

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4764864号
(P4764864)

(45) 発行日 平成23年9月7日 (2011.9.7)

(24) 登録日 平成23年6月17日 (2011.6.17)

(51) Int.Cl.	F I
G 0 6 Q 50/00 (2006.01)	G 0 6 F 17/60 1 5 0
G 0 6 Q 10/00 (2006.01)	G 0 6 F 17/60 1 7 0 A

請求項の数 11 (全 14 頁)

(21) 出願番号	特願2007-285998 (P2007-285998)	(73) 特許権者	500257300
(22) 出願日	平成19年11月2日 (2007.11.2)		ヤフー株式会社
(65) 公開番号	特開2009-116440 (P2009-116440A)		東京都港区赤坂9丁目7番1号
(43) 公開日	平成21年5月28日 (2009.5.28)	(74) 代理人	100089118
審査請求日	平成21年6月17日 (2009.6.17)		弁理士 酒井 宏明
		(72) 発明者	森 有紀子
			東京都港区六本木六丁目10番1号 ヤフー株式会社内
		審査官	関 博文
		(56) 参考文献	特開2006-227965 (JP, A)
			)
			特開2002-198997 (JP, A)
			)
			最終頁に続く

(54) 【発明の名称】 情報伝播抽出装置および情報伝播抽出方法

(57) 【特許請求の範囲】

【請求項1】

ネットワークを流通する単語の一定期間ごとの注目度指標値を算出する算出手段と、
前記算出手段によって算出された注目度指標値が所定の閾値以上となる単語の注目期間を検出する検出手段と、

前記検出手段によって検出された注目期間内に実際に流通した個々の単語データの流通元の位置情報を取得する取得手段と、

前記取得手段によって取得された位置情報に基づいて単語が注目された位置を決定する決定手段と、

前記決定手段によって決定された位置を含む注目期間情報を送信する送信手段と
を有することを特徴とする情報伝播抽出装置。

10

【請求項2】

前記取得手段は、

ユーザが投稿する日記形式のウェブログに記載されて流通した単語について、個々のウェブログデータを投稿したユーザの位置情報を取得することを特徴とする請求項1記載の情報伝播抽出装置。

【請求項3】

前記取得手段は、

ウェブログに対応付けて記憶されたユーザのプロバイダ情報またはユーザの接続元アドレス情報からユーザの位置情報を推定することを特徴とする請求項2記載の情報伝播抽出

20

装置。

【請求項 4】

前記取得手段は、

ウェブログに記載された文面からユーザの位置情報を推定することを特徴とする請求項 2 記載の情報伝播抽出装置。

【請求項 5】

前記決定手段は、

前記取得手段によって取得されたすべての位置情報のうち所定割合以上の位置情報が共通して示す最小の地域を単語の注目位置と決定することを特徴とする請求項 1 記載の情報伝播抽出装置。

10

【請求項 6】

ユーザが注目期間情報の把握を希望する単語を指定する単語情報を受信する受信手段をさらに有し、

前記算出手段は、

前記受信手段によって受信された単語情報が示す単語の注目度指標値を算出することを特徴とする請求項 1 記載の情報伝播抽出装置。

【請求項 7】

前記算出手段は、

t f / i d f 法による単語の出現頻度の評価結果に基づいて注目度指標値を算出することを特徴とする請求項 1 記載の情報伝播抽出装置。

20

【請求項 8】

前記取得手段は、

インターネットの検索サイトにおける検索によって流通した単語について、検索を行ったユーザの位置情報を取得することを特徴とする請求項 1 記載の情報伝播抽出装置。

【請求項 9】

前記取得手段は、

インターネットの掲示板サイトに対する書き込みによって流通した単語について、書き込みを行ったユーザの位置情報を取得することを特徴とする請求項 1 記載の情報伝播抽出装置。

【請求項 10】

30

コンピュータが実行する情報伝播抽出方法であって、前記コンピュータが、
ネットワークを流通する単語の一定期間ごとの注目度指標値を算出する算出ステップと

、

前記算出ステップにて算出された注目度指標値が所定の閾値以上となる単語の注目期間を検出する検出ステップと、

前記検出ステップにて検出された注目期間内に実際に流通した個々の単語データの流通元の位置情報を取得する取得ステップと、

前記取得ステップにて取得された位置情報に基づいて単語が注目された位置を決定する決定ステップと、

前記決定ステップにて決定された位置を含む注目期間情報を送信する送信ステップとを有することを特徴とする情報伝播抽出方法。

40

【請求項 11】

コンピュータによって実行される情報提供プログラムであって、前記コンピュータに、ネットワークを流通する単語の一定期間ごとの注目度指標値を算出する算出ステップと

、

前記算出ステップにて算出された注目度指標値が所定の閾値以上となる単語の注目期間を検出する検出ステップと、

前記検出ステップにて検出された注目期間内に実際に流通した個々の単語データの流通元の位置情報を取得する取得ステップと、

前記取得ステップにて取得された位置情報に基づいて単語が注目された位置を決定する

50

決定ステップと、

前記決定ステップにて決定された位置を含む注目期間情報を送信する送信ステップと
を実行させることを特徴とする情報提供プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報伝播抽出装置および情報伝播抽出方法に関し、特に、ネットワークの利用を通じて世間の関心事の特性をより詳細に分析する際に必要となる情報を提供する情報伝播抽出装置および情報伝播抽出方法に関する。

【背景技術】

【0002】

近年、インターネットにおいては、ブログと呼ばれる日記形式のウェブサイトが数多く公開されている。ブログは、ウェブログの略称であり、ユーザが自由に画像や文章をウェブ上に公開することができることから、手軽な情報発信の手段として多くのユーザに重宝されている。そして、ブログには、ユーザの関心事や関心事に対する意見などが掲載されることが多いため、例えばあるニュースが席巻すれば、このニュースに言及するブログ数が増加し、世間の関心事の趨勢が如実に反映される。このような傾向は、ブログ以外にも、例えばインターネット上の掲示板などで見られ、話題性のある単語の出現頻度が一時的に増加することがある。

【0003】

そこで、例えば非特許文献1に記載されたように、ブログや掲示板などのインターネット上の複数のドキュメントにおける単語の出現頻度を示すバースト度を算出し、バースト度が高い単語を検出することにより、注目されている話題を自動的に抽出することが検討されている。この非特許文献1に記載された手法によれば、所望の期間に話題となっている（すなわち、バーストしている）単語を容易に見出すことができ、例えば流行に即した商品開発などに供することができる。

【0004】

【非特許文献1】藤木稔明、南野朋之、鈴木泰裕、奥村学、「document streamにおけるburstの発見」、情報処理学会研究報告 自然言語処理、2004-NL-160-(13) pp.85-92、2004年3月

【発明の開示】

【発明が解決しようとする課題】

【0005】

上述したように、非特許文献1の手法によれば、所望の期間に注目された話題を抽出することが可能となる。ところで、一般に、注目される話題の中には、国や地方などの地域限定で注目されるものがある。また、ある地域で注目された話題が、次第に近隣の地域へ伝搬していくことなどもあり、単に話題が注目された期間を特定するだけでは、十分に世間の関心事の分析・予測を行うことができないという問題がある。

【0006】

具体的には、例えばある地方において飲食店が新たに開店する場合、当該地方においてはこの飲食店の開店が注目されると考えられるが、国や世界全体として注目されることはなく、「飲食店の開店」という話題の出現頻度が高くなる期間（バースト）は埋没してしまうことがある。また、例えば「石油価格の上昇」など、世界全体の関心事については、国ごとの影響の度合いに応じて話題が伝搬していくと考えられるが、従来の手法ではバーストの伝搬性を把握することはできない。

【0007】

本発明はかかる点に鑑みてなされたものであり、ネットワークの利用を通じて世間の関心事の特性をより詳細に分析する際に必要となる情報を提供することができる情報伝播抽出装置および情報伝播抽出方法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】

【0008】

上記課題を解決するために、本発明に係る情報伝播抽出装置は、ネットワークを流通する単語の一定期間ごとの注目度指標値を算出する算出手段と、前記算出手段によって算出された注目度指標値が所定の閾値以上となる単語の注目期間を検出する検出手段と、前記検出手段によって検出された注目期間内に実際に流通した個々の単語データの流通元の位置情報を取得する取得手段と、前記取得手段によって取得された位置情報に基づいて単語が注目された位置を決定する決定手段と、前記決定手段によって決定された位置を含む注目期間情報を送信する送信手段とを有する構成を採る。

【0009】

この構成によれば、同一の単語が注目されている注目期間のそれぞれに関して、この単語が注目されていた国や地方などの位置を特定することができ、単語が注目された期間の情報と位置の情報を関係づけて提供することができる。すなわち、ネットワークの利用を通じて世間の関心事の特性をより詳細に分析する際に必要となる情報を提供することができる。

10

【0010】

また、本発明に係る情報伝播抽出装置は、上記構成において、前記取得手段は、ユーザが投稿する日記形式のウェブログに記載されて流通した単語について、個々のウェブログデータを投稿したユーザの位置情報を取得する構成を採る。

【0011】

この構成によれば、多くのユーザが同一の単語をブログに記載した場合に、この単語が注目されていると判断し、主にどの地域のユーザが単語に注目してブログを投稿したかを特定することができる。

20

【0012】

また、本発明に係る情報伝播抽出装置は、上記構成において、前記取得手段は、ウェブログに対応付けて記憶されたユーザのプロバイダ情報またはユーザの接続元アドレス情報からユーザの位置情報を推定する構成を採る。

【0013】

この構成によれば、ブログと対応付けて記憶された比較的に正確なユーザの位置情報を取得することができる。

【0014】

また、本発明に係る情報伝播抽出装置は、上記構成において、前記取得手段は、ウェブログに記載された文面からユーザの位置情報を推定する構成を採る。

30

【0015】

この構成によれば、ユーザの位置情報に関する情報がブログに対応付けて記憶されていない場合でも、ユーザの位置情報を推定することができる。

【0016】

また、本発明に係る情報伝播抽出装置は、上記構成において、前記決定手段は、前記取得手段によって取得されたすべての位置情報のうち所定割合以上の位置情報が共通して示す最小の地域を単語の注目位置と決定する構成を採る。

【0017】

この構成によれば、同一の単語を流通させた多くのユーザの位置情報に共通する最小の地域が単語の注目位置となり、単語が注目されている地域を容易かつ正確に特定することができる。

40

【0018】

また、本発明に係る情報伝播抽出装置は、上記構成において、ユーザが注目期間情報の把握を希望する単語を指定する単語情報を受信する受信手段をさらに有し、前記算出手段は、前記受信手段によって受信された単語情報が示す単語の注目度指標値を算出する構成を採る。

【0019】

この構成によれば、ユーザが注目期間や注目位置の情報取得を希望する単語について、

50

注目期間や注目位置を得ることができ、ユーザの希望に応じた情報提供を行うことができる。

【0020】

また、本発明に係る情報伝播抽出装置は、上記構成において、前記算出手段は、 $t f / i d f$ 法による単語の出現頻度の評価結果に基づいて注目度指標値を算出する構成を採る。

【0021】

この構成によれば、単に出現頻度が高い単語の注目度指標値を大きくするのではなく、普遍的に使用されることによる出現頻度の上昇を考慮して、確実に注目されている単語のみの注目度指標値を大きくすることができる。

10

【0022】

また、本発明に係る情報伝播抽出装置は、上記構成において、前記取得手段は、インターネットの検索サイトにおける検索によって流通した単語について、検索を行ったユーザの位置情報を取得する構成を採る。

【0023】

この構成によれば、多くのユーザが同一の単語を検索した場合に、この単語が注目されていると判断し、主にどの地域のユーザが単語に注目して検索したかを特定することができる。

【0024】

また、本発明に係る情報伝播抽出装置は、上記構成において、前記取得手段は、インターネットの掲示板サイトに対する書き込みによって流通した単語について、書き込みを行ったユーザの位置情報を取得する構成を採る。

20

【0025】

この構成によれば、多くのユーザが同一の単語を掲示板に書き込んだ場合に、この単語が注目されていると判断し、主にどの地域のユーザが単語に注目して掲示板への書き込みを行ったかを特定することができる。

【0026】

また、本発明に係る情報伝播抽出方法は、ネットワークを流通する単語の一定期間ごとの注目度指標値を算出する算出ステップと、前記算出ステップにて算出された注目度指標値が所定の閾値以上となる単語の注目期間を検出する検出ステップと、前記検出ステップにて検出された注目期間内に実際に流通した個々の単語データの流通元の位置情報を取得する取得ステップと、前記取得ステップにて取得された位置情報に基づいて単語が注目された位置を決定する決定ステップと、前記決定ステップにて決定された位置を含む注目期間情報を送信する送信ステップとを有するようにした。

30

【0027】

また、本発明に係る情報提供プログラムは、コンピュータによって実行される情報提供プログラムであって、前記コンピュータに、ネットワークを流通する単語の一定期間ごとの注目度指標値を算出する算出ステップと、前記算出ステップにて算出された注目度指標値が所定の閾値以上となる単語の注目期間を検出する検出ステップと、前記検出ステップにて検出された注目期間内に実際に流通した個々の単語データの流通元の位置情報を取得する取得ステップと、前記取得ステップにて取得された位置情報に基づいて単語が注目された位置を決定する決定ステップと、前記決定ステップにて決定された位置を含む注目期間情報を送信する送信ステップとを実行させるようにした。

40

【0028】

これらによれば、同一の単語が注目されている注目期間のそれぞれに関して、この単語が注目されていた国や地方などの位置を特定することができ、単語が注目された期間の情報と位置の情報をと関係づけて提供することができる。すなわち、ネットワークの利用を通じて世間の関心事の特性をより詳細に分析する際に必要となる情報を提供することができる。

【発明の効果】

50

【0029】

本発明によれば、ネットワークの利用を通じて世間の関心事の特性をより詳細に分析する際に必要となる情報を提供することができる。

【発明を実施するための最良の形態】

【0030】

本発明の骨子は、ユーザがネットワーク上に流通させたキーワードとユーザの位置情報とを関連付け、キーワードに関する話題が注目された場合に、このキーワードを流通させたユーザの位置情報に基づいて、キーワードが注目された地域に関する情報を提供することである。以下、本発明の一実施の形態について、図面を参照して詳細に説明する。なお、以下においては、ユーザがブログを投稿することによってネットワーク上にキーワードを流通させる場合について説明するが、本発明はこれに限定されず、例えばインターネット上の掲示板に対する書き込みや検索サイトに対する検索キーワードの入力などもキーワードの流通に該当する。

10

【0031】

図1は、本発明の一実施の形態に係る通信システムの概要を示すブロック図である。同図に示す通信システムは、ユーザ端末100-1～100-mがネットワークNを介してバースト情報提供装置200およびブログサーバ300に接続される構成となっている。

【0032】

ユーザ端末100-1～100-mは、それぞれユーザが操作する端末であり、ネットワークNに接続されている。ユーザ端末100-1～100-mは、ユーザが例えばブログを投稿する際に、入力されたブログデータをネットワークNを介してブログサーバ300へ送信する。また、ユーザ端末100-1～100-mは、ユーザが所望のキーワードを入力すると、このキーワードをネットワークNを介してバースト情報提供装置200へ送信した後、キーワードの注目度に関するバースト情報をバースト情報提供装置200から受信して表示する。

20

【0033】

バースト情報提供装置200は、ユーザ端末100-1～100-mからキーワードを受信すると、キーワードを含むブログデータをブログサーバ300から取得し、キーワードのバースト度を算出する。そして、バースト情報提供装置200は、キーワードのバースト度が高くバーストが発生している場合には、バーストの発生期間および発生位置を示すバースト情報をユーザ端末100-1～100-mへ送信する。また、バースト情報提供装置200は、キーワードのバースト度が低くバーストが発生していない場合には、バースト度の変遷を示すバースト情報をユーザ端末100-1～100-mへ送信する。なお、バースト情報提供装置200の構成および動作については、後に詳述する。

30

【0034】

ブログサーバ300は、ユーザ端末100-1～100-mからブログデータが送信されると、このブログデータを記憶・更新する。そして、ブログサーバ300は、ユーザ端末100-1～100-mからブログの閲覧要求が受信されると、指定されたブログデータをユーザ端末100-1～100-mへ送信する。さらに、ブログサーバ300は、バースト情報提供装置200によってブログの検索が実行されると、検索ヒットしたブログデータをバースト情報提供装置200へ出力する。

40

【0035】

図2は、本実施の形態に係るバースト情報提供装置200の要部構成を示すブロック図である。同図に示すバースト情報提供装置200は、情報のバーストが伝播する状態を抽出する情報伝播抽出装置を含んでいる。具体的には、バースト情報提供装置200は、検索要求受信部201、検索部202、時系列整列部203、バースト度算出部204、バースト判定部205、位置推定部206、バースト位置解析部207、結果ページ生成部208、およびバースト情報送信部209を有している。

【0036】

検索要求受信部201は、ユーザ端末100-1～100-mから送信された検索要求

50

を受信する。すなわち、検索要求受信部 201 は、ユーザがバースト情報の取得を希望するキーワードを含む検索要求を受信する。

【0037】

検索部 202 は、ブログサーバ 300 に記憶された全ブログを検索して、検索要求に含まれるキーワードが記載されているブログを選別する。そして、検索部 202 は、選別された（すなわち、検索ヒットした）ブログのブログデータをブログサーバ 300 から取得する。ブログデータには、ユーザが投稿した日時ごとの 1 つ以上の記事のデータが含まれている。なお、本実施の形態においては、検索部 202 が直接ブログサーバ 300 内を検索するものとしたが、ブログサーバ 300 の代わりにブログのインデックスを保持するデータベースを検索するようにしても良い。この場合、データベースには、定期的なクロール
10

【0038】

時系列整列部 203 は、検索部 202 によって取得されたブログデータを時系列順に整列する。すなわち、ブログは日記形式のウェブサイトであるため、ブログデータにはそれぞれの記事が投稿された日時が記録されており、時系列整列部 203 は、各記事のブログデータの投稿日時を比較しながら、時系列順に整列する。

【0039】

バースト度算出部 204 は、時系列順に整列されたブログデータにおけるキーワードの出現頻度を用いて、例えば 1 日などの期間単位でキーワードのバースト度を算出する。具体的には、バースト度算出部 204 は、例えば tf/idf (term frequency/inverse document frequency) 法などを用いて、ブログデータにおけるキーワードの出現頻度を統計的に評価し、キーワードに関する 1 日ごとのバースト度を算出する。このとき、キーワードの出現頻度が特定の期間において高くなっていると、その期間におけるキーワードのバースト度が高くなる。すなわち、バースト度は、キーワードの注目度を表す指標値となっている。注目度の指標値を求める際に tf/idf 法を用いることにより、単に出現頻度が高いキーワードが注目されていることになるのではなく、普遍的に使用される単語ではないにもかかわらず出現頻度が高いキーワードが注目されていることになる。
20

【0040】

バースト判定部 205 は、期間ごとのバースト度と所定の閾値とを比較し、バースト度が所定の閾値以上である期間においてキーワードのバーストが発生していると判定する。換言すれば、バースト判定部 205 は、キーワードのバースト度が他に比べて突出している期間をキーワードが注目された期間と判定する。したがって、バースト判定部 205 は、例えば 1 日ごとに断続的にバースト度が所定の閾値以上となっていれば、1 日単位のバーストが複数回発生したと判定し、例えば 1 週間連続してバースト度が所定の閾値以上となっていれば、1 週間連続するバーストが 1 回発生したと判定する。そして、バースト判定部 205 は、それぞれのバーストに対応する期間をバースト位置解析部 207 へ通知する。なお、バースト判定部 205 は、キーワードについてのバーストが発生していなければ、全期間におけるバースト度の変遷をバースト位置解析部 207 へ出力する。
30

【0041】

位置推定部 206 は、検索部 202 によって取得されたブログデータそれぞれを投稿したユーザの位置を推定する。具体的には、位置推定部 206 は、ブログデータにユーザが加入するプロバイダや投稿時に使用されたユーザ端末 100-1~100-m の IP アドレスなどが含まれていれば、これらの情報からユーザの居住地もしくはユーザがブログを投稿した国や地方などの位置情報を特定する。また、位置推定部 206 は、ユーザが加入するプロバイダや IP アドレスが不明であれば、ブログデータの文面における言葉遣いや地域を示唆する単語などからユーザの位置情報を推定する。すなわち、位置推定部 206 は、ブログデータの文面に対する形態素解析などを実行し、文章中の方言や固有名詞などから、ブログを投稿したユーザの位置情報を推定する。
40

10

20

30

40

50

【0042】

バースト位置解析部207は、バースト判定部205からバーストの発生期間が通知されると、各バーストに属するブログデータの位置情報を解析して、バーストが発生した位置を決定する。具体的には、バースト位置解析部207は、バーストが発生した期間にキーワードが記載されたブログデータの位置情報を位置推定部206から取得し、所定割合以上のブログデータに共通する最小の地域においてバーストが発生したと決定する。すなわち、例えばバーストに属するブログデータの8割以上が東京都のユーザによって投稿されていれば、このバーストの位置を東京都と決定し、例えばバーストに属するブログデータが日本各地のユーザによって投稿されていれば、このバーストの位置を日本と決定する。そして、バースト位置解析部207は、各バーストの発生期間および発生位置を結果ページ生成部208へ出力する。なお、バースト位置解析部207は、全期間におけるバースト度の変遷がバースト判定部205から出力された場合には、このバースト度の変遷を結果ページ生成部208へ出力する。

10

【0043】

結果ページ生成部208は、バースト位置解析部207からバーストの発生期間および発生位置が出力されると、地図上に各バーストの発生期間および発生位置を表示する結果ページを生成する。具体的には、結果ページ生成部208は、例えばバーストが狭い地域から広い地域へ徐々に広がっているような場合には、それぞれのバーストの発生位置が地図上で区別可能に囲まれた結果ページを生成する。また、結果ページ生成部208は、例えばバーストが複数の地域を順に伝搬しているような場合には、それぞれのバーストの発生位置が発生順に矢印などで結ばれた結果ページを生成する。なお、結果ページ生成部208は、バースト位置解析部207からバースト度の変遷が出力された場合には、このバースト度の変遷を例えば折れ線グラフとして表示する結果ページを生成する。

20

【0044】

バースト情報送信部209は、結果ページ生成部208によって生成された結果ページをキーワード送信元のユーザ端末100-1~100-mへ送信する。バースト情報送信部209が送信する結果ページには、バーストの発生期間および発生位置を視覚的に示すバースト情報か、またはバースト度の変遷を視覚的に示すバースト情報が含まれている。

【0045】

次いで、上記のように構成されたバースト情報提供装置200の動作について、図3に示すフロー図を参照して、具体的に例を挙げながら説明する。なお、以下の説明においては、ユーザ端末100-1がキーワードに関するバースト情報の取得を目的とした検索要求を送信したものとする。

30

【0046】

ユーザ端末100-1から送信された検索要求は、ネットワークNを介してバースト情報提供装置200の検索要求受信部201によって受信される(ステップS101)。受信された検索要求には、ユーザがバースト情報の取得を希望するキーワードが含まれているため、検索部202によって、ブログサーバ300に記憶された全ブログが検索され、キーワードが記載されたブログのブログデータが取得される(ステップS102)。そして、このブログデータは、時系列整列部203および位置推定部206へ出力される。

40

【0047】

ブログデータが位置推定部206へ出力されると、位置推定部206によって、各ブログデータに付加された、ユーザが加入するプロバイダやユーザ端末100-1のIPアドレスの情報から、各ブログデータを投稿したユーザの位置情報(例えば居住地やブログ投稿場所)が推定される(ステップS103)。また、プロバイダやIPアドレスの情報がブログデータに付加されていなければ、位置推定部206によって、ブログデータの各記事の文面が解析され、文中の例えば方言や固有名詞などからユーザの位置情報が推定される。

【0048】

同時に、ブログデータが時系列整列部203へ出力されると、時系列整列部203によ

50

って、各ブログデータの記事が時系列に整列される（ステップS104）。そして、バースト度算出部204によって、1日などの期間ごとにキーワードのブログデータにおける出現頻度が評価され、期間ごとのバースト度が算出される（ステップS105）。これにより、例えば図4に示すように、1日ごとのバースト度の変遷が得られる。図4においてバースト度が高くなっている期間は、主にブログデータ中のキーワードの出現頻度が他の期間に比べて高くなっている期間である。そして、ブログデータ中の出現頻度が高いということは、キーワードが多くユーザによって注目された結果、キーワードに関連する話題に言及するブログが多く投稿されているということに他ならない。つまり、バースト度が高い期間には、多くのユーザがキーワードに関心を持っていることになる。

【0049】

10

こうしてブログデータが投稿された全期間にわたるバースト度が算出されると、バースト判定部205によって、バーストの発生の有無を判定する対象月日が設定される（ステップS106）。すなわち、ここでは、例えば全期間の初日が対象月日に設定され、バースト判定部205によって、対象月日のバースト度が所定の閾値以上であるか否かが判定される（ステップS107）。この結果、バースト度が所定の閾値以上であれば（ステップS107Yes）、対象月日においてキーワードがバースト状態にあると判定される（ステップS108）。一方、バースト度が所定の閾値未満であれば（ステップS107No）、対象月日においてキーワードが非バースト状態にあると判定される（ステップS109）。

【0050】

20

対象月日に関するバースト状態・非バースト状態の判定が完了すると、引き続きバースト判定部205によって、バースト度が算出されている全期間についてバースト状態・非バースト状態の判定が終了したか否かが判断され（ステップS110）、ここでは、初日に関する判定が完了したのみであるため（ステップS110No）、新たにバーストの発生の有無を判定する対象月日が設定される（ステップS106）。すなわち、ここでは、全期間の2日目が対象月日に設定され、再びバースト判定部205によって、2日目においてキーワードのバーストが発生しているか否かが判定される。

【0051】

このようにバースト判定部205によるバースト発生の有無の判定が繰り返され、全期間に関する判定が終了すると（ステップS110Yes）、2日間以上連続してバースト状態と判定された場合には、バースト状態と判定された期間が1つのバーストとしてまとめられ、全期間中のバーストが検出される。具体的には、例えば図5に示すように、バースト度が所定の閾値以上となる「バースト#1」および「バースト#2」がそれぞれバーストとして検出される。

30

【0052】

なお、本実施の形態においては、バースト判定部205が1日単位でバースト度が所定の閾値以上であるか否かを判定することにより、対象月日においてキーワードがバースト状態にあるか否かの判定を繰り返した。しかし、図5に示す「バースト#1」および「バースト#2」を検出するためには、バースト判定部205は、バースト度が所定の閾値以上となる期間を全期間から抽出するようにしても良い。また、バースト判定部205によってバーストが1つも検出されない場合には、全期間にわたるバースト度の変遷がそのままバースト位置解析部207を介して結果ページ生成部208へ出力され、バースト度の変遷を例えば折れ線グラフにして示す結果ページが生成され、バースト情報送信部209からユーザ端末100-1へ送信される。

40

【0053】

バースト判定部205によってバーストが検出されると、それぞれ検出されたバーストに対応する期間がバースト位置解析部207へ通知される。そして、バースト位置解析部207によって、それぞれのバーストに対応する期間に投稿されたブログデータを特定され、バーストごとのブログデータを投稿したユーザの位置情報が解析され（ステップS111）、バーストごとの位置が決定される。具体的には、バースト位置解析部207によ

50

って、バーストに属するブログデータを投稿したユーザの位置情報が位置推定部206から取得され、位置情報によって示される地域のうち所定割合以上のユーザに共通する最小の地域がバーストの位置と決定される。

【0054】

バースト位置の決定について具体例を挙げると、例えば図6に示すように、「バースト#1」が2007年10月25日に発生しており、「バースト#1」に属するブログデータのうち所定割合以上のブログデータの位置情報が「日本国東京都」であれば、「バースト#1」の位置は「東京都」と決定される。同様に、「バースト#2」が2007年11月3日から2007年11月7日に発生しており、「バースト#2」に属するブログデータのうち所定割合以上のブログデータの位置情報が関東地方に該当する都県であれば、「バースト#2」の位置は「関東地方」と決定される。さらに、例えば、バーストに属するブログデータの位置情報が日本各地に分散していれば、このバーストの位置は「日本」となる。

10

【0055】

バースト位置解析部207によってバーストごとの位置が決定されると、それぞれのバースト位置が結果ページ生成部208へ出力され、結果ページ生成部208によって、バースト位置を視覚的に確認可能な結果ページが生成される（ステップS112）。すなわち、例えば地図上にそれぞれのバースト位置を囲んで表示したり、それぞれのバースト位置間を矢印で結んで表示したりする結果ページが生成される。

【0056】

20

具体例を挙げると、例えば図7に示すように、「東京●●タウン」というキーワードのバーストが3つ検出された場合、結果ページ生成部208によって、それぞれのバーストのバースト位置を日本地図上に囲んでバーストの期間とともに表示する結果ページが生成される。このような結果ページによれば、「東京●●タウン」というキーワードが最初に東京都近辺でバースト状態となり、その後関東地方近辺でバースト状態となり、続いて本州・四国・九州近辺でバースト状態となったということが把握可能となる。したがって、このような結果ページを閲覧するユーザは、例えば「東京●●タウン」に関する話題が東京から発信され、徐々に周囲の地域に広がっていったなどと推測することが可能となる。

【0057】

また、例えば図8に示すように、「石油」というキーワードのバーストが3つ検出された場合、結果ページ生成部208によって、それぞれのバースト位置を世界地図上で結んでバーストの期間とともに表示する結果ページが生成される。このような結果ページによれば、「石油」というキーワードが最初にサウジアラビアなどの中東付近でバースト状態となり、その後アメリカ付近でバースト状態となり、続いて日本でバースト状態となったということが把握可能となる。したがって、このような結果ページを閲覧するユーザは、例えば「石油」の価格上昇が最初に中東の社会情勢に影響を与え、この影響がアメリカ、日本へと連鎖したなどと推測することが可能となる。

30

【0058】

結果ページ生成部208によって結果ページが生成されると、生成された結果ページは、バースト情報送信部209から検索要求送信元のユーザ端末100-1へ送信される（ステップS113）。そして、ユーザは、ユーザ端末100-1に表示される結果ページを閲覧し、キーワードに関するバースト情報を得ることができる。すなわち、キーワードについてバーストが発生していれば、それぞれのバーストの期間と位置を視覚的に確認することができ、キーワードについてバーストが発生していなければ、全期間にわたるキーワードのバースト度の変遷を折れ線グラフなどで確認することができる。そして、ユーザは、これらのバースト情報を、例えば商品開発のための需要予測や株取引などに利用することができる。

40

【0059】

以上のように、本実施の形態によれば、ブログにおけるキーワードの出現頻度からキーワードのバースト度を算出し、バースト度が所定の閾値以上となる期間をキーワードのバ

50

ーストとして検出し、それぞれのバーストに属するブログのユーザの位置情報を解析してバーストごとの位置を決定する。そして、バーストの期間のみではなく、バーストごとの位置をユーザへ提供するため、ユーザは、キーワードに関する話題の伝搬状況など、バーストの発生位置に関する詳細な情報を得ることができる。換言すれば、バースト情報提供装置は、ネットワークの利用を通じて世間の関心事の特性をより詳細に分析する際に必要となる情報を提供することができる。

【0060】

なお、上記一実施の形態においては、ブログにおけるキーワードの出現頻度を基にしてバースト度を算出するものとしたが、例えば検索サイトにおける検索キーワードの検索頻度や掲示板サイトへの書き込みにおけるキーワードの出現頻度などを基にしてバースト度を算出しても良い。さらに、ブログ、検索サイト、および掲示板サイトを2つ以上組み合わせてバースト度を算出しても良い。これらの場合には、検索サイトや掲示板サイトへアクセスしたユーザのIPアドレスなどからユーザの位置情報を推定すれば良い。また、上記一実施の形態において説明した情報伝播抽出方法をコンピュータが実行可能な形式で記述した情報提供プログラムを生成し、この情報提供プログラムをコンピュータに実行させることにより本発明の情報伝播抽出装置および情報伝播抽出方法を実現することも可能である。このとき、情報提供プログラムをコンピュータが読み取り可能な記録媒体に記憶させ、記録媒体を用いてコンピュータに情報提供プログラムを導入することも可能である。

【産業上の利用可能性】

【0061】

本発明は、ネットワークの利用を通じて世間の関心事の特性をより詳細に分析する際に必要となる情報を提供する場合に適用することができる。

【図面の簡単な説明】

【0062】

【図1】一実施の形態に係る通信システムの概要を示すブロック図である。

【図2】一実施の形態に係るバースト情報提供装置の要部構成を示すブロック図である。

【図3】一実施の形態に係るバースト情報提供装置の動作を示すフロー図である。

【図4】一実施の形態に係るバースト度の変遷の例を示す図である。

【図5】一実施の形態に係るバースト判定の例を示す図である。

【図6】一実施の形態に係るバースト位置解析結果の例を示す図である。

【図7】一実施の形態に係る結果ページの例を示す図である。

【図8】一実施の形態に係る結果ページの他の例を示す図である。

【符号の説明】

【0063】

- 200 バースト情報提供装置
- 201 検索要求受信部
- 202 検索部
- 203 時系列整列部
- 204 バースト度算出部
- 205 バースト判定部
- 206 位置推定部
- 207 バースト位置解析部
- 208 結果ページ生成部
- 209 バースト情報送信部
- 300 ブログサーバ

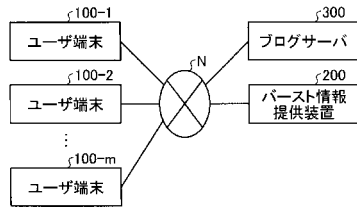
10

20

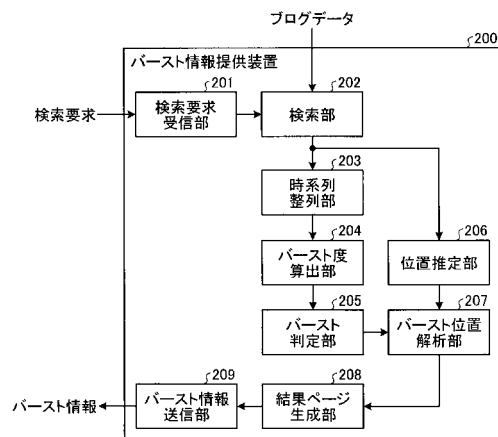
30

40

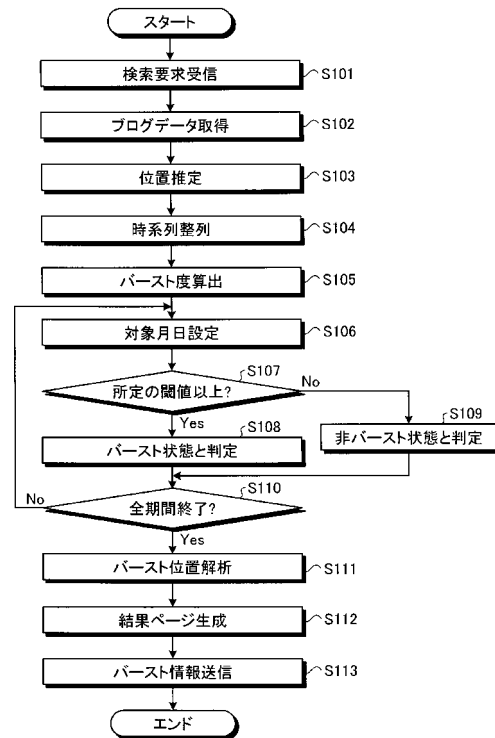
【図 1】



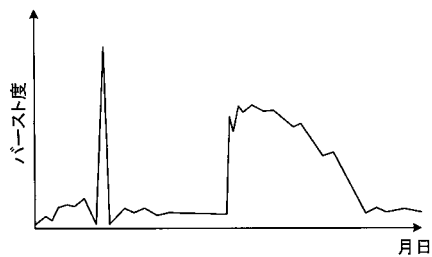
【図 2】



【図 3】



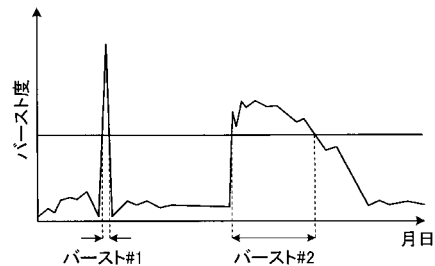
【図 4】



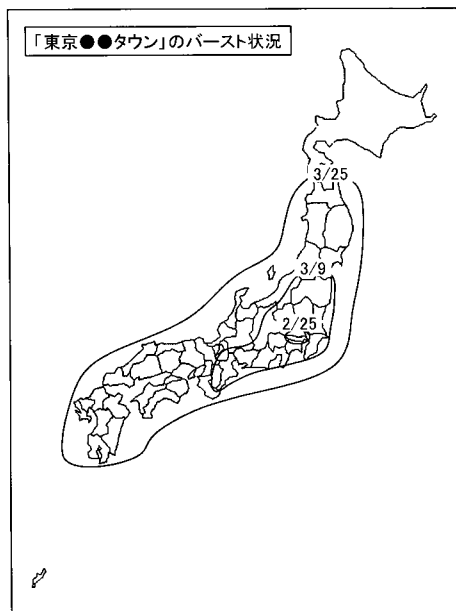
【図 6】

バースト	月日	対象ブログURL	位置	バースト位置
バースト #1	2007/10/25	http://AAA.BBB.CCC	日本国東京都	東京
		http://BBB.CCC.DDD	日本国東京都	
		⋮	⋮	
		http://CCC.DDD.EEE	日本国大分県	
バースト #2	2007/11/3 ～ 2007/11/7	http://DDD.EEE.FFF	日本国神奈川県	関東地方
		http://EEE.FFF.GGG	日本国茨城県	
		⋮	⋮	
		http://FFF.GGG.HHH	日本国東京都	
		⋮	⋮	

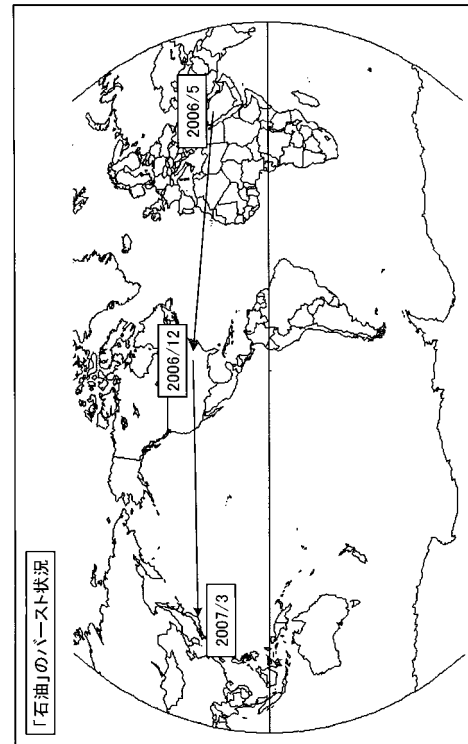
【図 5】



【図 7】



【図 8】



フロントページの続き

(58)調査した分野(Int.Cl., DB名)

G06Q 10/00-50/00