特　集

ウェブアクセスログとその利活用

Web Access Logs and its practical use

大塚 真吾　東京大学 生産技術研究所
Shingo Otsuka　Institute of Industrial Science, The University of Tokyo
otsuka@tkl.iis.u-tokyo.ac.jp

喜連川 優　（同　上）
Masaru Kitsuregawa　kitsure@tkl.iis.u-tokyo.ac.jp

keywords: Web access logs, Access logs mining, Access logs analysis, Web usage mining

1. は じ め に

インフラ網の急速な普及により，誰もが気軽にインターネットへアクセスできるようになった．その結果として，サイバー空間上における人々の行動が実世界に影響を与え始めた．例えば，検索エンジンやポータルサイトの検索語の解析から今後の流行を発見することが可能となり，実際に検索サイトなどでは検索数が急上昇した検索語を「注目キーワード」として公開している．また，テレビ番組や雑誌の記事などではジャンル別の検索語ランキングや急上昇した検索語の集計を行い，その結果が公表されている．さらに，独自の集計でウェブサイトのトラフィック情報を公表している Alexa の登場により，SEO/SEM (Search Engine Optimization/Marketing) の観点からもユーザのウェブページ閲覧情報に注目が集まっている．

ウェブページの閲覧情報であるウェブアクセスログから人々の行動パターンや傾向を抽出することは重要であり，現在までにリコメンテーションシステムやショッピングサイトおける購買促進など数多くの研究が行われている．ユーザの行動パターンの抽出にはユーザの特定が重要となるが，多くの研究では Cookie の情報やウェブページ閲覧者の IP アドレスをもとに特定を行っている．これに対し近年，テレビ視聴率調査と同様，統計的に偏りなく抽出されたユーザ（パネル）を対象に URL 履歴の収集を行うウェブ視聴率調査が登場し，このデータを用いることでユーザ毎にウェブページの閲覧情報を解析することが可能となった．各ユーザが閲覧したウェブページの変遷履歴は「クリックストリーム (click stream)」と呼ばれている．

このように，現在ウェブアクセスログは注目されており，様々な分野で研究や試みが行われている．そこで，本稿では，最近のウェブアクセスログの研究事例やアクセスログの動向について述べる．

2. アクセスログに関する研究事例

アクセスログを用いた研究は現在まで数多く行われておりその目的も様々である．この節では，ここ数年の間に発表された論文を中心に研究の紹介を行う[*1]．

2·1 ユーザやウェブページのクラスタリングに関する研究

アクセスログから類似するユーザの行動パターンを発見し，ユーザのグループ化やユーザのパーソナライゼーションを行う研究は購買促進や新規顧客の獲得などビジネスに結び付くため，現在までに様々な研究が行われている．また，ユーザが訪れたページの閲覧情報をもとにウェブページやページで売られている商品のクラスタリングを行う研究も行われている．その成果は amazon のリコメンテーションシステムに代表されるように，現在では様々な商用ツールが存在している．文献 [Eirinaki 03] では，パーソナライゼーションについて詳細な説明や各ツールの比較などを行っている．

文献 [Murata 04] ではウェブ視聴率調査データからユーザとそのユーザが閲覧したウェブページの URL 中の term 情報からユーザのグループ化を行う手法について述べている[*2]．

文献 [Zeng 02] ではアクセスログからユーザとそのユーザが閲覧したウェブページの組合せをもとにユーザとウェブページの有向 2 階層グラフを作成し，サイン係数を用いた類似度計算をもとにウェブページとユーザの両方のクラスタリングを行う手法を提案している．また，クラスタ化したユーザとウェブページの相関の抽出法についても論じている．UC Berkeley のコンピュータサイエン

*1　現在，アクセスログに関連する研究内容の分類について体系だったものが存在しないため，ここでは筆者らが独自に分類を行った．

*2　例えば，あるユーザが Yahoo! Japan で天気予報を見た場合，URL(http://weather.yahoo.co.jp/weather/) から，そのユーザは weather や yahoo と関連性があるといった情報を抽出できる．

ス科でコンピュータを利用している学生や教職員のウェブページ閲覧情報を用いた実験結果から，従来の手法 (k-means 法) は授業の内容に関係なく「PPT スライドの集合」などページの形態に着目したクラスタを生成するのに対して，提案手法では「モバイル関連のページ群」のようなあるトピックに関連したページ群をクラスタとして生成できたと述べている．

2·2 検索語のクラスタリングに関する研究

検索エンジンやポータルサイトのアクセスログの解析からユーザが入力した検索語と閲覧されたウェブページの組合せを大量に得ることが可能なため，これを用いて検索語のクラスタリングを行うことが可能である．最近では，Google サジェスト，goo サジェスト，Yahoo!の入力補助機能，など新しいサービスが提供され始めた．

文献 [Beeferman 00] では Lycos の 1 日分のアクセスデータ (50 万レコード) から，検索語とその直後に閲覧されたウェブページの組合せをもとに 2 部グラフを作成し，これ用いて検索語のクラスタリングを行う手法を提案している．また，サーチエンジンにおける応用例についても述べている．

文献 [Wen 02] ではオンライン百科事典の Encarta のサーバのログを用いて，Encarta サイト内でユーザが入力した検索語をクラスタリングする手法について述べている．論文中では以下の 4 つのクラスタリング手法を提案している．

(1) 複数の検索語を同時に入力したときの語の共起確率を利用
(2) 検索語を入力したすぐ後に閲覧したウェブページの共起確率を利用
(3) 1 の手法と 2 の手法の組合せ．
(4) 1 手法と 2 の手法にウェブサイトの階層構造を反映させたもの[*3]

無作為に抽出した検索語 20,000 語を対象に実験を行い，各手法により生成された 100 個のクラスタについて手作業で再現率と適合率を評価した結果，4 の手法が良いと述べている．

文献 [大久保 98] では NTT DIRECTORY で入力された検索ログを用いて，例えば「桜と花見」のようなある一定の期間では関連語 (同義語) となる検索語の発見から，ユーザの情報ニーズを抽出する方法について述べている．ユーザの検索要求が時間と共に変遷する例として，桜の花が咲く前の期間では「桜」は「桜前線や開花」など桜の咲き始める時期との関連が強いのに対して，桜の開花後は「造幣局や北海道」など，桜の名所との関連が強いという結果を示している．

*3 Encarta サイトは百科事典であるため，サイトの構造は分類ごとに階層化されている (論文中の説明では 4 階層)．論文では階層内のページ同士の類似度は高いため，クラスタリングの精度が向上すると述べている．

我々も検索語のクラスタリングに関する研究を行っており，NTT が運営している「i タウンページ」のアクセスログを用いて，サイトを利用する人の検索サポートを行う手法を提案している [Ohura 02]．例えば，出張先のホテルを検索したときの該当結果がない場合に結果を 0 件にするのではなく，サウナや民宿で登録されているページを表示させることを目的としたシステムの提案を行っている．

また，ウェブ視聴率調査データを用いた関連語の抽出法について研究を行っている [大塚 05]．この研究では [Toyoda 01] の方法でウェブコミュニティ[*4]を作成し，検索語とその直後に閲覧されたウェブコミュニティの組合せから，関連語の抽出を行う手法を提案している．また，ユーザが閲覧したウェブページのテキスト情報から形態素解析を用いて名詞を取り出し，これをもとに関連語の抽出を行う手法も提案している．

2·3 ユーザの行動パターン抽出に関する研究

ポータルや大学などウェブページを多く保持しているサイトを運営しているサイトのアクセスログやウェブ視聴率調査データを用いて，ユーザの行動パターンを抽出する試みが行われている．

文献 [Ali 03] ではサイト内におけるユーザの最良パス（Golden Path）を求める手法について述べている．車のポータルサイトのデータを用いて，トップページから問合せフォームページまでの最良パスの発見例を示している．

文献 [Jin 04] では大学のサーバのアクセスログと国営不動産会社のログを用いて，潜在的なユーザ行動の抽出方法について述べている．また，応用例として「安い物件を探しているユーザの行動パターンから安い物件のリストを作成する.」などウェブページ (データ) のグループ化についても述べている．

文献 [齋藤 06] ではウェブ視聴率調査データを用いてユーザの興味を把握するために，ユーザが入力した検索語と閲覧したサイト名を利用した可視化ツールの提案を行っており，ツールの利用例から新車や国内旅行について調べてる行動が把握できると述べている．

文献 [山田 05] ではユーザの興味の変遷パターンを抽出する手法の提案とその可視化について論じている．可視化ツールから，車の購入を検討しているユーザや健康に興味があるユーザの行動パターンを発見できたと述べている．

我々も文献 [Pramudiono 02] において，アクセスログを用いてサイトを利用するユーザの行動パターンを抽出する方法を提案している．i タウンページを用いた実験結果から「駅名を秋葉原にして検索するユーザはレストラ

*4 本稿では，ウェブコミュニティを「同じトピックに関心をもつ人々や組織によって作成されたウェブページの集合」と定義している [村田 03].

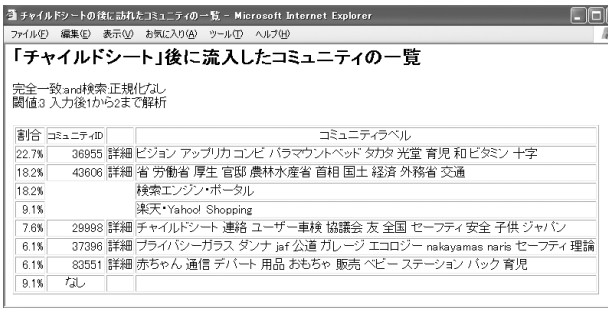

図 1 「チャイルドシート」と入力した後に流入したコミュニティの表示例

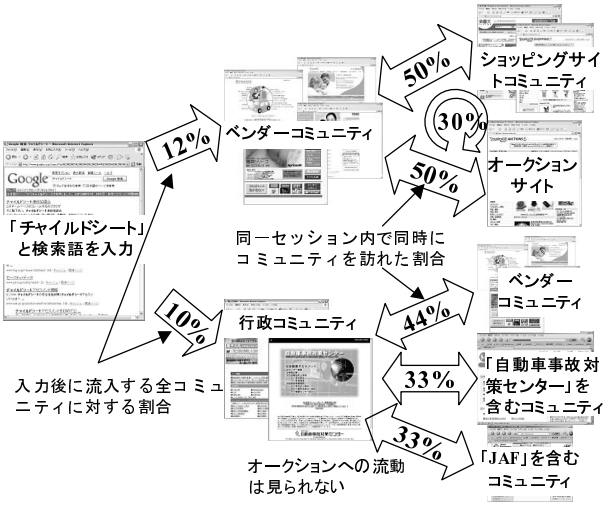

図 2 「チャイルドシート」と入力したユーザの行動

ンを探している場合が多い.」や「住所を北海道のある地名にして夜に検索をする人はガソリンスタンドを探している場合が多い」などの行動パターンの発見できた.

また，ウェブ視聴率調査データを用いてユーザの大域的な行動を抽出する研究を行っている [大塚 03, Otsuka 04]．データ中にある検索語と予め作成したウェブコミュニティ([Toyoda 01]) を用いてシステムの構築を行っている.「チャイルドシート」と入力した後に流入したコミュニティの解析を行った結果を図 1 に示す．解析結果は流入が多いコミュニティ順に表示される．1 番多いコミュニティ(ID:36955) のラベル[*5]からチャイルドシートベンダーに関連するコミュニティであることが推測できる．流入が 2 番目に多いコミュニティ(ID:43606) のラベルからこのコミュニティは行政関連であることが容易に想像できる．また，4 番目は楽天などのショッピングサイトである．

この結果とその他の解析機能を用いることで，我々は

*5 各々のコミュニティに含まれるページに対して張られたリンクのアンカータグの解析から，充分に正確ではないもののコミュニティの内容を表す単語群（コミュニティラベル）を自動的に抽出できており，これにより，解析者はコミュニティに含まれる個々のウェブページを閲覧することなくコミュニティの概要を把握できる.

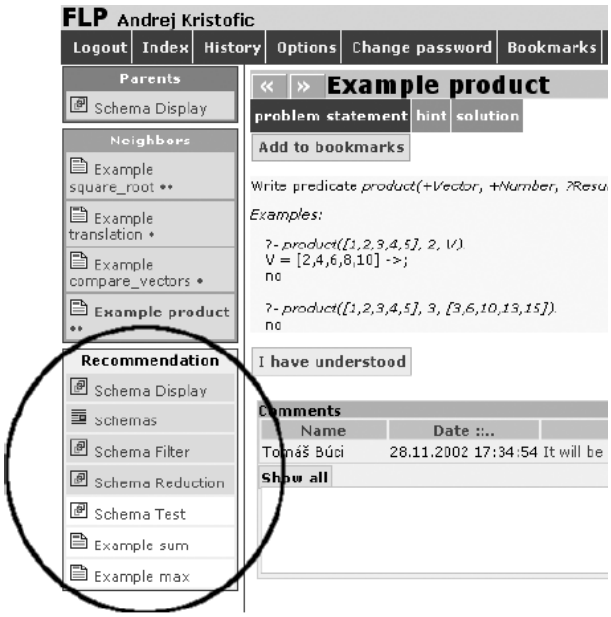

図 3 教材推薦システムの例

図 2 に示すような大域的なユーザの行動パターンを抽出できた[*6]．チャイルドシートの使用期間は短いためオークションなどで中古品を探すユーザが多く，同時にチャイルドシートベンダーとショッピングサイトで性能と販売価格の調査を行う傾向がある．一方，行政関連のコミュニティを訪れるユーザはベンダーや JAF などを含むコミュニティを訪れることから，チャイルドシートの安全性などの調査が目的だと推測できる.

2·4 その他の研究

文献 [Kristofic 05] では大学で運用しているウェブを使った教育システムのアクセスログから，相関ルールやシーケンシャルパターン抽出を用いて，学生に対して学習項目を推薦するシステムの実装について述べている．プログラミングコースのアクセスログ 3 年分を用いた結果，図 3 に示すように教材の推薦が可能であると述べている.

文献 [Tan 02] では大学のサーバのアクセスログからユーザの行動パターンを抽出する方法や，検索エンジンのロボットなどによるアクセスをログ中から排除する方法について論じている．また，e-commerce サイトのアクセスログを用いて，ウェブページ間の相関を発見するツールを提案している.

文献 [Fang 04] では大学のアクセスログ 3ヶ月分を用いて，大学の portal ページの最適化方法について論じている.

*6 図 2 自体はユーザの全体的な挙動をまとめて概観するために人手で描いたものであるが，個々の流動例えばチャイルドシートを検索語として入力した後 12% の割合でベンダーコミュニティへ，また 10% の割合で行政関連のコミュニティへ流出するという解析結果は本システムにより直接得ることができる.

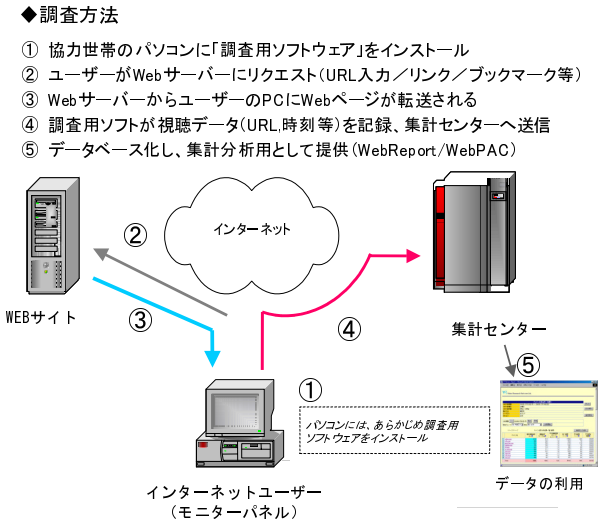

図 4 ウェブ視聴率調査データ収集法の概要

文献 [Cooley 03] では 1 日 60 万件のリクエストがある e-commerce サイトのアクセスログを解析するために作成したシステムについて詳細に述べている．また，頻繁に訪れたページ群の解析手法や販売している商品のクラスタリング手法について述べている．

3. ウェブアクセスログの動向

初めにテレビ視聴率調査データについて述べる．次にユーザが指定したサイトに関するトラフィック情報を提供する Alexa について述べる．最後に一般に公開されているウェブアクセスログの紹介を行う．

3·1 ウェブ視聴率調査データ

米国における本格的なインターネット視聴率調査は 1990 年代半ばから始められた．1990 年代後半には Media Metrix 社，NetRatings 社など複数の調査会社が独自の調査結果を公表していたが，その後，企業買収や合併など業界の再編成が行われ 2006 年現在では，Nielsen//NetRatings 社と comScore Networks 社の comScore Media Metrix 部門がウェブ視聴率調査の 2 大勢力となっている．

一方，日本においては 1999 年に日本リサーチセンターと Hypertak 社が共同でウェブ視聴率調査システムである JAR(Japan Access Rating) を開始したのを皮切りに，その半年後に日経 BP 社，2000 年に入り新たに Nielsen//NetRatings 社，ビデオリサーチインタラクティブ社[*7]，MediaMetrix が参入した．その後，日経 BP 社とビデオリサーチインタラクティブ社との業務提携，JAR のサービス終了，Media Metrix の解体[*8]などを経て，2006 年現在でサービスを行っているのは Nielsen//NetRatings 社とビデオリサーチインタラクティブ社の 2 社となった．また，両社は 2004 年に業務提携を発表しインターネット視聴率データの共有化を行った．したがって，2006 年 4 月現在の日本におけるウェブ視聴率調査データは 1 つのみである．

表 1 ウェブ視聴率調査データの一部

UserID	AccessTime	RefSec	URL
1	2002/9/30 00:00:00	4	http://www.tkl.iis.u-tokyo.ac.jp/welcome_j.html
2	2002/9/30 00:00:00	6	http://www.jma.go.jp/JMA_HP/jma/index.html
3	2002/9/30 00:00:00	8	http://www.kantei.go.jp/
4	2002/9/30 00:00:00	15	http://www.google.co.jp/
1	2002/9/30 00:00:04	6	http://www.tkl.iis.u-tokyo.ac.jp/Kilab/Welcome.html
5	2002/9/30 00:00:04	3	http://www.yahoo.co.jp/
6	2002/9/30 00:00:05	54	http://weather.crc.co.jp/
2	2002/9/30 00:00:06	11	http://www.data.kishou.go.jp/maiji/
3	2002/9/30 00:00:08	34	http://www.kantei.go.jp/new/kousikiyotei.html
5	2002/9/30 00:00:07	10	http://search.yahoo.co.jp/bin/search?p=%C5%B7%B5%A4
1	2002/9/30 00:00:10	300	http://www.tkl.iis.u-tokyo.ac.jp/Kilab/Members/members-j.html

検索語を含むURL

3·2 ウェブ視聴率調査データ抽出方法

日本におけるウェブ視聴率調査データ収集法の概要を図 4 に示し，その調査方法を以下に示す．

- インターネット視聴率調査会社が所有する全国のインターネットユーザーの調査協力サンプル（パネル）[*9]により視聴されたウェブページの情報を収集・集計．
- パネルがインターネット利用に使用するパソコンに調査用ソフトウェアをインストールし，視聴状況をリアルタイムで収集．

このように収集されたパネルログは表 1 に示すように，パネル ID, ウェブページにアクセスした時刻, ウェブページを閲覧した時間, アクセスしたウェブページの URL, などから構成されている．パネル ID とはパネル全員に対してユニークに割り当てた ID であり，個々のパネルが特定できる．また，検索エンジンサイトなどで入力された検索語の情報を保持している．

2006 年 4 月現在，日本におけるパネルの数は約 15,000 人である．Nielsen//NetRatings 社は世界 12ヶ国と 1 つの地域（香港）で数万規模のパネル数を保持している．

一方，comScore Networks 社は全く異なった手法でウェブ視聴率調査を行っている．同社はインターネットプロバイダ事業を行っており，パソコンに調査用ソフトウェアをインストールする見返りとして，

- サーバサイドのウイルスチェックを無料化
- インターネットの高速化
- 宝くじのプレゼント

*7 当時の社名はビデオリサーチネットコム社

*8 Media Metrix 社は 2000 年に Jupiter Communications 社と合併し，Jupiter Media Metrix 社となったが，経営難により 2001 年には Nielsen//NetRatings 社と合併を発表するが，米司法省の反対により合併は白紙となる．その後，北米のインターネット視聴率調査部門を comScore Networks 社へ，オンライン広告調査部門と欧州のインターネット視聴率調査部門を Nielsen//NetRatings 社へ売却し，その他の部門はハイテク関連情報のサイト運営などを行っている INT Media Group の事業部門となった．

*9 パネルはエリア別のインターネット利用率を基に，乱数で発生させた電話番号に対して自動的に電話をかける調査方法 (RDD(Random Digit Dialing) と呼ばれている.) を使って無作為抽出した世帯を決定しパネルの依頼を行う．

図 5 「Livedoor のトラフィックデータ」

などのサービスを行うことで，200 万人を超えるユーザを獲得している．その中から RDD 方式により無作為抽出したユーザをパネルとして視聴率調査データを作成している．

3·3 Alexa

米オンライン書店大手の Amazon.com の子会社である Alexa Internet 社 [Ale] はウェブサイト（ドメイン）毎にトラフィック情報を公開している．このサイトでは，

- 指定したサイト*10の閲覧率 (100 万人のうちどのくらい人が閲覧したか)
- 指定したサイトを閲覧したユーザの平均閲覧ページ数
- 指定したサイトの世界的なトラフィックランキング*11

などの情報を，今日，最近 1 週間の平均値，最近 3ヶ月の平均値，で見ることができる．さらに，過去 3ヶ月，6ヶ月，1 年，2 年,2001 年末から現在まで，の推移をグラフで見る事ができる．例として www.livedoor.com の現在までの閲覧率を図 5 に示す．図の横軸は年月を示し縦軸は 100 万人のうちどれだけの人がこのサイトを閲覧しているかを表している．グラフから 2004 年の 6 月頃から閲覧者（アクセス数）が増加していることがわかる．この時期はこのサイトを運営している企業がプロ野球球団を買収を表明した時期と同じであり，現実世界の動きとアクセス数が関連している事がわかる．その他にも，現実世界の動きと関連してアクセス数が上下している．

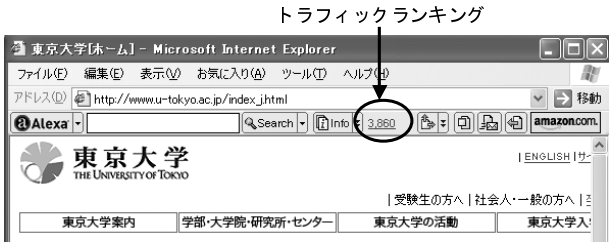

図 6 Alexa ツールバー

また，ウェブページの他にもブラウザのツールバーである「Alexa ツールバー」をインストールすることで，図 6 のようにユーザが閲覧しているページの世界ランキングなどを表示させることができる．Alexa ツールバーはブラウザの閲覧履歴を Alexa Internet 社へ自動的に送信する代わりに閲覧しているサイトのランキングや関連するページ情報の提供を行っている．これらの統計データは Alexa ツールバーをインストールしたユーザのウェブページ閲覧履歴をもとに計算されているため，データの信頼性について疑問視する意見もある．しかし，Alexa ツールバーを利用しているユーザーは世界で数百万人おり，このような大規模なウェブアクセスデータを一般公開しているサイトは他になく，あくまでも目安（サンプル）として利用すれば有用なデータであるといえる．

また，Alexa が有名になる前は，SEM/SEO に関連したサイトではコンテンツの最適化により検索エンジンの上位に自社サイトが掲載されることを評価基準としていたが，最近では，Alexa のランキング向上を評価基準とするサイトも少なくない．例えば，図 5 の一番下のテキストフィールドにサイトの URL を入れることで，競合他社のサイトと比較を行うことができるため，SEO 対策の結果例を競合他社との比較で評価することが可能である．

3·4 The Internet Traffic Archive

ウェブアクセスログを用いて研究を行う場合，実データを利用した検証は重要であるがウェブ視聴率調査データは一般に高価であり誰でも簡単に利用できるデータではない．また，Alexa の統計データの詳細は公表されていないため，ウェブアクセスログの研究に用いることは難しい．そこで，一般に公開されているウェブアクセスログデータである The Internet Traffic Archive[Tra] について紹介する．これは ACM SIGCOMM がスポンサーをしており，NASA などのトラフィックやページ数が多いウェブサーバのアクセスログや，大学のコンピュータを利用している学生や教職員のウェブページ閲覧情報が公開されている．

*10 表示可能なトラフィックデータはトップレベルドメイン単位のみであるため，例えば，http://www.iis.u-tokyo.ac.jp の場合は u-tokyo.ac.jp の解析結果となる．また，各 URL 毎のトラフィックデータは公開されていない．

*11 サイト閲覧率と平均閲覧ページ数を基に地理的な条件（地域によってサンプルユーザ数の数が異なるため）を加味してランキングを決めている．

4. お わ り に

本稿では，ウェブアクセスログを用いた研究事例の紹介とアクセスログの動向について述べた．インターネットの一般への普及により，ウェブ上での流行や評判が実世界に対して影響を与え初めている．そのため，ここ数年でインターネットは重要な役割を担うメディアとして捕らえられ，実際にインターネット広告費は年々増加し，既にラジオ広告費を超えている．今後は広告の効率的な配信やマーケティング戦略などアクセスログに対するより高度な解析手法が望まれる．

また，行政サービスなどの公的機関でも電子化が進んでおり，アクセスログを用いて悪意を持ったユーザの事前特定を行う研究など，今後はさらにアクセスログの解析技術に対するニーズが高まると予想される．

このように，アクセスログに関する研究の重要性は増しているが，研究成果は国内外問わずそれほど多く公開されていない．今後，この分野の研究発展に期待する．

謝　　辞

本原稿を執筆にあたり協力頂いた東芝ソリューション株式会社 SI 技術開発センター 平井潤様に，また，実験で利用したデータの提供に御協力頂いた株式会社ビデオリサーチインタラクティブに感謝します．

◇ 参 考 文 献 ◇

[Ale] Alexa Internet http://www.alexa.com/

[Ali 03] Ali, K. and Ketchpel, S.: Golden Path Analyzer: Using Divide-and-Conquer to Cluster Web Clickstreams, in *The 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD2003)* (2003)

[Beeferman 00] Beeferman, D. and Berger, A.: Agglomerative Clustering of Search Engine Query Log, in *The 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD2000)* (2000)

[Cooley 03] Cooley, R.: The Use of Web Structure and Content to Identify Subjectively Interesting Web Usage Patterns, *ACM Transactions on Internet Technology (ACM TOIT)*, Vol. 3, No. 2, pp. 99–116 (2003)

[Eirinaki 03] Eirinaki, M. and Vazirgiannis, M.: Web Mining for Web Personalization, *ACM Transactions on Internet Technology (ACM TOIT)*, Vol. 3, No. 1, pp. 1–27 (2003)

[Fang 04] Fang, X. and Sheng, O. R.: LinkSelector: A Web Mining Approach to Hyperlink Selection for Web Portals, *ACM Transactions on Internet Technology (ACM TOIT)*, Vol. 4, No. 2, pp. 209–237 (2004)

[Jin 04] Jin, X., Zhou, Y., and Mobasher, B.: Web Usage Mining Based on Probabilistic Latent Semantic, in *The 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD2004)* (2004)

[Kristofic 05] Kristofic, A. and M.Bielikova, : Improving Adaptation in Web-Based Educational Hypermedia by means of Knowledge Discovery, in *HYPERTEXT 2005 Sixteenth ACM Conference on Hypertext and Hypermedia (HT'05)* (2005)

[村田 03] 村田 剛志:Web コミュニティ, 情報処理, Vol. 44, No. 7, pp. 702–706 (2003)

[Murata 04] Murata, T.: Discovery of User Communities from Web Audience Measurement Data, in *The 2004 IEEE/WIC/ACM International Conference on Web Intelligence (WI2004)* (2004)

[大久保 98] 大久保 雅且, 杉崎 正之, 井上 孝史, 田中 一男：WWW検索ログに基づく情報ニーズの抽出, 情報処理学会論文誌, Vol. 39, No. 7, pp. 2250–2258 (1998)

[Ohura 02] Ohura, Y., Takahashi, K., Pramudiono, I., and Kitsuregawa, M.: Experiments on Query Expansion for Internet Yellow Page Services Using Web Log Mining, in *The 28th International Conference on Very Large Data Bases (VLDB2002)* (2002)

[大塚 03] 大塚 真吾, 豊田 正史, 喜連川 優：ウェブコミュニティを用いた大域 Web アクセスログ解析法の一提案, 情報処理学会論文誌：データベース, Vol. 44, No. SIG18(TOD20), pp. 32–44 (2003)

[Otsuka 04] Otsuka, S., Toyoda, M., Hirai, J., and Kitsuregawa, M.: Extracting User Behavior by Web Communities Technology on Global Web Logs, in *Proc. of 15th International Conference on Database and Expert Systems Applications (DEXA'2004)*, pp. 957–968 (2004)

[大塚 05] 大塚 真吾, 豊田 正史, 喜連川 優：大域ウェブアクセスログを用いた関連語の発見法に関する一考察, 情報処理学会論文誌：データベース, Vol. 46, No. SIG18(TOD26), pp. 82–92 (2005)

[Pramudiono 02] Pramudiono, I., Shintani, T., Takahashi, K., and Kitsuregawa, M.: User Behavior Analysis of Location Aware Search Engine, in *The Third International Conference on Mobile Data Management (MDM'02)* (2002)

[齋藤 06] 齋藤 皓太, 村田 剛志：Web 閲覧者の関心キーワードの抽出と巡回行動の可視化, 電子情報通信学会 WI2 研究会資料, pp. 141–146 (2006)

[Tan 02] Tan, P. and Kumar, V.: Mining Association Patterns in Web Usage Data, *International Conference on Advances in Infrastructure for e-Business, e-Education, e-Science, and e-Medicine on the Internet* (2002)

[Toyoda 01] Toyoda, M. and Kitsuregawa, M.: Creating a Web Community Chart for Navigating Related Communities, in *Conference Proceedings of Hypertext 2001*, pp. 103–112 (2001)

[Tra] The Internet Traffic Archive http://ita.ee.lbl.gov/

[Wen 02] Wen, J., Nie, J., and Zhang, H.: Query Clustering Using User Logs, *ACM Transactions on Information Systems (ACM TOIS)*, Vol. 20, No. 1, pp. 59–81 (2002)

[山田 05] 山田 和明, 中小路 久美子, 上田 完次：Web ユーザの行動履歴解析のためのデータマイニング, 電子情報通信学会 WI2 研究会資料, pp. 59–64 (2005)

[Zeng 02] Zeng, H., Chen, Z., and Ma, W.: A Unified Framework for Clustering Heterogeneous Web Objects, in *The Third International Conference on Web Information Systems Engineering (WISE'02)* (2002)

〔担当委員：××○○〕

19YY 年 MM 月 DD 日　受理

著 者 紹 介

大塚　真吾

1996 年千葉工業大学工学部情報工学科卒．2002 年同大大学院工学研究科博士後期課程修了．博士（工学）．同年，東京大学生産技術研究所 学術研究支援員．現在，同大同研究所 産学官連携研究員 特任助手．ログマイニング，テキスト処理，ウェブマイニングに興味を持つ．現在，Web インテリジェンスとインタラクション時限専門委員会専門委員，情報処理学会，日本データベース学会 各会員．

喜連川　優（正会員）

1978 年東京大学工学部卒．1983 年同大大学院工学系研究科情報工学博士課程了．工学博士．同年同大生産技術研究所講師．現在，同教授．2003 より同所戦略情報融合国際研究センター長．データベース工学，並列処理，Ｗｅｂマイニングに関する研究に従事．現在，日本データベース学会理事，情報処理学会，電子情報通信学会 各フェロー．平成 11-14 年 ACM SIGMODJapan Chapter Chair，平成 9,10 年電子情報通信学会データ工学研究専門委員会委員長．VLDB Trustee(97-02), IEEE ICDE,PAKDD,WAIM などステアリング委員, IEEE データ工学国際会議 (ICDE2005) General Chair.