(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2012-138027
(P2012-138027A)
(43)公開日　平成24年7月19日(2012.7.19)

(51)Int.Cl.		F I			テーマコード（参考）
G06F 17/30	(2006.01)	G06F	17/30	340Z	5B075
		G06F	17/30	170A	

審査請求 有　請求項の数 9　ＯＬ　（全 20 頁）

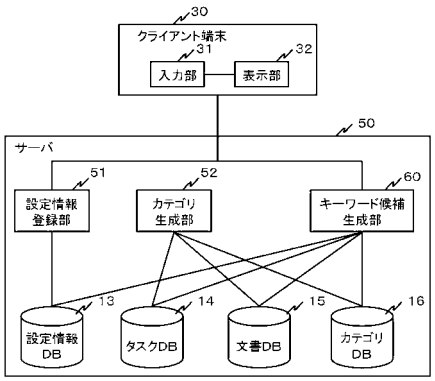
(21)出願番号　特願2010-291331 (P2010-291331)
(22)出願日　平成22年12月27日 (2010.12.27)

(71)出願人　000003078
株式会社東芝
東京都港区芝浦一丁目１番１号
(71)出願人　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　100149803
弁理士　藤原　康高
(72)発明者　藤田　慎一
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　高知尾　勝彦
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

最終頁に続く

(54)【発明の名称】情報検索システム、検索キーワード提示方法、およびプログラム

(57)【要約】　（修正有）
【課題】単語間の関連を保持する辞書や、ユーザが入力した検索キーワードと参照文書を記録するログ情報などを用いず、所望する情報を検索可能な検索キーワードの候補をユーザに提示できる情報検索システムを提供する。
【解決手段】情報検索システムは、検索対象の文書データを格納する文書データベースと検索対象の文書データに対するクラスタリング分析によるカテゴリ分類に関する情報を含むカテゴリ情報を格納するカテゴリデータベースとを備える。また、情報検索システムは、ユーザによって入力される検索キーワードを含む検索条件に基づいて文書データベースから文書データを検索し、検索した文書データがカテゴリのうちのどのカテゴリに属するかを判定し、検索した文書データが属すると判定されたカテゴリ、もしくは当該カテゴリの文書データから、キーワード候補を抽出し、抽出したキーワード候補をユーザに提示する。
【選択図】図２



【特許請求の範囲】
【請求項１】
検索対象の文書データを格納する文書データベースと、
カテゴリに前記検索対象の文書データのうちのどの文書データが属するかを管理するカテゴリ情報を格納するカテゴリデータベースと、
ユーザによって入力される検索キーワードを含む検索条件に基づいて前記文書データベースから文書データを検索する文書検索部と、
前記カテゴリデータベースを参照して、前記文書検索部によって検索された文書データが前記カテゴリのうちのどのカテゴリに属するかを判定するカテゴリ判定部と、
前記カテゴリ判定部によって前記検索された文書データが属すると判定されたカテゴリ、もしくは、当該カテゴリに属する文書データから、キーワード候補を抽出するキーワード候補抽出部と、
前記キーワード候補抽出部によって抽出された前記キーワード候補をユーザに提示する出力部と、
を備える情報検索システム。
【請求項２】
前記キーワード候補抽出部は、前記カテゴリ判定部によって前記検索された文書データが属すると判定されたカテゴリの文書データに含まれる単語のうち、前記検索キーワード以外の単語の出現頻度を計算し、前記出現頻度があらかじめ設定された所定数以上である前記単語を前記キーワード候補として抽出する請求項１に記載の情報検索システム。
【請求項３】
前記検索条件は前記文書データが前記文書データベースに登録された日時の範囲を指定するための検索期間情報を含み、
前記キーワード候補抽出部は、前記カテゴリ判定部によって前記検索された文書データが属すると判定されたカテゴリの文書データから、前記検索期間情報に基づいて文書データの抽出を行い、当該抽出された文書データにおいて、検索キーワード以外の単語における出現頻度を計算し、前記出現頻度があらかじめ設定された所定数以上である前記単語を前記キーワード候補として抽出する請求項１乃至請求項２のいずれか１項に記載の情報検索システム。
【請求項４】
前記キーワード候補抽出部は、前記カテゴリ判定部によって前記検索された文書データが属すると判定されたカテゴリの名称を前記キーワード候補として抽出する請求項１乃至請求項３のいずれか１項に記載の情報検索システム。
【請求項５】
前記キーワード候補抽出部は、前記カテゴリ判定部によって前記検索された文書データが属すると判定されたカテゴリに含まれる文書データに対して前記クラスタリング分析を行い、
当該クラスタリング分析の結果、新たに作成されたカテゴリが、前記検索された文書データを含む場合に、当該作成されたカテゴリの名称を前記キーワード候補として抽出する請求項１乃至請求項４のいずれか１項に記載の情報検索システム。
【請求項６】
前記キーワード候補抽出部は、前記検索された文書データが属する前記カテゴリに含まれる文書データから、前記検索された文書データと関連付けられた文書データを抽出し、当該抽出された文書データから、キーワード候補を抽出する請求項１乃至請求項４のいずれか１項に記載の情報検索システム。
【請求項７】
前記文書検索部は、あらかじめ定められた周期が到来した場合に、前記文書データを検索する請求項１乃至請求項６のいずれか１項に記載の情報検索システム。
【請求項８】
ユーザによって入力される検索キーワードを含む検索条件に基づいて文書データベース

から文書データを検索するステップと、
　カテゴリデータベースを参照して、検索された前記文書データがどのカテゴリに属するかを判定するステップと、
　検索された前記文書データが属すると判定されたカテゴリ、もしくは、当該カテゴリに属する文書データから、キーワード候補を抽出するステップと、
　抽出された前記キーワード候補をユーザに提示するステップと、
を備える検索キーワード提示方法。
【請求項９】
　検索対象の文書データを格納する文書データベースと、前記検索対象の文書データに対するクラスタリング分析によるカテゴリ分類に関する情報を含むカテゴリ情報を格納するカテゴリデータベースとを備える情報検索システムのプログラムであって、
　コンピュータに、
　ユーザによって入力される検索キーワードを含む検索条件に基づいて文書データベースから文書データを検索する機能と
　前記検索された文書データが前記カテゴリのうちのどのカテゴリに属するかを判定する機能と、
　前記検索された文書データが属すると判定されたカテゴリ、もしくは、当該カテゴリの文書データから、キーワード候補を抽出する機能と、
抽出した前記キーワード候補をユーザに提示する機能と、を実現させるプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
　本発明の実施形態は情報検索システム、検索キーワード提示方法、およびプログラムに関する。
【背景技術】
【０００２】
従来、文書が大量に蓄積されているデータベースよりユーザが所望している情報を見つける検索装置がある。
【０００３】
この検索装置において、検索に使用するキーワードはユーザが指定するものであるため、キーワードを指定するユーザ自身が知らない単語は検索キーワードとして使用することができない。これにより、ユーザは必要とする情報を取得できない場合がある。もしくは、ユーザが検索で使用するキーワードが廃れた場合にも、ユーザは必要とする情報を取得できない場合がある。
【０００４】
これらの問題に対し、単語と単語に関連する関連単語とを対応付ける関連辞書などを用いて検索を実現する装置がある。また、複数のカテゴリに分類した文書情報に対して検索を行い、参照要求した文書の識別情報をログ情報から取得することで、ユーザが入力した検索キーワードとカテゴリの適合度から検索キーワードを提示することを可能とする装置もある。
【０００５】
しかしながら、上述したような従来技術では、検索キーワードと関連付いている単語を探すような場合に、単語間の関連性に関する情報を保持する関連辞書が必要である。そのため、辞書を作成するための労力や費用の発生、新たに出てくる単語を辞書に適用するといったメンテナンスが発生するというような問題がある。
【０００６】
また、カテゴリに分類された文書の中からユーザが検索で用いたキーワードを元にどの文書を参照したかを知るにはログ情報とログ情報の解析が必要である。したがって、このログ情報を取得する手段と、取得したログ情報を解析する手段が必要になるという問題がある。

【先行技術文献】
【特許文献】
【０００７】
【特許文献１】特開２００８－７０９９１号公報
【発明の概要】
【発明が解決しようとする課題】
【０００８】
本発明が解決しようとする課題は、単語間の関連を保持する辞書、又はユーザが入力した検索キーワードと参照した文書を記録するログ情報を用いずに、ユーザが所望する情報を検索可能な検索キーワードの候補をユーザに提示することを可能とする情報検索システム、検索キーワード提示方法、およびプログラムを提供することである。
【課題を解決するための手段】
【０００９】
実施形態の情報検索システムは、検索対象の文書データを格納する文書データベースと、検索対象の文書データに対するクラスタリング分析によるカテゴリ分類に関する情報を含むカテゴリ情報を格納するカテゴリデータベースとを備える。実施形態の情報検索システムは、ユーザによって入力される検索キーワードを含む検索条件に基づいて文書データベースから文書データを検索し、検索された文書データがカテゴリのうちのどのカテゴリに属するかを判定し、検索された文書データが属すると判定されたカテゴリ、もしくは、当該カテゴリの文書データから、キーワード候補を抽出し、抽出したキーワード候補をユーザに提示する。
【図面の簡単な説明】
【００１０】
【図１】実施形態の情報検索システムの全体構成を示す図。
【図２】実施形態の情報検索システムの機能構成を示す図。
【図３】実施形態の情報検索システムの検索条件設定画面の一例を示す図。
【図４】実施形態の情報検索システムの処理手順の一例を示すフローチャート。
【図５】実施形態の情報検索システムのキーワード候補生成部の機能ブロック図。
【図６】実施形態の情報検索システムのキーワード候補生成部によるキーワード候補生成処理の一例を示すフローチャート。
【図７】実施形態の情報検索システムのカテゴリ解析結果の概念の一例を示す図。
【図８】実施形態の情報検索システムのカテゴリ名取得部によるキーワード候補抽出処理の一例を示すフローチャート。
【図９】実施形態の情報検索システムの新カテゴリ名取得部によるキーワード候補抽出処理の一例を示すフローチャート。
【図１０】実施形態の情報検索システムの新カテゴリ名取得部によるキーワード候補抽出処理の概念図。
【図１１】実施形態の情報検索システムの出現頻度解析部によるキーワード候補抽出処理の一例を示すフローチャート。
【図１２】実施形態の情報検索システムの出現頻度解析部によるＨＴＭＬカテゴリからのキーワード候補抽出処理の概念図。
【図１３】実施形態の情報検索システムの期間限定出現頻度解析部によるキーワード候補抽出処理の一例を示すフローチャート。
【図１４】実施形態の情報検索システムのＫＬ連携解析部によるキーワード候補抽出処理の一例を示すフローチャート。
【図１５】実施形態の情報検索システムのＫＬ連携解析部によるＨＴＭＬカテゴリからのキーワード候補抽出処理の概念図。
【図１６】実施形態の情報検索システムの検索条件設定画面の一例を示す図。
【発明を実施するための形態】
【００１１】

以下、本実施形態の情報検索システムについて図面を参照して説明する。

【００１２】

図１は、本実施形態に係る検索キーワード提示方法を用いた情報検索システムの構成例を示すブロック図である。

【００１３】

本実施形態の情報検索システムは、図１に示すように、クライアント端末からの入力に基づいて提示キーワードを作成するサーバコンピュータ（以下、「サーバ」という）５０にローカルエリアネットワーク（Ｌｏｃａｌ　Ａｒｅａ　Ｎｅｔｗｏｒｋ：ＬＡＮ）等のネットワーク２０を介してクライアントコンピュータ（以下、「クライアント端末」という）３０が複数台接続されたサーバクライアントシステムを想定する。サーバ５０およびクライアント端末３０は、例えば、一般的なパーソナルコンピュータである。

【００１４】

サーバ５０はコンピュータ１０と記憶装置１１を備える。コンピュータ１０は、ハードディスクドライブなどの記憶装置１１と接続されている。この記憶装置１１は、コンピュータ１０によって実行されるアプリケーションプログラムである本実施形態の情報検索システムのプログラム１２を記憶している。すなわち、記憶装置１１は、プログラム１２を記憶する記憶媒体として機能する。

【００１５】

クライアント端末３０上では、コンピュータ１０を利用するクライアントソフトウェアが動作する。クライアントソフトウェアは、例えばブラウザである。複数のクライアント端末３０は、ネットワーク２０を介してコンピュータ１０と接続されている。なお、図１には、２台のクライアント端末３０以外のクライアント端末３０は省略されている。

【００１６】

図２は本実施形態の情報検索システムの機能構成の一例を示すブロック図である。

【００１７】

図２に示すように、本実施形態の情報検索システムのクライアント端末３０は、入力部３１と表示部３２とを備える。入力部３１は例えばキーボードやマウスであり、表示部３２は例えば液晶ディスプレイである。

【００１８】

ユーザは入力部３１を用いて、ユーザが注目している文書（検索したい文書）に関するキーワード（以下、検索キーワードという）や、ユーザが文書データを検索する検索範囲などを含む検索条件をサーバに入力する。

【００１９】

なお、ユーザによる設定情報の入力は、例えば、表示部３２に表示される、文書データを検索するための検索条件を設定する画面である検索条件設定画面を用いて行う。

【００２０】

入力部３１から入力された検索条件はサーバ５０が有する設定情報データベース１３に設定情報として登録される。サーバ５０は、この設定情報に基づいて周期的に文書検索を行い、検索結果の文書に基づいて、ユーザに検索キーワードの候補（以下、キーワード候補という）を提示する。具体的には、表示部３２であるディスプレイに、この検索結果の文書に関するキーワード候補が表示される。すなわち、表示部３２はキーワード候補をユーザに提示するための出力部として機能する。なお、サーバ５０が文書検索もしくはキーワード候補抽出を行う周期は、あらかじめ設定される。

【００２１】

本実施形態の情報検索システムのサーバ５０は、設定情報登録部５１と、カテゴリ生成部５２と、キーワード候補生成部６０とを備えて構成される。本実施形態において、設定情報登録部５１、カテゴリ生成部５２、およびキーワード候補生成部６０はそれぞれ、図１に示されるコンピュータ１０が記憶装置１１に格納されているプログラム１２を実行することにより実現されるものとする。このプログラム１２は、コンピュータ読み取り可能な記憶媒体に予め格納して頒布可能である。また、このプログラム１２が、ネットワーク

２０を介してコンピュータ１０にダウンロードされても構わない。
【００２２】
また、サーバ５０は、設定情報データベース（以下、ＤＢという）１３、タスクＤＢ１４、文書ＤＢ１５、及びカテゴリＤＢ１６を含む。
【００２３】
設定情報ＤＢ１３には、ユーザが入力した検索条件が設定情報として記憶されている。設定情報は例えば、設定情報の識別子、検索キーワードなどを含む。
【００２４】
ここで、図３を参照して、設定情報の入力、及び設定情報の登録について説明する。図３は検索条件設定画面１００の一例である。
【００２５】
図３に示すように、検索条件設定画面１００は、「設定条件名」１０１、「検索キーワード」１０２、「推奨キーワード」１０３、「キーワード検索範囲」１０４、「タスク名」１０５、および「期間」１０６の項目を有する。図３の検索条件設定画面１００では、設定条件名１０１は「コンピュータ最新動向」、検索キーワード１０２は、「インターネット」もしくは「アーキテクチャ」、キーワード検索範囲１０４は、「タイトル」、タスク名１０５は「最新ネットワーク技術動向」、期間１０６は、「３日前から」と入力されている。
【００２６】
ユーザがこれらの項目を入力し、「設定ボタン」１０７を入力部３１であるマウスでクリックすると、サーバ５０にこれらの入力内容が入力される。入力された情報は、後述する設定情報登録部５１によって設定情報として設定情報ＤＢ１３に登録される。なお、本実施形態では、設定情報の識別子をユーザが入力する設定条件名１０１としているが、設定情報登録部５１によって、入力された設定情報の識別子である設定情報ＩＤが発行されるとしてもよい。
【００２７】
タスクＤＢ１４には、タスクの識別子であるタスクＩＤ、タスク名、タスクに関する説明などを含むタスク情報が記憶されている。なお、タスクとは、クラスタリング分析（以下、クラスタリングという）によりカテゴリを生成する単位である。すなわち、タスクＤＢ１４に格納されているタスク情報はクラスタリング分析を行う際に用いられる。なお、クラスタリングを実行する際の対象となる文書データは、ある１つのタスクに属しているため、あるカテゴリに属している文書は全て同じタスクに属している文書となる。
【００２８】
文書ＤＢ１５には、文書の識別子である文書ＩＤ、文書のタイトル、文書データの本文、文書のナレッジリンク（以下、ＫＬという）情報、文書の投稿日時、文書の属しているタスクのＩＤなどを含む文書データが記憶されている。
【００２９】
なお、ＫＬ情報とは、文書ＤＢ１５に格納されている複数の記事の相互間の関係を示す。すなわち、ＫＬとは異なる２つの文書間の関係を表す情報である。なお、ＫＬは、１つのタスク内に属している文書間のみでなく、別々のタスク内に属している文書間の情報としても生成される。すなわち、あるタスク以外のタスクに属している文書に含まれるキーワードを取得するには、あるタスクに属している文書とＫＬにより結びついている他のタスクに属している文書を取得し、その文書からキーワードを取得する。なお、文書ＤＢ１５には、そのほかに、文書の作成者名を含んでもよいし、あらかじめユーザがラベルを付与しても良い。
【００３０】
すなわちＫＬ情報は、例えばＫＬ情報毎に文書データの相互間の関係の元となる文書を識別する元文書ＩＤ 、当該文書相互間の関係の先となる文書を識別する先文書ＩＤ、及び当該元文書ＩＤと先文書ＩＤとの関係が示されている。
【００３１】

カテゴリＤＢ１６は、例えばカテゴリ名などのカテゴリに関する情報と、当該カテゴリに属する文書データなどを含むカテゴリ情報が記憶されている。本実施形態の情報検索システムにおいては、カテゴリＤＢ１６に格納されたカテゴリは、後述するカテゴリ生成部５２によって、タスクＤＢ１４と文書ＤＢ１５を参照して作成されるとする。

【００３２】

なお、本実施形態において、これらのＤＢは、記憶装置１１に格納される。

【００３３】

設定情報登録部５１は、ユーザが検索条件設定画面にて設定した、文書を検索するための条件情報を設定情報として設定情報ＤＢ１３に登録する。

【００３４】

カテゴリ生成部５２は、タスクＤＢ１４と文書ＤＢ１５とを参照して、どのタスクにどの文書が属しているかという情報を取得する。

【００３５】

そしてカテゴリ生成部５２は、取得した情報に基づいて、タスク毎にクラスタリング分析を実施し、カテゴリを生成する。カテゴリ生成部５２は、生成されたカテゴリに関する情報と、カテゴリにどの文書が属しているかという情報とをカテゴリ情報としてカテゴリＤＢ１６に登録する。なお、このカテゴリ生成処理はあらかじめ設定された期間ごとに定期的に行うものとする。

【００３６】

キーワード候補生成部６０は、設定情報ＤＢ１３から取得した設定情報に基づいて、タスクＤＢ１４と文書ＤＢ１５とから文書データを取得する。そして、キーワード候補生成部６０は、カテゴリＤＢ１６のカテゴリ情報に基づいて、取得した文書データがどのカテゴリに属しているかを判定する。

【００３７】

キーワード候補生成部６０は、上述したカテゴリ判定の結果、取得された文書が特定数以上属しているカテゴリのカテゴリ名、取得された文書が特定数以上属し、かつカテゴリが新規に作成されたカテゴリのカテゴリ名、取得された文書が特定数以上属しているカテゴリ内の文書内において出現頻度の高い単語、取得された文書が特定数以上属しているカテゴリ内の文書であって、期間１０６によって指定された期間内に生成されている文書内において出現頻度の高い単語、又は、取得された文書が特定数以上属しているカテゴリ内で取得された文書とＫＬで関連付いている文書内において出現頻度の高い単語などをキーワード候補として抽出する。抽出されたキーワード候補は、表示部３２の検索条件設定画面における推奨キーワード１０３欄に表示される。　なお、抽出されるキーワード候補はユーザが入力した検索キーワード１０２以外の語句である。

【００３８】

すなわち、キーワード候補生成部６０は、ユーザが入力した検索キーワードに基づいて検索された文書データを検索可能なキーワードをキーワード候補として抽出する。これらの抽出されたキーワード候補は、ユーザが所望する情報を取得する精度を向上できる。

【００３９】

ここで、図４を参照して、本実施形態の情報検索システムの動作について説明する。図４は、本実施形態の情報検索システムにおけるユーザに検索キーワードを提示する処理（以下、検索キーワード提示処理という）の手順の一例を示すフローチャートである。

【００４０】

まず、入力部３１によって、設定情報が入力される（ステップＳ１）。入力された設定情報は、ネットワーク２０を介してクライアント端末３０からサーバ５０に送信される。

【００４１】

サーバ５０が設定情報を受信すると、設定情報登録部５１が設定情報ＤＢ１３に受信した設定情報を登録する（ステップＳ２）。

【００４２】

あらかじめ設定された周期が到来すると（ステップＳ３がＹｅｓ）、キーワード候補生

成部６０が、設定情報ＤＢ１３に登録された設定情報に基づいて文書ＤＢ１５からキーワードを抽出し、キーワード候補を生成する（ステップＳ４）。抽出されたキーワード候補の一覧は、表示部３２に表示された検索条件設定画面１００の推奨キーワード１０３に表示される（ステップＳ５）。これにより、検索キーワード提示処理が終了する。なお、推奨キーワード１０３に表示されたキーワード候補をユーザが選択することで、選択されたキーワード候補が検索キーワード１０２に追加されるようにしても良い。また、この場合、追加されたキーワードは設定情報ＤＢ１３に更新されるようにしても良い。

【００４３】

なお、あらかじめ設定された周期が到来していない場合は（ステップＳ３がＮｏ）、その周期が来るまで待機する。

【００４４】

ここで、図５乃至図１５を参照して、図４のステップＳ４におけるキーワード候補生成部６０によるキーワード候補の生成（以下、キーワード候補生成処理という）について、詳しく説明する。

【００４５】

図５はキーワード候補生成部６０の機能ブロック図である。図５に示すように、キーワード候補生成部６０は、注目文書検索部６１、カテゴリ判定部６２、キーワード候補抽出部６３を備える。

【００４６】

注目文書検索部６１は、特定の周期が到来すると、設定情報ＤＢ１３より設定情報を取得し、タスクＤＢ１４と文書ＤＢ１５を参照して、設定情報ＤＢ１３に登録されている検索キーワード１０２、及び検索対象の期間１０６、検索対象のタスク名に合致する文書を抽出する。注目文書検索部６１は、この抽出結果である文書データをカテゴリ判定部６２に送信する。

【００４７】

カテゴリ判定部６２は、受信した文書のデータに基づいて、カテゴリＤＢ１６よりカテゴリ情報を取得する。そして、カテゴリ判定部６２は、取得したカテゴリ情報と、注目文書検索部６１から受信した検索結果である文書の一覧とに基づいて、検索結果に含まれている当該文書それぞれがどのカテゴリに属しているかを解析する。解析結果である各文書がどのカテゴリに属しているかという情報はキーワード候補抽出部６３に送信される。

【００４８】

キーワード候補抽出部６３は、出現頻度解析部６４、期間限定出現頻度解析部６５、カテゴリ名取得部６６、新カテゴリ名取得部６７、およびＫＬ連携取得部６８を備える。キーワード候補抽出部６３は、カテゴリ判定部６２から受信した解析結果に基づいて、キーワード候補を抽出する。そして、抽出されたキーワード候補の一覧は表示部３２の検索条件設定画面１００の推奨キーワード１０３に表示される。以下、キーワード候補抽出部６３が備える、出現頻度解析部６４、期間限定出現頻度解析部６５、カテゴリ名取得部６６、新カテゴリ名取得部６７、およびＫＬ連携取得部６８について説明する。

【００４９】

出現頻度解析部６４は、キーワード抽出部６３がカテゴリ判定部６２から受信した解析結果に基づいて、検索結果の文書を含んでいる各カテゴリ内に属している文書における、検索キーワード１０２に設定されている単語以外で出現頻度の高い単語をキーワード候補として抽出する。出現頻度解析部６４は、抽出したキーワード候補をキーワード候補抽出部６３に送信する。

【００５０】

期間限定出現頻度解析部６５は、キーワード抽出部６３がカテゴリ判定部６２から受信した解析結果から、検索結果の文書を含んでいる各カテゴリ内に属していて、期間１０６によって指定された期間内で生成された文書において、検索キーワード１０２に設定されている単語以外で出現頻度の高い単語をキーワード候補として抽出する。期間限定出現頻度解析部６５は、抽出したキーワード候補をキーワード候補抽出部６３に送信する。

【００５１】
カテゴリ名取得部６６は、キーワード抽出部６３がカテゴリ判定部６２から受信した解析結果から、カテゴリ内に含まれる文書数をカテゴリごとに計測する。カテゴリ名取得部６６は計測結果に基づき、特定数以上の注目文書を含むカテゴリのカテゴリ名をキーワード候補として抽出する。カテゴリ名取得部６６は、抽出したキーワード候補をキーワード候補抽出部６３に送信する。
【００５２】
新カテゴリ名取得部６７は、キーワード抽出部６３がカテゴリ判定部６２から受信した解析結果から、カテゴリ内の検索結果の文書数をカテゴリごとに計測する。新カテゴリ名取得部６７は計測結果に基づき、特定数以上の検索結果の文書を含むカテゴリの中でクラスタリング分析を行う。そして、新カテゴリ名取得部６７はクラスタリング分析により新たに作成されたカテゴリのなかで、検索結果の文書が含まれるカテゴリのカテゴリ名を検索キーワード候補として抽出する。新カテゴリ名取得部６７は、抽出したキーワード候補をキーワード候補抽出部６３に送信する。
【００５３】
ＫＬ連携取得部６８は、キーワード抽出部６３がカテゴリ判定部６２から受信した解析結果から、検索結果の文書を含んでいる各カテゴリ内に属していて、検索結果の文書とＫＬにより関連付いている文書内において、検索キーワードに設定されている単語以外で出現頻度の高い単語を検索キーワード候補として抽出する。ＫＬ連携取得部６８は、抽出したキーワード候補をキーワード候補抽出部６３に送信する。
【００５４】
なお、単語の頻出頻度の高さや、カテゴリが含む検索結果の文書の数の多さやを判定するための閾値は、あらかじめシステム構築時などに設定されているとする。また、ユーザが設定可能なようにしてもよい。また、閾値は文書ＤＢ１５に登録されたデータ量に対する検索結果の文書のデータ量の割合などでも良い。もしくは閾値を設定するのではなく、一番検索結果の文書が多く属するカテゴリを抽出するとしても良い。
【００５５】
また、ＫＬとは、異なる２つの文書間の因果関係を表す情報であり、ある文書から見た他の文書への参照とその説明から成る。また、ＫＬには因果関係に沿った方向性があり、入力としての元文書から出力としての先文書へという関係を保持している。ＫＬは、例えば「Ｄｅｓｃｒｉｂｅ」、「Ｆｏｌｌｏｗ」、「Ｒｅｖｉｓｅ」、「Ｓｕｍ」、「Ｃｏｎｓｕｌｔ」という種類（以下、クラスという）を有する。
【００５６】
なお、Ｄｅｓｃｒｉｂｅとは、説明的な関係にあり、元文書と先文書が共存しないと文脈的には理解できないほど強い結びつきを持っていることを意味するＬｉｎｋである。例えば、添付ファイルとそれを説明するメッセージがこの関係に当てはまる。Ｆｏｌｌｏｗとは、ある文書を受けてそれに補足／返答／反論するなど、文脈的な情報を追加していくこと全般を意味するＬｉｎｋである。Ｒｅｖｉｓｅとは、ある文書を改訂して置き換えている関係を意味するＬｉｎｋである。Ｓｕｍとは、ある文書をまとめたり、要約したりする関係を意味するＬｉｎｋである。Ｃｏｎｓｕｌｔとは、ある文書を参考にしたという関係を意味するＬｉｎｋである。
【００５７】
続いて、図６乃至図１５を参照して、図４のステップＳ４におけるキーワード候補生成部６０によるキーワード候補生成処理について説明する。
【００５８】
図６は本実施形態のキーワード候補生成部６０によるキーワード候補生成処理の一例を示すフローチャートである。
【００５９】
キーワード候補生成部６０が設定情報を取得すると、注目文書検索部６１が、この設定情報に基づいて文書を抽出する（ステップＳ１１）。具体的には、注目文書検索部６１は

、設定情報に登録されている検索キーワード１０２及び検索対象の期間１０６、検索対象のタスク名と合致する文書をタスクＤＢ１４と文書ＤＢ１５から抽出する。条件に該当し、抽出された文書を注目文書とする。そして注目文書検索部６１は、抽出した注目文書データはカテゴリ判定部６２に送信する。なお、設定情報の取得は、設定情報ＤＢ１３に登録された設定情報を取得してもよいし、入力部３１から入力された設定情報を直接受信してもよい。

【００６０】

続いて、カテゴリ判定部６２は、カテゴリＤＢ１６を参照して注目文書検索部６１から受信した注目文書データのカテゴリを解析する（ステップＳ１２）。具体的には、カテゴリ判定部６２は、カテゴリＤＢ１６からカテゴリ情報を取得する。そして、取得したカテゴリ情報と注目文書検索部６１から受信し注目文書データの一覧とに基づいて、カテゴリ判定部６２は、検索結果の注目文書ごとに、どのカテゴリに属しているかを解析する。カテゴリ判定部６２による解析結果はキーワード候補抽出部６２に送信される。なお、各注目文書がどのカテゴリに属しているかという情報をカテゴリ解析結果と呼ぶものとする。

【００６１】

図７にカテゴリ解析結果の概念図を示す。図７に示すように、本実施形態の情報検索システムにおいては、カテゴリＤＢ１６に「ＨＴＭＬ」１６１、「ネットワーク」１６２、および「無線ＬＡＮ」１６３のカテゴリが存在する。ＨＴＭＬ１６１のカテゴリには、文書データ１５１-1乃至１５１-5が属している。そのうち、当該カテゴリに属すると判定される文書データは、斜線の文書である注目文書データ１５１-1乃至１５１-3である。また、ネットワーク１６２のカテゴリには、文書データ１５２-1乃至１５２-3が属している。そのうち、当該カテゴリに属すると判定される文書データは、ドットの文書である注目文書データ１５２-1である。また、無線ＬＡＮ１６３のカテゴリには、文書データ１５３-1乃至１５３-6が属している。そのうち、当該カテゴリに属すると判定される文書データはない。

【００６２】

キーワード候補抽出部６３が解析結果を受信すると、キーワード候補抽出部６３によって、ステップＳ１３乃至ステップＳ１７のキーワード候補抽出処理が行われる。

【００６３】

まず、キーワード候補抽出部６３のカテゴリ名取得部６６がキーワード候補の抽出を行う（ステップＳ１３）。次に、キーワード候補抽出部６３の新カテゴリ名取得部６７がキーワード候補の抽出を行う（ステップＳ１４）。続いて、キーワード候補抽出部６３の出現頻度解析部６４がキーワード候補の抽出を行う（ステップＳ１５）。続いて、キーワード候補抽出部６３の期間限定出現頻度解析部６５がキーワード候補の抽出を行う（ステップＳ１６）。最後に、キーワード候補抽出部６３のＫＬ連携解析部６８がキーワード候補の抽出を行う（ステップＳ１７）。

【００６４】

キーワード候補抽出処理が終了すると、キーワード候補抽出部６３は抽出されたキーワード候補を全てクライアント端末２０に送信する（ステップＳ１８）。これにより、キーワード候補生成処理は終了する。

【００６５】

ここで、図８乃至図１５を参照して、図６のステップＳ１２乃至ステップＳ１７のそれぞれの処理を詳しく説明する。

【００６６】

まず、図８を参照して、図６のステップＳ１３におけるカテゴリ名取得部６６によるキーワード候補抽出処理について説明する。図８はカテゴリ名取得部６６によるキーワード候補抽出処理の一例を示すフローチャートである。

【００６７】

まず、カテゴリ名取得部６６は、カテゴリ判定部６２によるカテゴリ解析結果を受信する（ステップＳ１３１）。受信したカテゴリ解析結果に基づいて、カテゴリ名取得部６６

は、検索結果の注目文書群のカテゴリごとの個数を集計する（ステップＳ１３２）。

【００６８】

集計した結果、注目文書が特定数以上含まれている場合（ステップＳ１３３がＹｅｓ）、該当するカテゴリのカテゴリ名をキーワード候補と判定する（ステップＳ１３４）。ステップＳ１３４でキーワード候補と判定された全てのカテゴリ名をキーワード候補抽出部６３に送信し（ステップＳ１３５）、処理を終了する。

【００６９】

本実施形態では、注目文書が３以上含まれているカテゴリのカテゴリ名を抽出するとする。すなわち、「ＨＴＭＬ」が抽出される。なお、集計した結果、注目文書が特定数以上含まれていない場合（ステップＳ１３３がＮｏ）、キーワードの候補は抽出せず、処理を終了する。

【００７０】

本実施形態では、カテゴリ名をキーワード候補とすることで、検索キーワードと関連付いたキーワードの抽出を可能とする。

【００７１】

図９を参照して、図６のステップＳ１４における新カテゴリ名取得部６７によるキーワード候補抽出処理について説明する。図９は新カテゴリ名取得部６７によるキーワード候補抽出処理の一例を示すフローチャートである。

【００７２】

まず、新カテゴリ名取得部６７は、カテゴリ判定部６２によるカテゴリ解析結果を受信する（ステップＳ１４１）。受信したカテゴリ解析結果に基づいて、新カテゴリ名取得部６７は、検索結果の注目文書群のカテゴリごとの個数を集計する（ステップＳ１４２）。

【００７３】

集計した結果、注目文書が特定数以上含まれているカテゴリが存在する場合（ステップＳ１４３がＹｅｓ）、新カテゴリ名取得部６７は、当該カテゴリに含まれる文書データのクラスタリング分析を行い、新たなカテゴリ（以下、新カテゴリという）を生成する（ステップＳ１４４）。

【００７４】

新カテゴリが生成されると、新カテゴリ名取得部６７は、新カテゴリの中に、注目文書データが含まれている新カテゴリがあるか否かを判定する（ステップＳ１４５）。注目文書データが含まれている新カテゴリがある場合（ステップＳ１４５がＹｅｓ）対象の新カテゴリ名（以下、新カテゴリ名という）をキーワード候補と判定し（ステップＳ１４６）、当該キーワード候補をキーワード候補抽出部６３に送信して（ステップＳ１４７）、処理を終了する。

【００７５】

なお、集計した結果、注目文書が特定数以上含まれているカテゴリが存在しない場合（ステップＳ１４３がＮｏ）、キーワードの候補は抽出せず、処理を終了する。また、注目文書データが含まれている新カテゴリがない場合（ステップＳ１４５がＮｏ）も、キーワードの候補は抽出せず、処理を終了する。

【００７６】

図１０を参照して、本実施形態の情報検索システムにおける新カテゴリ名取得部６７によるキーワード候補抽出処理の一例を具体的に説明する。なお、本実施形態では、あらかじめ定められたカテゴリ毎に属する文書数の閾値は３である。

【００７７】

したがって、図１０に示すように、ＨＴＭＬ１６１に属する文書データに関して処理が行われる。すなわち、ＨＴＭＬ１６１に属する文書データにクラスタリングが行われ、カテゴリが生成される。ここでは、新カテゴリとして、タグ１６４とハイパーテキスト１６５とが生成されている。このタグ１６４には注目文書データが、属しているため、本実施形態ではタグ１６４がキーワード候補として抽出される。

【００７８】

次に図１１を参照して、図６のステップＳ１５における出現頻度解析部６４によるキーワード候補抽出処理について説明する。図１１は出現頻度解析部６４によるキーワード候補抽出処理の一例を示すフローチャートである。

【００７９】

まず、出現頻度解析部６４は、カテゴリ判定部６２によるカテゴリ解析結果を受信する（ステップＳ１５１）。受信したカテゴリ解析結果に基づいて、出現頻度解析部６４は、検索結果の注目文書が属しているカテゴリ毎に、当該カテゴリ内の文書データにおいて出現する頻度の高い単語を抽出する。

【００８０】

すなわち、出現頻度解析部６４は、検索結果の注目文書が属しているカテゴリ毎に、当該カテゴリ内の文書データに含まれる各単語のうち、検索キーワード１０２に設定されている単語以外の単語についての出現頻度を集計する（ステップＳ１５２）。

【００８１】

集計した結果、出現頻度が特定数以上の単語が存在する場合（ステップＳ１５３がＹｅｓ）、これらの単語をキーワード候補と判定し（ステップＳ１５４）、当該キーワード候補をキーワード候補抽出部６３に送信して（ステップＳ１５５）、処理を終了する。

【００８２】

集計した結果、出現頻度が特定数以上の単語が存在しない場合（ステップＳ１５３がＮｏ）、出現頻度解析部６４はキーワードの候補は抽出せず、処理を終了する。

【００８３】

図１２に、出現頻度解析部６４によるＨＴＭＬ１６１カテゴリからのキーワード候補抽出処理の概念図を示す。図１２に示すように、出現頻度解析部６４はＨＴＭＬ１６１カテゴリに属する文書データのタイトルに含まれる単語を抽出する。そのうち、出現数があらかじめ定めた数値より多いものであって、検索キーワード１０２である「インターネット」を除く単語をキーワード候補とする。すなわち、網掛けの「グリッドコンピュータ」という単語がキーワード候補として抽出される。

【００８４】

検索キーワード１０２が属しているカテゴリにおける出現頻度の高い単語は、ユーザが所望する情報との関連性が高いと考えられる。このため、これらの単語をキーワード候補としてユーザに提示することで、ユーザが所望する情報を取得する精度を上げることができる。

【００８５】

図１３を参照して、図６のステップＳ１６における期間限定出現頻度解析部６５によるキーワード候補抽出処理について説明する。図１３は期間限定出現頻度解析部６５によるキーワード候補抽出処理の一例を示すフローチャートである。

【００８６】

まず、期間限定出現頻度解析部６５は、カテゴリ判定部６２によるカテゴリ解析結果を受信する（ステップＳ１６１）。受信したカテゴリ解析結果に基づいて、期間限定出現頻度解析部６５は、検索結果の注目文書が属しているカテゴリ毎に、当該カテゴリ内の文書データにおいて、設定情報の期間１０６によって設定された期間内に生成された、すなわち、文書ＤＢ１５に登録された文書を抽出する（ステップＳ１６２）。すなわち、期間限定出現頻度解析部６５は、対象のカテゴリ内の文書データを文書の登録日時によって絞り込む。

【００８７】

ステップＳ１６２において抽出された文書データにおいて、期間限定出現頻度解析部６５は、検索キーワード１０２に設定されている単語以外の単語において出現する頻度の高い単語を抽出する。すなわち、期間限定出現頻度解析部６５は、検索結果の注目文書が属しているカテゴリ毎に、当該カテゴリ内の文書データに含まれる各単語の出現頻度を集計する（ステップＳ１６３）。

【００８８】

集計した結果、出現頻度が特定数以上の単語が存在する場合（ステップＳ１６４がＹｅｓ）、これらの単語をキーワード候補と判定し（ステップＳ１６５）、当該キーワード候補をキーワード候補抽出部６３に送信して（ステップＳ１６６）、処理を終了する。

【００８９】

集計した結果、出現頻度が特定数以上の単語が存在しない場合（ステップＳ１６４がＮｏ）、期間限定出現頻度解析部６５はキーワードの候補は抽出せず、処理を終了する。

【００９０】

検索対象となる文書データを指定した期間１０２で絞り込むことで、最新の文書を対象にして検索キーワードの候補を抽出することが可能となる。したがって、これらのキーワード候補をユーザに提示することで、ユーザが所望する情報を取得する精度を上げることができる。

【００９１】

図１４を参照して、図６のステップＳ１７におけるＫＬ連携解析部６８によるキーワード候補抽出処理について説明する。図１４はＫＬ連携解析部６８によるキーワード候補抽出処理の一例を示すフローチャートである。

【００９２】

まず、ＫＬ連携解析部６８は、カテゴリ判定部６２によるカテゴリ解析結果を受信する（ステップＳ１７１）。受信したカテゴリ解析結果に基づいて、ＫＬ連携解析部６８は、検索結果の注目文書が属しているカテゴリ内の文書データにおいて、当該注目文書とナレッジリンクにより関連付いている文書を抽出する（ステップＳ１７２）。このナレッジリンクに関する情報は文書ＤＢ１５から抽出される。

【００９３】

ステップＳ１６２において抽出された文書データにおいて、ＫＬ連携解析部６８は、検索キーワード１０２に設定されている単語以外の出現する頻度の高い単語を抽出する。すなわち、ＫＬ連携解析部６８は、抽出結果の文書データが属しているカテゴリ毎に、当該カテゴリ内の文書データに含まれる検索キーワード１０２に設定されている単語以外の単語の各出現頻度を集計する（ステップＳ１７３）。

【００９４】

集計した結果、出現頻度が特定数以上の単語が存在する場合（ステップＳ１７４がＹｅｓ）、これらの単語をキーワード候補と判定し（ステップＳ１７５）、当該キーワード候補をキーワード候補抽出部６３に送信して（ステップＳ１７６）、処理を終了する。

【００９５】

集計した結果、出現頻度が特定数以上の単語が存在しない場合（ステップＳ１７４がＮｏ）、ＫＬ連携解析部６８はキーワードの候補は抽出せず、処理を終了する。

【００９６】

図１５に、ＫＬ連携解析部６８によるＨＴＭＬ１６１カテゴリからのキーワード候補抽出処理の概念図を示す。図１５に示すように、ＫＬ連携解析部６８は、ＨＴＭＬ１６１に属する文書データに、ＫＬによって関連付けられた文書データを抽出する（点線の文書）。ＫＬ連携解析部６８は、これらの抽出された文書データのタイトルに含まれる単語であって、あらかじめ定めた数値以上含まれている単語をキーワード候補として抽出する。この際、検索キーワード１０２に入力された単語は除外される。

【００９７】

上述したように、ＫＬ連携解析部６８は、ユーザが指定した検索キーワード１０２を含む文書データと関連する文書データが属するカテゴリにおいて、特定数以上の出現頻度のある単語をキーワード候補として抽出する。すなわち、本実施形態は、ユーザが検索した文書と関連する文書における出現頻度の高い単語をキーワード候補としてユーザに提示することで、ユーザが所望する情報を取得する精度を上げることを可能とする。

【００９８】

図１６に本実施形態の情報検索システムにおけるキーワード候補生成処理後に表示部３２に表示される画面の一例を示す。

【００９９】
図１６に示すように、本実施形態の情報検索システムは、キーワード候補生成処理後、検索条件設定画面１００の推奨キーワード１０３に、キーワード候補抽出部６３によって抽出されたキーワード候補を表示する。本実施形態においては、「ＨＴＭＬ」と「グリッドコンピューティング」とが表示される。
【０１００】
すなわち、本実施形態の情報検索システムは、このようにユーザが入力した検索キーワードに基づいて、検索キーワードの候補を提示することによって、ユーザが指定した検索キーワードが廃れた場合など、ユーザが入力した検索キーワード以外のよりユーザの注目している文書を検索できるキーワードをユーザに提示することができる。これにより、ユーザ新たに検索キーワードを設定する必要をなくすことを可能とする。また、ユーザが知らない最新の検索キーワードを取得することを可能とする。
【０１０１】
故に、本実施形態の情報検索システムは、検索において必要とされる最新の情報を取得することが可能となるため、業務効率を向上させることができる。また、ユーザが所望する情報を検索する精度を向上する検索キーワードをユーザに提示することを可能とする。
【０１０２】
なお、本実施形態の情報検索システムは、検索キーワードと関連付いている単語を探すような場合に、単語間の関連性に関する情報を保持する関連辞書の作成やメンテナンスなどの作業を行う必要が無い。また本実施形態の情報検索システムは、ユーザが検索で用いたキーワードを元にどの文書を参照したかを解析するためのログ情報の取得やログ情報の解析などの処理をする必要も無い。
【０１０３】
なお、ここでは本発明の実施形態を説明したが、この実施形態は例として提示したものであり、発明の範囲を限定することは意図していない。この新規な実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。この実施形態やその変形は、発明の範囲や要旨に含まれるとともに、特許請求の範囲に記載された発明とその均等の範囲に含まれる。
【０１０４】
例えば、キーワード候補抽出処理における、単語の出現頻度が高いかを判定する際に用いられる特定数は、本実施形態の情報検索システム構築時に設定してもいいし、サーバの管理者によって設定できるようにしてもよい。
また、例えば、本実施形態の情報検索システムにおける文書検索は、ユーザの指定したタイミングで行われるようにしても良い。
【０１０５】
また、キーワード候補抽出処理は、キーワード候補抽出部６３が有する各部が順次行ってもよいし、同時に行っても良い。なお、キーワード候補抽出部６３が有する各部がキーワード候補抽出処理を行う順番は特に限定されない。また、キーワード候補抽出処理はすべてを行わず、いくつかの処理の組み合わせでも良い。
【０１０６】
また、キーワード候補抽出部６３によるキーワード候補抽出処理において、期間限定出現頻度解析部６５における単語の出現頻度の集計は行わず、出現頻度解析部６４による単語の出現頻度の集計データを用いても良い。
【０１０７】
また、キーワード候補抽出部６３によるキーワード候補抽出処理における単語の出現頻度の集計は、出現頻度解析部６４、期間限定出現頻度解析部６５、及びＫＬ連携解析部６７それぞれで行わず、例えば、出現頻度解析部６４などのどれか一つの部による単語の出現頻度の集計データをその他の部においても用いるようにしても良い。
【０１０８】
同様に、キーワード候補抽出部６３によるキーワード候補抽出処理における各カテゴリ

に含まれている注目文書の集計は、カテゴリ名取得部６６、新カテゴリ名取得部６７それぞれで行わず、例えば、カテゴリ名取得部６６による計データを新カテゴリ名取得部６７においても用いるようにしても良い。

【０１０９】

また、検索条件はその検索の周期と対応させて、複数登録することが可能である。

【０１１０】

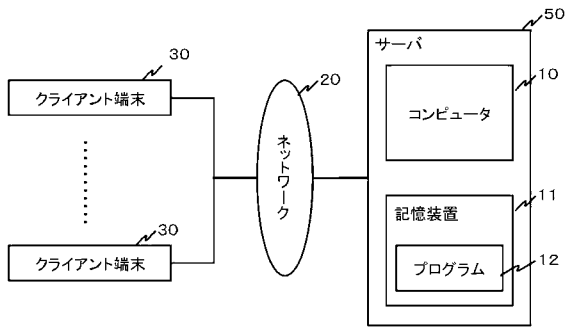
また、設定情報には、文書データの作成者や、文書データのラベルなどを設定してもよい。

【符号の説明】

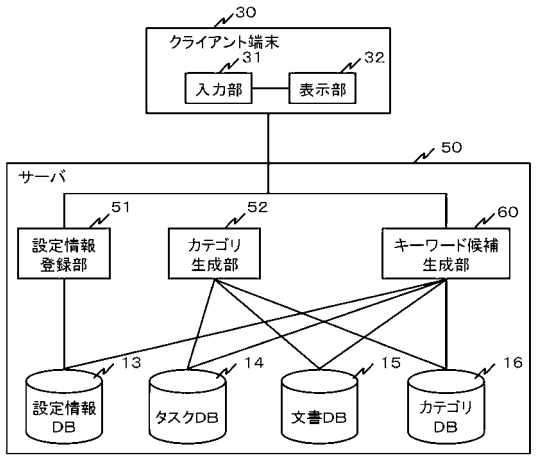
【０１１１】

１０…コンピュータ、１１…記憶装置、１２…プログラム、１３…設定情報ＤＢ、１４…タスクＤＢ、１５…文書ＤＢ、１６…カテゴリＤＢ、２０…ネットワーク、３０…クライアントコンピュータ、３１…入力部、３２…表示部、５０…サーバ、５１…設定情報登録部、５２…カテゴリ生成部、６０…キーワード候補生成部、６１…注目文書検索部、６２…カテゴリ判定部、６３…キーワード候補抽出部、６４…出現頻度解析部、６５…期間限定出現頻度解析部、６６…カテゴリ名取得部、６７…新カテゴリ名取得部、６８…ナレッジリンク連携取得部

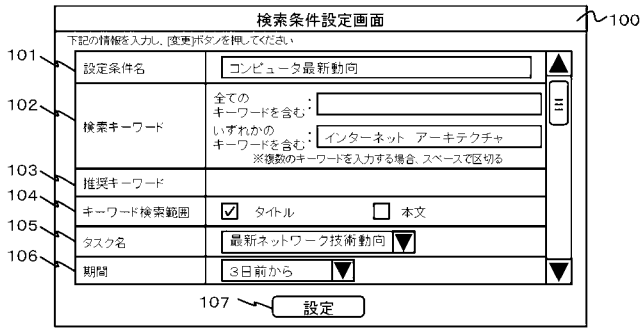
【図１】



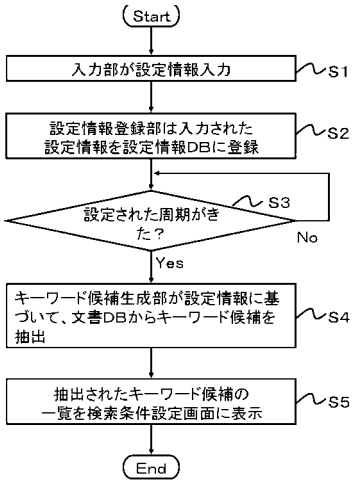
【図２】



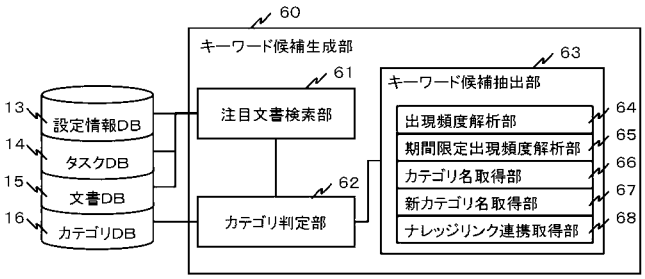
【図３】


検索条件設定画面
100
下記の情報を入力し、[変更]ボタンを押してください
101
設定条件名
コンピュータ最新動向
102
検索キーワード
全ての
キーワードを含む：
いずれかの
キーワードを含む：
インターネット アーキテクチャ
※複数のキーワードを入力する場合、スペースで区切る
103
推奨キーワード
104
キーワード検索範囲
タイトル
本文
105
タスク名
最新ネットワーク技術動向
106
期間
3日前から
107
設定

【図４】


Start
入力部が設定情報入力
S1
設定情報登録部は入力された
設定情報を設定情報DBに登録
S2
設定された周期がき
た？
S3
No
Yes
キーワード候補生成部が設定情報に基
づいて、文書DBからキーワード候補を
抽出
S4
抽出されたキーワード候補の
一覧を検索条件設定画面に表示
S5
End

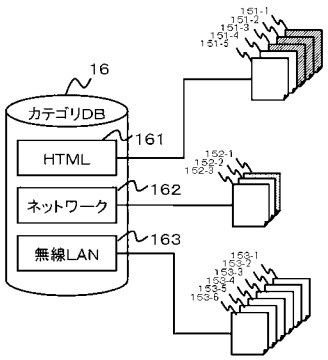
【図５】


60
キーワード候補生成部
61
注目文書検索部
62
カテゴリ判定部
63
キーワード候補抽出部
出現頻度解析部
期間限定出現頻度解析部
カテゴリ名取得部
新カテゴリ名取得部
ナレッジリンク連携取得部
64
65
66
67
68
13
設定情報DB
14
タスクDB
15
文書DB
16
カテゴリDB

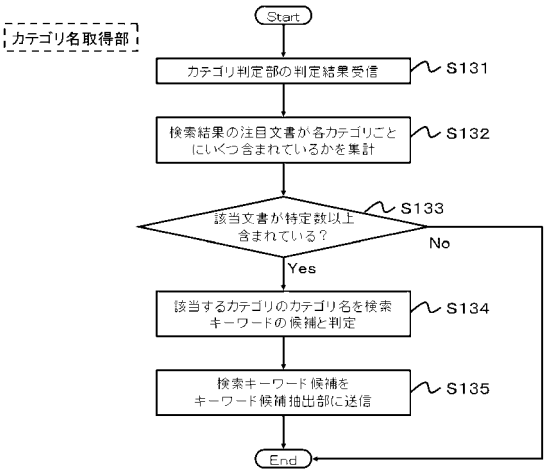
【図６】


Start
注目文書検索部は設定情報に
基づいて注目文書を抽出
S11
カテゴリ判定部はカテゴリDBを
参照して注目文書のカテゴリを判定
S12
カテゴリ名取得部が
キーワード候補を抽出
S13
新カテゴリ名取得部が
キーワード候補を抽出
S14
出現頻度解析部が
キーワード候補を抽出
S15
期間限定出現頻度解析部が
キーワード候補を抽出
S16
ナレッジリンク連携解析部が
キーワード候補を抽出
S17
キーワード候補抽出部は、キーワード
候補全てをクライアント端末に送信
S18
End

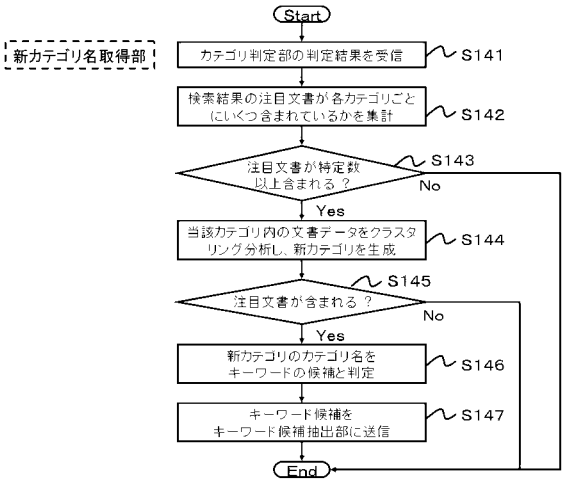
【図７】


16
カテゴリDB
161
HTML
162
ネットワーク
163
無線LAN
151-1
151-2
151-3
151-4
151-5
152-1
152-2
152-3
153-1
153-2
153-3
153-4
153-5
153-6

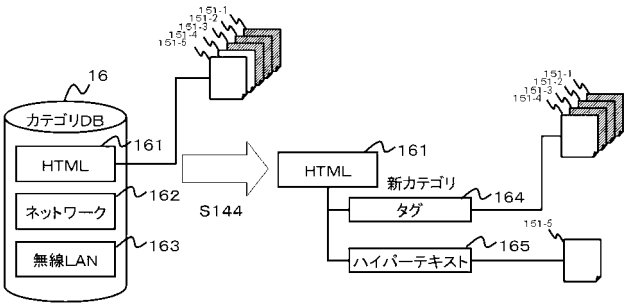
【図８】


カテゴリ名取得部
Start
カテゴリ判定部の判定結果受信
S131
検索結果の注目文書が各カテゴリごと
にいくつ含まれているかを集計
S132
該当文書が特定数以上
含まれている？
S133
No
Yes
該当するカテゴリのカテゴリ名を検索
キーワードの候補と判定
S134
検索キーワード候補を
キーワード候補抽出部に送信
S135
End

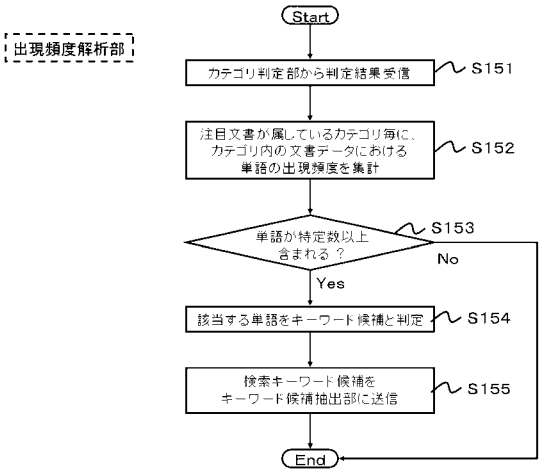
【図９】


新カテゴリ名取得部
Start
カテゴリ判定部の判定結果を受信
S141
検索結果の注目文書が各カテゴリごと
にいくつ含まれているかを集計
S142
注目文書が特定数
以上含まれる ？
S143
No
Yes
当該カテゴリ内の文書データをクラスタ
リング分析し、新カテゴリを生成
S144
注目文書が含まれる ？
S145
No
Yes
新カテゴリのカテゴリ名を
キーワードの候補と判定
S146
キーワード候補を
キーワード候補抽出部に送信
S147
End

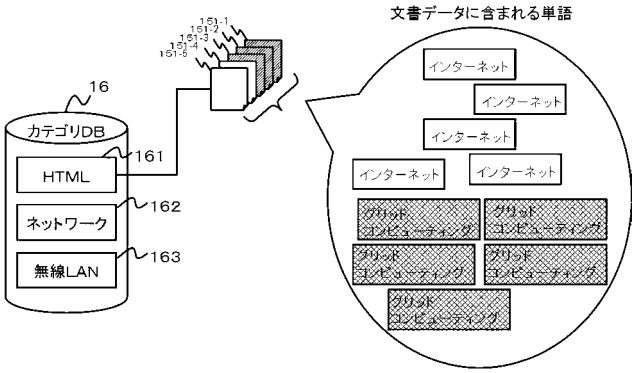
【図１０】


16
カテゴリDB
161
HTML
162
ネットワーク
163
無線LAN
151-1
151-2
151-3
151-4
151-5
S144
161
HTML
新カテゴリ
164
タグ
165
ハイパーテキスト

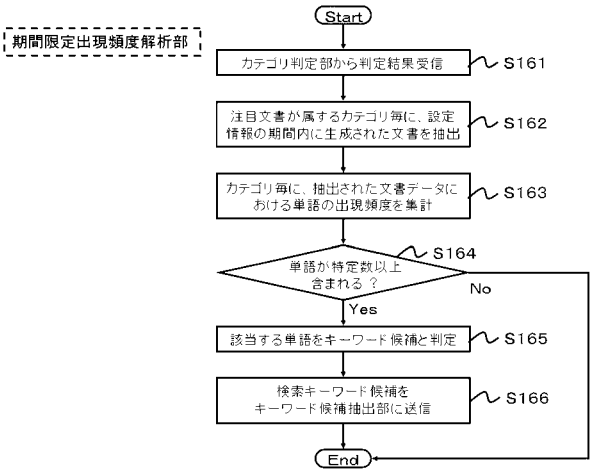
【図１１】



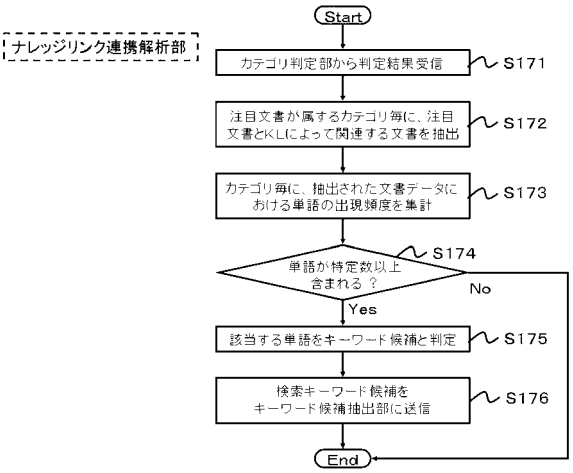
【図１２】



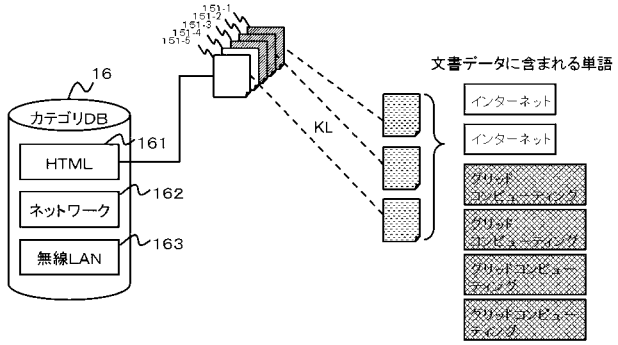
【図１３】



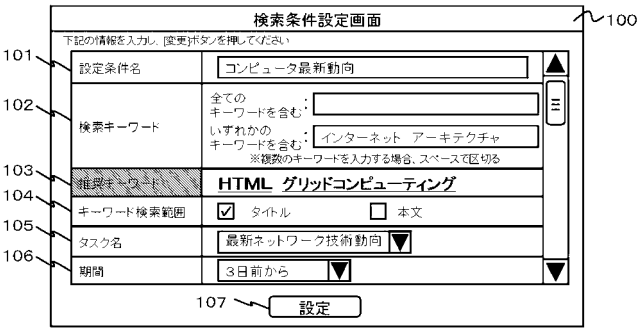
【図１４】



【図１５】


16
カテゴリDB
161
HTML
162
ネットワーク
163
無線LAN
151-1
151-2
151-3
151-4
151-5
KL
文書データに含まれる単語
インターネット
インターネット
グリッドコンピューティング
グリッドコンピューティング
グリッドコンピューティング
グリッドコンピューティング

【図１６】


検索条件設定画面
100
下記の情報を入力し、[変更]ボタンを押してください
101
設定条件名
コンピュータ最新動向
102
検索キーワード
全ての
キーワードを含む：
いずれかの
キーワードを含む：
インターネット アーキテクチャ
※複数のキーワードを入力する場合、スペースで区切る
103
推奨キーワード
HTML グリッドコンピューティング
104
キーワード検索範囲
タイトル
本文
105
タスク名
最新ネットワーク技術動向
106
期間
3日前から
107
設定

フロントページの続き

(72)発明者　佐々木　淳哉
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

Ｆターム(参考) 5B075 ND03　QM10　UU40