

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6154072号
(P6154072)

(45) 発行日 平成29年6月28日(2017.6.28)

(24) 登録日 平成29年6月9日(2017.6.9)

(51) Int.Cl. F I
G06F 17/30 (2006.01) G O 6 F 17/30 2 2 O Z
 G O 6 F 17/30 1 7 O A

請求項の数 5 (全 17 頁)

(21) 出願番号	特願2016-523487 (P2016-523487)	(73) 特許権者	000004226 日本電信電話株式会社 東京都千代田区大手町一丁目5番1号
(86) (22) 出願日	平成27年5月25日(2015.5.25)	(74) 代理人	110002147 特許業務法人酒井国際特許事務所
(86) 国際出願番号	PCT/JP2015/064939	(72) 発明者	折原 慎吾 東京都千代田区大手町一丁目5番1号 日 本電信電話株式会社内
(87) 国際公開番号	W02015/182559	(72) 発明者	熊谷 充敏 東京都千代田区大手町一丁目5番1号 日 本電信電話株式会社内
(87) 国際公開日	平成27年12月3日(2015.12.3)	(72) 発明者	安部 哲哉 東京都千代田区大手町一丁目5番1号 日 本電信電話株式会社内
審査請求日	平成28年6月2日(2016.6.2)		
(31) 優先権主張番号	特願2014-111596 (P2014-111596)		
(32) 優先日	平成26年5月29日(2014.5.29)		
(33) 優先権主張国	日本国(JP)		

最終頁に続く

(54) 【発明の名称】 情報分析システム、情報分析方法及び情報分析プログラム

(57) 【特許請求の範囲】

【請求項1】

ネットワーク上に投稿された発言の集合であるスレッドについて、当該スレッドに含まれる発言の重要度を、当該発言に関するデータである発言データに基づいて、当該発言ごとに分析する発言分析部と、

前記スレッドが、予め設定された複数のカテゴリのいずれに属するかを、当該スレッドに関するデータであるスレッドデータに基づいて分析するスレッド分析部と、

前記発言ごとに、当該発言と、当該発言の重要度と、当該発言が含まれるスレッドのカテゴリとを対応づけて所定の記憶部に格納する格納部と

を備えることを特徴とする情報分析システム。

10

【請求項2】

前記スレッド分析部は、更に、前記スレッドデータに基づいて、前記スレッドの重要度を分析し、

前記格納部は、前記発言の重要度及び前記スレッドの重要度に基づいて、当該発言の総合重要度を算出し、算出した総合重要度と、前記発言と、当該発言が含まれるスレッドのカテゴリとを前記発言ごとに対応づけて格納することを特徴とする請求項1に記載の情報分析システム。

【請求項3】

前記発言分析部は、更に、前記発言の内容が予め設定された複数のカテゴリのいずれに属するかを示す発言カテゴリを、前記発言データに基づいて、前記発言ごとに分析し、

20

前記格納部は、更に、前記発言ごとの発言カテゴリを、当該発言ごとに対応づけて格納することを特徴とする請求項 1 又は 2 に記載の情報分析システム。

【請求項 4】

情報分析システムによって実行される情報分析方法であって、

ネットワーク上に投稿された発言の集合であるスレッドについて、当該スレッドに含まれる発言の重要度を、当該発言に関するデータである発言データに基づいて、当該発言ごとに分析する発言分析工程と、

前記スレッドが、予め設定された複数のカテゴリのいずれに属するかを、当該スレッドに関するデータであるスレッドデータに基づいて分析するスレッド分析工程と、

前記発言ごとに、当該発言の重要度と、当該発言が含まれるスレッドのカテゴリとを対応づけて所定の記憶部に格納する格納工程と

を含むことを特徴とする情報分析方法。

10

【請求項 5】

ネットワーク上に投稿された発言の集合であるスレッドについて、当該スレッドに含まれる発言の重要度を、当該発言に関するデータである発言データに基づいて、当該発言ごとに分析する発言分析ステップと、

前記スレッドが、予め設定された複数のカテゴリのいずれに属するかを、当該スレッドに関するデータであるスレッドデータに基づいて分析するスレッド分析ステップと、

前記発言ごとに、当該発言の重要度と、当該発言が含まれるスレッドのカテゴリとを対応づけて所定の記憶部に格納する格納ステップと

をコンピュータに実行させることを特徴とする情報分析プログラム。

20

【発明の詳細な説明】

【技術分野】

【0001】

本発明の実施形態は、情報分析システム、情報分析方法及び情報分析プログラムに関する。

【背景技術】

【0002】

近年、情報処理技術の高速化及び低コスト化や、インターネット技術等のネットワーク技術の普及等に伴って、ネットワーク上に膨大な量の情報（電子データ）が蓄積されている。例えば、インターネット上の掲示板（電子掲示板）には、利用者によって投稿された発言が時系列順にまとめられている。

30

【0003】

ここで、利用者が所望の情報を得るために、各種情報を推薦する情報推薦技術が提案されている。例えば、掲示板から特徴ベクトルを生成して特徴を分析する技術がある。この技術では、例えば、利用者によってキーワードが入力されると、入力されたキーワードに対応する特徴を有する掲示板が利用者に提示される。

【先行技術文献】

【特許文献】

【0004】

【特許文献 1】特開 2010 - 231471 号公報

40

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、従来技術では、ネットワーク上に投稿された情報から有益な発言を分析することが難しいという問題があった。例えば、掲示板から特徴ベクトルを生成して特徴を分析する技術では、掲示板に意味の無い発言が多く含まれる場合に、有益な発言が意味の無い発言に埋もれてしまっていた。

【0006】

そこで、この発明は、ネットワーク上に投稿された情報から有益な発言を分析すること

50

を目的とする。

【課題を解決するための手段】

【0007】

実施形態に係る情報分析システムは、発言分析部と、スレッド分析部と、格納部とを備える。発言分析部は、ネットワーク上に投稿された発言の集合であるスレッドについて、当該スレッドに含まれる発言の重要度を、当該発言に関するデータである発言データに基づいて、当該発言ごとに分析する。スレッド分析部は、前記スレッドが、予め設定された複数のカテゴリのいずれに属するかを、当該スレッドに関するデータであるスレッドデータに基づいて分析する。格納部は、前記発言ごとに、当該発言と、当該発言の重要度と、当該発言が含まれるスレッドのカテゴリとを対応づけて所定の記憶部に格納する。

10

【発明の効果】

【0008】

本願の開示する技術の一つの態様によれば、ネットワーク上に投稿された情報から有益な発言を分析することができるという効果を奏する。

【図面の簡単な説明】

【0009】

【図1】図1は、第1の実施形態に係る情報分析システムの構成の一例を示す図である。

【図2】図2は、第1の実施形態に係るスレッドデータ記憶部に記憶されるスレッドデータの一例を示す図である。

【図3】図3は、第1の実施形態に係る情報分析システムにおける処理の流れを示すフローチャートである。

20

【図4】図4は、第1の実施形態による効果を説明するための図である。

【図5】図5は、第2の実施形態に係る情報分析システムにおける処理の流れを示すフローチャートである。

【図6】図6は、第2の実施形態による効果を説明するための図である。

【図7】図7は、第3の実施形態に係る情報分析システムにおける処理の流れを示すフローチャートである。

【図8】図8は、第3の実施形態による効果を説明するための図である。

【図9】図9は、情報分析プログラムを実行するコンピュータを示す図である。

【発明を実施するための形態】

30

【0010】

以下に添付図面を参照して、この発明に係る情報分析システム、情報分析方法及び情報分析プログラムの実施形態を詳細に説明する。なお、以下の実施形態により本発明が限定されるものではない。

【0011】

[第1の実施形態]

以下の実施形態では、第1の実施形態に係る情報分析システムの構成、情報分析システムにおける処理の流れを順に説明し、最後に第1の実施形態による効果を説明する。

【0012】

[情報分析システムの構成]

40

図1は、第1の実施形態に係る情報分析システム100の構成の一例を示す図である。図1に示すように、情報分析システム100は、通信処理部110、記憶部120、機能部130、及び制御部140を有する。また、情報分析システム100は、インターネットに接続されている。

【0013】

通信処理部110は、ネットワーク上における装置との間でやり取りされ各種情報に関する通信を制御する。例えば、通信処理部110は、後述の収集部131の制御によりインターネット上の各種電子掲示板(BBS(Bulletin Board System))サイトにアクセスする。

【0014】

50

記憶部 1 2 0 は、図 1 に示すように、スレッドデータ記憶部 1 2 1 及び分析結果記憶部 1 2 2 を有する。記憶部 1 2 0 は、例えば、R A M (Random Access Memory)、フラッシュメモリ (Flash Memory) 等の半導体メモリ素子、又は、ハードディスク、光ディスク等の記憶装置などである。

【 0 0 1 5 】

スレッドデータ記憶部 1 2 1 は、インターネット上の B B S サイトから収集されたスレッドデータを記憶する。ここで、スレッドデータは、ネットワーク上に投稿された発言の集合であるスレッドに関するデータである。例えば、スレッドデータ記憶部 1 2 1 は、収集部 1 3 1 によって収集されたスレッドデータを記憶する。また、例えば、スレッドデータ記憶部 1 2 1 は、分析部 1 3 2 によって参照される。

10

【 0 0 1 6 】

図 2 は、第 1 の実施形態に係るスレッドデータ記憶部 1 2 1 に記憶されるスレッドデータの一例を示す図である。図 2 に示すように、例えば、スレッドデータ記憶部 1 2 1 に記憶されるスレッドデータ 1 0 は、タイトル 1 1、発言 1 2、及び発言 1 3 を含む。ここで、タイトル 1 1 は、スレッドのタイトルである。また、発言 1 2、1 3 は、B B S の利用者によって投稿された発言である。発言 1 2、1 3 は、発言順序、発言者名、発言日時、及び発言内容を含む。図 2 に示す例では、発言 1 2 の発言順序が「1」であり、発言者名が「A 1」であり、発言日時が「A 2」であり、発言内容が「A 3」である場合を示す。

【 0 0 1 7 】

分析結果記憶部 1 2 2 は、分析結果を記憶する。例えば、分析結果記憶部 1 2 2 は、後述の分析部 1 3 2 によって分析された分析結果を記憶する。分析結果記憶部 1 2 2 に記憶される分析結果は、例えば、後述の格納部 1 3 3 によって格納される。また、分析結果記憶部 1 2 2 に記憶される分析結果は、利用者の要求に応じて出力される。

20

【 0 0 1 8 】

図 1 の説明に戻る。機能部 1 3 0 は、収集部 1 3 1、分析部 1 3 2、及び格納部 1 3 3 を有する。ここで、機能部 1 3 0 は、各処理を受け持つところであり、実際にはソフトウェア (の 1 コンポーネント) またはミドルウェアとして実現される。また、制御部 1 4 0 は、通信処理部 1 1 0、記憶部 1 2 0、及び機能部 1 3 0 の動作を制御し、情報分析システム 1 0 0 の動作を司るもので、実際には C P U (Central Processing Unit) や M P U (Micro Processing Unit) 等の集積回路等で実現される。

30

【 0 0 1 9 】

収集部 1 3 1 は、ネットワーク上に投稿された情報からスレッドデータを収集する。例えば、収集部 1 3 1 は、インターネット上の B B S サイトを巡回し、B B S サイトからスレッドデータを収集する。そして、収集部 1 3 1 は、収集したスレッドデータをスレッドデータ記憶部 1 2 1 に格納する。

【 0 0 2 0 】

なお、収集部 1 3 1 は、H T M L (HyperText Markup Language) タグやスクリプト、広告等の文字情報以外の情報がスレッドに含まれる場合には、それらの情報を取り除いた上でスレッドデータ記憶部 1 2 1 に格納する。また、収集部 1 3 1 によって巡回される巡回対象のサイトは、予め U R L (Uniform Resource Locator) 等のリストとして指定されていても良いし、収集するごとに指定されても良い。

40

【 0 0 2 1 】

分析部 1 3 2 は、収集部 1 3 1 によって収集されたスレッドデータを分析する。図 1 に示すように、分析部 1 3 2 は、発言切り出し部 1 3 2 A、発言分析部 1 3 2 B、及びスレッド分析部 1 3 2 C を有する。

【 0 0 2 2 】

発言切り出し部 1 3 2 A は、スレッドデータに含まれる発言をスレッドデータから切り出し、発言に関するデータである発言データを生成する。ここで、発言データとは、例えば、「発言順序」、「発言者名」、「発言日時」、「発言内容」、及び「付加情報」が対応づけられたデータである。「発言順序」、「発言者名」、「発言日時」、及び「発言内

50

容」は、スレッドデータから対応する箇所が抜き出されたデータである。また、「付加情報」は、「文字数」、「返信フラグ」、及び「返信数」が対応づけられたデータである。「文字数」は、発言内容の文字数に対応する。「返信フラグ」は、該当する発言が別の発言への返信であるか否かを表すフラグである。例えば、該当する発言の発言内容が「> (数字)」から始まる場合には、その発言が返信であることを示す返信フラグ「1」が付与される。一方、「> (数字)」から始まらない場合には、その発言が返信ではないことを示す返信フラグ「0」が付与される。なお、(数字)は、返信対象の発言順序を表す。「返信数」は、該当する発言への返信の数を表す。例えば、発言内容が「> (数字)」から始まる発言の数が、その数字の発言番号の返信数に対応する。

【0023】

例えば、発言切り出し部132Aは、スレッドデータ記憶部121に記憶されたスレッドデータを取得する。そして、発言切り出し部132Aは、取得したスレッドデータに含まれる発言をそれぞれ切り出し、発言データをそれぞれ生成する。発言切り出し部132Aは、生成したそれぞれの発言データを発言分析部132Bに送る。また、発言切り出し部132Aは、取得したスレッドデータをスレッド分析部132Cに送る。

【0024】

一例として、発言切り出し部132Aが図2に示すスレッドデータ10を取得した場合を説明する。発言切り出し部132Aは、スレッドデータ10から発言12の発言順序「1」、発言者名「A1」、発言日時「A2」、及び発言内容「A3」を切り出す。そして、発言切り出し部132Aは、発言内容「A3」の文字数「A4」を計数する。また、発言切り出し部132Aは、発言内容「A3」が「> (数字)」から始まるか否かに応じて、返信フラグ「A5 (A5は1又は0)」を生成する。また、発言切り出し部132Aは、スレッドデータ10に含まれる発言のうち、「> 1」から始まる発言の数「A6」を計数する。この結果、発言切り出し部132Aは、発言順序「1」、発言者名「A1」、発言日時「A2」、発言内容「A3」、文字数「A4」、返信フラグ「A5」、及び返信数「A6」が対応づけられたデータを、発言12の発言データとして生成する。このように、発言切り出し部132Aは、スレッドデータ10に含まれる全ての発言について、発言データを生成する。

【0025】

発言分析部132Bは、スレッドに含まれる発言の重要度を、発言データに基づいて、発言ごとに分析する。例えば、発言分析部132Bは、発言データから発言の重要度(スコア)を算出するための機械学習エンジンを用いて、発言切り出し部132Aによって生成された発言データから発言の重要度を分析する。そして、発言分析部132Bは、分析結果として、発言の重要度を格納部133に送る。なお、この場合、機械学習エンジンには、重要な発言の特徴ベクトルと重要でない発言の特徴ベクトルとを予め学習させておく。学習させる方法としては、例えば、代表的なBBSサイトからいくつかのスレッドを選択し、スレッドに含まれる発言を重要な発言と重要でない発言とに人手で分類して学習させる方法が挙げられる。

【0026】

具体的には、発言分析部132Bは、発言データを機械学習エンジンにかけるために、各発言データを特徴ベクトルにそれぞれ変換する。ここで、発言分析部132Bは、例えば、発言データのうち、数値データと文字列データとをそれぞれ個別に変換する。例えば、発言分析部132Bは、発言データのうち、発言順序、発言日時、文字数、返信フラグ、及び返信数等の数値データを、そのまま特徴ベクトルの値とする。なお、これに限らず、発言分析部132Bは、数値データに対して何らかの数値演算処理を施して、特徴ベクトルの値としても良い。

【0027】

また、発言分析部132Bは、発言者名及び発言内容等の文字列データに対しては、形態素解析、n-gram、若しくは区切り文字を用いて、特徴ベクトルに変換する。例えば、発言分析部132Bは、形態素解析を用いた場合には、文字列データを品詞によって

10

20

30

40

50

分割し、それらを特徴ベクトルへ変換する。このような形態素解析には、例えばオープンソースの *Mecab* 等のライブラリを利用することができる。例えば、発言分析部 132 B は、文字列データが「*Twitter* (登録商標) の使い方が、まだ、よくわからん。」である場合には、形態素解析を用いて、「*Twitter* / の / 使い方 / が / 、 / まだ / 、 / よく / わから / ん / 。

【0028】

また、発言分析部 132 B は、*n-gram* を用いた場合には、文字列データを先頭から 1 文字ずつずらしながら *n* 文字の組を作り、それらを特徴ベクトルへ変換する。例えば、発言分析部 132 B は、文字列データが「*Twitter* の使い方が、まだ、よくわからん。」である場合には、*n-gram* ($n=3$) を用いて、「*Twi* / *wit* / *itt* / *tte* / *ter* / *er* の / *r* の使 / の使 / …」と分割する。

10

【0029】

また、発言分析部 132 B は、区切り文字を用いた場合には、文字列データを別途定めた区切り文字 (空白やカンマ “,” など) によって分割し、それらを特徴ベクトルへ変換する。例えば、発言分析部 132 B は、文字列データが「*Twitter* の使い方が、まだ、よくわからん。」であり、区切り文字が句読点 “、” である場合には、区切り文字を用いて、「*Twitter* の使い方が / まだ / よくわからん。」と分割する。なお、一般に、形態素解析は日本語の文章に、空白区切りは英語に適用されることが多い。

【0030】

そして、発言分析部 132 B は、このようにして要素に分割された文字列データを特徴ベクトルに変換する。特徴ベクトルの変換手法としては、例えば、各要素の出現回数をそのまま特徴ベクトルとする方式、回数によらず出現するか否かを 1 又は 0 に対応させる方式等がある。これらは使用する機械学習エンジンのライブラリが対応する手法であれば、どのようなものを用いても構わない。

20

【0031】

そして、発言分析部 132 B は、各発言データから変換した各特徴ベクトルを、機械学習エンジンに与え、各発言の重要度を算出する。ここで使用する機械学習エンジンは種類を問わないが、例えば、オープンソースの *Jubatus* (登録商標) 等を用いることができる。そして、発言分析部 132 B は、各発言データに対して、機械学習エンジンで算出された各発言の重要度を付与し、発言データと発言の重要度とをそれぞれ合わせて格納部 133 に送る。

30

【0032】

なお、上述した発言分析部 132 B の処理はあくまで一例に過ぎない。例えば、発言分析部 132 B は、機械学習エンジンによる分析の精度を向上させるために、入力される各発言データに対して、不要文字の除去、文字種の統一、及び特定グループの単語の置き換え等の前処理を行っても良い。具体的には、発言分析部 132 B は、不要文字の除去として、例えば、文字列データに対して、余計な空白や言語処理の障害となる URL 等の削除を行う。また、例えば、発言分析部 132 B は、文字種の統一として、文字列データに使用されている文字について、英大文字小文字やいわゆる半角全角の統一を行う。また、例えば、発言分析部 132 B は、特定グループの単語の置き換えとして、ガラケー、スマホ、スマートフォン、及びケータイという単語を携帯電話という単語に置き換える。また、例えば、特徴ベクトルの基となる文字列データは、発言内容のみを用いても良いし、“発言者と発言内容” のように、いくつかの要素を連結した文字列データを用いても良い。

40

【0033】

スレッド分析部 132 C は、スレッドが、予め設定された複数のカテゴリのいずれに属するかを、スレッドデータに基づいて分析する。例えば、スレッド分析部 132 C は、スレッドデータからスレッドのカテゴリを分析するための機械学習エンジンを用いて、発言切り出し部 132 A から受け付けたスレッドデータからスレッドが属するカテゴリを分析する。そして、スレッド分析部 132 C は、分析結果として、スレッドが属するカテゴリを格納部 133 に送る。なお、この場合、機械学習エンジンには、いくつかのスレッドと

50

そのカテゴリを予め学習させておく。学習させる方法としては、例えば、代表的な BBS サイトからいくつかのスレッドを選択し、スレッドが属するカテゴリを手入力して学習させる方法が挙げられる。

【0034】

具体的には、スレッド分析部 132C は、スレッドデータを機械学習エンジンにかけるために、スレッドデータを特徴ベクトルに変換する。ここで、スレッド分析部 132C は、例えば、スレッドデータに含まれる文字列データ、例えば、タイトルと、各発言の発言内容とを連結した文字列データに対して、形態素解析、n-gram、若しくは区切り文字を用いて、特徴ベクトルに変換する。なお、形態素解析、n-gram、及び区切り文字を用いて特徴ベクトルに変換する処理の説明は、上述した説明と同様であるので省略する。

10

【0035】

そして、スレッド分析部 132C は、スレッドデータから変換した特徴ベクトルを機械学習エンジンに与え、スレッドが属するカテゴリを決定する。ここで使用する機械学習エンジンは種類を問わないが、例えば、オープンソースの Juba tus 等を用いることができる。そして、スレッド分析部 132C は、機械学習エンジンで決定されたスレッドのカテゴリを格納部 133 に送る。

【0036】

なお、上述したスレッド分析部 132C の処理はあくまで一例に過ぎない。例えば、スレッド分析部 132C は、機械学習エンジンによる分析の精度を向上させるために、入力される各発言データに対して、不要文字の除去、文字種の統一、及び特定グループの単語の置き換え等の前処理を行っても良い。また、例えば、スレッド分析部 132C は、特徴ベクトルの基となる文字列データとして、タイトルと各発言の発言内容以外にも、発言者等の他の要素を組み合わせ用いても良い。また、スレッド分析部 132C は、文字列データに限らず、発言数等の数値データを用いて特徴ベクトルを求めても良い。

20

【0037】

格納部 133 は、発言ごとに、当該発言と、発言の重要度と、発言が含まれるスレッドのカテゴリとを対応づけて分析結果記憶部 122 に格納する。例えば、格納部 133 は、発言ごとに、発言データ及びその重要度を発言分析部 132B から受け付ける。また、格納部 133 は、スレッドが属するカテゴリをスレッド分析部 132C から受け付ける。そして、格納部 133 は、発言分析部 132B から受け付けた発言データ及びその重要度の組に、スレッド分析部 132C から受け付けたスレッドのカテゴリを発言のカテゴリとして付与し、分析結果記憶部 122 に格納する。

30

【0038】

なお、格納部 133 によって分析結果記憶部 122 に格納された情報は、例えば、「カテゴリ」をキーとして検索されて、「重要度」のスコア順にソートされて利用者に提示される。また、特定のスコア（重要度）以上のデータのみが提示されるようにしても良い。

【0039】

[情報分析システムによる処理]

図 3 は、第 1 の実施形態に係る情報分析システムにおける処理の流れを示すフローチャートである。

40

【0040】

図 3 に示すように、情報分析システム 100 の収集部 131 は、インターネット上の BBS サイトからスレッドデータを収集する（ステップ S101）。収集部 131 は、収集したスレッドデータをスレッドデータ記憶部 121 に格納する。

【0041】

続いて、発言切り出し部 132A は、スレッドデータから発言を切り出し、発言データを生成する（ステップ S102）。例えば、発言切り出し部 132A は、スレッドデータ記憶部 121 に記憶されたスレッドデータを取得して、取得したスレッドデータに含まれる発言をそれぞれ切り出し、発言データをそれぞれ生成する。

50

【 0 0 4 2 】

そして、発言分析部 1 3 2 B は、スレッドに含まれる発言の重要度を発言ごとに分析する（ステップ S 1 0 3）。例えば、発言分析部 1 3 2 B は、発言データから発言の重要度（スコア）を算出するための機械学習エンジンを用いて、発言切り出し部 1 3 2 A によって生成された発言データから発言の重要度を分析する。

【 0 0 4 3 】

そして、スレッド分析部 1 3 2 C は、スレッドのカテゴリを分析する（ステップ S 1 0 4）。例えば、スレッド分析部 1 3 2 C は、スレッドデータからスレッドのカテゴリを分析するための機械学習エンジンを用いて、発言切り出し部 1 3 2 A から受け付けたスレッドデータからスレッドが属するカテゴリを分析する。

10

【 0 0 4 4 】

そして、格納部 1 3 3 は、発言ごとに、発言データ、発言の重要度、及びスレッドのカテゴリを対応づけて分析結果記憶部 1 2 2 に格納する（ステップ S 1 0 5）。

【 0 0 4 5 】

なお、上述した処理手順は、必ずしも上述した順序で実行されなくても良い。例えば、ステップ S 1 0 3 の処理とステップ S 1 0 4 の処理は、上述した順序とは逆の順序で実行されても良いし、並行処理として実行されても良い。また、例えば、収集済みのスレッドデータが存在する場合には、ステップ S 1 0 1 の処理は実行されなくても良い。

【 0 0 4 6 】

[第 1 の実施形態の効果]

20

上述してきたように、第 1 の実施形態に係る情報分析システム 1 0 0 は、ネットワーク上に投稿された発言の集合であるスレッドについて、スレッドに含まれる発言の重要度を、発言データに基づいて、発言ごとに分析する。そして、情報分析システム 1 0 0 は、スレッドが、予め設定された複数のカテゴリのいずれに属するかを、スレッドデータに基づいて分析する。そして、情報分析システム 1 0 0 は、発言ごとに、発言と、発言の重要度と、発言が含まれるスレッドのカテゴリとを対応づけて所定の記憶部に格納する。このため、情報分析システム 1 0 0 は、ネットワーク上に投稿された情報から有益な発言を分析することができる。

【 0 0 4 7 】

図 4 は、第 1 の実施形態による効果を説明するための図である。図 4 に示すように、例えば、情報分析システム 1 0 0 において、収集部 1 3 1 は、インターネット上の B B S サイトからスレッドデータを収集し（S 1 0）、収集したスレッドデータを発言切り出し部 1 3 2 A へ送る（S 1 1）。続いて、発言切り出し部 1 3 2 A は、スレッドデータから発言を切り出して発言データを生成し、生成した発言データを発言分析部 1 3 2 B へ送る（S 1 2）。そして、発言分析部 1 3 2 B は、発言の重要度を発言ごとに分析し、発言データと重要度の組を格納部 1 3 3 へ送る（S 1 3）。続いて、発言切り出し部 1 3 2 A は、スレッドデータをスレッド分析部 1 3 2 C へ送る（S 1 4）。そして、スレッド分析部 1 3 2 C は、スレッドのカテゴリを分析し、分析したカテゴリを格納部 1 3 3 へ送る（S 1 5）。そして、格納部 1 3 3 は、発言ごとに、発言データ、重要度、及びカテゴリを対応づけて分析結果記憶部 1 2 2 に格納する（S 1 6）。このように、情報分析システム 1 0 0 は、収集したスレッドデータに含まれる発言ごとに、発言データと、重要度と、カテゴリとを対応づけて蓄積する。このため、情報分析システム 1 0 0 は、例えば、意味の無い発言が多く含まれる場合であっても、有益な発言が意味の無い発言に埋もれることなく検索可能な状態で、分析結果を蓄積することができる。また、情報分析システム 1 0 0 は、スレッド単位でカテゴリの分析を行うことにより、発言単位で分析する場合と比較して分析対象とする情報量を増加させるので、発言の属するカテゴリを精度良く分析することができる。この結果、情報推薦技術において、利用者は、スレッド単位で抽出・推薦される場合と比べて、無意味な発言が除かれた、真に有用な発言のみを得ることが可能となる。

30

40

【 0 0 4 8 】

[第 2 の実施形態]

50

第1の実施形態では、情報分析システム100が発言ごとに重要度を分析する場合を説明したが、実施形態はこれに限定されるものではない。例えば、情報分析システム100は、スレッドの重要度についても分析し、発言の重要度とスレッドの重要度とを統合する場合であっても良い。そこで、第2の実施形態では、情報分析システム100が、スレッドの重要度についても分析し、発言の重要度とスレッドの重要度とを統合する場合を説明する。なお、以下の説明では、第1の実施形態と共通する構成及び処理については、説明を省略する。

【0049】

図1を用いて、第2の実施形態に係る情報分析システム100の構成について説明する。第2の実施形態に係る情報分析システム100は、図1に示した情報分析システム100と同様の構成を備えるが、スレッド分析部132C及び格納部133における処理が一部相違する。

10

【0050】

スレッド分析部132Cは、第1の実施形態において説明した処理に加えて、スレッドデータに基づいて、スレッドの重要度を分析する。例えば、スレッド分析部132Cは、スレッドデータからスレッドの重要度を算出するための機械学習エンジンを用いて、スレッドの重要度を分析する。そして、スレッド分析部132Cは、分析結果として、スレッドが属するカテゴリに加えて、スレッドの重要度を格納部133に送る。なお、この場合、機械学習エンジンには、重要なスレッドの特徴ベクトルと重要でないスレッドの特徴ベクトルとを予め学習させておく。学習させる方法としては、例えば、代表的なBBSサイトからいくつかのスレッドを選択し、重要なスレッドと重要でないスレッドとを手で分類して学習させる方法が挙げられる。

20

【0051】

具体的には、スレッド分析部132Cは、スレッドデータを機械学習エンジンにかけるために、スレッドデータを特徴ベクトルに変換する。ここで、スレッド分析部132Cは、例えば、スレッドデータに含まれる文字列データ、例えば、タイトルと、各発言の発言内容とを連結した文字列データに対して、形態素解析、n-gram、若しくは区切り文字を用いて、特徴ベクトルに変換する。なお、形態素解析、n-gram、及び区切り文字を用いて特徴ベクトルに変換する処理の説明は、上述した説明と同様であるので省略する。

30

【0052】

そして、スレッド分析部132Cは、スレッドデータから変換した特徴ベクトルを機械学習エンジンに与え、スレッドの重要度を算出する。ここで使用する機械学習エンジンは種類を問わないが、例えば、オープンソースのJubatus等を用いることができる。そして、スレッド分析部132Cは、機械学習エンジンで算出されたスレッドの重要度を格納部133に送る。

【0053】

なお、上述したスレッド分析部132Cの処理はあくまで一例に過ぎない。例えば、スレッド分析部132Cは、文字列データに限らず、スレッドにおける発言数等の数値データを用いて特徴ベクトルを求めても良い。

40

【0054】

格納部133は、発言の重要度及びスレッドの重要度に基づいて、発言の総合重要度を算出する。そして、格納部133は、算出した総合重要度と、発言と、発言が含まれるスレッドのカテゴリとを発言ごとに対応づけて格納する。

【0055】

例えば、格納部133は、発言ごとに、発言データ及び発言の重要度を発言分析部132Bから受け付ける。また、格納部133は、スレッドが属するカテゴリ及びスレッドの重要度をスレッド分析部132Cから受け付ける。そして、格納部133は、発言分析部132Bから受け付けた発言の重要度と、スレッド分析部132Cから受け付けたスレッドの重要度とに基づいて、総合重要度を発言ごとに算出する。ここで、総合重要度の算出

50

方法は、例えば、発言の重要度及びスレッドの重要度の和であっても良いし、積、平均、その他の算術演算による値であっても良い。そして、格納部 1 3 3 は、発言データ、総合重要度、及びカテゴリを対応づけて、発言ごとに分析結果記憶部 1 2 2 に格納する。

【 0 0 5 6 】

図 5 は、第 2 の実施形態に係る情報分析システムにおける処理の流れを示すフローチャートである。なお、図 5 に示すステップ S 2 0 1 ~ S 2 0 3 の処理は、図 3 のステップ S 1 0 1 ~ S 1 0 3 の処理と同様であるので、説明を省略する。

【 0 0 5 7 】

図 5 に示すように、情報分析システム 1 0 0 のスレッド分析部 1 3 2 C は、スレッドのカテゴリ及びスレッドの重要度を分析する（ステップ S 2 0 4）。例えば、スレッド分析部 1 3 2 C は、第 1 の実施形態において説明した処理に加えて、スレッドデータからスレッドの重要度を算出するための機械学習エンジンを用いて、スレッドの重要度を分析する。そして、スレッド分析部 1 3 2 C は、分析結果として、スレッドが属するカテゴリに加えて、スレッドの重要度を格納部 1 3 3 に送る。

10

【 0 0 5 8 】

そして、格納部 1 3 3 は、発言の重要度とスレッドの重要度とから、発言の総合重要度を算出する（ステップ S 2 0 5）。例えば、格納部 1 3 3 は、発言の重要度及びスレッドの重要度の和、積、平均、その他の算術演算による値等を総合重要度として算出する。そして、格納部 1 3 3 は、発言データ、発言の総合重要度、及びスレッドのカテゴリを対応づけて、発言ごとに分析結果記憶部 1 2 2 に格納する（ステップ S 2 0 6）。

20

【 0 0 5 9 】

なお、図 5 に示した処理手順は、必ずしも上記の順序で実行されなくても良い。例えば、ステップ S 2 0 3 の処理とステップ S 2 0 4 の処理は、上述した順序とは逆の順序で実行されても良いし、並行処理として実行されても良い。

【 0 0 6 0 】

このように、第 2 の実施形態に係る情報分析システム 1 0 0 は、スレッドの重要度についても分析し、発言の重要度とスレッドの重要度とを統合する。これにより、情報分析システム 1 0 0 は、スレッドの重要度を加味した上で発言ごとの重要度を算出することができるので、より有益な発言を分析することができる。

【 0 0 6 1 】

図 6 は、第 2 の実施形態による効果を説明するための図である。図 6 に示すように、例えば、情報分析システム 1 0 0 において、収集部 1 3 1 は、インターネット上の B B S サイトからスレッドデータを収集し（S 2 0）、収集したスレッドデータを発言切り出し部 1 3 2 A へ送る（S 2 1）。続いて、発言切り出し部 1 3 2 A は、スレッドデータから発言を切り出して発言データを生成し、生成した発言データを発言分析部 1 3 2 B へ送る（S 2 2）。そして、発言分析部 1 3 2 B は、発言の重要度を発言ごとに分析し、発言データと重要度の組を格納部 1 3 3 へ送る（S 2 3）。続いて、発言切り出し部 1 3 2 A は、スレッドデータをスレッド分析部 1 3 2 C へ送る（S 2 4）。そして、スレッド分析部 1 3 2 C は、スレッドのカテゴリ及びスレッドの重要度を分析し、分析したカテゴリ及びスレッドの重要度を格納部 1 3 3 へ送る（S 2 5）。そして、格納部 1 3 3 は、発言の重要度とスレッドの重要度とから、発言の総合重要度を算出し、発言データ、発言の総合重要度、及びカテゴリを対応づけて、発言ごとに分析結果記憶部 1 2 2 に格納する（S 2 6）。これにより、第 2 の実施形態に係る情報分析システム 1 0 0 は、より有益な発言を分析することができる。

30

40

【 0 0 6 2 】

[第 3 の実施形態]

第 1 及び第 2 の実施形態では、情報分析システム 1 0 0 がスレッドのカテゴリを分析する場合を説明したが、実施形態はこれに限定されるものではない。例えば、情報分析システム 1 0 0 は、発言の内容をカテゴリ分類した発言カテゴリについても分析し、それぞれの発言に付与する場合であっても良い。なお、発言カテゴリとは、例えば、「好意的」、

50

「敵対的」、「賞賛」、「冒とく」といった発言者の感情を推測するものや、「引用」、「提案」、「質問」、「回答」といった発言の種類を推測するもの等が含まれる。すなわち、発言カテゴリは、発言の内容が、予め設定された複数のカテゴリのいずれに属するかを示す情報である。

【 0 0 6 3 】

そこで、第3の実施形態では、情報分析システム100が、発言カテゴリを分析し、それぞれの発言に付与する場合を説明する。なお、以下の説明では、第2の実施形態と共通する構成及び処理については、説明を省略する。

【 0 0 6 4 】

図1を用いて、第3の実施形態に係る情報分析システム100の構成について説明する。第3の実施形態に係る情報分析システム100は、第2の実施形態に係る情報分析システム100と同様の構成を備えるが、発言分析部132B及び格納部133における処理が一部相違する。

10

【 0 0 6 5 】

発言分析部132Bは、第1の実施形態において説明した処理に加えて、発言カテゴリを、発言データに基づいて、発言ごとに分析する。例えば、発言分析部132Bは、発言データから発言カテゴリを導出するための機械学習エンジンを用いて、発言カテゴリを分析する。そして、発言分析部132Bは、分析結果として、発言の重要度に加えて、発言カテゴリを格納部133に送る。なお、この場合、機械学習エンジンには、いくつかの発言と発言カテゴリとの組を予め学習させておく。学習させる方法としては、例えば、代表的なBBSサイトからいくつかのスレッドを選択し、スレッドに含まれる発言を「好意的」、「敵対的」、「賞賛」、「冒とく」、「引用」、「提案」、「質問」、「回答」といったカテゴリに人手で分類して学習させる方法が挙げられる。

20

【 0 0 6 6 】

具体的には、発言分析部132Bは、第1の実施形態に係る処理において変換された特徴ベクトルを機械学習エンジンに与え、発言ごとに発言カテゴリを算出する。ここで使用する機械学習エンジンは種類を問わないが、例えば、オープンソースのJubatus等を用いることができる。そして、発言分析部132Bは、機械学習エンジンで算出されたスレッドの重要度を格納部133に送る。

【 0 0 6 7 】

格納部133は、発言ごとの発言カテゴリを、当該発言ごとに対応づけて格納する。例えば、格納部133は、発言データ、総合重要度、スレッドのカテゴリ、及び発言カテゴリを対応づけて、発言ごとに分析結果記憶部122に格納する。

30

【 0 0 6 8 】

図7は、第3の実施形態に係る情報分析システムにおける処理の流れを示すフローチャートである。なお、図7に示すステップS301～S302の処理は、図3のステップS101～S102の処理と同様であるので、説明を省略する。

【 0 0 6 9 】

図7に示すように、情報分析システム100の発言分析部132Bは、発言の重要度及び発言カテゴリを分析する(ステップS303)。例えば、発言分析部132Bは、発言データから発言カテゴリを導出するための機械学習エンジンを用いて、発言カテゴリを分析する。そして、発言分析部132Bは、分析結果として、発言の重要度に加えて、発言カテゴリを格納部133に送る。

40

【 0 0 7 0 】

続いて、スレッド分析部132Cは、スレッドのカテゴリ及びスレッドの重要度を分析する(ステップS304)。例えば、発言分析部132Bは、第1の実施形態において説明した処理に加えて、スレッドデータからスレッドの重要度を算出するための機械学習エンジンを用いて、スレッドの重要度を分析する。そして、発言分析部132Bは、分析結果として、スレッドが属するカテゴリに加えて、スレッドの重要度を格納部133に送る。

50

【 0 0 7 1 】

そして、格納部 1 3 3 は、発言の重要度とスレッドの重要度とから、発言の総合重要度を算出する（ステップ S 3 0 5）。例えば、格納部 1 3 3 は、発言の重要度及びスレッドの重要度の和、積、平均、その他の算術演算による値等を総合重要度として算出する。そして、格納部 1 3 3 は、発言データ、発言の総合重要度、発言カテゴリ、及びスレッドのカテゴリを対応づけて、発言ごとに分析結果記憶部 1 2 2 に格納する（ステップ S 3 0 6）。

【 0 0 7 2 】

なお、図 7 に示した処理手順は、必ずしも上記の順序で実行されなくても良い。例えば、ステップ S 3 0 3 の処理とステップ S 3 0 4 の処理は、上述した順序とは逆の順序で実行されても良いし、並行処理として実行されても良い。

10

【 0 0 7 3 】

このように、第 3 の実施形態に係る情報分析システム 1 0 0 は、発言カテゴリを分析し、それぞれの発言に付与する。これにより、情報分析システム 1 0 0 は、スレッドのカテゴリに加え、発言ごとの発言カテゴリも踏まえて分析するので、より有益な発言を分析することができる。

【 0 0 7 4 】

図 8 は、第 3 の実施形態による効果を説明するための図である。図 8 に示すように、例えば、情報分析システム 1 0 0 において、収集部 1 3 1 は、インターネット上の B B S サイトからスレッドデータを収集し（S 3 0）、収集したスレッドデータを発言切り出し部 1 3 2 A へ送る（S 3 1）。続いて、発言切り出し部 1 3 2 A は、スレッドデータから発言を切り出して発言データを生成し、生成した発言データを発言分析部 1 3 2 B へ送る（S 3 2）。そして、発言分析部 1 3 2 B は、発言の重要度及び発言カテゴリを発言ごとに分析し、発言データ、発言の重要度、及び発言カテゴリの組を格納部 1 3 3 へ送る（S 3 3）。続いて、発言切り出し部 1 3 2 A は、スレッドデータをスレッド分析部 1 3 2 C へ送る（S 3 4）。そして、スレッド分析部 1 3 2 C は、スレッドのカテゴリ及びスレッドの重要度を分析し、分析したカテゴリ及びスレッドの重要度を格納部 1 3 3 へ送る（S 3 5）。そして、格納部 1 3 3 は、発言の重要度とスレッドの重要度とから、発言の総合重要度を算出し、発言データ、発言の総合重要度、発言カテゴリ、及びスレッドのカテゴリを対応づけて、発言ごとに分析結果記憶部 1 2 2 に格納する（S 3 6）。これにより、第 3 の実施形態に係る情報分析システム 1 0 0 は、より有益な発言を分析することができる。

20

30

【 0 0 7 5 】

なお、第 3 の実施形態では、第 2 の実施形態に係る情報分析システム 1 0 0 が、更に、発言カテゴリを分析し、それぞれの発言に付与する場合を説明したが、これに限定されるものではない。例えば、第 1 の実施形態に係る情報分析システム 1 0 0 が、発言カテゴリを分析し、それぞれの発言に付与する場合であっても良い。すなわち、第 3 の実施形態において、スレッドの重要度を算出する処理は実行されなくても良い。

【 0 0 7 6 】

[システム構成等]

また、図示した各装置の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。すなわち、各装置の分散・統合の具体的形態は図示のものに限られず、その全部または一部を、各種の負荷や使用状況などに応じて、任意の単位で機能的または物理的に分散・統合して構成することができる。例えば、格納部 1 3 3 は、分析部 1 3 2 と統合しても良いし、発言分析部 1 3 2 B 若しくはスレッド分析部 1 3 2 C と統合しても良い。更に、各装置にて行なわれる各処理機能は、その全部または任意の一部が、C P U および当該 C P U にて解析実行されるプログラムにて実現され、あるいは、ワイヤードロジックによるハードウェアとして実現され得る。

40

【 0 0 7 7 】

また、本実施例において説明した各処理のうち、自動的におこなわれるものとして説明

50

した処理の全部または一部を手動的に行うこともでき、あるいは、手動的におこなわれるものとして説明した処理の全部または一部を公知の方法で自動的にこなうこともできる。この他、上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種のデータやパラメータを含む情報については、特記する場合を除いて任意に変更することができる。

【0078】

[処理対象]

また、上述した第1～第3の実施形態では、処理対象としてBBSのスレッドデータが分析される場合を説明したが、実施形態はこれに限定されるものではない。例えば、情報分析システム100は、電子メールがスレッド形式(所定の電子メールに対する返信が連なって表示される形式)で表示されたものや、SNS(Social Networking Service)等における所定の投稿に対して返信された一連の発言についても処理対象とすることができる。なお、電子メールを対象とする場合、返信の判定や返信数のカウントには、発言冒頭の「>(数字)」に着目する方法に代えて、電子メールの「In-Reply-To」ヘッダに着目して返信の判定や返信数のカウントを行うことができる。他の処理対象についても同様に、処理対象に応じた返信の判定法や返信数のカウント方法を規定することができる。

10

【0079】

[プログラム]

また、上記実施形態において説明した情報分析システム100が実行する処理をコンピュータが実行可能な言語で記述したプログラムを作成することもできる。例えば、第1の実施形態に係る情報分析システム100が実行する処理をコンピュータが実行可能な言語で記述した情報分析プログラムを作成することもできる。この場合、コンピュータが情報分析プログラムを実行することにより、上記実施形態と同様の効果を得ることができる。さらに、かかる情報分析プログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録された情報分析プログラムをコンピュータに読み込ませて実行することにより上記第1の実施形態と同様の処理を実現してもよい。以下に、図1に示した情報分析システム100と同様の機能を実現する情報分析プログラムを実行するコンピュータの一例を説明する。

20

【0080】

図9は、情報分析プログラムを実行するコンピュータ1000を示す図である。図9に例示するように、コンピュータ1000は、例えば、メモリ1010と、CPU1020と、ハードディスクドライブインタフェース1030と、ディスクドライブインタフェース1040と、ネットワークインタフェース1070とを有し、これらの各部はバス1080によって接続される。

30

【0081】

メモリ1010は、図9に例示するように、ROM(Read Only Memory)1011及びRAM(Random Access Memory)1012を含む。ROM1011は、例えば、BIOS(Basic Input Output System)等のブートプログラムを記憶する。ハードディスクドライブインタフェース1030は、図9に例示するように、ハードディスクドライブ1031に接続される。ディスクドライブインタフェース1040は、図9に例示するように、ディスクドライブ1041に接続される。例えば磁気ディスクや光ディスク等の着脱可能な記憶媒体が、ディスクドライブに挿入される。

40

【0082】

ここで、図9に例示するように、ハードディスクドライブ1031は、例えば、OS1091、アプリケーションプログラム1092、プログラムモジュール1093、プログラムデータ1094を記憶する。すなわち、上記のいずれかのプログラムは、コンピュータ1000によって実行される指令が記述されたプログラムモジュールとして、例えばハードディスクドライブ1031に記憶される。

【0083】

50

また、上記実施形態で説明した各種データは、プログラムデータとして、例えばメモリ 1010 やハードディスクドライブ 1031 に記憶される。そして、CPU 1020 が、メモリ 1010 やハードディスクドライブ 1031 に記憶されたプログラムモジュール 1093 やプログラムデータ 1094 を必要に応じて RAM 1012 に読み出し、各手順を実行する。

【0084】

なお、情報分析プログラムに係るプログラムモジュール 1093 やプログラムデータ 1094 は、ハードディスクドライブ 1031 に記憶される場合に限られず、例えば着脱可能な記憶媒体にそれぞれ記憶され、ディスクドライブ等を介して CPU 1020 によって読み出されてもよい。あるいは、情報分析プログラムに係るプログラムモジュール 1093 やプログラムデータ 1094 は、ネットワーク（LAN (Local Area Network)、WAN (Wide Area Network) 等）を介して接続された他のコンピュータにそれぞれ記憶され、ネットワークインタフェース 1070 を介して CPU 1020 によって読み出されてもよい。

10

【0085】

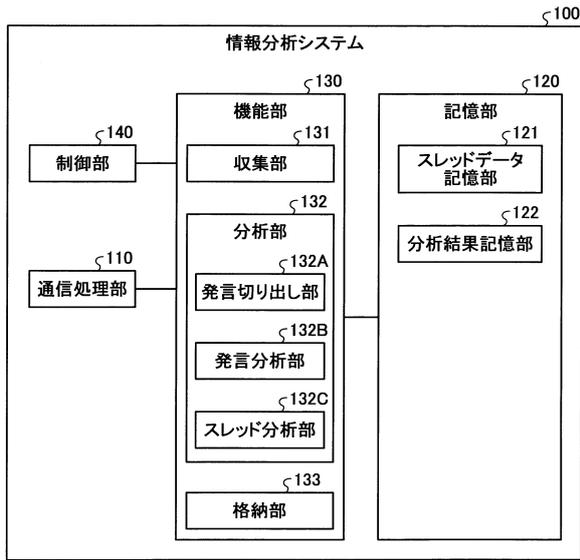
これらの実施形態やその変形は、本願が開示する技術に含まれると同様に、特許請求の範囲に記載された発明とその均等の範囲に含まれるものである。

【符号の説明】

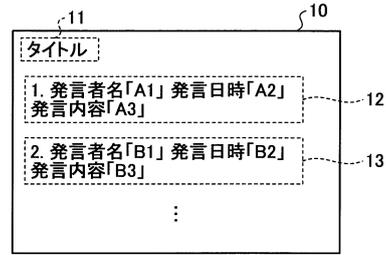
【0086】

100	情報分析システム	20
110	通信処理部	
120	記憶部	
121	スレッドデータ記憶部	
122	分析結果記憶部	
130	機能部	
131	収集部	
132	分析部	
132A	発言切り出し部	
132B	発言分析部	
132C	スレッド分析部	30
133	格納部	
140	制御部	

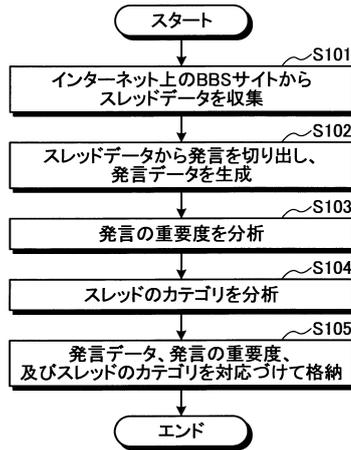
【図1】



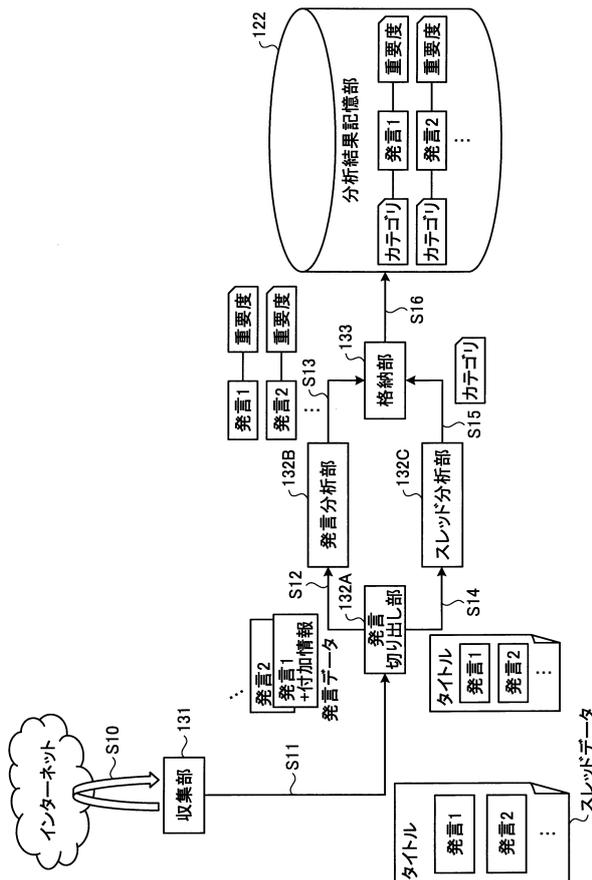
【図2】



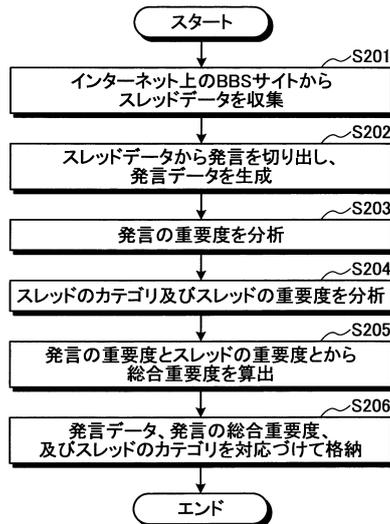
【図3】



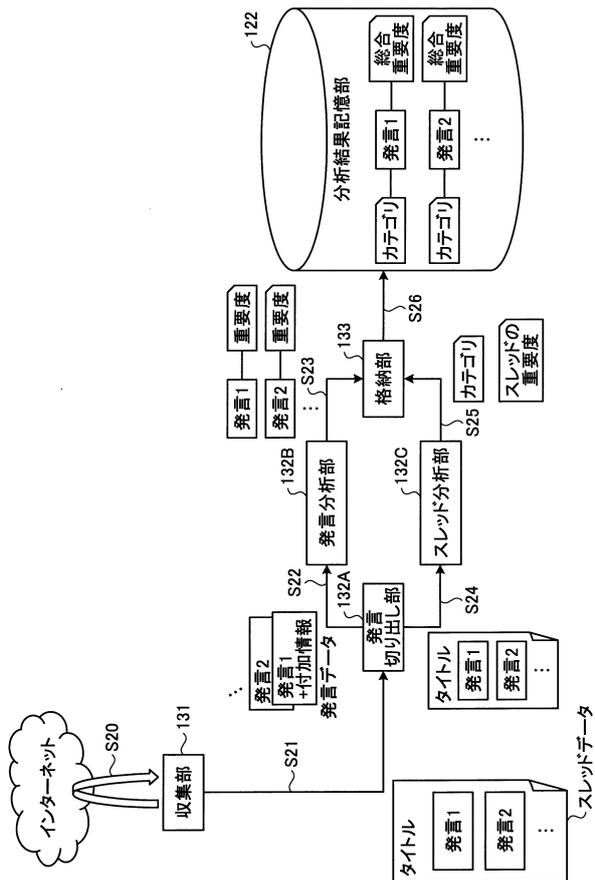
【図4】



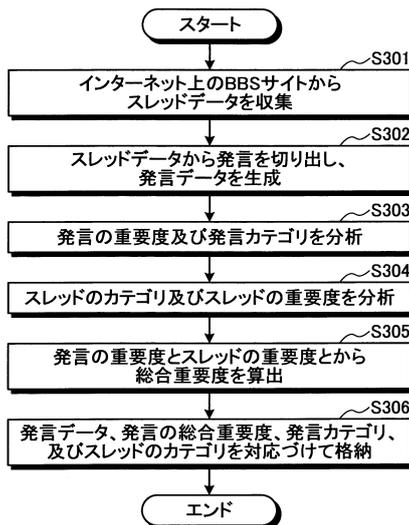
【図5】



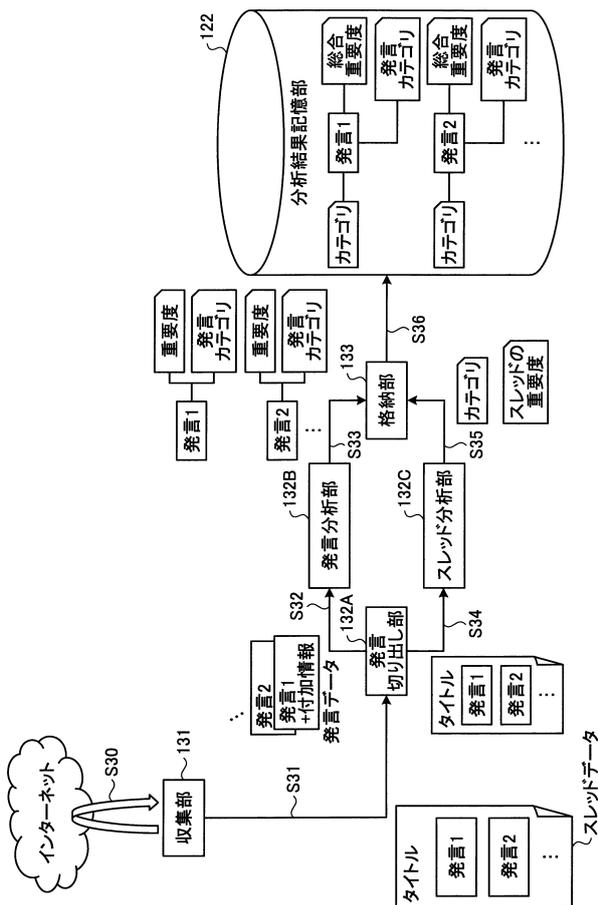
【図6】



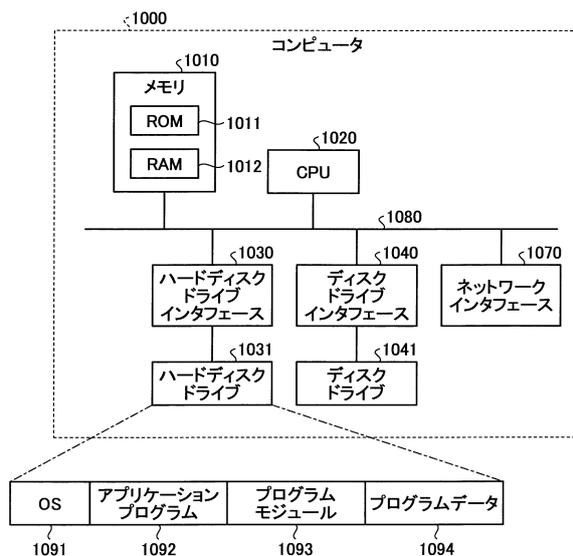
【図7】



【図8】



【図9】



フロントページの続き

審査官 小太刀 慶明

- (56)参考文献 特開2008-176721(JP,A)
特開2007-328610(JP,A)
特開2007-58755(JP,A)
国際公開第2003/046764(WO,A1)
米国特許出願公開第2012/0095952(US,A1)
米国特許出願公開第2007/323412(US,A1)

- (58)調査した分野(Int.Cl., DB名)
- | | |
|------|-------|
| G06F | 17/30 |
| G06Q | 50/00 |
| G06F | 13/00 |