

# Weblog 記事を用いた社会事象に関する関心动向の分析

\*福原知宏<sup>†1</sup> 村山敏泰<sup>†1</sup> 西田豊明<sup>†2</sup>

<sup>†1</sup> 科学技術振興機構 社会技術研究開発センター(社会人参加)  
〒100-0004 東京都千代田区大手町 1-1-2 リそな・マルハビル 18F

<sup>†2</sup> 京都大学大学院情報学研究科  
〒606-8501 京都市左京区吉田本町  
連絡先 : fukuhara@ristex.jst.go.jp

## Abstract:

Weblog 記事の収集・解析を行うシステム: KANSHIN を開発し, 社会事象に関する関心动向の分析を行った. 本論文ではこのシステムを用いた分析結果として(1)社会的関心のパターン, (2)個人の関心动向, (3)社会的関心と実世界データとの相関について述べる.

## 1. はじめに

Weblog 記事を用いた社会事象に関する社会の関心动向について述べる. 近年, JR 西日本の脱線事故や 2004 年に発生した地震, 台風など, 我々の身の回りには社会や生活に係る様々な問題が存在している. これらの問題に対して社会全体が取り組んでいくためには, 問題の発生メカニズムと対策についての科学的検討に加え, 人々が日頃どのような問題に関心を持ち, ひとたび問題が発生した際にどう振る舞ったかという社会科学的検討も重要である. 本研究では後者の観点から Weblog 記事を用いた社会の関心动向の分析を行っている. Weblog は近年 World Wide Web 上で普及しつつある情報発信ツールであり, 多くの人々が Weblog 上で情報発信を行っている. 筆者らは Weblog 記事の自動収集・解析を行うシステム: KANSHIN を開発し, 近年の社会事象に関する分析を行った. 以下ではシステムの概要について述べ(2 節), 本システムを用いて得られた分析結果として Weblog に見られる関心パターン(3 節), 個人の関心动向(4 節), 社会的関心と気象データとの関係(5 節)について述べる.

## 2. Weblog 記事を用いた関心解析システム: KANSHIN

Weblog 記事を用いた関心解析システム: KANSHIN について述べる. Figure 1 にシステムの全体像を示す. 本システムは Weblog サイトから RSS 及び Atom ファイルを定期的に収集し, 解析する. システムは日本と中国の Weblog サイトから収集した記事を記事テーブルに格納すると共に, 記事に対して形態素解析を行い, 名詞と形容詞を抽出して索引テーブルに格納する. 形態素解析には

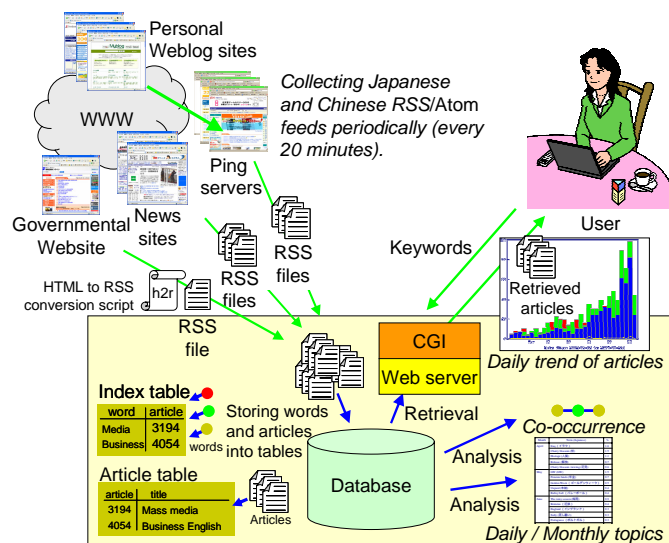


Figure 1. システム全体像

ChaSen<sup>1</sup> (日本語の場合)と ICTCLAS<sup>2</sup> (中国語の場合)を用いた。システムは1日に2万件(日本語),2千件(中国語)の記事を収集しており,これまでに1千万件(日本語),46万件(中国語)の記事を収集している。システムには次の機能がある。(1)記事検索機能:利用者の指定したキーワードで記事を検索し,記事数の推移を示すグラフと記事リストを提供する。(2)共起語検索機能:利用者の指定したキーワードの共起語を検索する。(3)話題検出機能:ある日の話題やある月に特徴的なキーワードを話題語として検出し,利用者にメールで通知する。

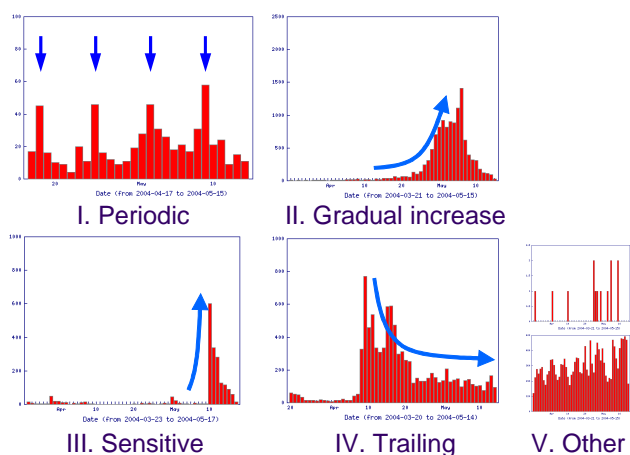


Figure 2. Weblogに見られる関心パターン

### 3. Weblog における社会的関心：社会的関心の分類

Weblogに見られる社会的関心を5つのパターンに分類した。Figure 2にパターンの一覧を示す。社会的関心のパターンは(1)周期型,(2)漸次増加型,(3)突発型,(4)関心持続型,(5)その他である。以下,それぞれについて述べる。

#### 3.1 周期型：“冬のソナタ”,“給料日”など

周期型パターンは周期的に記事数が増える出現するパターンである。このパターンには(1)テレビ番組の名前(“冬のソナタ”,“エンタの神様”など),(2)生活に関連する周期的な出来事(“授業”,“週末”,“給料日”など)が該当する。長期的観点では“夏休み”,“オリンピック”もこれに該当する。

#### 3.2 漸次増加型：“GW”,“台風”,“クリスマス”など

漸次増加型パターンは徐々に記事数が増えるパターンである。Figure 2のパターンIIのグラフは2004年4月から5月前半にかけての“GW”を含む記事数の推移である。この図に見られるように記事数は連休に向かって徐々に増加し,連休が過ぎると共に徐々に減少している。このように,人々が事前にある出来事の内容を把握しており,その出来事について強い関心を持っている場合,漸次増加型パターンが生じると考えられる。この他の例として“台風”,“選挙”,“夏”などが存在する。

#### 3.3 突発型：“Winny”,“地震”など

突発型パターンはある日を境に急激に記事数が増加するパターンである。このパターンは,ある出来事に対して人々がその存在を予期しておらず,かつその出来事に強い関心を抱いている場合に生じる。Figure 2のパターンIIIのグラフは2004年4月から5月前半にかけての“Winny”を含む記事数の推移である。グラフ中のピークは2004年5月10日のWinny 開発者逮捕を受けた人々の反応であり,人々がこの問題に強い関心を持っていたことが分かる。このパターンに属する他の話には“地震”がある。

<sup>1</sup> <http://chasen.naist.jp/> (2005-06-02 確認)

<sup>2</sup> <http://mtgroup.ict.ac.cn/~zhp/ICTCLAS.htm> (2005-06-02 確認)

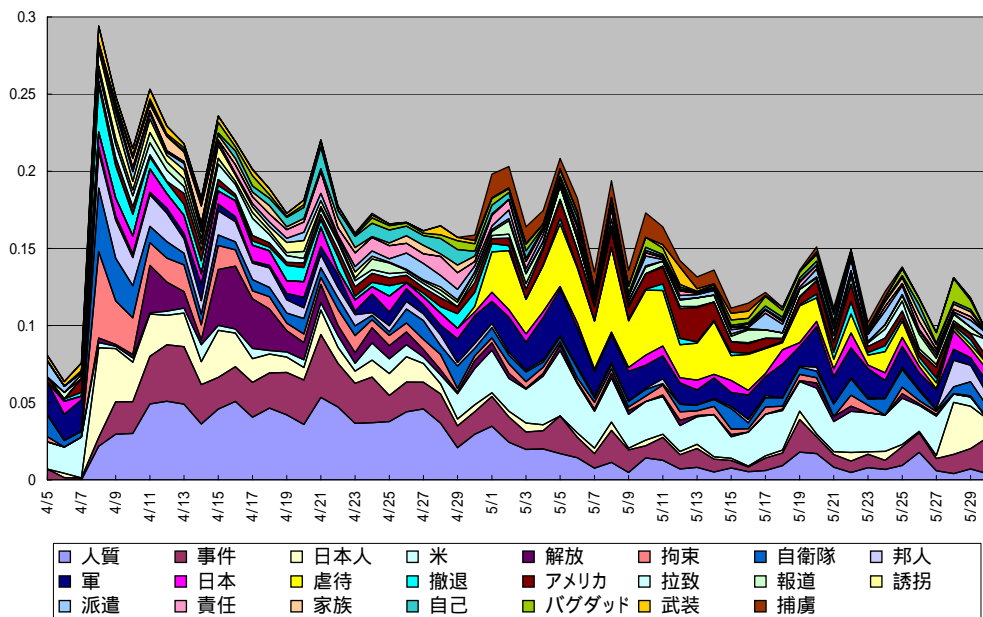


Figure 3. “イラク”に対する共起語の推移 (2004/04/05 から 2004/05/30 まで)

### 3.4 関心持続型：“イラク”，“ニッポン放送”，“JR” など

ある出来事が発生した後に、人々の関心が持続するパターンである。Figure 2 のパターン IV のグラフは“イラク”を含む記事数の推移である。2004 年 4 月に発生したイラク人質事件により、人々の関心は急激に高まり、持続的な関心となったことが分かる。Figure 3 に“イラク”と共起する単語の推移を示す。y 軸は Dice 係数による共起度である。社会の関心は当初、“人質”，“事件”，“日本人”，“自衛隊”などに当てられていたが、時間経過とともに人質の“解放”（4 月 11 日，15 日付近），自己責任（4 月 23 日付近），米軍による捕虜虐待事件（5 月 1 日付近），日本人襲撃事件（5 月 28 日付近）と推移していったことが分かる。このパターンに該当する他の語には“ニッポン放送”，“JR”があげられる。

### 3.4 その他：“日記”，“原発”，“水不足” など

上に挙げたパターンのいずれにも該当しないパターンである。これに該当する語は(1)日常的に使われている語(“日記”，“今日”，“ブログ”など)や，(2)人々が現時点において関心を示していない語(“食糧問題”，“水不足”，“海洋汚染”)などである。ただし後者に属する語でも一旦問題が顕在化すると突発型や関心持続型に移行する例も見られる。

## 4. 個人の関心動向に関する分析

個人が社会問題に対してどのような関心を持っているかという観点から，ここでは小泉首相のメールマガジン<sup>3</sup>を用いて個人の関心動向を分析した。Figure 4 に小泉首相の関心動向を示す。x 軸は記事の発行日であり，y 軸は記事に含まれる単語数である。ここでは“SARS”，“BSE”，“(鳥)インフルエンザ”，“地震”，“テロ”，“台風”，“津波”という単語を用いた。この図から，小泉首相の関心はその時々での社会の出来事に応じて高まる半面，“テロ”以外の問題に対しては関心が継続していない事が分かる。

<sup>3</sup> <http://www.kantei.go.jp/jp/m-magazine/> (2005-06-02 確認)

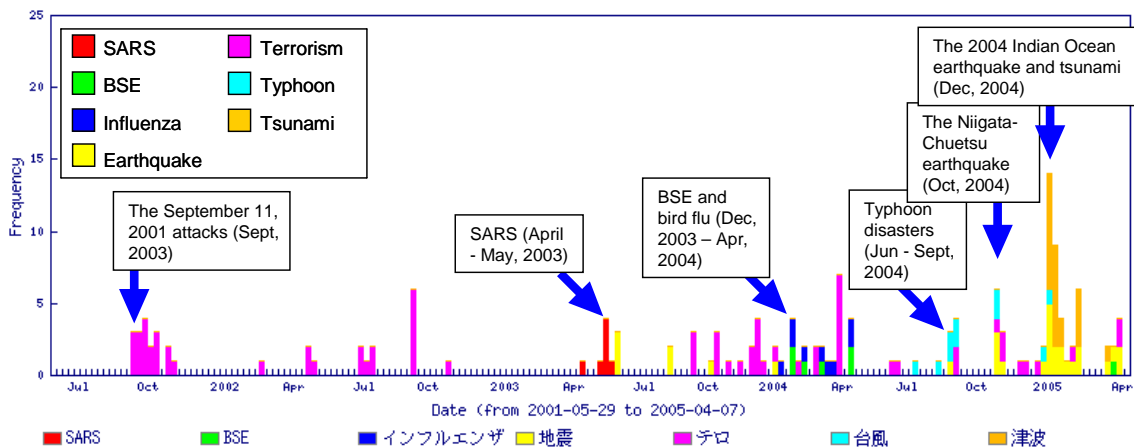


Figure 4. 社会問題に関する小泉首相の関心動向

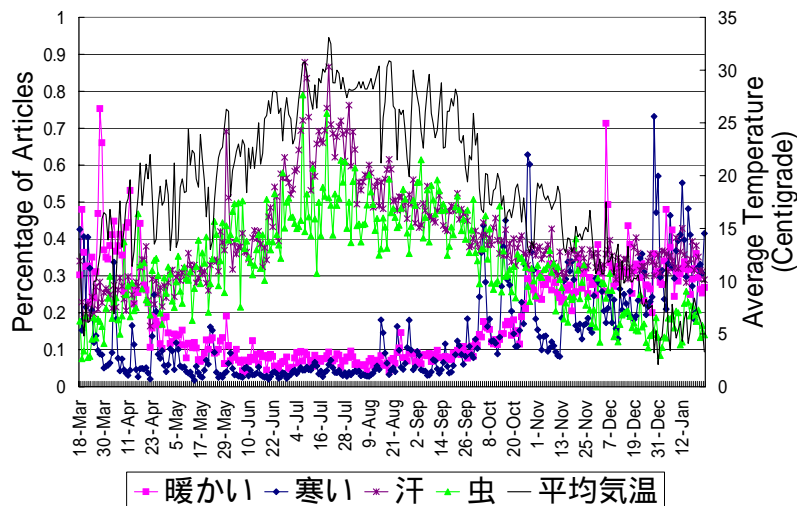


Figure 5. 東京の日中平均気温と Weblog 中に出現する単語の関係

## 5. 社会的関心と実世界データとの関係

Weblog に見られる関心が実世界のどのような刺激の影響を受けているかという観点から，ここでは気象データと Weblog に出現する単語との相関関係について調査した．Figure 5 に東京(大手町)の日中平均気温と単語(“暖かい”，“寒い”，“汗”，“虫”)を含む記事数の関係を示す．この図に見られるように，ある種の語は気温と相関関係を持って出現していることが分かる．

## 6. まとめ

Weblog を用いた関心解析システムと，本システムを用いた社会事象に関する関心分析の結果について述べた．Weblog の出現と Weblog 上での情報発信活動の活発化により，人々の社会問題についての反応や関心を容易に把握できるようになった．本研究の今後の課題は，(1)システムの多言語への拡張と多国間での関心比較，(2)他の実世界データ(新聞報道量や各種統計データ)を用いた関心分析である．