

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6675868号
(P6675868)

(45) 発行日 令和2年4月8日(2020.4.8)

(24) 登録日 令和2年3月13日(2020.3.13)

(51) Int.Cl.

F 1

G06F 16/906	(2019.01)	G06F 16/906	16/906
G06F 16/954	(2019.01)	G06F 16/954	
G06F 13/00	(2006.01)	G06F 13/00	5 6 0 C
G06Q 30/02	(2012.01)	G06Q 30/02	3 1 2

請求項の数 11 (全 16 頁)

(21) 出願番号

特願2015-249682 (P2015-249682)

(22) 出願日

平成27年12月22日 (2015.12.22)

(65) 公開番号

特開2017-117068 (P2017-117068A)

(43) 公開日

平成29年6月29日 (2017.6.29)

審査請求日

平成30年12月21日 (2018.12.21)

(73) 特許権者 390002761

キヤノンマーケティングジャパン株式会社
東京都港区港南2丁目16番6号

(73) 特許権者 592135203

キヤノンITソリューションズ株式会社
東京都港区港南2丁目16番6号

(74) 代理人 100096091

弁理士 井上 誠一

(72) 発明者 蔵満 琢麻

東京都品川区東品川2丁目4番11号 キ
ヤノンITソリューションズ株式会社内

審査官 後藤 彰

最終頁に続く

(54) 【発明の名称】情報処理装置、情報処理方法、及びプログラム

(57) 【特許請求の範囲】

【請求項 1】

特定の対象に関する1以上の投稿を抽出する抽出手段と、
抽出された各投稿を投稿の種別ごとに集計する集計手段と、
投稿の種別ごとの集計情報を表示する表示手段と、
を備え、

前記投稿の種別には、少なくとも、投稿の拡散を目的とする投稿と、特定の対象に関する記事を紹介することを目的とする投稿とを含むことを特徴とする情報処理装置。

【請求項 2】

前記表示手段は、投稿の種別ごとの投稿数を時系列グラフで表示することを特徴とする請求項1に記載の情報処理装置。

【請求項 3】

前記時系列グラフに、投稿に関する注釈を付与する注釈付与手段を更に備えることを特徴とする請求項2に記載の情報処理装置。

【請求項 4】

前記注釈の内容は、投稿の種別毎に抽出した話題と当該話題ごとの件数であって、投稿数の多い投稿の種別順に表示したものであることを特徴とする請求項3に記載の情報処理装置。

【請求項 5】

10

20

特定の対象に関する 1 以上の投稿を抽出する抽出手段と、
抽出された各投稿を投稿の種別ごとに集計する集計手段と、
投稿の種別ごとの投稿数を時系列グラフで表示する表示手段と、
前記時系列グラフに、投稿に関する注釈を付与する注釈付与手段と、
を備え、
前記注釈の内容は、投稿の種別毎に抽出した話題と当該話題ごとの件数を表示したもの
である
ことを特徴とする情報処理装置。

【請求項 6】

前記注釈の内容は、投稿の種別毎に抽出した話題と当該話題ごとの件数を投稿数の多い 10
投稿の種別順に表示したものである
ことを特徴とする請求項 5 に記載の情報処理装置。

【請求項 7】

投稿数に基づいてイベント発生期間を取得するイベント取得手段を、更に備え、
前記注釈付与手段は、前記イベント発生期間に対応した前記時系列グラフ上に注釈を付
与する
ことを特徴とする請求項 3 乃至 6 のいずれか 1 項に記載の情報処理装置。

【請求項 8】

情報処理装置の抽出手段が、特定の対象に関する 1 以上の投稿を抽出する抽出ステップ 20
と、
前記情報処理装置の集計手段が、抽出された各投稿を投稿の種別ごとに集計する集計ス
テップと、
前記情報処理装置の表示手段が、集計した投稿の種別ごとの集計情報を表示する表示ス
テップと、
を含み、
前記投稿の種別には、少なくとも、投稿の拡散を目的とする投稿と、特定の対象に関す
る記事を紹介することを目的とする投稿とを含む
ことを特徴とする情報処理方法。

【請求項 9】

コンピュータを、 30
特定の対象に関する 1 以上の投稿を抽出する抽出手段と、
抽出された各投稿を投稿の種別ごとに集計する集計手段と、
投稿の種別ごとの集計情報を表示する表示手段と、
して機能させるためのプログラムであって、
前記投稿の種別には、少なくとも、投稿の拡散を目的とする投稿と、特定の対象に関す
る記事を紹介することを目的とする投稿とを含む
ことを特徴とするプログラム。

【請求項 10】

情報処理装置の抽出手段が、特定の対象に関する 1 以上の投稿を抽出するステップと、
情報処理装置の集計手段が、抽出された各投稿を投稿の種別ごとに集計するステップと 40
、
情報処理装置の表示手段が、投稿の種別ごとの投稿数を時系列グラフで表示するステッ
プと、
情報処理装置の注釈付与手段が、前記時系列グラフに、投稿に関する注釈を付与するス
テップと、
を含み、

前記注釈の内容は、投稿の種別毎に抽出した話題と当該話題ごとの件数を表示したもの
である
ことを特徴とする情報処理方法。

【請求項 11】

コンピュータを、
特定の対象に関する1以上の投稿を抽出する抽出手段と、
抽出された各投稿を投稿の種別ごとに集計する集計手段と、
投稿の種別ごとの投稿数を時系列グラフで表示する表示手段と、
前記時系列グラフに、投稿に関する注釈を付与する注釈付与手段と、
して機能させるためのプログラムであって、
前記注釈の内容は、投稿の種別毎に抽出した話題と当該話題ごとの件数を表示したもの
である
ことを特徴とするプログラム。

【発明の詳細な説明】

10

【技術分野】

【0001】

本発明は、Twitterなどのマイクロブログから投稿を抽出する技術に関する。

【背景技術】

【0002】

Twitterなどのマイクロブログは、いつでも誰でも書き込むことが可能であるため、部分的ではあるものの、世の中の声がテキストデータとしてリアルタイムに集まる場所とみることができる。近年では、速報性が高い情報源としてマイクロブログを活用しているユーザや企業も多い。

【0003】

20

しかし、人手で確認できる投稿の量には限界があり、日々増加する投稿を全て確認することは困難であるため、非特許文献1に示されるように、投稿の集合に対してなんらかの機械的な処理を施し、全体を俯瞰して評判や話題を抽出する研究が盛んにおこなわれている。

【0004】

例えば、図1に示すように、特定の商品名を含む投稿を抽出し、投稿数の時系列推移を可視化することにより、商品に対してなんらかのイベントが発生したタイミングを把握することができる。また、図2に示すように、投稿中に含まれる各単語の頻度など、統計情報を参照することにより、全ての投稿を読まずとも、どのような話題が数多く発生しているかを大まかに把握できる。

30

【0005】

前述の技術を組み合わせ、話題が急増した期間を確認して該当する期間の話題を可視化することが可能となる。しかし、投稿数が急増したすべての期間を人手で調査する作業には非常に手間がかかる。また、イベントが発生した期間においては、拡散された投稿や、イベントに関連する単語が頻出する傾向があり、イベントに関連しない話題や評価を把握しづらくなる問題もある。

【0006】

グラフが急増・急減した原因分析を分かりやすくする手法として、特許文献1では、イベントの発生を表す単語（「発売」「発表」「開催」「発覚」など）と時間表現とを含む投稿から特定の対象に発生したイベントのスケジュール情報を自動的に抽出し、時系列のグラフ上に注釈付する手法が提案されている。

40

【先行技術文献】

【特許文献】

【0007】

【特許文献1】WO2009/101954

【非特許文献】

【0008】

【非特許文献1】奥村 学、「ソーシャルメディアを対象としたテキストマイニング」電子情報通信学会基礎・境界ソサイエティ Fundamental Review Vol.6 No.4 pp.285-293, 2013.

50

【非特許文献 2】乾 孝司、他「テキストを対象とした評価情報の分析に関する研究動向」自然言語処理、言語処理学会、2006年7月、Vol. 13、No. 3、pp. 201 - 241

【発明の概要】

【発明が解決しようとする課題】

【0009】

特許文献1に示される技術では、各期間において発生しているイベントの一部を時系列グラフ上に提示することが可能にはなるが、提示するイベントに関する投稿が対象の期間中においてどの程度の割合を占めるかを示すことができない。すなわち、提示されているイベント以外の話題が投稿数の急増に寄与している可能性もあることから、システムの利用者はより詳細な分析が必要かどうかを注釈付された情報のみからは判断できず、投稿数が急増した原因分析の労力削減という観点において十分であるとはいえない。

10

【0010】

本発明は、前述の問題点に鑑みてなされたもので、投稿に対して、より詳細な分析作業を必要とするか否かの判断材料をシステム利用者に提供し、システム利用者の作業負荷軽減を図ることを目的とする。

【課題を解決するための手段】

【0011】

前述した課題を解決するための第1の発明は、特定の対象に関する1以上の投稿を抽出する抽出手段と、抽出された各投稿を投稿の種別ごとに集計する集計手段と、投稿の種別ごとの集計情報を表示する表示手段と、を備え、前記投稿の種別には、少なくとも、投稿の拡散を目的とする投稿と、特定の対象に関する記事を紹介することを目的とする投稿とを含むことを特徴とする情報処理装置である。

20

第1の発明によれば、抽出された投稿を、投稿の種別ごとに集計し、投稿の種別ごとの集計情報として表示する。これにより、投稿の種別ごとに集計結果を把握することができるため、投稿に対して、より詳細な分析作業を必要とするか否かの判断材料をシステム利用者に提供し、システム利用者の作業負担軽減を図ることが可能となる。また、抽出された投稿を、投稿の拡散を目的とする投稿と特定の対象に関する記事を紹介することを目的とする投稿とに分類して分析することが可能となる。

【0013】

30

第1の発明において、前記表示手段は、投稿の種別ごとの投稿数を時系列グラフで表示することが望ましい。このとき、前記時系列グラフに、投稿に関する注釈を付与する注釈付与手段を更に備えることが望ましい。これによって、注釈付きの時系列グラフ上において、投稿の構成（種別）を把握することが容易になり、更に詳細な分析を行う必要があるか否かを判断するための時間を短縮することができる。また、前記注釈の内容は、投稿の種別毎に抽出した話題と当該話題ごとの件数であって、投稿数の多い投稿の種別順に表示したものであることが望ましい。これによって、時系列グラフ上において、主な話題と話題ごとの投稿数を併せて注釈付与することで、投稿数が増えた原因分析が容易となる。

【0014】

40

第2の発明は、特定の対象に関する1以上の投稿を抽出する抽出手段と、抽出された各投稿を投稿の種別ごとに集計する集計手段と、投稿の種別ごとの投稿数を時系列グラフで表示する表示手段と、前記時系列グラフに、投稿に関する注釈を付与する注釈付与手段と、を備え、前記注釈の内容は、投稿の種別毎に抽出した話題と当該話題ごとの件数を表示したものであることを特徴とする情報処理装置である。

第2の発明によれば、抽出された投稿を、投稿の種別ごとに集計し、投稿の種別ごとの投稿数を時系列グラフで表示し、この時系列グラフに、投稿に関する注釈を付与する。また、注釈の内容は、投稿の種別毎に抽出した話題と当該話題ごとの件数を表示したものとする。これによって、注釈付きの時系列グラフ上において、投稿の種別を把握することができなり、更に詳細な分析を行う必要があるか否かを判断するための時間を短縮することができる。また、時系列グラフ上において、主な話題と話題ごとの投稿数を併せて注釈

50

付与することで、投稿数が増えた原因分析が容易となる。

また、前記注釈の内容は、投稿の種別毎に抽出した話題と当該話題ごとの件数を投稿数の多い投稿の種別順に表示したものであることが望ましい。

【0015】

また、第1、第2の発明において、投稿数に基づいてイベント発生期間を取得するイベント取得手段を、更に備え、前記注釈付与手段は、前記イベント発生期間に対応した前記時系列グラフ上に注釈を付与することが望ましい。これによって、投稿数が増加するイベント発生期間において投稿の構成（種別）を容易に把握することが可能となる。

【0016】

第3の発明は、情報処理装置の抽出手段が、特定の対象に関する1以上の投稿を抽出する抽出ステップと、前記情報処理装置の集計手段が、抽出された各投稿を投稿の種別ごとに集計する集計ステップと、前記情報処理装置の表示手段が、集計した投稿の種別ごとの集計情報を表示する表示ステップと、を含み、前記投稿の種別には、少なくとも、投稿の拡散を目的とする投稿と、特定の対象に関する記事を紹介することを目的とする投稿とを含むことを特徴とする情報処理方法である。

第4の発明は、情報処理装置の抽出手段が、特定の対象に関する1以上の投稿を抽出するステップと、情報処理装置の集計手段が、抽出された各投稿を投稿の種別ごとに集計するステップと、情報処理装置の表示手段が、投稿の種別ごとの投稿数を時系列グラフで表示するステップと、情報処理装置の注釈付与手段が、前記時系列グラフに、投稿に関する注釈を付与するステップと、を含み、前記注釈の内容は、投稿の種別毎に抽出した話題と当該話題ごとの件数を表示したものであることを特徴とする情報処理方法である。

【0017】

第5の発明は、コンピュータを、特定の対象に関する1以上の投稿を抽出する抽出手段と、抽出された各投稿を投稿の種別ごとに集計する集計手段と、投稿の種別ごとの集計情報を表示する表示手段と、して機能させるためのプログラムであって、前記投稿の種別には、少なくとも、投稿の拡散を目的とする投稿と、特定の対象に関する記事を紹介することを目的とする投稿とを含むことを特徴とするプログラムである。

第6の発明は、コンピュータを、特定の対象に関する1以上の投稿を抽出する抽出手段と、抽出された各投稿を投稿の種別ごとに集計する集計手段と、投稿の種別ごとの投稿数を時系列グラフで表示する表示手段と、前記時系列グラフに、投稿に関する注釈を付与する注釈付与手段と、して機能させるためのプログラムであって、前記注釈の内容は、投稿の種別毎に抽出した話題と当該話題ごとの件数を表示したものであることを特徴とするプログラムである。

【発明の効果】

【0018】

本発明によれば、投稿に対して、より詳細な分析をする必要があるか否かの判断材料をシステム利用者に提供することができるようになる。

【図面の簡単な説明】

【0019】

【図1】投稿数の時系列推移の可視化例を示す図

40

【図2】投稿中に含まれる各単語の頻度を示す図

【図3】マイクロブログ分析システム100の機能構成を示す図

【図4】情報処理装置101のハードウェア構成の一例を示すブロック図

【図5】投稿データ120の一例を示す図

【図6】投稿集計データ130の一例を示す図

【図7】投稿取得処理のフローチャート

【図8】投稿集計処理のフローチャート

【図9】投稿可視化部113が生成した投稿数日別推移グラフの一例を示す図

【図10】投稿可視化部113が生成した時系列グラフの一例を示す図

【図11】図10の注釈の全内容を示す図

50

【発明を実施するための形態】**【0020】**

以下、図面を参照して、本発明の実施形態を詳細に説明する。

【0021】

図4は、本発明の実施形態に係わる情報処理装置101のハードウェア構成の一例を示すブロック図である。

図4において、201はCPUで、システムバス204に接続される各デバイスやコントローラを統括的に制御する。また、ROM203あるいは外部メモリ211には、CPU201の制御プログラムであるBIOS(Basic Input / Output System)やオペレーティングシステムプログラム(以下、OS)や、各サーバあるいは各PCの実行する機能を実現するために必要な各種プログラム等が記憶されている。
10

【0022】

202はRAMで、CPU201の主メモリ、ワークエリア等として機能する。CPU201は、処理の実行に際して必要なプログラム等をROM203あるいは外部メモリ211からRAM202にロードして、該ロードしたプログラムを実行することで各種動作を実現するものである。

【0023】

また、205は入力コントローラで、入力装置209等からの入力を制御する。206はビデオコントローラで、液晶ディスプレイ等のディスプレイ装置210への表示を制御する。なお、ディスプレイ装置は、液晶ディスプレイに限らず、CRTディスプレイなどであっても良い。これらは必要に応じてクライアントが使用するものである。
20

【0024】

207はメモリコントローラで、ポートプログラム、各種のアプリケーション、フォントデータ、ユーザファイル、編集ファイル、各種データ等を記憶するハードディスク(HD)や、フレキシブルディスク(FD)、あるいはPCMCIカードスロットにアダプタを介して接続されるコンパクトフラッシュ(登録商標)メモリ等の外部メモリ211へのアクセスを制御する。

外部メモリ211には、後述するマイクロブログ分析システム100の各機能をCPU201により実行するプログラムが格納されている。このプログラムを汎用のコンピュータにインストールすることによって、本発明に係る情報処理装置101を得ることができる。
30

【0025】

208は通信I/Fコントローラで、ネットワーク(例えば、図1に示したLAN400)を介して外部機器と接続・通信するものであり、ネットワークでの通信制御処理を実行する。例えば、TCP/IPを用いた通信等が可能である。

【0026】

なお、CPU201は、例えばRAM202内の表示情報用領域へアウトラインフォントの展開(ラスタライズ)処理を実行することにより、ディスプレイ装置210上の表示を可能としている。また、CPU201は、ディスプレイ装置210上の不図示のマウスカーソル等でのユーザ指示を可能とする。
40

ハードウェア上で動作する各種プログラムは、外部メモリ211に記録されており、必要に応じてRAM202にロードされることによりCPU201によって実行されるものである。

なお、全ての装置がこれらの構成を備えているわけではなく、必要なものを備えていればよい。

【0027】

本発明は、マイクロブログにおいて投稿が急増する際(イベント発生時)には、イベントの情報に関するWebページのURLが投稿に含まれる可能性が高いことや、マイクロブログユーザの関心を集めた投稿が引用されて拡散されやすいことなど、マイクロブログの性質を利用して話題の抽出、および、可視化を行う。以下、図面を参照して、本発明
50

の実施形態を詳細に説明する。

【0028】

図3は、本発明の実施形態におけるマイクロブログ分析システム100の機能構成を示す図である。

マイクロブログ分析システム100は、投稿取得部111と、投稿集計部112と、投稿可視化部113と、投稿データ120と、投稿集計データ130とを備える。

【0029】

投稿取得部111は、マイクロブログから特定の対象に関する投稿を取得し、投稿データ120に格納する。投稿取得部111における投稿取得処理については、後に例を用いて詳しく説明する。

10

【0030】

投稿集計部112は、投稿データ120に保存された投稿を参照し、投稿データの時系列グラフを生成するために必要な各種集計処理を行い、集計結果を投稿集計データ130に格納する。投稿集計部112における集計処理については、後に例を用いて詳しく説明する。

【0031】

投稿可視化部113は、投稿集計データ130を参照し、投稿データの時系列グラフ生成を行う。時系列グラフの生成方法については、後に例を用いて詳しく説明する。

【0032】

図5は、投稿データ120の一例であり、投稿を一意に識別する投稿IDと、投稿日時、投稿文に加え、投稿取得部111が別途付与する記事タイトルと、拡散対象IDとを投稿ごとに保持する。記事タイトル、および、拡散対象IDについては、後の投稿取得処理の説明において詳しく説明する。

20

【0033】

図6は、投稿集計データ130の一例である。投稿集計データ130は、全投稿集計データ131、記事投稿集計データ132、拡散投稿集計データ133、意見投稿集計データ134から成る。各データの詳細は、後の投稿集計処理の説明において詳しく説明する。

【0034】

次に、投稿取得部111における投稿取得処理、投稿集計部112における集計処理、投稿可視化部113における時系列グラフの生成方法について、順に例を用いて説明する。

30

【0035】

まず、投稿取得部111における投稿取得処理について、図5、図7を用いて説明する。図7は、本発明の実施形態における投稿取得処理のフローチャートである。以下、図7の処理に従って、「製品A」に関する投稿をマイクロブログから取得する例を示す。

【0036】

ステップS101において、情報処理装置101のCPU201(投稿取得部111)は、分析対象である「製品A」を投稿文中に含む投稿をマイクロブログから取得する。以後の処理を説明するため、ここでは下表に示す5件の投稿(投稿A～投稿E)を含む投稿の集合を取得したものとする。

40

なお、ステップS101によって、本発明の「マイクロブログから、特定の対象に関する1以上の投稿を抽出する抽出手段」の一例が構成されている。

【0037】

【表1】

表. マイクロブログから取得した「製品A」に関する投稿(一部抜粋)

投稿ID	投稿日時	投稿者	投稿文	引用
投稿A	2015/10/04 10:01:15	投稿者A	Z社が製品Aを今年中にリリースするらしい。 http://zsha.example.com/news/1001.html http://zsha.example.com/news/1002.html	
投稿B	2015/10/04 10:30:00	投稿者B	Z社が新製品の提供時期を発表！ 進化した製品Aの驚きの性能とは！？ http://cnews.example.com/docs/59812.html	10
投稿C	2015/10/04 10:32:02	投稿者C	RT@投稿者B: Z社が新製品の提供時期を 発表！進化した製品Aの驚きの性能とは！？ http://cnews.example.com/docs/59812.html	投稿B
投稿D	2015/10/04 10:32:03	投稿者C	製品A欲しいな…。性能3倍とかすごい。	
投稿E	2015/10/05 15:51:52	投稿者B	RT@投稿者C: 製品A欲しいな…。 性能3倍とかすごい。	投稿D

【0038】

ステップS101で取得した各投稿に対して、CPU201(投稿取得部111)は、ステップS102～ステップS107の処理を繰り返し実行する。ステップS102～ステップS107の処理は、投稿に含まれるURLや、他の投稿を引用した投稿に対して付加情報を取得し、投稿データ120に保存する処理である。

【0039】

まず、ステップS103において、CPU201(投稿取得部111)は、投稿文にURLが含まれるか否かを判定する。CPU201(投稿取得部111)は、投稿文にURLが含まれていれば、処理をステップS104に移行する。例えば、投稿AはURLを投稿文中に含むため、CPU201(投稿取得部111)は、ステップS104の処理を実行する。

【0040】

ステップS104において、CPU201(投稿取得部111)は、URLが示すWebページのタイトルをインターネット経由で取得する。ここで、取得したタイトルを「記事タイトル」と呼ぶ。なお、投稿文に複数のURLが含まれている場合、CPU201(投稿取得部111)は、すべてのURLに対する記事タイトルを取得する。例えば、投稿Aにおいては、URL「<http://zsha.example.com/news/1001.html>」に対する記事タイトル「製品Aを今冬発売予定」と、URL「<http://zsha.example.com/news/1002.html>」に対する記事タイトル「製品Aの性能表」をそれぞれ取得する。

【0041】

次に、ステップS105において、CPU201(投稿取得部111)は、投稿が他の投稿を引用したものであるかを判定する。CPU201(投稿取得部111)は、投稿が他の投稿を引用したものであれば、処理をステップS106に移行する。例えば、投稿Cは投稿Bを引用した投稿であるため、ステップS106の処理を実行する。ここで、説明を簡略化するため、各投稿が引用した投稿は、マイクロブログから投稿を取得する時点で取得可能な情報として説明する。

【0042】

ステップS106において、CPU201(投稿取得部111)は、引用した投稿の投

10

20

30

40

50

稿IDを取得する。ここで取得した投稿IDを「拡散対象ID」と呼ぶ。

【0043】

次に、ステップ107において、CPU201(投稿取得部111)は、投稿IDと、投稿日時と、投稿文と、ステップ104で取得した記事タイトルと、ステップ106で取得した拡散対象IDとを、投稿データ120に追加する。ここで、記事タイトルが存在する投稿を「記事投稿」(Web上の記事を紹介することを目的とする投稿)、拡散対象IDが存在する投稿を「拡散投稿」(投稿の拡散を目的とする投稿)、いずれも存在しない投稿を「意見投稿」と呼ぶ。例えば、投稿A、投稿Bは記事投稿、投稿Cは記事投稿かつ拡散投稿、投稿Dは意見投稿、投稿Eは拡散投稿である。

【0044】

以上、投稿取得部111における投稿取得処理について、例を用いて説明した。なお、説明の簡略化のため、ステップS105について、各投稿が引用した投稿は、マイクロログから投稿を取得する時点で取得可能な情報として説明したが、取得不可能な環境においても、投稿文の内容や投稿日時の前後関係により、引用対象の投稿を機械的に取得することが可能である。

【0045】

次に、投稿集計部112における投稿集計処理について、図5、図6、図8を用いて説明する。図8は、本発明の実施形態における投稿集計処理のフローチャートである。以下、図8の処理に従って、「製品A」に関する「2015年10月4日」の投稿を集計する例を示す。

10

【0046】

まず、ステップS201において、情報処理装置101のCPU201(投稿集計部112)は、集計対象期間に投稿された投稿を投稿データから取得する。例えば、集計対象期間が2015年10月4日の1日間である場合、図5の投稿データ120において投稿A、投稿B、投稿C、投稿Dは取得対象であり、投稿Eは取得対象外である。

20

【0047】

次に、ステップS201で取得した各投稿に対して、CPU201(投稿集計部112)は、ステップS202～ステップS209の処理を繰り返し実行する。ステップS202～ステップS209の処理は、投稿集計データ130を更新する処理である。

【0048】

30

まずステップS203において、CPU201(投稿集計部112)は、投稿データに記事タイトルが存在するか否かを判定する。CPU201(投稿集計部112)は、投稿データに記事タイトルが含まれていれば、処理をステップS204に移行する。

【0049】

ステップS204において、CPU201(投稿集計部112)は、記事投稿集計データ132を更新する。記事投稿集計データ132は、集計対象期間、記事タイトル、投稿数から成る。ここで、投稿数は、各集計期間において各記事タイトルが含まれる投稿の数である。CPU201(投稿集計部112)は、記事投稿集計データ132を参照し、投稿データと同じ集計対象期間・記事タイトルであるレコードが存在すれば、該当するレコードの投稿数に1を加算し、該当するレコードが存在しなければ、投稿数が1である新たなレコードを登録する。

40

【0050】

次にステップS205において、CPU201(投稿集計部112)は、投稿データに拡散対象IDが存在するか否かを判定する。CPU201(投稿集計部112)は、投稿データに拡散対象IDが含まれていれば、処理をステップS206に移行する。

【0051】

ステップS206において、CPU201(投稿集計部112)は、拡散投稿集計データ133を更新する。拡散投稿集計データ133は、集計対象期間、拡散対象ID、投稿数から成る。ここで、投稿数は、各集計期間において各拡散対象IDが含まれる投稿の数である。CPU201(投稿集計部112)は、拡散投稿集計データ133を参照し、投

50

稿データと同じ集計対象期間・拡散対象IDであるレコードが存在すれば、該当するレコードの投稿数に1を加算し、該当するレコードが存在しなければ、投稿数が1である新たなレコードを登録する。

【0052】

次にステップS207において、CPU201（投稿集計部112）は、投稿データが意見投稿（記事タイトルも拡散対象IDも存在しない）であるか否かを判定する。CPU201（投稿集計部112）は、投稿データが意見投稿であれば、処理をステップS208に移行する。

【0053】

ステップS208において、CPU201（投稿集計部112）は、意見投稿集計データ134を更新する。意見投稿集計データ134は、集計対象期間、意見、投稿数から成る。ここで、「意見」は、投稿文から自然言語処理を用いて抽出する文字列である。説明を簡略化するため、ここでは投稿文に含まれる形容詞や動詞を投稿データの意見として取り扱う（形容詞や動詞を抽出することは公知の技術で可能である）。例えば、図5の投稿データ120における投稿Dの投稿文「製品A欲しいな...。性能3倍とかすごい。」から、「欲しい」「すごい」などの形容詞を抽出する。CPU201（投稿集計部112）は、意見投稿集計データ134を参照し、抽出した各意見に対して、同じ集計対象期間のレコードが存在すれば、該当するレコードの投稿数に1を加算し、該当するレコードが存在しなければ、投稿数が1である新たなレコードを登録する。

【0054】

次にステップS209において、CPU201（投稿集計部112）は、全投稿集計データ131を更新する。全投稿集計データ131は、投稿数、記事投稿数、拡散投稿数、意見投稿数から成る。投稿集計部112は、全投稿集計データ131を参照し、同じ集計対象期間のレコードが存在すれば、該当するレコードにおいて、投稿数に1を加算し、該当するレコードが存在しなければ、投稿数が1である新たなレコードを登録する。また、投稿データが記事投稿であれば記事投稿数に1を加算し、拡散投稿であれば拡散投稿に1を加算し、意見投稿であれば意見投稿数に1を加算する。

【0055】

以上、投稿集計部112における投稿集計処理について、例を用いて説明した。なお、説明の簡略化のため、ステップS208について、形容詞や動詞の原型を投稿データの意見として取り扱う方法を説明したが、非特許文献2で紹介されるような、評価情報抽出技術の出力結果を用いてもよい。

なお、図7（投稿取得処理）のステップS102～ステップS107および図8（投稿集計処理）のステップS201～ステップS209によって、本発明の「抽出された各投稿を投稿の種別ごとに分類し集計する集計手段」「前記集計手段は、抽出された投稿を、少なくとも、投稿の拡散を目的とする投稿と、Web上の記事を紹介することを目的とする投稿と、に分類し集計する」の一例が構成されている。投稿の種別は、拡散投稿かつ記事投稿、拡散投稿を除く記事投稿、記事投稿を除く拡散投稿、意見投稿を含む。

【0056】

次に、投稿可視化部113における投稿データの可視化方法について、図1、図6、図9、図10を用いて説明する。図9、および、図10は、投稿可視化部113が投稿集計データ130を参照して生成する時系列グラフであり、システムの利用者はWebブラウザなどを通じて当該グラフを閲覧、操作することができる。図11は、図10の注釈の全内容を示した図である。

以下、時系列グラフの構成について説明する。

【0057】

情報処理装置101のCPU201（投稿可視化部113）は、全投稿集計データ131を参照し、集計対象期間ごとの投稿数を、積上げ棒グラフにより可視化する。図9は、投稿可視化部113が生成した製品Aに対する2015年10月の投稿数日別推移グラフである。投稿可視化部113は、拡散投稿を除く記事投稿の投稿、記事投稿かつ拡散投稿

10

20

30

40

50

の投稿数、記事投稿を除く拡散投稿の投稿数、意見投稿数をそれぞれ別の色で可視化する。なおこれによって、本発明の「投稿の種別ごとの集計情報を表示する表示手段」「前記表示手段は、投稿の種別ごとの投稿数を時系列グラフで表示する」の一例が構成されている。

なお、各投稿数は、全投稿集計データ131が保持する値を用いて、下記の式で算出できる。

【0058】

(数1)

$$= \text{記事投稿かつ拡散投稿の投稿数} = (\text{記事投稿数} + \text{拡散投稿数} + \text{意見投稿数}) - \text{投稿数}$$

10

$$\text{拡散投稿をのぞく記事投稿の投稿数} = \text{記事投稿数} -$$

$$\text{記事投稿をのぞく拡散投稿の投稿数} = \text{拡散投稿数} -$$

$$\text{意見投稿数} = \text{意見投稿数}$$

【0059】

また、CPU201(投稿可視化部113)は、投稿集計データ130を参照し、システムの利用者によるクリック操作で展開する注釈を時系列グラフ上に付与する。図10は、システムの利用者が、「10月4日」、「10月10日」、「10月20日」における棒グラフをクリックし、それぞれの注釈を展開した後の時系列グラフを表している。なおこれによって、本発明の「前記時系列グラフに、投稿に関する注釈を付与する注釈付与手段」の一例が構成されている。

20

以下、投稿可視化部113における注釈に記載する内容の取得方法について説明する。

【0060】

注釈は、イベントが発生したと考えられる箇所(イベント発生期間)に付与される。

イベント発生期間とは、具体的には以下の方法により求めることが可能である。

【0061】

まず、CPU201(投稿可視化部113)は、分析期間内における単位期間あたりの投稿数を降順でソートし、下位半分の投稿数の平均値を取得する(平常時平均値)。そして、CPU201(投稿可視化部113)は、投稿数がN以上かつ平常時平均値のM倍以上である箇所をイベント発生期間と判断する。なおこれによって、本発明の「投稿数に基づいてイベント発生期間を取得するイベント取得手段」の一例が構成されている。

30

【0062】

CPU201(投稿可視化部113)は、各イベント発生期間において、投稿の種別ごとに特徴的な話題を抽出し、投稿数の多い投稿の種別順に抽出した話題を話題ごとの件数付で時系列グラフ上に注釈を付与する。なおこれによって、本発明の「前記注釈付与手段は、前記イベント発生期間に対応した前記時系列グラフ上に注釈を付与する」「前記注釈の内容は、投稿数の多い投稿の種別順に抽出した話題と、当該話題ごとの件数である」の一例が構成されている。

以下、記事投稿、拡散投稿、意見投稿の順に、話題の抽出方法を説明する。

【0063】

まず、記事投稿における話題の抽出方法について説明する。CPU201(投稿可視化部113)は、記事投稿集計データ132を参照し、集計対象期間における記事タイトルのうち、投稿数の上位L件を話題として抽出する。(Lは自然数)

40

【0064】

次に、拡散投稿における話題の抽出方法について説明する。CPU201(投稿可視化部113)は、拡散投稿集計データ133を参照し、集計対象期間における拡散対象IDのうち、投稿数の上位M件を抽出し、各拡散対象IDに対応する投稿を話題として抽出する。(Mは自然数)

【0065】

次に、意見投稿における話題の抽出方法について説明する。CPU201(投稿可視化部113)は、意見投稿集計データ134を参照し、集計対象期間における意見のうち、

50

投稿数の上位N件を話題として抽出する。(Nは自然数)

【0066】

以上で説明した、投稿可視化部113が出力する注釈付きの時系列グラフを参照することにより、システムの利用者は時系列グラフ上において、集計対象期間における投稿の構成を把握することが容易になり、更に詳細な分析を行う必要があるか否かを判断するための時間を短縮することができる。例えば、図10から、「10月10日に投稿数が急増している原因は、製品Aを紛失したユーザの投稿が拡散されているだけで、製品A自体に問題が発生したわけではない」ことが分かるため、当該期間に関する分析を省略するという判断をすることが可能となる。また、投稿の種別ごとに話題を集約することで、意見投稿のみを分析することが可能となり、イベントとは関係の無い評価(通常時の評判)を分析することが可能となる。

【0067】

また、「10月4日、10月20日を境に意見投稿数の割合が増加している」ことが分かるため、当該期間に関してより注力して分析する、という判断をすることが可能になる。

【0068】

以上、実施形態例を詳述したが、本発明は、例えば、システム、装置、方法、プログラムもしくは記憶媒体等としての実施態様をとることが可能であり、具体的には、複数の機器から構成されるシステムに適用しても良いし、また、一つの機器からなる装置に適用しても良い。

10

20

【0069】

なお、上述した各種データの構成及びその内容はこれに限定されるものではなく、用途や目的に応じて、様々な内容で構成されることは言うまでもない。

【0070】

また、本発明は、システム或いは装置にプログラムを供給することによって達成される場合にも適用できることは言うまでもない。この場合、本発明を達成するためのソフトウェアによって表されるプログラムを格納した記録媒体を該システム或いは装置に読み出すことによって、そのシステム或いは装置が、本発明の効果を享受することが可能となる。

【0071】

さらに、本発明を達成するためのソフトウェアによって表されるプログラムをネットワーク上のサーバ、データベース等から通信プログラムによりダウンロードして読み出すことによって、そのシステム或いは装置が、本発明の効果を享受することが可能となる。

30

【0072】

なお、上述した各実施形態及びその変形例を組み合わせた構成もすべて本発明に含まれるものである。

【符号の説明】

【0073】

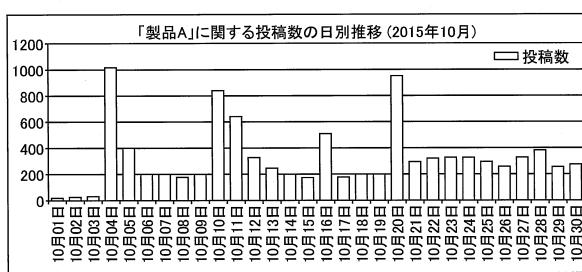
100マイクロブログ分析システム
101情報処理装置
111投稿取得部
112投稿集計部
113投稿可視化部
120投稿データ
130投稿集計データ
131全投稿集計データ
132記事投稿集計データ
133拡散投稿集計データ
134意見投稿集計データ
201CPU
202RAM

40

50

- 203 ROM
 204 システムバス
 209 入力装置
 210 ディスプレイ装置
 211 外部メモリ

【図1】

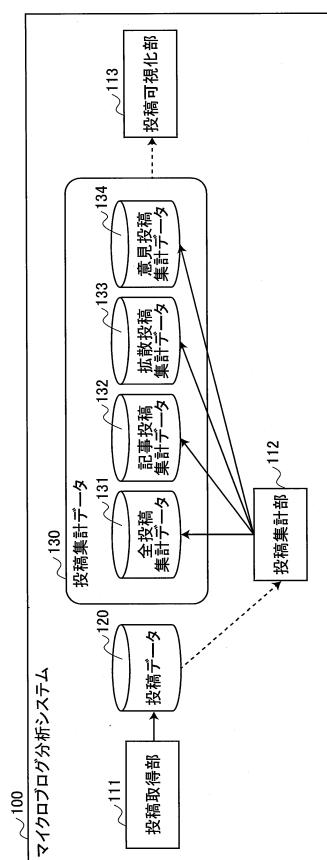


【図2】

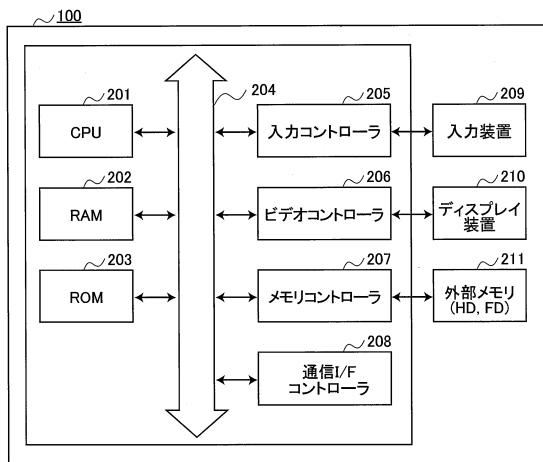
「製品A」単語ランキング (2015年10月4日)

名詞	投稿数
新製品	560
リリース	560
性能	481
モデル	353
New	322
本格派	331
形容詞	投稿数
新しい	353
すごい	86
欲しい	54
やばい	32
高い	29
強い	13
動詞	投稿数
発売する	421
発表する	401
搭載する	390
進化する	386
予約する	121
買う	64

【図3】



【図4】



【図5】

120 投稿データ			
投稿ID	投稿日時	投稿者	投稿文
投稿A	2015/10/04 10:01:15	投稿者A	Z社が製品Aを今冬発売予定。 http://zsha.example.com/nes/1001.html
投稿B	2015/10/04 10:30:00	投稿者B	Z社が新製品の提供時期を発表、進化した 製品Aの驚きの性能とは！？ http://cnnews.example.com/docs/58812.html
投稿C	2015/10/04 10:32:02	投稿者C	Z社が新製品の提供時期を 発表！進化した製品Aの驚きの性能とは！？ http://cnnews.example.com/docs/58812.html
投稿D	2015/10/04 10:32:03	投稿者C	製品A驚いな…。性能3倍 とかすごい。
投稿E	2015/10/05 15:51:52	投稿者B	RT@投稿者C: 製品A欲しいな…。性能3倍 とかすごい。

【図6】

130 投稿集計データ				
集計対象期間	投稿数	記事投稿数	拡散投稿数	意見投稿数
2015年10月1日	10	0	8	2
2015年10月2日	20	11	10	9
2015年10月3日	35	15	15	20
2015年10月4日	1021	863	712	127

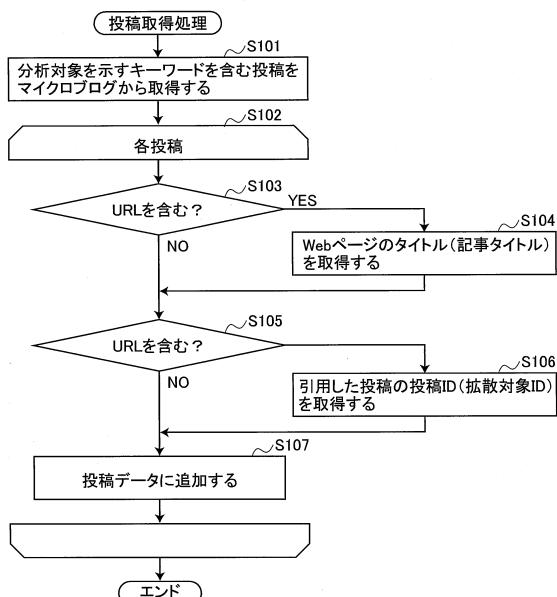
131 全投稿集計データ		
集計対象期間	記事タイトル	投稿数
2015年10月4日	製品Aを今冬発売予定	59
2015年10月4日	製品Aの性能表	4
2015年10月4日	Z社が新製品の提供時期を発表！ 進化した製品Aの驚きの性能とは！？	450

132 記事投稿集計データ		
集計対象期間	記事タイトル	投稿数
2015年10月4日	製品Aを今冬発売予定	59
2015年10月4日	製品Aの性能表	4
2015年10月4日	Z社が新製品の提供時期を発表！ 進化した製品Aの驚きの性能とは！？	450

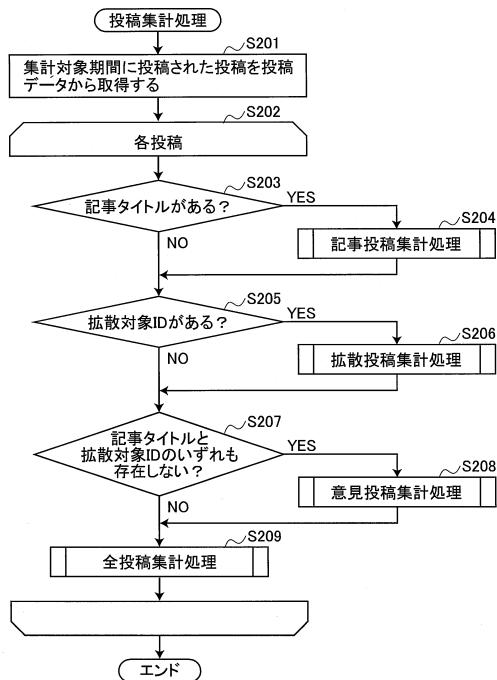
133 拡散投稿集計データ		
集計対象期間	拡散対象ID	投稿数
2015年10月4日	投稿A	4
2015年10月4日	投稿B	391

134 意見投稿集計データ		
集計対象期間	意見	投稿数
2015年10月4日	欲しい	32
2015年10月4日	高い	29
2015年10月5日	買う	49

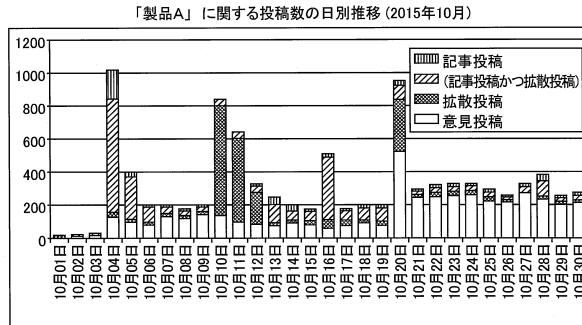
【図7】



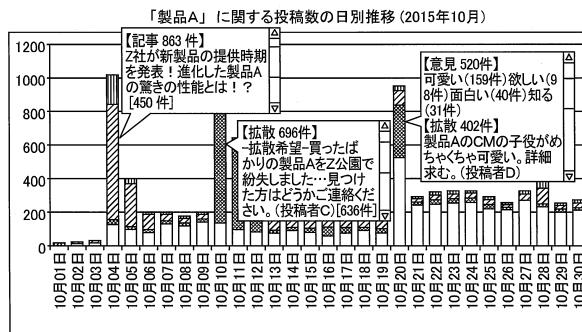
【図 8】



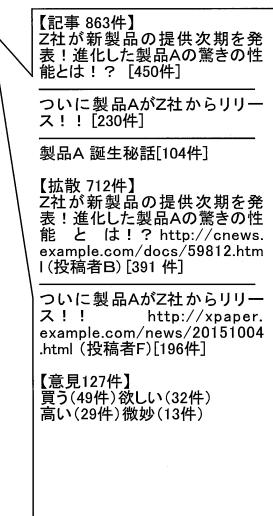
【図 9】



【図 10】



【図 11】



フロントページの続き

(56)参考文献 特開2013-033375(JP,A)
特開2012-079247(JP,A)
特開2005-032197(JP,A)
鳥海 不二夫，“震災時におけるTwitterのリツイート分析”，電子情報通信学会技術研究報告 AI2012-1-AI2012-7 人工知能と知識処理，日本，一般社団法人電子情報通信学会，2012年 6月14日，第112巻，第94巻，p.19-24

(58)調査した分野(Int.Cl., DB名)

G 06 F 16 / 906
G 06 F 16 / 954
G 06 F 13 / 00
G 06 Q 30 / 02