

SERPAnalyzer：社会調査支援のための SERP アーカイブからの特徴的ランク変動分析システム

中部文子^{†1} 渡辺 知恵美^{†1} 小山 直子^{†2}
館 かおる^{†2} 伊藤 一成^{†3}
深山 鷹一^{†3} 増永 良文^{†3}

商用検索エンジンによる検索結果ランキングは世の中の動きにあわせて日々変化している。我々は Web ページのランク変動を観察することで実世界で起こっている事象を発見・分析できるのではないかと考え、社会学者向けに Web ページのランク変動を収集・分析システム SERPWatcher, SERPAnalyzer を開発している。SERPWatcher は、7 つの商用検索エンジンにより検索結果ランキングを収集・提示しており、SERPAnalyzer は、そのデータを使って特徴的なランク変動の情報を自動抽出・提示している。ランクアップした箇所の時期やランク、期間を提示することで社会事象発生時期や規模、定着度を推定することができ、社会学者が社会概念形成のプロセスを探るにあたって、有益な情報を提示することができた。

SERPAnalyzer: An Analysis System for Noteworthy Ranking Transitions from SERP Archive for Social Survey Aid

FUMIKO NAKABE,^{†1} CHIEMI WATANABE,^{†1}
NAOKO OYAMA,^{†2} KAORU TACHI,^{†2} KAZUNARI ITO,^{†3}
YOICHI MIYAMA^{†3} and YOSHIFUMI MASUNAGA^{†3}

Recently, finding social concerns from Web is gaining an increasing interest. We focus on the Search Engine Result Pages (SERP) because we consider that the rank transitions of web pages are reflected social events and trend. Based on the consideration, we are developing SERP data archiving and visualization system: SERPWatcher and SERP analyzing system: SERPAnalyzer. SERPAnalyzer shows information of rank rising points. This system makes it easy and useful to social analysis from rank transitions.

1. はじめに

Web には、様々な主体が情報を発信し実世界での出来事や営為が写し込まれている。近年ウェブマイニングにより社会で起こっていることを読み解くことができるのではないかと期待が高まっているなか、我々は、社会の動きを映し出す 1 つの媒体として、SERP (Search Engine Result Page) に着目している。SERP とは検索エンジンにユーザがキーワードを入力したときにその結果として表示されるページのことである。

社会の動きに関する調査の中で、我々は事象に対する社会的な概念の動きに関する調査を取り上げる。概念および社会的概念に関して哲学的に様々な解釈が存在するが、本稿では事象に対する概念を「ある事象に対する解釈、意見、考え」とする。事象に対する概念は人によって様々であるが、その中で多くの人々によって注目され共有されているものや、一部の人々によってのみ注目されたり共有されたりしている概念もある。また、それらの概念は事件や政策、流行など社会で起こる出来事などの影響を受け日々変化する。このように社会を構成する人々によって共有される「事象に対する概念」の動きを、本稿では社会的な概念の動きと呼ぶこととする。

たとえばジェンダーフリーとは社会的、文化的に形成された男女の役割のことを表すが、人によって男女の固定概念は本来の個人の機会を縛ってきたものであるから開放すべきであるという概念を持つ人や、これまでの歴史や文化で培われてきた男性らしさ女性らしさを尊重すべきだという概念を持つ人もいる。このような概念がマスメディアによって取り上げられたり、政策によって注目されたりするなどして、多くの人の共感を得たり、異なる概念を持つ者同士が論争を行うことによってその趨勢を争ったりすることもある。

我々は、このような社会的な概念の動きを反映する指標として、SERP が親和性が高いのではないかと考えた。つまり検索キーワードを「事象」とし、検索結果に登場する各 Web ページの内容を事象に対する概念であると考えれば、SERP におけるランキングの遷移は、多くの人によって注目されている概念の遷移ととらえることができる。先ほど取り上げたジェンダーフリーに関しても、上位ページがどのような概念に基づいて書かれたページが多

†1 お茶の水女子大学大学院人間文化創成科学研究科
Department of Computer Science, Ochanomizu University

†2 お茶の水女子大学ジェンダー研究センター
Institute for Gender Studies, Ochanomizu University

†3 青山学院大学社会情報学部
School of Social Informatics, Aoyama Gakuin University

いのか、またそれがどのように遷移するのかについて観察することによって、実世界における社会的な概念の動きについて調査するヒントとなる。

我々が SERP に着目した理由の 1 つはランキングアルゴリズムにある。たとえば Google の検索結果ランキングを決める指標の 1 つとして用いられている PageRank⁴⁾ では、多くの良質なページから参照されるページは良質なページと見なされ、検索結果の上位にランキングされる⁵⁾。また、上位にランキングされるページが多く利用者の閲覧されることによって、ランキングの結果そのものがサイト内容の信用性や価値を代弁するものとなり、社会的な概念の形成に大きく影響する可能性も示唆されている⁷⁾。

そこで我々は、SERP におけるランキングの推移を観察および検索エンジンごとと比較したりランキングの推移を分析したりするシステムを提案する。

我々の提案するシステムは、ランキング履歴収集・提示を行う SERPWatcher¹⁾⁻³⁾、*1 とランキング履歴から社会調査にあたって注目すべき変動を抽出・提示する SERPAnalyzer からなる。SERPWatcher は、利用者が指定する検索キーワードに対して、7 つの検索エンジンの検索結果ランキングを定期的に収集し、ランキング履歴を提示している。ユーザは過去にどのような Web ページが何位にランクしてその後どのようにランクを変化させたかなどを確認することができる。さらに、社会的な概念の遷移を探るにあたり調査者はその手がかりとなるページに焦点を定める必要がある。人々の概念形成に影響を与えたページはどのようなものかを考えたとき、我々は、ランクアップしてからすぐにはランクダウンするか、あるいはしばらく浮上し続けるのかという情報が調査の材料の 1 つになるのではないかと考えた。そこで SERPAnalyzer では、SERPWatcher で蓄積したランキング履歴からランクアップしたページを抽出し、それらの「浮上開始日」、「浮上ランク(最大何位までランクアップしたか)」、「浮上期間」を提示した。本システムを使った分析を行った結果、ランク浮上したページ数の推移から、社会事象の発生時期を推測できることや、浮上ランク・浮上期間をもとに調査対象ページを絞り込んでいくことが容易にできるようになった。また、これを対話的に繰り返すことで検索キーワードに関する事象についてどのような出来事が起こったのか、またそれらのその後の定着したのかについて推測する手がかりが得られるようになった。

本稿の構成は以下のとおりである。まず 2 章で検索エンジンにおけるランキングの仕組みおよび、これを社会科学調査を支援するために必要なシステム要件について述べる。3 章

で SERPWatcher, 4 章で SERPAnalyzer について述べ、5 章で提案システムを使った社会分析を 3 つのケーススタディを用いて議論する。6 章で関連研究、7 章でまとめと今後の課題を述べる。

2. 検索エンジンのランク変動を利用した社会調査

本章では、検索エンジンのランキングの仕組みについて述べ、検索エンジンのランキングを社会調査支援へ利用するために必要な機能について述べる。

2.1 検索エンジンのランキング

本節では検索エンジンが採用する基本的なランキングアルゴリズムについて述べ、事象に対する社会的な概念の変遷を調査するために検索エンジンのランキングデータを利用する理由について述べる。

2011 年 3 月 7 日のデータ⁹⁾によれば、日本のインターネット利用比率は Yahoo! Japan が 55.6%, Google が 37.6% となっており、この 2 つの検索エンジンが多くの人に使われている。また、独自のディレクトリ型検索エンジンを有していた Yahoo! Japan が 2010 年 7 月に Google の検索エンジンを利用することを発表した¹⁰⁾ ことから、事実上、2011 年 3 月現在において日本国内で検索サービスを利用するほとんどの人が Google の検索エンジンを利用している。以降、Google を例に検索エンジンのランキングの基礎的な仕組みである PageRank について述べる。PageRank とは「多くの良質なページからリンクされているページは、やはり良質なページである」という考えに基づいて計算された Web ページの点数で、式 (1) で表される。

$$R'(u) = c \sum_{v \in B_u} \frac{R'(v)}{N_v} + (1-c)E(u) \quad (1)$$

ウェブページ u のランク $R'(u)$ は式 (1) で表される。 B_u は、 u にリンクしているページ集合、 N_v はページ v がリンクしているページ集合の要素数である。また、 $E(u)$ はランクの源となる Web ページ群中のあるベクトルであり c は正規化用定数である。各ウェブページは自分のランク (R_u) を持ち、他のページにリンクすると自分のランクを分配する、自分のランクは他のページからももらったランクの合計で決まる。また、検索エンジンでは PageRank のほかにランクを決める指標として、たとえばアンカーテキストを考慮したり、ページ内の単語の位置や大きさを考慮するなどが知られている(ただしどのようにランキング結果に反映させているかは公表されていない)。このように検索エンジンはページのリン

*1 <http://serp.si.aoyama.ac.jp>

ク関係やページの文面, その他の方法により Web ページを評価し, 「より多くの人から注目されていて, かつ信頼度の高いページ」を上位にランキングしているといえる。

我々が検索結果ページを社会分析に利用できると思った理由の 1 つはこのようなランキングの仕組みにある。つまり, 検索キーワードを「事象」とし, 検索結果に登場する各 Web ページの内容を事象に対する概念であると考えれば, 検索結果ランキングの遷移は, 多くの人によって注目されている概念の遷移ととらえることができるからである。また検索エンジンにおけるランキングデータの更新は近年非常にスピードアップしている。2010 年 6 月に Google が Caffeine という次世代インデックス技術を発表し, 従来より 50%新しいページをランキング結果に表示できるようになっている。これらの技術向上により, 最近では数時間前に作成された Web ページが検索結果として反映されるようになっている。検索エンジンは日々刻々変化する良質な Web ページをつねに求めることができるため, ランキングの変動を定期的に取得し観察することで, 多くの人によって注目される概念がどの時期にどのように変化したかを分析するための手がかりとすることができる。

一方で, 検索結果ランキングを提示すること自体が, 検索エンジンの利用者が事象に対して持つ概念に影響を与えている可能性も示唆されている⁷⁾。文献 6) によると, Google の検索結果画面の第 1 ページ目の 1 位から 7 位までのエリアはゴールドトライアングルと呼ばれ, 検索者の 50%から 100%の視認率があるといわれている。上位にランキングされれば, 利用者に認識してもらえるが, 下位にランキングされると視認率が大幅に下がり, 利用者へ認識されない恐れがある。検索エンジンで検索を行った人は事象に対しての情報やほかの人が持つ概念を調べるために検索を行う者も多い, それらのユーザが検索結果上位にランキングされるページを閲覧することにより, その Web ページに記述される「概念」に着目したり, そこからまた新たな概念が派生したりする可能性が高くなり, ランキングの結果そのものが社会的な概念の形成に大きく影響すると考えられる。

以上にあげた検索エンジンの結果ランキングの特徴は検索ランキングを社会調査支援に利用する際に留意すべき問題点をはらんでいる。それは検索エンジン会社のランキングの方針に社会分析そのものが影響を受けてしまうという点である。つまり検索エンジンは PageRank のような基本アルゴリズムは公開しているものの, そのほか検索エンジン最適化 (SEO) への対策方法や詳細なランキング方式については公開していない。SEO やスパムページなどに対して検索エンジンは日々対策を行っているが, その一方で検索エンジンが特定のページを基本的なランキングとは別に上位に位置づけやすくする可能性も否定できない。そのため単一の検索エンジンの検索結果だけでなく, 複数の検索エンジンの検索結果を収集し,

これらを比較分析することによって, より多面的に分析をするよう注意する必要がある。また, 必ずしも検索エンジンのランキング結果だけを見て, 実世界での社会的な概念の遷移を正確にすべて読み取ることができるわけではないことに注意したい。そのため, 社会学者がランキング結果の遷移を分析することによって社会的な概念の遷移に関する手がかりを見つけたら, その検索ランキングの要因を探る手立てとしてニュースやバックリンク数などを提供し, 社会学者が次のステップとして社会調査 (アンケートや統計調査など) を進められるよう支援する必要がある。

2.2 ランク変動から社会的な概念の変遷を読み解くために必要な機能

前節に述べたような検索エンジンの特徴から, 本システムで備えるべき基本機能を以下の 2 点にまとめた。

- A) ランク変動から事象に対する概念の遷移を発見するための手がかりを提供する。
- B) ランク変動の複数エンジン間での比較など多角的に分析する仕組みを提供する。

第 1 点目の「ランク変動から事象に対する概念の遷移を発見するための手がかりを提供する」ために具体的に以下の機能が必要であると考えられる。

- A1) ランキングデータの収集
調査したいキーワードに関するランキングの推移を取得するため, 定期的かつ長期的にランキングをアーカイブする。また単一の検索エンジンだけでなく複数の検索エンジンによるランキングを同時にアーカイブする必要がある。
- A2) ランキングの遷移を一望できる可視化機構
ある時点でのランキング結果を起点に, その前後でどのようにランキングの遷移が起こったかを表示する可視化機構によってアーカイブデータを一望できるようにする必要がある。
- A3) 特徴的なランク変動の抽出および分析機構
膨大なランキング遷移データの中から特徴的なランク変動を抽出し, ランク変動を分析する支援機能を提供する必要がある。

2 点目の「ランク変動の複数エンジン間での比較など多角的に分析する仕組みを提供する」ために必要な機能としては以下の点が考えられる。

- B1) ランキングの対話的および多角的な可視化
要件 A2 とも関連するが, ランク変動やある 1 日のランク一覧を複数の検索エンジンで比較するなど多角的かつ対話的な可視化機能を提供する必要がある。
- B2) ランク変動に関連する出来事やランク変動の原因を調査するためのリソースの提供

Web ページのバックリンク数やバックリンクリスト, データを収集した日のニュースアーカイブを提供することによって, ランク変動が何と関連するのか, より深い分析に進むことを助ける .

3. SERPWatcher : ランキングヒストリ収集と提示

我々の提案するシステムは, 大きく分けて 2 つの機能からなる . 1 つは, ランキングを収集してヒストリを提示する機能 SERPWatcher, もう 1 つは, 社会分析を補助するためにヒストリから社会事象を特徴づけるようなランク変動特徴抽出して提示する機能 SERPAnalyzer である (図 1). 本章では, ランキング収集・提示機能 SERPWatcher について述べる .

3.1 ランキングヒストリ収集

まず前章の要件 A1 としてあげたように, 各検索エンジンでのランキングヒストリを収集する必要がある . SERPWatcher のランキングヒストリ収集機能では, 図 2 に示す検索キーワード登録画面を通して, 利用者は検索キーワードとランキング収集曜日を登録することができる . 検索キーワードは, ユーザの興味のあるキーワードを自由に入力することが可能である . 現在, “貧困”, “ジェンダーフリー”, “選挙”, “小沢一郎” など社会的事象に関連するものがこれまでに約 50 件登録されている .

キーワードが登録されると指定した曜日に検索結果収集が始まる . 7 つの検索エンジン Google, Yahoo!, goo, Baidu, Infoseek, Bing によって検索結果のページを 1 位から 500 位まで収集し, データベースに格納する . 検索結果ランキングの収集は, Google に関しては Google Web Service API¹⁶⁾, Yahoo! Japan は, ウェブ検索 API¹⁷⁾, Bing, Baidu, goo, Infoseek, excite では API を提供していないため, HTML 文書を取得しランキング情報を抽出することにより行った . 501 位以下のランキングは収集しておらず, 500 位以内に現れなかったページはすべて 501 位と見なす . また, ランキングを構成する各 Web ページのページアーカイブと一部のバックリンク数も保存している . バックリンクの収集は, 取得可能である Yahoo! と Google の場合にのみ行っている . Yahoo! では, Yahoo! Site Explorer API¹⁷⁾ を利用した . Google ではバックリンク数取得の API を提供していないため, 検索語として “link:対象ページの url” と入力することでバックリンク数とリンクしているページを得ることができるサービス¹⁸⁾ を利用している . ただし, この方法で得られるリンク数は実際リンクされているよりも少なく表示されていると推測され, 正確なバックリンク数は取得できていないと思われる . しかし, そのページがどのようなサイトからリンクされているかを確認することはページの情報発信者がどのような人に興味を持たれている

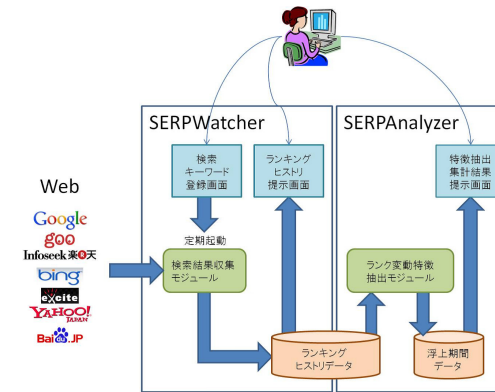


図 1 提案システムの概念図

Fig. 1 Process flow of the proposed system.



図 2 検索キーワード登録画面

Fig. 2 Interface for registering keywords.

いるかなどを推測する材料となるなど, 調査のうえでは有益であると考えている .

3.2 提示

次に, 前章の要件の A2 で述べたように, 収集したランキングの遷移を一望できる可視化機能が必要である . 本節では SERPWatcher による可視化結果の提示方法について述べる . 図 3 に, ランキングヒストリ提示画面の一例を示す . 収集したランキング結果は随時, ランキングヒストリ提示画面に反映され, そこから, 利用者はどの検索エンジン・検索キーワードで検索した結果がいつどんなランキングだったのかを確認することができる . こちらは, Google で “貧困” と検索したときの 2010 年 5 月 20 日 (基準日) のランキングを示している . 画面左部に利用者が設定した基準日のランキングとそれらのページタイトルを表示している . 中央部には, それらのページの基準日以前以後のランク値を表示している . ランク

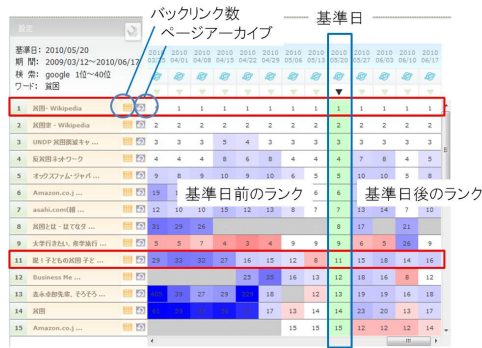


図 3 ランキング履歴提示画面

Fig. 3 Visualizing the ranking history by using SERPWatcher.

値は、基準日より上位であったか下位であったかに応じて背景色を設定しており、基準日のランクは緑色、基準日前後のランクについては、基準日と同じなら白、基準日より上位なら赤、下位なら青でその度合いが強いほど濃い。たとえば、基準日に 1 位であった「貧困-Wikipedia」が過去もその後も変わらず 1 位であったことが分かる。また、11 位の「脱!子どもの貧困子ども HAPPY 化計画」は 2 カ月程前はもっと低いランクであったが上昇し基準日以降にまた下降している様子が分かる。

また、図中 で示した 2 つのボタンを押すとそれぞれページアーカイブ、バックリンク数を確認することができる。これによりユーザは興味深いランキング遷移を発見したときにそれが何の事象と関連があるかを調査するきっかけとすることができる。これは 2.2 節で述べたシステム要件の B2 を満たす。

4. SERPAnalyzer

前章で述べた SERPWatcher によって、登録したキーワードに関する過去の検索結果ランキングを確認することができる。調査者は、そこから社会概念形成のプロセスを探るにあたり手がかりとなるページに焦点を定める必要がある。しかし、長期間ランキング履歴を蓄積した場合、そのデータは膨大になる。それらを SERPWatcher ですべて可視化した巨大な表から何かを読み解くのはユーザにとって大きな負担となる。

そこで、調査を進めていくうえで判断材料の 1 つとなるような情報を自動抽出し提示するシステム SERPAnalyzer を提案する。本システムが想定している社会調査とは、検索キー

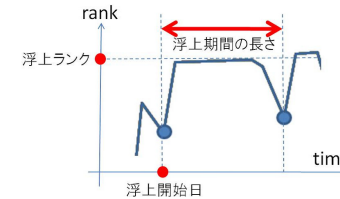


図 4 浮上期間から分かる社会的事象を特徴づける 3 つの指標

Fig. 4 Three measurements for an active period of the ranking transition.

ワードに関する事象の社会概念が形成されるプロセスを分析することであった。社会概念が形成される過程には、それに関する事象に人々の注目が集まっていく様子がみられると考えられる。2.2 節で述べたように、検索エンジンのランキングの性質として、多くの人に参照されたページが上位に浮上する可能性が高いことから、我々は、ランクが浮上するということが人々の注目度を推測する 1 つの指標となるのではないかと考えた。そこで、Web ページがランクアップする箇所を抽出し、それらがいつランクアップしたか、どのくらいの期間、最高で何位までランクアップしたかという情報を提示することにした。

4.1 節で、ランク変動からのランクアップする箇所の自動抽出について、4.2 節では、抽出した情報の提示について述べる。

4.1 浮上期間抽出

ランクアップしてからランクダウンするランク変動を浮上期間と呼び、その抽出について述べる。また、浮上期間からそのページがどのように注目されたかを特徴づける指標となる「浮上開始日」「浮上ランク」「浮上期間の長さ」が得られる(図 4)。「浮上開始日」から Web ページの内容に関連する出来事が注目を集め始めた時期の推測ができる。最も上昇したときのランク「浮上ランク」からはその話題のメジャー度が推測できる。ランクダウンするまでの「浮上期間の長さ」は注目を集め続けた期間であり、その話題の定着度を推測することができる。

Web ページのランク変動から、次の手順で浮上期間を取り出す。ランク変動は 2 次元平面上の点列 $P = \{P_i(date, rank) | i = 0 \dots n\}$ で表し、まず元のランク変動 $P = \{P_0, \dots, P_n\}$ に対して、区分最小二乗法¹⁹⁾を適用し、ランク変動を単純化する。区分最小二乗法は、2 次元平面上の点列をできるだけ少ない本数の直線で近似するアルゴリズムであり、これにより細かい変動が 1 本の線分に近似され、Web ページの長期にわたるランク変動のおおまかな流れをとらえることができる。区分最小二乗法適用後の点列を Q とする。最後に、 Q に


```

INPUT : Q=(q1,...,qn) //qiはqi.dateとqi.rankからなる
j=0
Start[j]=q0
For ( i = 0; i < n-1; i++){
//傾き-の後に傾き+になったとき、切り替わった地点から新しい浮上期間とする
If ( (qi+1.rank - qi.rank)/(qi+1.date - qi.date) < 0
&& (qi+2.rank - qi+1.rank)/(qi+2.date - qi+1.date) > 0 ){
End[j] = qi+1
Active_periods[j] = (Start[j],End[j])
j++
Start[j] = qi+2
}
}
OUTPUT : Active_periods
    
```

図 5 浮上期間抽出アルゴリズム

Fig. 5 Algorithm for extracting active periods from the ranking transition.

対し浮上期間抽出関数を適用する。図 5 に浮上期間抽出関数のアルゴリズムを示す。区分最小二乗法によりランク変動は直線近似されている。それらの直線を順に見ていき、傾きが正ならば浮上期間の始まりとし、その後直線の傾きが負になり再び正となった点までを 1 つの浮上期間とする。

4.2 提 示

抽出した浮上期間の提示とユーザの操作手順について述べる。図 6 に SERPAnalyzer の画面遷移をユーザの操作の流れに沿って示す。

「条件選択画面」(図 6 中 ①)は、分析対象とする Web ページを絞り込む画面である。検索エンジン、検索キーワード、さらに Web ページを絞り込むためのキーワードを自由に入力する。この自由入力のキーワードについては、5.3 節で詳細を述べる。これらの条件を設定し、ボタンを押すと「浮上頻度観察画面」(図 6 中 ②)に遷移する。この画面は、対象 Web ページすべてのランク変動から浮上期間を抽出し、浮上を開始したページ数を月ごとに集計したヒストグラムである。浮上開始ページ数が多い月などユーザが気になった月を画面下のプルダウンリストから選択してボタンを押すと、「浮上ランク・期間観察画面」(図 6 中 ③)に遷移する。この画面では前画面で選択した月に浮上開始した Web ページの浮上ランク・期間を観察することができる。また、各 Web ページのタイトルなど詳細情報を知りたい場合は、画面右で浮上ランクの範囲と浮上期間の長さの範囲を指定しボタンを押すと、画面下部に指定した範囲の Web ページのタイトルなどが表示される。さらに、各 Web ページのランク変動を確認したい際は、3 章で示したランキング履歴提示画面にリンクし、対象 Web ページのランク変動を確認することもできる。

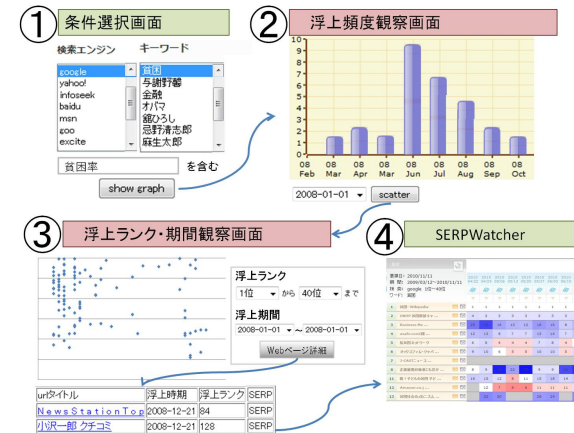


図 6 提案分析システムの画面と操作の流れ
Fig. 6 Process flow of SERPAnalyzer.

このように、抽出した浮上期間をから得られる「浮上開始日」「浮上ランク」「浮上期間」を集計し提示し、ユーザがランキングの特徴に応じて Web ページを絞り込んで観察することができるようになる。

4.3 サブキーワードによる Web ページの絞り込み

図 6 中の「1 条件選択画面」内にある、自由に入力する語による絞り込み機能について述べる。これは、検索エンジン・検索キーワードの条件に加え、タイトルやスニペットに含む語を指定して分析対象 Web ページをさらに絞る機能である。以降では、この語のことをサブキーワードと呼ぶことにする。たとえば、検索キーワード“貧困”で検索した際の結果ページには、貧困ビジネスに言及するページや、海外の貧困問題に言及するページ、など異なる話題が混在している。そこで、検索キーワード“貧困”のときにサブキーワードとして“貧困ビジネス”や“アメリカ”などと、調査したいトピックを象徴するような語を指定して分析対象ページを絞り、分析を行うことが可能になる。さらに、それらを比較したり、得られた結果をもとに対話的に分析を繰り返したりすることによって新しい発見が得られると考えられる。

5. 提案システムを使った社会分析例

3 章、4 章で述べた提案システム SERPWatcher, SREPAnalyzer の有用性を確かめる

ため、以下の検索キーワードを用いて分析を行った。本章では、その中から3つの分析例 Case Study1~3 を紹介する。CaseStudy1 では、検索キーワード“貧困”，サブキーワード“貧困率”を例に、浮上ページ数，浮上期間，浮上ランクがどのように分布するかを確認した。CaseStudy2 では、検索キーワード“選挙”を例に、サブキーワードにより対話的に分析を繰り返すことで事象を掘り下げていくことができることを示す。CaseStudy3 では、検索キーワード“ジェンダーフリー”を例に、分析により絞り込んでいった Web ページの内容を分析することで、検索キーワードに関する事象に関する話題の変遷を知る手がかりが得られることを示す。

5.1 Case Study1 : キーワード“貧困”に関する分析

本システムを使って Web ページの浮上ページ数，浮上ランク，浮上期間を観察するとどのような特徴が見つけ出せるかを知ることを目的に，複数の検索キーワード，サブキーワードを例に分析を行った。本節では，そのうちから検索キーワード“貧困”，サブキーワード“貧困率”の例を用いて，浮上ページ数の集計結果，浮上ランク・浮上期間がどのように分布するのかについて述べる。また，検索エンジン Google と Yahoo! による比較も行いどのような違いがあるか観察した。

Google で“貧困”と検索した際の結果ページ(2009年5月~2010年4月)を使用し，以下の(1)~(4)の手順で分析を行った。

- (1) 分析対象 Web ページの絞り込み
- (2) 浮上開始日の観察
- (3) 浮上ランク・浮上期間の長さの観察
- (4) Web ページの内容調査

(1)~(4)の詳細を以下に述べる。

- (1) 分析対象 Web ページの絞り込み

分析対象 Web ページの絞り込み条件は，検索エンジン，検索キーワードに加え，サブキーワードも指定できる。検索キーワード“貧困”の検索結果の Web ページには様々な話題が混在しているため，話題を絞って分析した方が分析を行いやすいと考え，サブキーワードとして“貧困率”を用いた。これは Google の「他のキーワード」を参考にした。「他のキーワード」とは，Google の検索結果画面の上部に，検索キーワードとともに検索されることの多い語として提示している語である。“貧困率”は，Google で“貧困”と検索すると，“他のキーワード”としてあげられていた語の1つである。

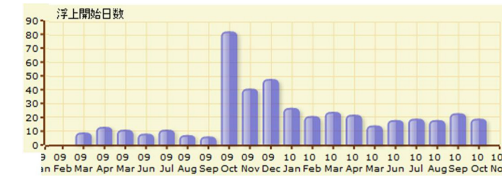


図7 浮上開始日の月ごと頻度
(検索エンジン: Google 検索キーワード: 貧困 サブキーワード: 貧困率)

Fig. 7 Frequencies for active periods (Search engine: Google, Keyword: Poverty, Sub-Keyword: Poverty Rate).

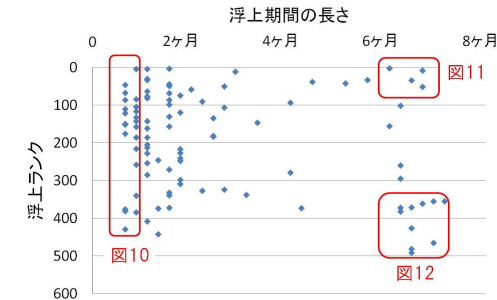


図8 浮上ランク・浮上期間の長さ(2009年10月浮上)

Fig. 8 The highest rank and the length of each active period which starts the period Oct. 2009.

- (2) 浮上開始日の観察

分析対象 Web ページから，浮上開始日の頻度を月ごとに集計した結果を図7に示す。

ここから，2009年9月までは浮上する Web ページが10件前後であるのに，2009年10月に80件以上の Web ページが浮上していることが分かる。この時期に，“貧困率”についてなんらかの出来事が起こったことが推測できる。そこで，2009年10月に浮上したページがどのようなページであったか調べるため，まずそれらの浮上ランク・浮上期間の長さを観察した。

- (3) 浮上期間のランク・長さの観察

図8に2009年10月に浮上したページの浮上ランクと浮上期間の長さの散布図を示す。縦軸は最も浮上したときのランク，横軸は浮上し始めてから下降するまでの期間である。上にあるほど浮上ランクは高く，右にあるほど浮上期間が長い Web ページである。この散布

17 SERPAnalyzer : 社会調査支援のための SERP アーカイブからの特徴的ランク変動分析システム

タイトル	種類
<貧困率> 政府として調査する方針固める 長妻厚生労働相(毎日新聞...	ニュース
子どもの貧困率調査へ 国内 Reuters	ニュース
母子、父子家庭の貧困率 OECD中ワースト1(エコノミックニュース...	ニュース
「貧困率」って?	ブログ
「相対的貧困率」について色々と考えてみる.....(4)「経済格差」は「世代...	ブログ
10/21(水) 貧困率ワースト4位。酔語酔吟 夢がたりウェブリブログ	ブログ
首相が相対的貧困率15.7%に「ひどい」- 政治ニュース..	ブログ
虹とモンスーン【反貧困ネットワークから】貧困率測定についての声明	ブログ
JavaScript++かも日記:【雑談】相対的貧困率って何?	ブログ
甚之介の剣道雑記帳: 疑問だらけの貧困率	ブログ
【調査】日本は7人に1人以上が貧困 日本の貧困率、15.7..	ブログ
村野瀬玲奈の秘書課広報室 日本の貧困率がやっと明らか	ブログ
時事ドットコム: 貧困率「ひどい数字」=鳩山首相	ブログ
厚生労働省: 子どもの貧困率調査へ 不況で問題深刻化受け - 47NEWS(よんな...	ニュース
<貧困率> 日本15.7% 先進国で際立つ高水準(毎日新聞社..	ニュース
貧困率: 政府として調査する方針固める 長妻厚生労働相 - 毎日.jp ..	ブログ
7人に1人も... 日本の貧困率が世界4位 - 政治・社会 - ZAKZAK	ニュース
えっ日本は世界5位の貧困国? 「貧困率」が示すものってなに R25	ブログ
山陰中央新報 - 日本の貧困率は15.7% 07年、98年以降で最悪	ブログ
日本の貧困率は15.7%/07年、98年以降で最悪? 四国新聞社	ニュース
えかわ珈琲店のブログ: 貧困率	ブログ
千葉発 幕張経由 社労士事務所のblog: 日本の「貧困率」15.7..	ブログ
大脳道場 NO.1398 新政府の「貧困率調査」への着手を歓迎する。調査..	ブログ

図9 浮上期間の短いページ

Fig. 9 The list of Web pages whose activate period's lengths are short.

図の特徴として、主に浮上期間の長さが短い Web ページ群と、浮上期間が長く浮上ランクが高い Web ページ群、浮上期間が長く浮上ランクが低い Web ページ群があることが分かる。次に、この3グループの内容の詳細を観察した。

(4) Web ページの内容調査

図 8 で特徴的であった 3 つの Web ページのグループについて内容を調べた。まず、浮上期間の長さが短いグループを見た。浮上期間の長さが短い順に 23 件取り出した Web ページのタイトルを図 9 に示す。これらは、内容はほとんどが 2009 年 10 月の朝日新聞に掲載された元厚生労働大臣の長妻昭氏が日本の相対的貧困率が 2007 年に 15.7%に達したことを伝えるニュースに言及するものであった。ページの種類は個人のブログやニュース記事が多くあった。ニュースが多い原因として、ニュースサイトの記事は、日々の出来事に応じて発生・消滅を繰り返しているのがそれがこのような結果に現れると考えられる。

次に、浮上期間が長く浮上ランクが上位であるグループを観察した。浮上期間の長さが 2 カ

タイトル	浮上開始日	浮上順位	浮上期間
厚生労働省: 相対的貧困率の公表..	2009-10-15	9	210
asahi.com(朝日新聞社): 日本 の 貧 困 率 ...	2009-10-15	3	189
池田信夫 blog: 「貧困率」についての誤解	2009-10-29	12	91
貧困率 - Wikipedia	2009-11-29	2	175

図 10 浮上期間の長さが長く浮上ランクが上位であるページ

Fig. 10 The list of Web pages whose activate period's lengths are long and their highest ranks are upper level.



図 11 「貧困率-Wikipedia」のランク変動 (2009 年 8 月 27 日 ~ 2010 年 4 月 29 日)

Fig. 11 Rank transition the Wikipedia Webpage whose title is “Poverty Rate” (From 2009/8/27 to 2010/04/29).

月以上かつ浮上ランクが 30 位以上という条件で取り出した Web ページを図 10 に示す。この 4 つは、相対的貧困率を算出したという内容の厚生労働省公式ページ、朝日新聞のニュースサイトが貧困率算出を伝えた記事、有名ブログが相対的貧困率の見方について考えを述べているブログ、Wikipedia の貧困率のページ、であった。なかでも 1 番目のページはこの話題の発端となった厚生労働省の公式発表の文書を掲載しているページで、多くの人と言及していた貧困率の話題の発端となるページである。浮上期間の長さが短いページとは性質が異なるものが取り出せたといえる。

また、「貧困率-Wikipedia」は、図 11 に示すように、2009 年 11 月以前は 13 位や圏外であったが 11 月ごろから次第に安定的に 3 位にランクするようになっており、貧困率の発表をきっかけに「貧困率」という言葉が一般的な言葉として扱われるようになった様子が見える。

最後に、浮上期間が長く浮上ランクが低い Web ページの内容を調査した。図 12 にそのタイトルリストを示す。ページの種類はニュースやブログが多くあったが、これまでに示した 2 つのグループとの違いは掲示板があることである。

タイトル	種類
[鳩山首相、貧困率を見て「大変ひどい数字。なんでこんな日本にしたの...]	掲示板
日本の貧困率15.7%	掲示板
日本の「貧困率15.7%」格差是正と経済全体の成長、どちらを優先す...	掲示板
日本の「貧困率」15.7%、OECD中4位 - BIGLOBEニュース	ニュース
日本の貧困率は15.7% 厚生省が初公表:イザ!	ニュース
タカマサのきまぐれ時評2 <貧困率>政府として調査する方針固める...	ブログ
「貧困率」に関する若干の問題点(1) - 松尾光太郎 de 海馬之玄関...	ブログ
<貧困率>日本15.7% 先進国で際立つ。子ども手当は所得制限の...	ブログ
日本の「貧困率」15.7%という数字? - メルベージュ森通信	ブログ

図 12 浮上期間が長く浮上ランクが低い Web ページの内容

Fig. 12 The list of Web pages whose activate period's lengths are long and their highest ranks are low level.

図 12 に示す 3 件の掲示板の内訳は「ニュース畑 もっと知りたい語りたい 投稿で育てるニュースの畑」や 2ちゃんねる^{*1}のニュース板、また 2ちゃんねるのニュース板をそのまま貼り付けているまとめのページであった。これらはすべて 1 件目にニュースが貼り付けてあり、それに対して数人の人が意見を投稿するというページであり、どれもページが発生してから意見投稿が 1, 2 カ月続いていた。低いランクで長く浮上しているページに掲示板が含まれていたのは、このように発生してからしばらく更新が繰り返される掲示板の性質が影響していると考えられる。

また、ここまでで述べた分析対象ページ：検索エンジン Google、検索キーワード“貧困”、サブキーワード“貧困率”の分析結果と、分析対象ページ：検索エンジン Yahoo!、検索キーワード“貧困”、サブキーワード“貧困率”の分析結果を比較した(図 13)。まず、共通している点として、浮上ページ数に関しては、どちらも 2009 年 10 月に最も浮上するページが多いことが分かる。また、浮上ランク・浮上期間の分布についても、短期間浮上するページ群、長期間上位で浮上するページ群、長期間下位で浮上するページ群の 3 つのグループがあることもあげられる。つまり、出来事が起こると浮上するページが増え、それらの中にはすぐランクダウンするものと浮上を続けるものが存在するという本質的な点では変わらない。

Google の場合と Yahoo! の場合で異なることとしては、まず、浮上ページ数が全体的に少ないことがあげられる。たとえば、それぞれ最も浮上ページ数の多かった 2009 年 10 月のその件数は Google では 100 件近くあるが、Yahoo! は 26 件である。また、浮上ランク・浮上期間の分布では Yahoo! が比較的短期浮上のページが少ない。ここから、Google が

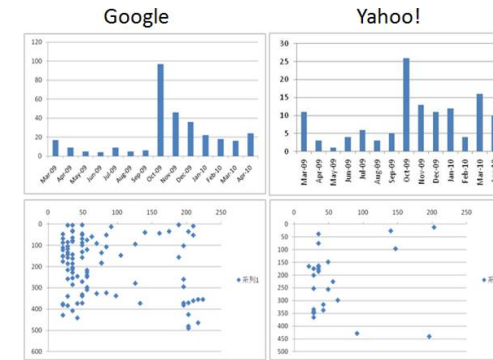


図 13 Google, Yahoo! での分析結果の比較

Fig. 13 The highest rank and the length of each active period by Google's SERP (Left) and by Yahoo!'s SERP (Right).

Yahoo! よりランキングのアップダウンが激しいことが分かる。また、長期間浮上するページの内容が異なっていた。Yahoo! の場合、浮上期間の短い Web ページ群には、ニュースやブログのほか、厚生労働省の貧困率発表のページも含まれていたことや、長期浮上のページには Yahoo!知恵袋などの Q&A サイトが含まれていることや、浮上期間が長く浮上ランクの低いページは、個人のブログであった。どのようなページがどのようなランク変動をするかは検索エンジンのアルゴリズムによると考えられる。しかし、長期に浮上するページはどの場合も比較的有名な人物によるページであったり、掲示板のように更新され続けるといった長期間浮上する可能性を持つページであった。

最後に、他の検索キーワード、サブキーワードによる分析結果についても触れておく。複数の条件で分析を行った結果、浮上ランク・浮上期間の分布が短期浮上、長期上位浮上、長期下位浮上の 3 グループにかたまって分布するケースは、検索キーワード“選挙”、サブキーワード“東京”、検索キーワード“北朝鮮”、サブキーワード“拉致”、検索キーワード“男女共同参画”、検索キーワード“フェミニズム”など結果のうち浮上件数が多かった月上旬 2 月分の分布をそれぞれ確認した際、ほとんどすべての例で確認された。しかし、分布がグループごとにかたまらず、均等に散らばる場合も少数だが存在した。このような例は、分析対象ページの収集期間の最後に近い日に浮上したページを観察した際に見られた。この原因は、浮上したページがその後浮上を続けたとしても最後のランキング収集日までの期間しか浮上したと表示できないことにより、多くのページが同じように短い期間しか浮上していな

*1 <http://www.2ch.net/>



図 14 検索キーワード“選挙”, サブキーワードなし 浮上ページ数

Fig. 14 Frequencies for active periods (Keyword: Election, No Sub-Keyword).

いかのように表示されてしまうことにある。

以上のことから、多くの分析結果は、浮上ページ数、浮上期間、浮上ランクの分布について以下のことがいえる。浮上ページ数は、検索キーワードに関する事象についてなんらかの出来事が起こった際に増加する。これにより、調査者は発生した出来事を調査するきっかけを得られる。また、浮上したページの浮上ランク・浮上期間の分布を見ると、多くの場合、短期間浮上、長期間上位で浮上、長期間下位で浮上の3グループにかたまって現れる。それにより調査者は、たとえば、長期浮上したページは注目度が高いページである可能性が高いなど、どのグループに属するかによって Web ページの性質を推測する手がかりを得られる。

5.2 Case Study2 : キーワード“選挙”に関する分析

分析結果から特徴が得られない場合でも、適当なサブキーワードを選び、対話的に分析を繰り返すことで、分析事象を絞り込み、特徴を発見することができる。本節では、検索キーワード“選挙”による分析を例にその様子を示す。まず、サブキーワードなしで、浮上ページ数を観察した。

図 14 から、特別に浮上するページが多い時期があるわけではないことが分かる。その中でも 2009 年 7 月や 2009 年 9 月は比較的に浮上するページが多いことが分かる。そこで、これらの月に浮上したページのタイトルや内容から、どのような話題が存在するかを確認することにした。図 15 は 2007 年 7 月に浮上したページを浮上期間が長い順に上位リストアップしたものである。ここから、衆議院議員総選挙に関するページや都議会議員選挙に関する話題があることが分かる。そこで次に、衆議院議員総選挙について調べを進めることにした。サブキーワードを“衆議院”として再び浮上ページ数を調べた様子を図 16 に示す。すると、先ほどより衆議院議員総選挙の行われた次の月である 2009 年 9 月に浮上したページが多いことが顕著に現れた。そこで、衆議院議員総選挙について調べてみると、2009 年 8 月 30 日に衆議院議員総選挙が行われたことが分かった。多くの人がこの選挙結果などに言及したことでこの時期に浮上ページが増えたと考えられる。それらのページのタイト

urlタイトル	浮上時期	浮上ランク	浮上期間(日数)
京都市 京都市選挙管理委員会事務局	2009-07-08	44	483
選挙管理委員会/トリネット/鳥取県公式サイト	2009-07-15	104	476
選挙についての紹介 - 総務省(選挙)	2009-07-01	68	469
東京都議会議員選挙/JCP特集	2009-07-08	128	462
岡山県ホームページ - 選挙管理委員会事務局	2009-07-29	29	462
グラフで語ろう総選挙2008 第45回衆議院議員総選挙特集	2009-07-15	262	455

図 15 検索キーワード“選挙”, サブキーワードなし 2009 年 7 月

Fig. 15 The Web pages which start activate periods Sep. 2009 (Keyword: Election, No Sub-Keyword).

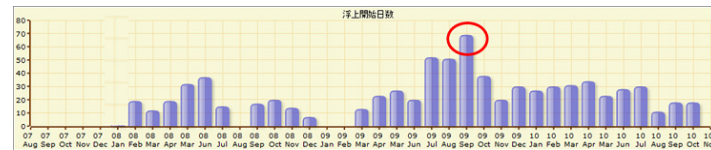


図 16 検索キーワード“選挙”, サブキーワード“衆議院” 浮上ページ数

Fig. 16 Frequencies for active periods (Keyword: Election, Sub-Keyword: House of Representatives).

urlタイトル	浮上時期	浮上ランク	浮上期間(日数)
ドラクエ9のすれ違い通信が今回の衆議院議員選挙に利用されている	2009-08-16	365	413
Twitterで選挙 衆議院議員 逢坂誠二の徒然日記/ウェブブログ	2009-08-30	57	77
Twitterを使った選挙活動が違法に - スラッシュドット・ジャパン	2009-08-23	163	28
ITで衆議院議員選挙を考える - 週末スペシャル:ITpro	2009-08-02	157	21

図 17 検索キーワード“選挙”, サブキーワード“衆議院” 2009 年 9 月浮上

Fig. 17 The Web pages which start activate periods Sep. 2009 (Keyword: Election, Sub-Keyword: House of Representatives).

ルを確認すると、図 17 に示すように、選挙活動の中での twitter の存在が話題になっていたり、ゲームの通信機能を利用して選挙の集会を呼びかけていたり、IT の力を駆使した選挙が広がってきているというコラムなどが目に入った。そこで、次にその中から twitter をサブキーワードとして選んで浮上ページ数を観察した(図 18)。すると、2008 年 2 月からのデータを対象としているにもかかわらず、2009 年 1 月からしか浮上ページがないという結果がでた。このことから、選挙の話題の中に twitter が出てきたのは 2009 年に入ってからということが推測できる。さらに、浮上数の多かった 2009 年 7 月のページを見てみると(図 19)、公示日から選挙日が終了するまでの間、選挙活動目的の文書などの更新は制限されるという公職選挙法に twitter は抵触するのかがという議論がなされている様子がみてとれ

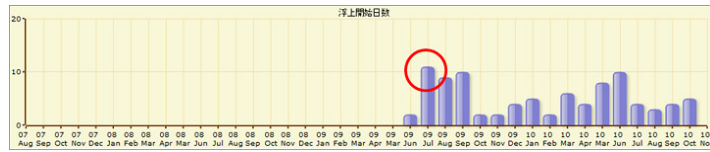


図 18 検索キーワード“選挙”，サブキーワード“twitter”

Fig. 18 Frequencies for active periods (Keyword: Election, Sub-Keyword: Twitter).

urlタイトル	浮上時期	浮上ランク	浮上期間(日数)
Twitterを使った選挙活動が違法に - スラッシュドットジャパン	2009-07-22	18	83
POLAR BEAR BLOG なしけしになる公職選挙法	2009-07-08	35	49
Twitter議員にYouTube選民 ネットと選挙、日本でも変化の機運	2009-07-22	41	49
日本の公職選挙法は、やっぱりTwitterのつぶやきも禁止だった	2009-07-29	174	42
Twitterと選挙 衆議院議員 逢坂誠二の徒然日記/ウェブブログ	2009-07-22	77	42
Going My Way 閣議で選挙運動への利用の可否を検討された Twitter	2009-07-22	172	42
Twitterを使った選挙活動が違法に - スラッシュドットジャパン	2009-07-29	64	35
選挙: 衆議院「つぶやき」は公職法違反 政府「ツイッター」禁止	2009-07-22	31	35
「Twitter」のつぶやきもダメ? 公職選挙法の逆効果 インターネット	2009-07-22	89	28
アコフ、2009年ネット選挙元年に - 米軍経団	2009-07-15	132	28
公職選挙法は、Twitterのつぶやきすら違法認定するかもしれない	2009-07-01	86	21

図 19 検索キーワード“選挙”，サブキーワード“twitter”

Fig. 19 The Web pages which start activate periods Sep. 2009 (Keyword: Election, Sub-Keyword: Twitter).

た．選挙に IT の力が活用され始めたこと，またそれがきっかけで起こってきた問題があるなど，そのキーワードの中で新しく出てきた概念について気づくことができた．

その他の分析として，検索キーワード“小沢一郎”で分析を行った結果，サブキーワードなしの結果から得られるページに頻出した，“起訴”や“辞任”などの語をサブキーワードとして指定すると，小沢一郎氏が検察審査会により起訴議決をされる時期や，秘書逮捕により引責辞任する時期に浮上ページが激増している様子が現れた．

以上の分析により，分析結果から特徴的な結果が得られない場合でも，結果に現れたページを眺め，頻出する語などをサブキーワードとして用い分析を行い，再び得られた結果を眺め，サブキーワードを指定し直して分析を行うというように対話的な分析を行うことで，分析対象事象を絞り込んでいくことができ，何らかの特徴を発見することができる．

5.3 Case Study3 : キーワード“ジェンダーフリー”に関する分析

本節では，浮上期間の情報をもとに絞り込んでいった Web ページの内容がどのようなものであるか，いつ作成されたページであるかを観察することで，その概念の移り変わりを知



図 20 サブキーワードとして利用した「他のキーワード」

Fig. 20 We use “Other Keyword” column on Google search page as sub-Keywords.

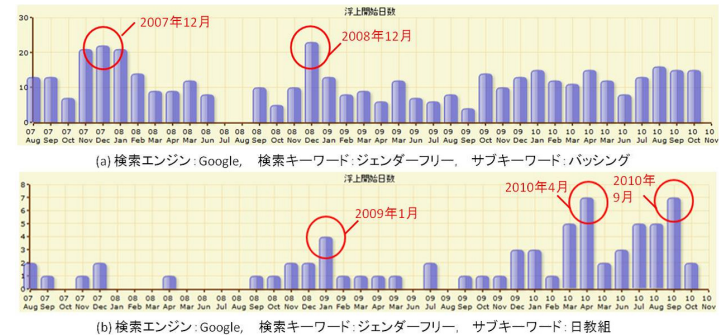


図 21 サブキーワード“バッシング”，“日教組”による検索結果ページの浮上開始日頻度

Fig. 21 Frequencies for active periods (Keyword: Gender Free, Sub-Keyword: Bashing (Top), “Nikkeiyouso” (Bottom)).

る手がかりを得られることを示す．例として，検索エンジン Google，検索キーワード“ジェンダーフリー”による検索結果ページを対象とした分析例を示す．サブキーワードとして，図 20 に示す Google の検索結果とともに提示される「他のキーワード」から“バッシング”，“日教組”という語を用いた．“バッシング”とは，ジェンダーフリーという概念に対して起こった攻防の際によく用いられた語である．“日教組”とは，日本教職員組合のことで，これも論争の際によくあげられる語である．以降では，これら 2 種類の分析例について述べる．

図 21 中 (a)，(b) に，サブキーワードが“バッシング”，“日教組”であるページ群の浮上開始日の頻度を示す．まず“バッシング”で浮上頻度が多かった 2007 年 12 月と 2008 年 12 月についてページの内容を調べた．内容の特徴として，書籍に関するページの割合が多いことがあげられる．たとえば 2007 年 12 月では 27 件中 19 件が，2008 年 12 月では 24 件中 9 件が書籍販売サイトでジェンダーフリーバッシングに関する本を販売しているページである．

図 22 に書籍販売サイトに載っていた書籍のタイトルと発行年月の一覧を示す．浮上開始日と比べて 1 年以上前に発行された書籍が多く浮上していることが分かる．

書籍タイトル	著者	発行年月
ジェンダー・フリートラブル	木村涼子	2005年12月
ジェンダーフリー・性教育バッシングここが知りたい50のQ&A	浅井善夫	2003年12月
バックラッシュ! なぜジェンダーフリーは叩かれたのか?	上野千鶴子	2006年6月
「ジェンダー」の危機を超える! -徹底討論! バックラッシュ	若菜 みどり	2006年8月
Q&A 男女共同参画/ジェンダーフリー・バッシング	日本女性学会ジェンダー研究会	2006年6月
ジェンダー・フリーは止まらない	上野千鶴子	2002年1月

図 22 書籍販売サイトで売られていた書籍

Fig. 22 Books which are referred at the site of online bookstore.

「荻上式ブログ ジェンダーフリーバッシングはいつ頃はじまったのか」*1によるとメディア上でのジェンダーフリーに関する論争が起きたのが 2002 年ごろであり、現在も論争は続いているものの、メディアやネット上での激しいやりとりは落ち着いているようである。図 22 からは、論争のわき起こった 2002 年以降にまとめられた書籍が何かのきっかけで引用されることでこのような書籍に関するページが多く浮上すると推測できる。

次に、サブキーワード“日教組”での結果の分析を行った。図 21 (b) からは、2009 年 1 月、2010 年 4 月、2010 年 9 月に浮上したページが多いことが分かる。図 23 にその各月に浮上した Web ページの内容の詳細を示す。

特徴的であったこととして、2009 年 1 月、2010 年 4 月に浮上したページはブログが多いが、その多くのブログ記事の作成日は浮上した時期の 1, 2 年前の 2007, 2008 年であることがあげられる。書籍に関するページの例と同様に、過去に話題にしていたことがたまたま引用されることによりこのようなことが起こると推測できる。もう 1 つ注目すべきことは、2010 年 9 月に浮上した Web ページの内容は、6 件中 4 件が、幸福実現党の局長がジェンダーフリー教育を批判するページや支持者がそのページを引用するページであった。また、これらはこれまで述べてきたページと比べて作成日の新しいページであった。このことから、2010 年夏ごろまではジェンダーフリーに関する目新しい話題がなかったなか、2010 年 9 月周辺で幸福実現党に関わりのある人々がジェンダーフリー批判に関して盛り上がったことをうかがい知ることができる。

検索キーワード“ジェンダーフリー”、サブキーワード“バッシング”、“日教組”における分析の結果、浮上開始日より 2, 3 年昔に作成された Web ページがランクアップする例が多く確認された。それは、過去にわき起こった話題が何かのきっかけにたびたび引用されて

*1 <http://d.hatena.ne.jp/seijotcp/20060112/p1>

タイトル (2009年1月浮上)	種類	浮上開始日	作成日
現役おばちゃん教師 日教組の何が悪いのよ!!	ブログ	2009年1月18日	2008年6月3日
日教組が推進するジェンダーフリー教育の弊害	ブログ	2009年1月19日	2008年10月5日
【ジェンダーフリー】女子中学生が男の前でも抵抗なく着替え	2chニュース板	2009年1月3日	2008年10月5日
【日教組】ジェンダーフリーすげー(@_@ !)着替えだけでなく、奮闘	ブログ	2009年1月1日	2008年12月27日
日本教職員組合(日教組) 関連情報の編集履歴	辞書サイト	2009年1月15日	不明
ジェンダーフリー&バックラッシュ騒動まとめ	ブログ	2009年1月18日	2007年10月以前
タイトル (2010年4月浮上)	種類	浮上開始日	作成日
現役おばちゃん教師 日教組の何が悪いのよ!!	ブログ	2010年1月18日	2008年6月3日
人権・平和教育、ジェンダーフリー教育に熱心な地域は学力が低い?	Q&Aサイト	2010年4月4日	2008年10月8日
男性への逆差別とジェンダーフリー教育、過激な性教育	ブログ	2010年4月18日	2010年4月17日
ぼやきつり 民主政権は“日教組政権”だ「WILL」より	ブログ	2010年4月25日	2009年7月5日
東京で実際に行われていた性教育	動画サイト	2010年4月25日	2007年10月19日
女子供の洗脳で男もバカになった①ジェンダーフリー 日本総フスケ化	ブログ	2010年4月18日	2010年4月8日
【教育】「日教組の強い地域、学力低い」と判明...	2chニュース板	2010年4月18日	2008年10月8日
タイトル (2010年9月浮上)	種類	浮上開始日	作成日
幸福実現党 多摩川後援会 ジェンダーフリー教育の事態	ブログ	2010年9月26日	2010年7月16日
男性への逆差別とジェンダーフリー教育、過激な性教育	ブログ	2010年9月26日	2010年4月17日
幸福実現党メールマガジン[ジェンダーフリー教育の問題点]	政党公式ページ	2010年9月26日	2010年9月24日
政・経・心の指針[ジェンダーフリー教育の問題点]	ブログ	2010年9月19日	2010年9月25日
オーバの幸福実現党応援宣言! 「ジェンダーフリー教育の問題点」	ブログ	2010年9月19日	2010年9月24日
【教育】「日教組の強い地域、学力低い」と判明...	2chニュース板	2010年9月12日	2008年10月8日

図 23 サブキーワード“日教組”による検索結果のうち

2009 年 1 月、2010 年 4 月、2010 年 9 月に浮上した Web ページ

Fig. 23 The list of Web pages whose activate periods are 2009/01 (Top), 2010/04 (Middle) and 2010/09 (Bottom).

いる結果と推測できる。これらのことから、ジェンダーフリーに関する話題は数年前からあまり変化がないと考えることができる。

このように、過去に作成されたページが再び浮上するという事は、多くの良質なページからリンクされるページが浮上するという検索エンジンのランキングの仕組みを顕著に表したものだといえる。検索エンジンによる検索結果ページのランク変動を利用することで過去の話題が現在の話題においてどの程度重要視されているかなどを測ることができる。

6. 関連研究

本章では、我々の研究にかかわりのある「Web 上の話題抽出・分類」「ランク変動解析」に関連した研究について述べる。

Web 上の話題を抽出・分析する Web コミュニティ抽出は現在までに多数提案されている。Web コミュニティ抽出手法には、主に Web ページ間のリンク構造を利用した手法とド

キュメントの類似度を考慮する手法がある。東京大学の豊田ら¹³⁾は2003年に、1999年から2002年の間に収集した日本のWebアーカイブからコミュニティの発展過程をリンク解析により詳細に分析し、コミュニティの分裂・合併や成長縮小などの特徴を発見している。早稲田大学の飯村ら¹⁴⁾は2009年にWebページ間のリンク構造を有効グラフと見なし、提案アルゴリズムにより、Webコミュニティの抽出を行っている。その結果既存アルゴリズムより適合率を向上させた。東京大学の藤村ら¹⁵⁾は2005年に肯定・否定への文書分類を基にしたWebからの評判抽出を行っている。あらかじめ収集した肯定的・否定的評判コーパスから統計的に評価表現を抽出し、評判分類に拡張する手法である。上記の関連研究における問題点として、分析のために大量のWebページを収集しなければならないこと、分析する際に適切でないスパムページなどを除去する必要があることがあげられる。これに対し我々は、自ら大量のページを収集する必要がなく分析する材料として適切なページを得られることから、検索エンジンによる結果ページを利用することにした。また、検索エンジンのランキングはより人々が注目している事柄に関するWebページを上位にランクさせることが考えられるので、世相を分析するのに適していると考えた。

検索エンジンの検索結果に着目した研究としては、早稲田大学の吉田ら¹²⁾の研究がある。これは、幅広い種類の語を検索キーワードとして得られる一般的なサイトのランク変動パターンを抽出し、サイトの生成日時や更新頻度について調査することでサイトの特徴を解析している。この研究は、検索エンジンの一般的なランク変動の傾向を分析しているのに対し、我々の研究は、社会・政治に関連する語を検索キーワードとして用い、キーワードごとのランキングの特徴を取り出し、傾向を発見するためにランク変動の収集・解析をしている。

7. まとめと今後の課題

商用検索エンジンによる検索結果ランキングの変動から社会事象を読み取ることを目的とした、社会科学者のためのランク変動分析システムSERPWatcher, SERPAnalyzerについて述べた。話題の発生時期, 規模, 定着度の推測する指標として, ランクが浮上した時期, そのランク, 浮上を続けた期間を示した。提案システムを用いたランク変動の分析から, 社会事象の発生や性質など読み解く手がかりが得ることができた。今後は, より適切に分析対象ページを絞り込む機能の追加や, 様々な検索キーワード・検索エンジンの組合せで分析を行い, システムに改良を加えていきたい。

謝辞 本研究は一部, 日本学術振興会平成19年度科学研究費補助金(基盤研究(B))(課題番号19300025)「社会科学の新しい研究方法論としての統合型ウェブマイニング環境の

開発研究」の補助を受けて行われた。

参 考 文 献

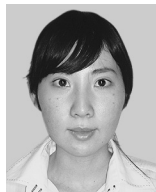
- 1) 増永良文, 渡辺知恵美, 伊藤一成, 小山直子, 深山鷹一, 舘かおる: 新しい社会調査法としての検索エンジン結果ページ群の自動収集・分析装置の開発—SERP Watcherの設計, DEIM 2009 D7-5 (Mar. 2009).
- 2) Masunaga, Y., Oyama, N., Watanabe, C., Ito, K., Tachi, K. and Miyama, Y.: SERPWatcher: A SERP Mining Tool as a Novel Social Survey Method in Sociology, *Database Systems for Advanced Applications (DASFAA2010)* (2010.04.02). (Excellent Demonstration Award).
- 3) Masunaga, Y., Oyama, N., Ito, K., Watanabe, C., Tachi, K. and Miyama, Y.: SERPWatcher: A Sophisticated Mining Tool utilizing Search Engine Results Pages (SERPs) for Social Change Discovery, *Symposium on Social Intelligence and Networking (SIN-10)* (In conjunction with the SocialCom2010) (2010).
- 4) The PageRank Citation Ranking: Bringing Order to the Web, January 29 (1998).
- 5) 西田圭介: Googleを支える技術—巨大システムの内側の世界, 技術評論社.
- 6) Research and Reports@Eyetools.
http://eyetools.com/research_google_eyetracking_heatmap.html
- 7) 舘かおる(編著): テクノ/バイオ・ポリテクス—科学・医療・技術の今, 作品社.
- 8) 盛山和夫: 社会調査法入門, 有斐閣ブックス.
- 9) ジーエーブプロ: 検索エンジン・ポータルサイト 日本でのシェア(利用率) (2011.03.07).
<http://www.webcreate.ga-pro.com/search.html>
- 10) ヤフー株式会社: Yahoo! JAPANの検索サービスにおけるグーグルの検索エンジンと検索連動型広告配信システムの採用, ならびにYahoo! JAPANからグーグルへのデータ提供について (2010.07.27). <http://pr.yahoo.co.jp/release/2010/0727a.html>
- 11) Google: Google search and search engine spam (2011.01.21).
<http://googleblog.blogspot.com/2011/01/google-search-and-search-engine-spam.html>
- 12) 吉田泰明, 平手勇宇, 山名早人: 商用検索エンジンにランキングされたサイトのランク変動パターン, DEIM 2009 A7-6 (Mar. 2009).
- 13) 豊田正史, 喜連川優: 日本のウェブアーカイブにおけるウェブコミュニティ発展過程の詳細分析, DEWS 2003 2-P-05 (Mar. 2003).
- 14) 飯村卓也, 平手勇宇, 山名早人: webページ間の関連性の伝播を用いたwebコミュニティ抽出手法の評価, DEIM 2009 B9-3 (Mar. 2009).
- 15) 藤村 滋, 豊田正史, 喜連川優: 文の構造を考慮した評判抽出手法, DEWS 2005 6C-i8 (2005).
- 16) Google: Google Web Search API. <http://code.google.com/intl/ja/apis/websearch/docs/>

- 17) Yahoo!: Site Explorer API. <http://developer.yahoo.com/search/siteexplorer/V1/inlinkData.html>
- 18) Google バックリンクページランク表示ツール. <http://exp.webnavisys.com/tool/php12.html>
- 19) Kleinberg, J. and Tardos, E. (著), 浅野孝夫, 浅野泰仁, 小野孝男, 平田富夫 (訳): アルゴリズムデザイン, 共立出版.
- 20) OpenFlashChart2. http://pullmonkey.com/projects/open_flash_chart2

(平成 22 年 12 月 21 日受付)

(平成 23 年 4 月 10 日採録)

(担当編集委員 高久 雅生)



中部 文子

2007 年お茶の水女子大学理学部情報科学科卒業。2011 年同大学大学院人間文化研究科博士後期課程修了。同年より大日本印刷株式会社 IPS 事業部。在学中は Web マイニングに関する研究に従事。



渡辺知恵美 (正会員)

お茶の水女子大学大学院人間文化創成科学研究科理学専攻講師。1998 年お茶の水女子大学理学部情報科学科卒業。2003 年同大学大学院人間文化研究科博士後期課程修了。博士 (理学)。同年奈良女子大学理学部情報科学科助手を経て、2005 年より現職。2011 年ジョージア工科大学客員研究員。データベース, Web マイニング等の研究に従事。ACM, 日本データベース学会各会員。



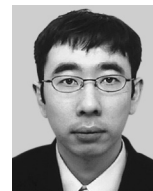
小山 直子

お茶の水女子大学生活科学部非常勤講師ほか。1987 年お茶の水女子大学大学院修士課程家政学研究科修了。2010 年度まで同大学ジェンダー研究センター客員研究員として、社会表象としてのジェンダー情報論の視点で Web 解析に従事。服飾史および視覚表象分析が専門。日本家政学会 (服飾史・服飾美学部会), 日本風俗史学会, 国際服飾学会, 日本データベース学会, 日本感性工学会各会員。



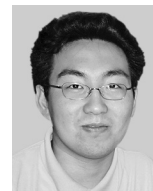
館 かおる

お茶の水女子大学大学院人間文化創成科学研究科教授。ジェンダー研究センター教授。お茶の水女子大学大学院人文科学研究科修了。日本学術会議連携会員。2000 年以降ジェンダー研究の立場からテクノロジーを研究。特にウェブ世界のジェンダー分析を試みている。著書・編著『母性から次世代育成力へ』新曜社, 『学校をジェンダーフリーに』明石書店, 『テクノ/バイオ・ポリテックスー科学・医療・技術のいま (21 世紀 COE ジェンダー研究のフロンティア第 4 巻)』作品社, 『女性とたばこの文化誌』世織書房等。



伊藤 一成 (正会員)

青山学院大学社会情報学部准教授。2005 年慶應義塾大学大学院理工学研究科後期博士課程修了。博士 (工学)。青山学院大学理工学部助手, 助教, 同社会情報学部助教を経て, 2010 年より現職。メディア情報処理に関する研究に従事。日本データベース学会, 人工知能学会, 電気情報通信学会各会員。



深山 鷹一 (正会員)

2005 年東京工科大学コンピュータサイエンス学部卒業。2007 年青山学院大学大学院理工学研究科博士前期課程修了。2011 年青山学院大学大学院社会情報学研究科博士後期課程単位取得退学。



増永 良文 (フェロー)

青山学院大学社会情報学部教授。1970 東北大学大学院工学研究科博士課程修了, 工学博士。データベースシステムやウェブマイニングの研究・開発に従事。日本データベース学会会長。お茶の水女子大学名誉教授。電子情報通信学会フェロー。DASFAA Outstanding Contributions Award 受賞。著書に『リレーショナルデータベース入門』(サイエンス社), 『データベース入門』(サイエンス社), 『コンピュータサイエンス入門—コンピュータ・ウェブ・社会』(サイエンス社), 訳書(監訳)に『オブジェクト指向データベース入門』(共立出版)等。
