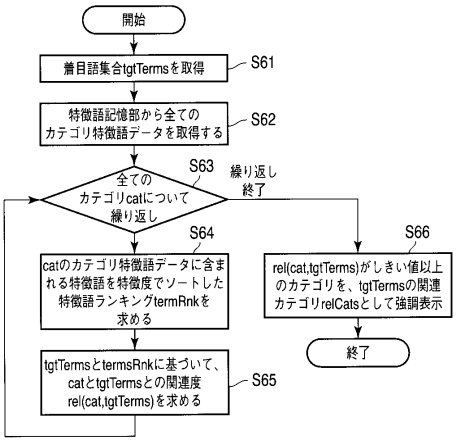
【図１２】

図12



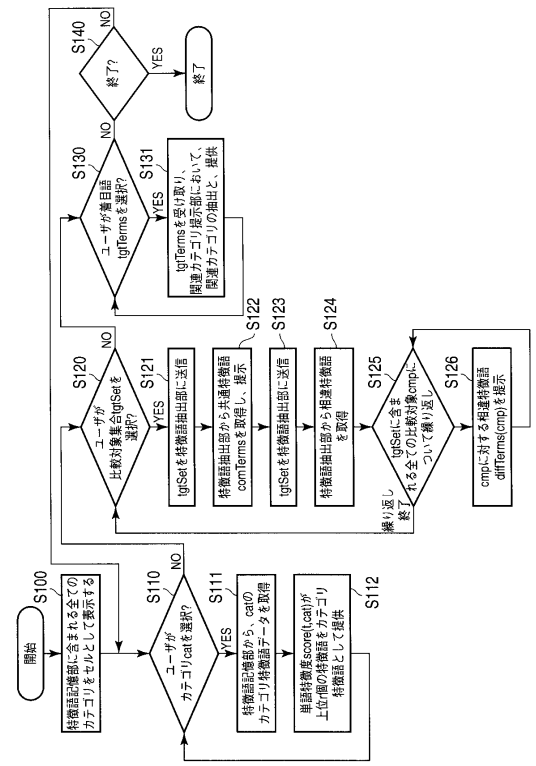
【図１３】

図13



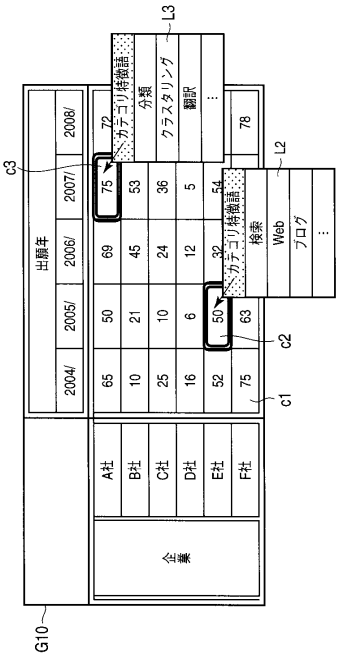
【図１４】

図 14



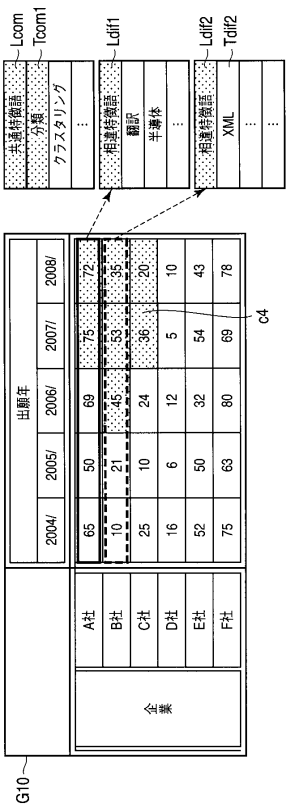
【図１５】

図 15



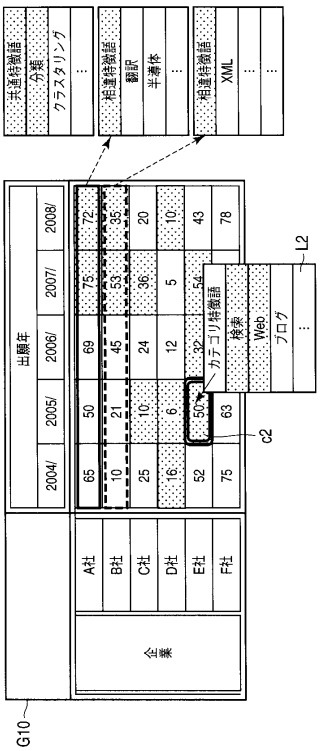
【図１６】

図 16



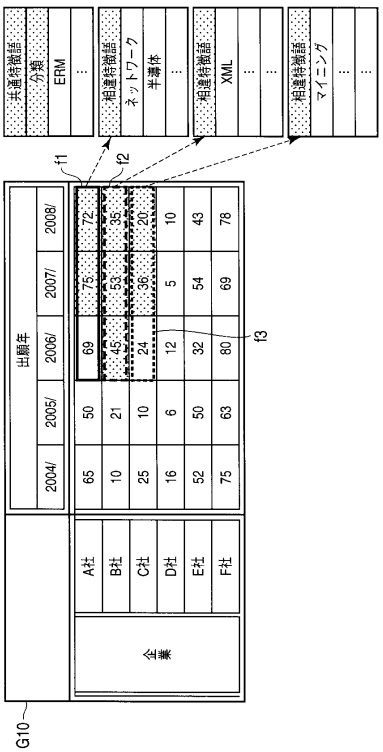
【図１７】

図 17



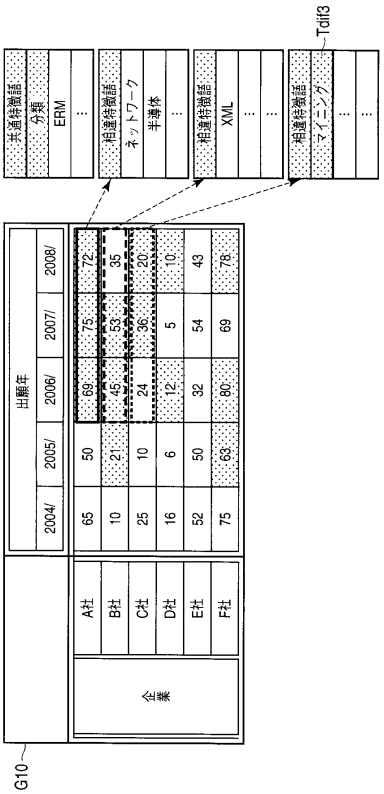
【図１８】

図18



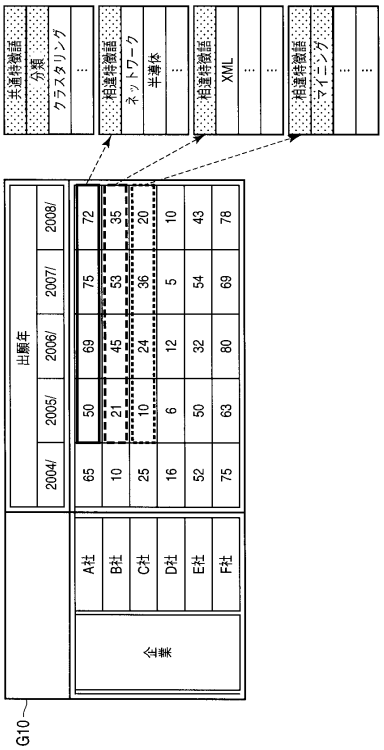
【図１９】

図19



【図２０】

図20



【図２１】

図21

G20

A社		出願年			
		2005	2006	2007	2008
特徴語	分類	10	25	50	31
	クラスタリング	21	45	53	35
	XML	10	24	26	20
	マイニング	123	101	201	93

【図２２】

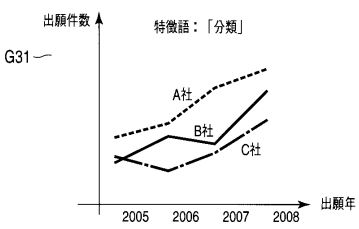
図22

G30

「分類」		出願年			
		2005	2006	2007	2008
企業	A社	10	13	21	26
	B社	4	10	9	23
	C社	5	4	7	12

【図２３】

図23



フロントページの続き

(74)代理人 100075672
弁理士 峰 隆司
(74)代理人 100095441
弁理士 白根 俊郎
(74)代理人 100084618
弁理士 村松 貞男
(74)代理人 100103034
弁理士 野河 信久
(74)代理人 100119976
弁理士 幸長 保次郎
(74)代理人 100153051
弁理士 河野 直樹
(74)代理人 100140176
弁理士 砂川 克
(74)代理人 100101812
弁理士 勝村 紘
(74)代理人 100124394
弁理士 佐藤 立志
(74)代理人 100112807
弁理士 岡田 貴志
(74)代理人 100111073
弁理士 堀内 美保子
(74)代理人 100134290
弁理士 竹内 将訓
(74)代理人 100127144
弁理士 市原 卓三
(74)代理人 100141933
弁理士 山下 元
(72)発明者 岩崎 秀樹
東京都港区芝浦一丁目１番１号 東芝ソリューション株式会社内
(72)発明者 後藤 和之
東京都港区芝浦一丁目１番１号 東芝ソリューション株式会社内
(72)発明者 松本 茂
東京都港区芝浦一丁目１番１号 東芝ソリューション株式会社内
(72)発明者 平 博司
東京都港区芝浦一丁目１番１号 東芝ソリューション株式会社内
(72)発明者 宮部 泰成
東京都港区芝浦一丁目１番１号 東芝ソリューション株式会社内

審査官 齊藤 貴孝

(56)参考文献 特開２００１－０９２８２５（ＪＰ，Ａ）
特開２００３－３４５８１０（ＪＰ，Ａ）
国際公開第２００６／１１５２６０（WO，Ａ１）
特開２００９－２８８９９９（ＪＰ，Ａ）
特開２００９－２９４９３８（ＪＰ，Ａ）
特開２００９－２９４９３９（ＪＰ，Ａ）
特表２００８－５２４７１２（ＪＰ，Ａ）

特開２００６－２１５６７５（ＪＰ，Ａ）
特開２００３－３４５８１１（ＪＰ，Ａ）
国際公開第２００７／０６９６６３（WO，Ａ１）
特開２００７－００４２３３（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１７／３０
Ｇ０６Ｑ　１０／００