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－

増永良文† 渡辺知恵美‡ 伊藤一成† 小山直子‡ 竹内純人† 深山鷹一† 舘かおる‡

† 青山学院大学 〒229-8558 神奈川県相模原市淵野辺 5-10-1
‡ お茶の水女子大学 〒112-8610 東京都文京区大塚 2-1-1

E-mail: † {masunaga, kaz}@si.aoyama.ac.jp, s_takeuchi@irc.aoyama.ac.jp, d8108008@cc.aoyama.ac.jp
‡ {watanabe.chimei, oyama.naoko, tachi.kaoru}@ocha.ac.jp

あらまし 我々は SERP Watcher を開発している．これは，ウェブの検索エンジン結果ページ(Search Engine Results Page, SERP)に現れるウェブページの順位の変動は，実世界の社会変動と関係しているのではないかという，我々のこれまでのウェブマイニング研究のもたらした知見に基づく．SERP Watcher はより正式には，Search Engine Results Page Ranking Change Watcher と称するが，利用者が指定した検索キーワードに関する SERP の順位変動を時間の経過とともに監視していき，その変動幅や変動パターンがあらかじめ指定された許容範囲を超えるとき，実世界で何らかの変動が起こっているかもしれないと判断して警告を発し，利用者に実際に社会調査を促す．つまり，SERP Watcher はこれまで社会科学の分野で知られているアンケート調査，インタビュー調査，あるいは実地調査に代わる新しい社会調査法になりうると考えられる．

キーワード SERP Watcher，ウェブマイニング，検索エンジン，SERP，社会科学，社会調査法

Development of an Automatic Collection and Analysis System of SERPs as a New Social Survey Method
－Design of SERP Watcher－

Yoshifumi MASUNAGA† Chiemi WATANABE‡ Kazunari ITO† Naoko OYAMA‡
Sumito TAKEUCHI† Yoichi MIYAMA† and Kaoru TACHI‡

† Aoyama Gakuin University 5-10-1 Fuchinobe,Sagamihara-shi, Kanagawa, 229-8558 Japan
‡ Ochanomizu University 2-1-1 Otsuka, Bunkyo-ku, Tokyo, 112-8610 Japan

E-mail: † {masunaga, kaz}@si.aoyama.ac.jp, s_takeuchi@irc.aoyama.ac.jp, d8108008@cc.aoyama.ac.jp
‡ {watanabe.chimei, oyama.naoko, tachi.kaoru}@ocha.ac.jp

Abstract We are developing SERP Watcher. This is based on the finding of our research on web mining that the change of the ranking of the web pages appeared in the search engine results page (SERP) relates to the social change of the real world. SERP Watcher is actually an abbreviation of SERP Ranking Change Watcher. It judges that some changes might have happened by the real world, and warns the user when the fluctuation band or the change pattern tolerance exceeded specified level. When warning is originated, the user actually surveys the society. In other words, SERP Watcher takes the place of the questionnaire survey, the interview investigation or the field survey known in the field of the social science, and it has been thought to be able to become a new social investigation method.

Keyword SERP Watcher, Web Mining, Search Engine, SERP, Social Science, Social Survey Method

1. はじめに
1.1. 研究･開発の目的

ウェブには実世界のさまざまな出来事が写し込まれている．実世界は時間の経過と共に時々刻々と変化しているので，その変化をウェブをマイニングすることにより掴まえることができるならば，それを手がかりとして，実世界で一体何が起こっているのかをタイムリーに知ることができるのではないかと考えられる．そこで我々は，実世界の出来事はそれに関連する検索キーワー

ドによる検索エンジン結果ページ（search engine results page, SERP）のランキングに変動を与えるという知見[5, 7, 10]に基づき，利用者（社会科学分野の研究者）が指定する検索キーワードに対して，さまざまな検索エンジンの SERP を定期的に収集し，SERP 順位の変動を分析し，その変容により警告を発して，利用者に調査を促すシステムを開発している．ここではそのシステム開発の現状を報告する．

1.2. 研究･開発の背景

ウェブ（World Wide 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．ウェブ 2.0 が提唱され，ウェブが有する潜在的可能性がますます顕在化されつつあるなか，筆者らは，ウェブコミュニティの分析研究[1～10]を通してウェブマイニングが社会科学の新しい研究方法論になりうる可能性を明らかにしてきた．端的に表現すれば，社会科学ではこれまで,アンケート調査,インタビュー調査,あるいは実地調査が主たる社会調査法として知られているが，これからはウェブマイニングがそれらに加えて，有望な研究方法論になりうるのではないか，という知見である．また，その研究過程で，検索サイト Google の SERP（ search engine results page，検索エンジン結果ページ）には， その表示順位に Google が公表している順位付けストラテジでは解明しがたい不可解さがあることを発見して，検索サイトの信用性（trustworthiness）に関する研究[5, 7, 10]も発表してきた．この一連の研究で， 我々が次に行わなければならないとは「ウェブマイニングは社会科学の新しい研究方法論」という，これまでの研究を通して得た発見を確たるものとすることである．我々は，このような研究を進めるには，単にデータベースエンジニアがウェブマイニングツールを構築して何かを検証しようとしても，ドメイン知識の欠落ゆえに， その真価を問えないことによりそれ以上研究が進捗しないことを認知し，特に社会学で活発に研究が行われているジェンダー分野に焦点を当てて，ジェンダーに関するドメイン知識を豊富に有する者を研究チームの主要メンバーとして擁して研究を遂行することにより，理工学の域を超えた研究成果を得ることに成功してきた．換言すれば，文理融合した研究体制を整えることにより，初めてウェブマイニングの分析結果や SERP Ranking の信頼性を的確に判断することが可能となり，研究が進展する一方，そこで得られた知見をエンジニアにフィードバックすることにより，真に有用な工学的進展が達成されたのである．

1.3. 関連する研究･開発

SERP の順位はウェブページをデザインしたときに，一番気になることである．ウェブページをデザインするにあたっては，ヒットしてほしい検索キーワードをソースコードの meta tag 部分で指定するのが普通であるが，その結果，それらのキーワードで検索をかけた場合に SERP での順位が大変気になることとなる．そのために，SEO (Search Engine Optimization)技術が開発されてきたことはよく知られている．

SERP に人々は敏感であるがゆえに，これまでいくつかのツールが開発されてきた．その典型が，SERP Checker[1]や SERP Rank[2]や Search Engine Position Checker[3]などのツールで，そのようなツールを提供するサイトにアクセスして，以下のようなインタラクションのもと，結果をうる：

【SERP Checker などが提供する機能】

１．　注目するウェブサイトの URL を指定する．ここに注目するとは，たとえば自分が作成したウェブサイトである．

２．　検索キーワードを指定する．そのサイトがそのキーワードで検索された場合，SERP の何番目に位置するかを知りたい．

３．　着目する検索エンジンを（一般には複数個）指定する．

４．　それらを入力指定して，チェックを実行させると，注目するウェブサイトが，指定した検索キーワードのもと,どの検索エンジンで何位で表示されるか,という結果の一覧が表示される．

上記のツール以外では，検索エンジンは世界にいくつかあることを念頭に，またそれぞれの検索エンジンの SERP 順位付けのストラテジが異なることを考慮に入れると，検索キーワードを入力した場合，ひとつの検索エンジンに着目するのではなく，複数の検索エンジンの SERP に留意した検索結果を得たいとする要求がある．このために，大別すると次の二つのアプローチのもとツールが作成，提供されてきた．

１．　統合型メタサーチエンジン

２．　非統合型メタサーチエンジン

統合型メタサーチエンジンは指定された検索キーワードによるそれぞれの検索エンジンの SERP 順位を何らかの正規化と重み付けのアルゴリズムで計算し，SERP を合成してそれを検索結果としてユーザに返す．たとえば，本を検索する場合に，いくつかのインターネット書店を一括して検索してそれらの結果を合成してユーザに返す例が挙げられる．非統合型メタサーチエンジンは，検索キーワードに対して，指定された複数の検索エンジンでの SERP を並列に返すよう機能す

[1] http://www.selfseo.com/serp_check.php
[2] http://www.serprank.com/
[3] http://www.webmaster-toolkit.com/search-engine-position-checker.shtml

る．このような機能は広義にはマッシュアップ（mushup）と考えられる．

検索エンジンが返す SERP は必ずしも利用者の検索意図に沿ったものではないとの観点から，SERP をできるだけ利用者の検索の意図に沿ったものとするため，文脈の指定，検索ログの利用などさまざまなアイディアのもと数多くの研究が行われている．ここでは，それらに一つ一つ言及することはしないが，このような研究も広義には検索エンジンをベースにしたマッシュアップの研究といえよう．

最近，Google 社が Google Trends[4]というサービスを開始した．これは検索キーワード（複数でもかまわない）を指定すると，過去にさかのぼり，そのキーワードが使われた頻度を時間軸上に表示してくれ，それと同期させる形で，Google ニュースの参照数から検索キーワードの使用頻度とニュースを関連付けてみることができるというサービスである．SERP Watcher と Google Trends は，ある検索キーワードに注目して指標を時系列的に追うところは同じであるが，次の点で本質的に異なる．

１．　Google Trends は検索キーワードが使用された頻度に着目しているのに対して，SERP Watcher は検索キーワードに関して Google を含む主要な検索エンジン毎で SERP 順位の変動に着目している点が異なる．

２．　Google Trends は現時点から過去を振り返り検索キーワードの使用頻度変化を見るのに対して，SERP Watcher は指定された検索キーワードについて SERP 順位に変動が起きた時点で利用者に社会調査を喚起する点が異なる．

最近，Google はそのトップページで「急上昇（検索キー)ワード」を 5 個表示するサービスを始めている．その他，新規ウェブページの立ち上がりやウェブページの更新あるいは移動を監視して報告してくれるサービスを有料で行う会社[5]もある．

以下の議論で明らかになるように，SERP Watcher は上記のいずれの研究ともその目的や手法を異にするものである．ここで，我々が開発している SERP Watcher と従来開発されてきた SERP Checker，あるいはメタサーチエンジンとの違いをまとめておくことは意味のあることと考えられるので，それを表 1 に示す．

1.4. 各章の内容の概要

以下，第 2 章で，SERP Watcher の設計を論じる．2.1 節で SERP Watcher の持つべき性質を詳しく規定する．2.2 節でそのシステム設計の概要を，2.3 節でそのデータベース機能を論じる．リレーショナルデータベース構成を基本とするものの，OLAP キューブを構成することによる多次元データベース構成が有力な分析機能になることを示す．2.4 節では SERP Watcher のアラーム機能について論じる．第 3 章では，これまで前実験として収集してきたいくつかのキーワードごとにとられてきた SERP の順位変動を分析して，アラート機能を実現するための考察を行う．第 4 章は，まとめと今後の課題である．

表 1　関連研究のまとめ―各種ウェブツールの比較―
Table 1. Summary of the Related Works -Comparison of Various Web Tools-

ツールの名称／タイプ	機能	目的
SERP Watcher	検索キーワード，アラート条件を指定，アラートが来ると具体的な社会調査を開始する	社会科学での新しい社会調査法
SERP Checker あるいは SERP Rank	ウェブページの URL，検索キーワード，検索エンジンを指定して，そのウェブページの SERP における順位を返す	SEO がらみ
メタサーチエンジン	複数の検索エンジンの SERP をまとめて表示	SERP の精度向上
検索エンジン精度向上ツール	検索エンジンの SERP に手を加えて精度を向上させる	SERP の精度向上
Google Trends	検索キーワードの使用履歴をみる	検索キーワードの使用動向の分析

2. SERP Watcher―その設計―

2.1. SERP Watcher とは

SERP Watcher とは，以下に示す事項を考慮して実現される SERP 順位変動監視ツールである．

１．　同じ検索キーワードを入力しても，検索エンジンが異なると SERP 自体が一般には異なる．これは，検索エンジンのタイプ（ロボット型かディレクトリ型か)，同じ検索エンジンでも異なるアーカイブ（＝収集したウェブページ群)にアクセスしたことによるずれ，などが原因である．したがって，ツール構築の我々の目的が「新しい社会調査法」の提供にあるならば，一つの検索エンジンに固執するのではなくて，異なる検索エンジンについては皆，その SERP に注目しなければ，社会の変動を見逃してしまうことにつながりかねず，原則として全検索エンジンの SERP を監視の対象とする．

２．　しかしながら，本研究では，まずはロボット型検索エンジンを用いた検索サイトの SERP を監視することとする．なぜならば，社会変動の予兆となるような SERP の変化は，「ウェブサイトの登録」を前提とするディレクトリ型検索エンジンが捕捉するよりも時間的に前に，登録もされていないウェブサイトのウェブページをロボットが拾ってきて，SERP の変動として現れるのではないか，という考えによる．

[4] http://www.google.co.jp/trends
[5] http://searchenginewatch.com

３．　ここで，ロボット型検索エンジンの相関関係を考慮しておくことが，重複したデータを収集しないためにも，また SERP 分析のためにも重要である．**図 1** に α SEO が示す 2008 年 12 月のロボット型検索エンジンの相関図を示す．

図 1　ロボット型検索エンジンの相関図[6]

Fig 1. Correlation Diagram of Search Engine Robots

図からわかるように，独立した検索エンジンを有する検索サイトは，Google, Yahoo! Japan, msn, freshEYE, ASK.jp である．ちなみに，Infoseek 楽天は Google から「powered by」されている．このようにどこかの検索サイトから SERP を提供してもらっている検索サイトは原則として SERP Watcher の対象としなくてもよい．しかしながら，我々がこれまで行った研究で，Google がその提携サイトに提供した SERP が Google 本体が使用した SERP と異なったものであるとしか結論の仕様がない事例を見つけており[5, 7, 10]，Google, Yahoo! Japan, msn, freshEYE, ASK.jp に加えて，Infoseek 楽天と excite も監視すべき検索サイトに加える．また，サービスを開始した「百度」も監視の対象とする．

４．　社会の変動は連続して監視される SERP 順位の変動に現れるであろうという知見をもとにツールは開発されねばならない．つまり，我々のこれまでの研究から，たとえば，我が国の男女共同参画社会基本法の制定をめぐって，ジェンダー擁護派とジェンダーバッシング派がそれらが擁するウェブサイトの SERP 順位を争って極めてめまぐるしく変動し，ジェンダーバッシング派が勝利した時点からは，その変動がぴたりと治まったという事例に遭遇している．

５．　だからこそ，Watcher は指定された検索キーワードのもと，複数検索エンジンの SERP を定期的に収集し分析して，何らかの異変を検出したならば直ちに利用者（ここでまず想定しているのは，社会の変動を鋭敏にキャッチしたいと考えている社会科学の研究者）にその異変を知らせる．

【SERP Watcher の提供する機能】

[6] http://www.alphaseo.jp/seo-report/081210_143352.html　より引用

１．　検索キーワードを指定する．検索キーワードの選定は，このキーワードによる何らかの検索エンジンの SERP が社会変動をキャッチすることがあるのではないか，と利用者が推測して取り上げたということである．

２．　SERP Watcher のアラート条件を指定する．条件としては，default 条件と（それに優先する）ユーザ指定の 2 種類を設定可能とする．

３．　それらを入力指定して，あとは SERP Watcher の実行に任せると，それは定期的にさまざまな検索エンジンで指定された検索キーワードの SERP を監視し，アラート条件が「真（true)」となった時点で，利用者に警告を発する．

2.2. SERP Watcher のシステム設計

これまでの議論から，SERP Watcher は次の部分機能を有しないといけない．

１．　検索キーワード指定機能

２．　アラート条件指定機能

３．　SERP データベース構築機能

４．　SERP データマイニング機能

ここに，4. の SERP データマイニング機能は，3. の機能のもとで格納されていくストリームデータを対象に，2. で指定されたアラート条件が満たされていくかどうかを SERP 収集のタイミングで毎回検証する機能をいう．**図 2** に SERP Watcher の概念図を示す．

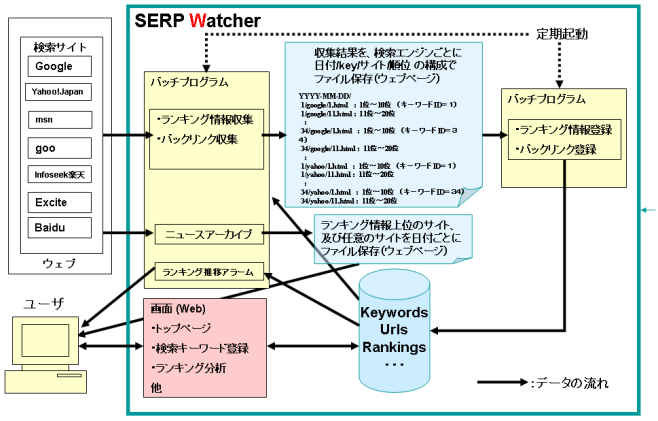

図 2 SERP Watcher の概念図

Fig 2. Conceptual Diagram of SERP Watcher

2.3. SERP Watcher のデータベース設計

SERP Watcher が取り扱うデータを一元的に管理するためのデータベースを設計する．管理しなければならないデータ群は次のとおりである．

１．　利用者データ

２．　検索キーワードデータ

３．　（SERP の）収集日データ

４．　SERP データ

５．　（SERP に現れる）ウェブページデータ

使用者データ，検索キーワードデータ，SERP の収集日データ，そして SERP データの 4 項組で一つの意味のあるデータとなる．その SERP に現れる各ウェブページについて，URL，タイトル，スニペット，バックリンク（の数と個々の URL）を収集・格納する必要がある．**図 3** に設計した SERP Watcher データベースのリレーショナルデータベーススキーマを示す．

ここで，SERP Watcher データベースの性質について論じておくことは意味がある．2 つの特徴がある．

１．　SERP は定期的に収集される．

２．　収集された SERP データは蓄積される一方で，更新されることはない．

つまり，SERP Watcher データベースは一種のストリームデータであり，なおかつ更新されることはないので，OLAP (Online Analytical Processing)処理を行うことにより，多次元的な分析が可能となる．OLAP には ROLAP と MOLAP という 2 つの実現方式があるが，SERP Watcher ではデータはリレーショナルデータベースとして組織化されているので，リレーショナルデータベースに格納されたデータを直接検索・集計し、結果をクライアント側で多次元データに構成して視覚化する ROLAP 方式で OLAP を実現することができる．

別の問題としては，SERP データは蓄積される一方なので，どのようなデータを不要と判断し，どのようなデータを捨てていくかという問題も生じる．この問題は，定期的に収集する SERP データは時系列データであるとみなせば，ストリームデータマイニングで研究されてきたデータ抽象化の方法を導入することも考えられる．

2.4. SERP Watcher のアラーム機能

指定された検索キーワードで定期的に SERP を収集していくとき，SERP 順位の時系列的変化にユーザが指定した変化が生じたときに，アラームが鳴る機能を「アラーム機能」と呼んでいる．SERP 順位の変動がどのような条件を満たしたときにアラームが鳴るようにするのか，自由に指定できるようにするために，「アラーム記述言語」を策定する必要がある．この際，考慮すべき点が 2 つある．

１．　SERP は一般に複数の検索エンジンで収集されているので，単一の検索エンジンの SERP 順位の変動だけに注目してアラームを鳴らすのか，そうではなくいくつかの検索エンジンで同時にアラーム条件が「真」となった場合にアラームを鳴らすのか，いずれの場合も条件を利用者が指定できる記述能力を有すること．（利用者指定条件）

２．　利用者が何も指定をしない場合でも，ある種の条件が満たされれば，アラームが鳴るという default 機能を持たすこと．（Default 条件）

2.5. SERP Watcher の分析･表示機能

アラームが鳴り，利用者は SERP 順位の変動を確認しつつ，さらにその変動の原因を調査するために，SERP Watcher が有するべき分析と表示機能を論じる．明らかにこの機能は収集された SERP Watcher データベースがリレーショナルデータベースとしての基本構成であることを前提として論じるのか,そうではなく,（ROLAP に基づく）OLAP データベースとして論じるのかで使用できる演算子に違いが出るので，その結果機能にも違いを生じる．端的にいえば，次の違いである．

１．　リレーショナルデータベースの基本構成の場合,リレーショナルデータベース言語 SQL で分析を行う．

２．　OLAP の場合，多次元データベース構成になるので，ダイス（dice），スライス（slice），ドリルダウン（drilldown）といった多次元分析に特有な演算を使える．

もう一つ大事な機能として，アラームがなった時点で，利用者に次の仕事ができるようなインタフェースを備えておくことが大事である．これらの機能は蓄積した SERP データを直接触るものではないが，SERP 順位の変動の原因を理解する上で，欠かせない機能であると考えられる．特に 2 つの機能を挙げる．

１．　現時点のウェブページ分析を行える機能

２．　過去に遡り，たとえば新聞記事などを閲覧できる機能

3. SERP Watcher のデータ可視化機能

SERP アーカイブを検索した結果を OLAP に基づいて可視化する機能をシステムに埋め込んだ．これにより社会科学の研究者が発想・分析することを支援する **図 4 と図 5** にシステムのスクリーンショットを示す．

検索エンジン，日付，登録されている検索キーワードなどの中からどのディメンションを軸に設定するかの選択行為（ダイシング）は画面上部のラジオボタンやプルダウンメニューから指定して組み合わせる．これにより様々な視点や角度からの分析，集計が可能となる．例えば，図 4 では Google での「ジェンダー」の検索順位が時間とともにどのように推移を示したかを表示している．一方図 5 では，2008 年 12 月 19 日での「ジェンダー」の検索順位がそれぞれの検索エンジンで何位であったかを一覧表示している．

他にもスライシング,ドリリングなどの一連の OLAP 操作をサポートしている．

また，図 4 では 2008 年 12 月 19 日を基点とした表示結果,図 5 では Google を基軸として展開しているが，別の日付や検索エンジン名をクリックして指定することにより，その日付や検索エンジンで得た順位を帰順点として再読み込するような機能も有する．

さらに，基順点よりも順位が高くなると背景色を赤く，低くなると青くなるようにし,その順位差に濃淡を付けている．これにより時系列的な推移や他要素との比較判別がより容易になる．

この機能は視認性に優れており，社会科学の研究者にとって優しいユーザインタフェースであると同時に，強力な分析ツールとなる． なぜなら， 順位の入れ替わりに際しては社会科学的に大変興味深い事象が関連していることが多いからである．

rank	title	2008/12/19	2008/12/21	2008/12/22	2008/12/23	2008/12/24	2008/12/25	2008/12/26	2008/12/29	2008/12/30	2008/12/31	2009/01/01	2009/01/02
1	ジェンダーフリーとは ～Q&Aですぐわかる！～	1	1	1	1	1	1	1	1	1	1	1	1
2	恐るべしジェンダーフリー教育	2	3	2	2	2	2	2	2	2	2	2	2
3	ジェンダーフリー	3	2	3	6	6	6	6	6	6	6	5	5
4	ジェンダーフリーとは - はてなキーワード	4	5	6	3	3	3	3	3	3	3	6	6
5	ぶっ飛ばせジェンダーフリー 真の男女共同"家族・社会"をめ	5	4	5	5	5	5	5	5	5	5	4	4

図 4 多次元データの可視化例（横軸：日時，縦軸：Web ページ，対象：順位）

Fig 4. A Visualization Example of Multidimensional Data (Horizontal axis: Date, Vertical axis: Web Pages, Subject: Rank)

rank	title	google	excite	Yahoo!	infoseek	baidu	msn	goo
1	ジェンダーフリーとは ～Q&Aですぐわかる！～	1	2	2	1	14	5	1
2	恐るべしジェンダーフリー教育	2	7	7	6	28	14	2
3	ジェンダーフリー	3	4	4	3	4	1	7
4	ジェンダーフリーとは - はてなキーワード	4	11	11	9	2	7	5
5	ぶっ飛ばせジェンダーフリー 真の男女共同"家族・社会"をめざそう ...	5	9	9	12	15	6	4

図 5 多次元データの可視化例（横軸：検索エンジン，縦軸：Web ページ，対象：順位）

Fig 5. A Visualization Example of Multidimensional Data (Horizontal axis: Search Engines, Vertical axis: Web Pages, Subject: Rank)

4. まとめと今後の課題

検索エンジン技術を用いた社会科学の多角的な調査を支援し，新たな発見を支援するシステムを開発することで，ウェブマイニングが社会科学の新しい研究方法論になることを実証する目的で， SERP Watcher の構築を論じた．

今後の課題として，SERP Watcher のアラーム機能の分析，ROLAP による SERP 順位変動分析機能の設計，それに基づくユーザインタフェースの実現方式の精査があげられる．

【謝辞】

本研究は一部，日本学術振興会平成 20 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．

文　献

[1] 増永良文，小山直子:ジェンダー関連 Web サイトのコミュニティ分析とポータルサイト構築―Web コミュニティの関連性から見たグローバル化―，「グローバル化とジェンダー規範」に関する研究報告書， pp.101-122， お茶の水女子大学, 2002 年 3 月．

[2] 小山直子， 増永良文：Companion-を用いたジェンダー関連 Web コミュニティの詳細分析，夏のデータベースワークショップ（DBWS2004）会議録，7A-3， 2004 年 7 月．

[3] 増永良文， 小山直子：Web マイニングツールを用いたジェンダー関連 Web コミュニティの通時的分析，日本データベース学会 Letters Vol.3, No.3, pp.21-24, 2004 年 12 月．

[4] Naoko Oyama, Yoshifumi Masunaga, Kaoru Tachi: A Diachronic Analysis of Gender-related Web Communities using a HITS-based Mining Tool, Frontiers of WWW Research and Development --APWeb2006, LNCS3841, Springer, pp.355-366, January 2006.

[5] 小山直子， 増永良文， 舘かおる： ウェブ検索ポータルサイトの信用性と透過性―検索キーワード「ジェンダーフリー」を通して見るウェブの世界―、DEWS2006（ 電子情報通信学会 17 回データ工学ワークショップ／ 第 4 回日本データベース学会年次大会）会議録, ISSN 1347-4413, 1B-i7, 8p.， 2006 年 3 月．

[6] 小山直子： 社会現象の分析手法としてのウェブマイニング、日本のデータベース研究最前線第 22 回，月刊Ｄ Ｂマガジン 2006 年 6 月号，翔泳社．

[7] 増永良文， 小山直子:キーワード“ジェンダーフリー”を通してみる検索サイト Google の信用性と透明性，日本データベース学会 Letters, Vol.5, No.2, pp.105-108, 2006 年 9 月．

[8] 増永良文：コンピュータサイエンス入門―コンピュータ・ウェブ・社会―（本），第 14 章ウェブと社会，サイエンス社，2008 年 1 月．

[9] 石川 沙織，渡辺 知恵美，小山 直子，舘 かおる，増永 良文：検索エンジン技術を用いた社会科学の多角的調査支援システムの開発， DEWS2008 A1-5， 2008 年 3 月．

[10] Naoko Oyama and Yoshifumi Masunaga: On the Trustworthiness and Transparency of a Web Search Site examined using “Gender-equal” as a Search Keyword, Proceedings of APWeb2008, LNCS, Springer, April 2008.

※ 次ページに図 3（Fig 3）を掲載

keywords

Attribute	Type	Null	Key	Default	Extra
id	int(11)	NO	PRI	NULL	auto_increment
keyword	varchar(256)	YES		NULL	
user_id	int(11)	YES		NULL	
wday0	tinyint(1)	YES		NULL	収集曜日（日）
:					:
wday6	tinyint(1)	YES		NULL	収集曜日（土）

urls

Attribute	Type	Null	Key	Default	Extra
id	int(11)	NO	PRI	NULL	auto_increment
title	text	YES		NULL	
abst	text	YES		NULL	
url1	text	YES		NULL	
url2	text	YES		NULL	
user_id	int(11)	YES		NULL	
wday0	tinyint(1)	YES		NULL	収集曜日（日）
:					:
wday6	tinyint(1)	YES		NULL	収集曜日（土）

users

Attribute	Type	Null	Key	Default	Extra
id	int(11)	NO	PRI	NULL	auto_increment
login	varchar(255)	YES		NULL	ユーザID
email	varchar(255)	YES		NULL	
crypted_password	varchar(40)	YES		NULL	
salt	varchar(40)	YES		NULL	
created_at	datetime	YES		NULL	
updated_at	datetime	YES		NULL	
remember_token	varchar(255)	YES		NULL	
remember_token_expires_at	datetime	YES		NULL	

rankings

Attribute	Type	Null	Key	Default	Extra
id	int(11)	NO	PRI	NULL	auto_increment
snap	datetime	YES		NULL	
rank	int(11)	YES		NULL	
engine	varchar(128	YES		NULL	
keyword_id	int(11)	YES		NULL	
url_id	int(11)	YES		NULL	

backlinks

Attribute	Type	Null	Key	Default	Extra
id	int(11)	NO	PRI	NULL	auto_increment
snapdate	date	NO	MUL		
target_url_id	int(11)	NO			
url_id	int(11)	NO			
created_at	datetime	YES		NULL	
updated_at	datetime	YES		NULL	

archives

Attribute	Type	Null	Key	Default	Extra
id	int(11)	NO	PRI	NULL	auto_increment
snapdate	date	NO	MUL		
url_id	int(11)	NO			
arckind	varchar(32)	YES		NULL	アーカイブの種類
created_at	datetime	YES		NULL	
updated_at	datetime	YES		NULL	

Arrows represent foreign keys.

図 3 SERP Watcher のデータベーススキーマ

Fig 3. Database Schema of SERPWatcher