女性向けフリーマガジン発行サイトにおけるユーザの行動分析

大塚　真吾† 　高久　雅生† 　喜連川　優†† 　宮崎　収兄†††

† 独立行政法人 物質・材料研究機構　〒 305–0047 茨城県つくば市千現 1–2–1
†† 東京大学 生産技術研究所　〒 153–8505 東京都目黒区駒場 4–6–1
††† 千葉工業大学 情報科学部 情報工学科　〒 275–0016 千葉県習志野市津田沼 2–17–1
E-mail: †{otsuka.shingo,takaku.masao}@nims.go.jp, ††kitsure@tkl.iis.u-tokyo.ac.jp,
†††miyazaki-n@mz.cs.it-chiba.ac.jp

あらまし　本稿では実世界と Web 空間における行動の関連性を調査するために，都内の女性 OL を対象としたフリーマガジンと連動するサイトのアクセスログの分析を行った．分析結果から，ユーザが Web ページを閲覧した曜日や時間と実世界におけるユーザの行動との間に関連性がある特徴的な行動を発見することができた．
キーワード　アクセスログ解析，Web 行動調査，Web 空間と実世界

A Study for Analysis of User Behavior in The Free Magazine Site for Women

Shingo OTSUKA†, Masao TAKAKU†, Masaru KITSUREGAWA††, and Nobuyoshi MIYAZAKI†††

† Scientific Information Office, National Institute for Materials Science
Sengen 1–2–1, Tsukuba-shi, Tsukiba, 305–0047 Japan
†† Institute of Industrial Science, The University of Tokyo
Komaba 4–6–1, Meguro-ku, Tokyo, 153–8505 Japan
††† Department of Computer Science, Faculty of Information and Computer Science, Chiba Institute of Technology
Tsudanuma 2–17–1, Narashino-shi, Chiba, 275–0016 Japan
E-mail: †{otsuka.shingo,takaku.masao}@nims.go.jp, ††kitsure@tkl.iis.u-tokyo.ac.jp,
†††miyazaki-n@mz.cs.it-chiba.ac.jp

Abstract　In order to investigate relationships of user's behavior between web and real world, we have analyzed web access logs of a web site of a free magazine for targeting working women in Tokyo. From the results, we have observed characteristic behaviors, such as the web pages access time of day and of day of week are related to users' actions or events in real world.
Key words　Web access logs analysis, Investigation of behavior on the web, Relationship between web and real world

1. はじめに

インターネットの爆発的な普及により，自分の興味がある情報の探索，購入予定の商品やレストラン，旅館の評判探索など，ユーザが望む情報を Web 空間から手軽に入手できるようになった．

Web 空間におけるユーザ行動を支援するための研究も盛んに行われており．例えば，ユーザの行動抽出，ある分野に関する評判の自動抽出，検索支援，商品の推薦などあらゆる研究が行われている．さらに，あるサイトにおけるユーザの行動分析を詳細に解析することで，Web ページのアクセス数向上や Web 広告について統合的にサポートするビジネスも誕生している．

ユーザの行動分析をするためにはユーザの特性を把握する必要があるが，既存の研究では，

- ある特定の内容に特化したサイトのアクセスログ
- 検索エンジンやポータルサイトなど不特定多数のユーザがアクセスを行うサイトのアクセスログ
- 各検索エンジンが提供している検索ランキングやスニペットなど検索結果に関連する情報
- 一般に公開されている大学の proxy サーバや NASA な

どのアクセスログ
などを利用しており，Web 空間におけるユーザの行動に限定したものである．

アクセスログを用いた研究は，例えば，Web ページやユーザのクラスタリング [2], [12]，ユーザの行動抽出 [3], [10]，検索支援 [1], [4], [9], [11]，閲覧行動支援 [5] など数多くの研究が行われている．これらの研究は主に Web 空間におけるユーザの行動支援を目的としているため，実世界の行動との関連に関しては議論されていない．実世界と Web 空間の両者に関連する研究としては，宇陀らの図書館サービスの利用者を対象とした利用状況の調査 [8] や，柴田らのライフログを用いた子供の事故防止対策に関する研究 [6] などが行われている．

また，上記の研究で用いられている主なログデータとしては

- 自サイトまたは共同研究している企業のアクセスログデータ
- The Internet Traffic Archive で公開されているアクセスログデータ [7]
- インターネット視聴率調査会社が提供するウェブサイト視聴データ (パネルログ)(注1)
- Overture が提供するクエリログ (キーワードアドバイスツール)(注2)

などがあるが，基本的には不特定多数のユーザを対象にしたデータのため，我々が用いた年齢層や性別が偏ったアクセスログとはその性質が異なる．

しかし，ユーザはある周期で睡眠や食事を取り，多くの人は仕事をしているため，Web 空間における行動は実世界と何らかの関連性を持っていると考えられる．本稿では 30 代前後の働く女性をターゲットとしたフリーマガジンと連動するサイトのアクセスログを解析することで，実世界と Web 空間における行動の関連性を見出すことを試みる．

2. フリーマガジンとアクセスログデータの詳細

まず，今回我々が解析を行ったサイト(注3)と連動しているフリーマガジン「Well」(図 1) についての説明を行う．この雑誌は偶数月の 20 日に発行されており，以下の方法で毎号約 30 万部が配布されている．

- 朝と夕方に，丸の内，銀座，六本木などで街頭配布 (発行日から一週間程度)
- 地下鉄の駅に設置されているラック (約 20 箇所)
- コンビニに設置されているラック (約 750 箇所)
- 企業配布 千代田・中央・港区約 4,000 オフィス（約 7 万人定期読者）
- 都内飲食店，スクール，美容院，など設置協力店 (約 500 店舗)

雑誌の内容は銀座・丸の内・青山・六本木など山手線の内側

図 1　フリーマガジン「Well」のトップページ

をメインターゲットにしたものになっており，雑誌で紹介されている店舗は以下の 8 つに分類されている．

（1） グルメ＆ドリンク
（2） ビューティー
（3） アミューズメント＆レジャー
（4） キャリア＆ライフ
（5） ショッピング＆ピックアップ
（6） ヘルス＆クリニック
（7） ヘア＆メイク
（8） リラクゼーション

この雑誌に掲載された店舗情報はすぐに連動するサイトに掲載される仕組みとなっている．このサイトでは今のところユーザ登録などの仕組みを導入していないため，アクセスしたユーザの年齢や男女比などを正確に掴むことはできないが，サイト内で使われた検索語やサイトに流入するために使われた検索語から 20-30 代の女性が大多数を占めていると考えられる．

今回解析を行ったアクセスログは 2004 年 2 月から 2008 年 12 月まで(注4)であり，データ量は約 8Gbyte であった．ここから，html,htm,php へのファイルアクセスのみを抽出し，さらに Web クローラーなど自動巡回ツールからのアクセスを取り除くクリーニング処理を行うことで，最終的には 250 万アクセスとなった．また，アクセス元の IP 情報の種類は約 55 万であった．1 人で複数の IP アドレスを使用していたり，アクセス毎に IP アドレスが変更されることがあるため，ユーザ数の詳細については分からないが，少なくとも数千から数万人のアクセスがあったと考えている．

今回解析を行ったアクセスログは Apache サーバのログのため，アクセスしたユーザの IP アドレス，アクセス時間，アクセスされたページ，リファラー，ブラウザなどの情報を含んでいる．また，サイト内には自前の検索機能があるため，アクセスされたページを解析することで，サイト内で用いた検索語を

(注1)：ネットレイティングスやビデオリサーチインタラクティブが有償で提供するデータ
(注2)：2007 年 5 月に正式サポートを終了
(注3)：http://my-well.net
(注4)：途中，2005 年 1 月-5 月と 2006 年 7 月-12 月はデータが欠落している．

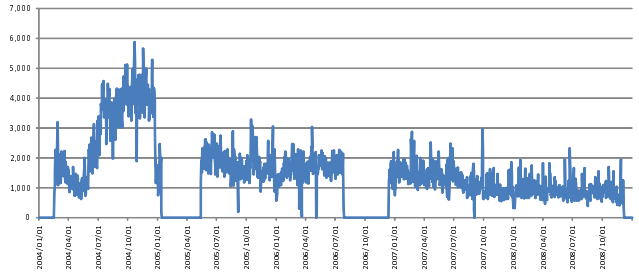

図 2　アクセス数の変遷 (その 1)

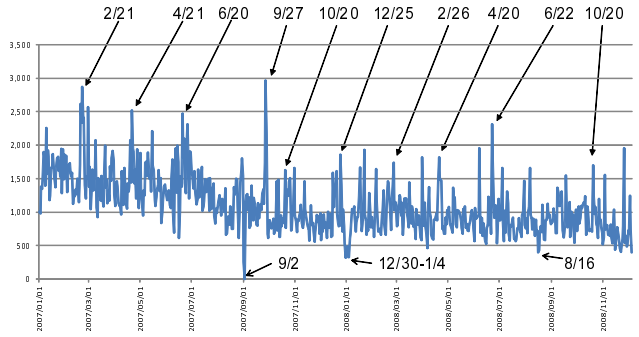

図 3　アクセス数の変遷 (その 2)

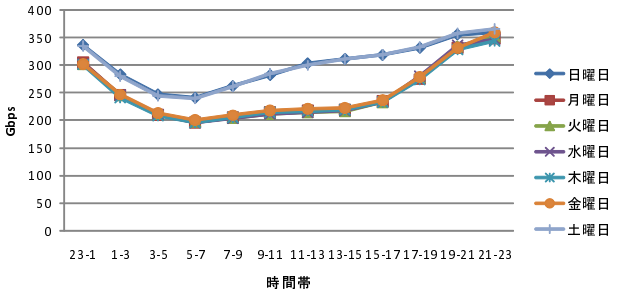

図 4　総務省が公表している Web アクセスの状況

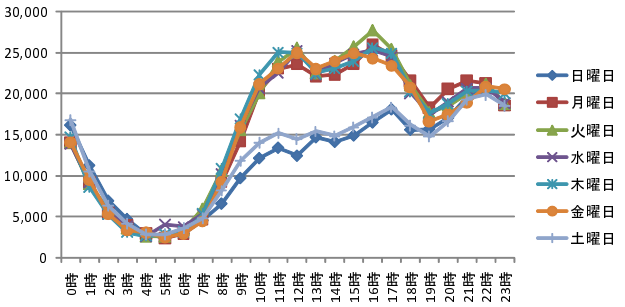

図 5　フリーマガジンと連動するサイトのアクセス状況

抽出することができる．さらに，各店舗は Well コードと呼んでいる ID を割当てており，サイト内の各店舗ページはこの番号を元に自動生成される仕組みになっている．Well コードからジャンルを特定することが可能なため，今回の解析ではジャンル毎にアクセス状況を調べることが可能となる．

なお，本サイトでは携帯電話に対応したページも存在するが PC 版のページと比較してアクセス数が少なく，また，PC 版のページは画像などが多いため，携帯電話からのアクセスには適していない．したがって，本稿では携帯電話からのアクセスについては議論しない．

3. 解 析 結 果

アクセスログから得られる情報を元に今回の解析では以下の調査を行った．

（1） 日ごとのアクセス数の変遷
（2） アクセスした曜日や時間帯とアクセス数の関連
（3） 検索エンジンやポータルサイトのシェア
（4） サイト内で使われて検索語
（5） サイトに流入するために使われた検索語

1,2 については Well コードを元にジャンル毎に集計した店舗ページのアクセスについても解析を行った．

3.1 日ごとのアクセス数の変遷

2004 年 2 月から 2008 年 12 月までの日ごとのアクセス数の変遷を図 2 に示す．ここ 2 年ほどの 1 日当たりのアクセス数は 1,000 件程度で安定している．次に期間を直近 2 年に限定したものを図 3 に示す．アクセスが乱高下していることがわかる．図中でアクセス数が極端に多い日を見てみると，概ねフリーマガジンの発行日である偶数月の 20 日付近になっており，最新号を入手した読者が同時に Web ページの閲覧を行っていることがわかる．12/25 のアクセス数増加については，クリスマスに関連した食事会などのイベントが原因であると推測される．9/27 のアクセス数増加についての原因は現在調査中である．また，アクセス数が極端に少ない日についてはお盆や正月となっている．9/2 については停電によるサーバ停止が原因である．

3.2 アクセスされた曜日や時間帯とアクセス数の関係

まず，解析結果を示す前に比較データとして，一般に公開されているデータとして総務省が提供する「我が国のインターネットにおけるトラヒックの集計・試算」(注5) を元に 2008 年 5 月の 1 週間のアクセス状況を図 4 に示す．このデータはブロードバンド (DSL,FTTH) 契約者から総務省が協力を要請した協力 ISP6 社 7 ネットワークへのトラフィック状況であり，単位は Gbps となっている(注6)．我々が提示するデータは全てアクセス数となっているため，単純に比較することはできないが，アクセスの傾向については比較すること可能である．トラフィックの傾向としては夜間や土日が多く，平日の日中は少ないことがわかる．

一方，我々のデータの解析結果を図 5 に示す．以後全てのグラフの縦軸はアクセス数である．アクセス数の傾向としては，平日の日中が多く土日や深夜は少ない．また，平日の昼や 19 時に急激に減少していることから，昼休みや仕事帰りの時間帯では Web ページをあまり見ていないことがわかる．

3.3 ジャンルによるアクセス傾向の比較

前節で述べたように，ユーザが閲覧した店舗ページの URL

(注5)：http://www.soumu.go.jp/s-news/2008/080829_9.html
(注6)：ただし，このデータはブロードバンドを対象としており，ネットゲームなどでのデータのやり取りも含まれていると考えられるため，Web ページのアクセス状況とは異なるが，ここでは誰でも入手可能なデータの例として比較対象とした．

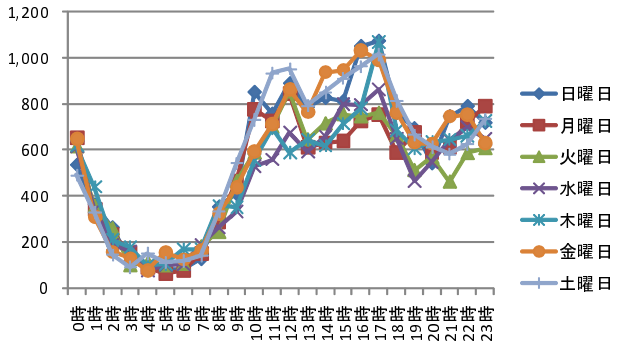

図 6　グルメ＆ドリンクのアクセス状況

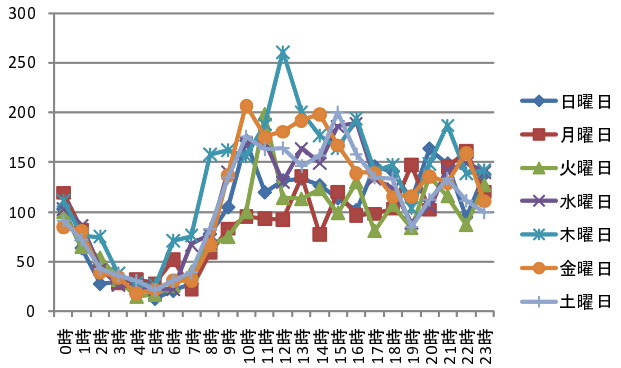

図 7　ヘルス＆クリニックのアクセス状況

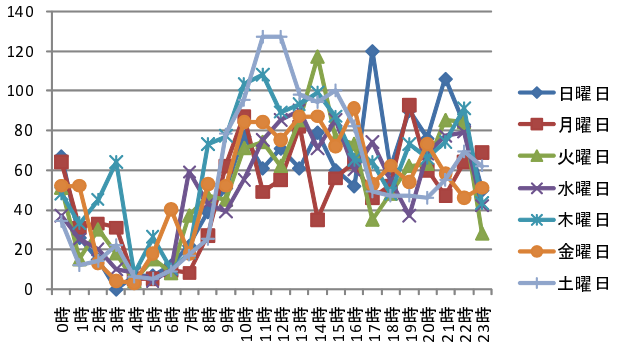

図 8　ヘア＆メイクのアクセス状況

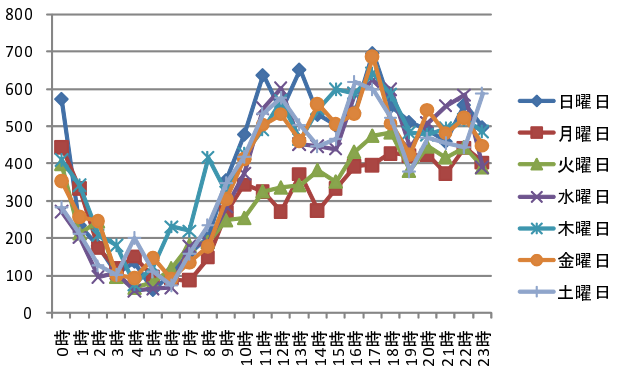

図 9　リラクゼーションのアクセス状況

からそのページが属するジャンルの識別が可能なため，各ジャンル毎にアクセスされた曜日と時間帯について解析を行った．ここでは，特徴的な傾向が見られた，グルメ＆ドリンク，ヘルス＆クリニック，ヘア＆メイク，リラクゼーションについての結果を示す．

3.3.1　グルメ＆ドリンク

ジャンル「グルメ＆ドリンク」のアクセス状況を図 6 に示す．アクセスの傾向としては，木曜日，金曜日，土曜日など週の後半のアクセスが多く，反対に月曜日，火曜日が少ない．また，週の後半の午前と 16 時頃のアクセスが多い傾向が見られるのは，夜に行われる飲み会の場所を探している行動と推測される．

3.3.2　ヘルス＆クリニック

ジャンル「ヘルス＆クリニック」のアクセス状況を図 7 に示す．アクセスの傾向としては，グルメ＆ドリンクに似ているが，午前や午後の早い時間のアクセスが多い．これは，病院やエステなどの閉店時間が比較的早いことが影響していると思われる．

3.3.3　ヘア＆メイク

ジャンル「ヘア＆メイク」のアクセス状況を図 8 に示す．アクセスの傾向としては，どの曜日のアクセス数も乱高下しているため傾向が掴み難いが，木曜日と土曜日の午前中にアクセスが多い．また，金曜日の日中のアクセスも比較的多いことから，週末に行われることが多い飲み会や結婚式のために美容院を探す行動と推測される．

3.3.4　リラクゼーション

ジャンル「リラクゼーション」のアクセス状況を図 8 に示す．アクセスの傾向としては，水曜日から日曜日までのアクセスが多く，帰宅前の 16-17 時のアクセスが多い．おそらく週の後半で溜まってきた疲れを仕事帰りや週末に癒すためにお店のチェックしている人が多いと推測される．

3.3.5　ジャンルの比較

各ジャンルのアクセス時間の比較を図 10,11 に，曜日の比較を図 12,13 に示す．図 11,13 についてはアクセス数が多い 2 つのジャンルを除いたグラフである．アクセスされたジャンルはグルメ＆ドリンク，リラクゼーションの順でアクセス数が多い．上記で述べたジャンル以外の特徴としては，

- ほとんどのジャンルでは月曜日，火曜日のアクセス数が他の曜日と比べて少ない．
- ショッピング＆ピックアップは深夜を除いてアクセス数の差が少ない．
- レジャー＆アミューズメントは夜のアクセスが多い．
- ビューティは午前中や午後の早い時間のアクセスが多い．

などの行動が挙げられる．

3.4　ポータルサイトのシェア

アクセスログのリファラー情報を解析した結果，サイト内のページ以外では検索エンジンなどのポータルサイトからのアクセスが上位を占めていた．そこで，サイトに流入したポータルサイトの推移について調査を行った．アクセスが多い上位 5 サイトの結果を図 14 に示す．解析では Yahoo!や Google，MSN，goo など予め対象となるポータルサイト(注7)を決めておき，こ

(注7)：実験では，Yahoo! Japan, MSN, Google, goo, Excite, Infoseek,

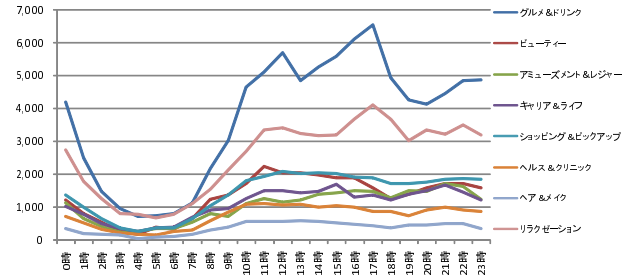

図 10　ジャンルごとの時間の比較 (全ジャンル)

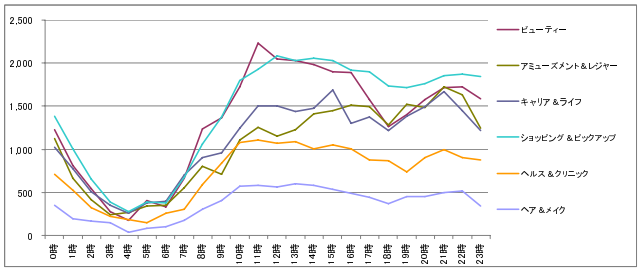

図 11　ジャンルごとの時間の比較 (グルメ&ドリンクとリラクゼーション以外)

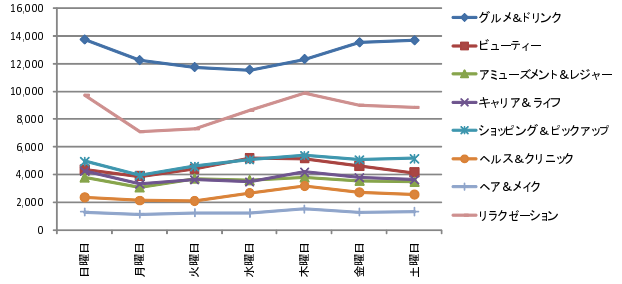

図 12　ジャンルごとの曜日の比較 (全ジャンル)

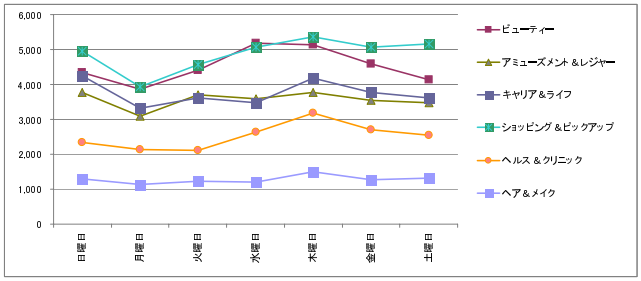

図 13　ジャンルごとの曜日比較 (グルメ&ドリンクとリラクゼーション以外)

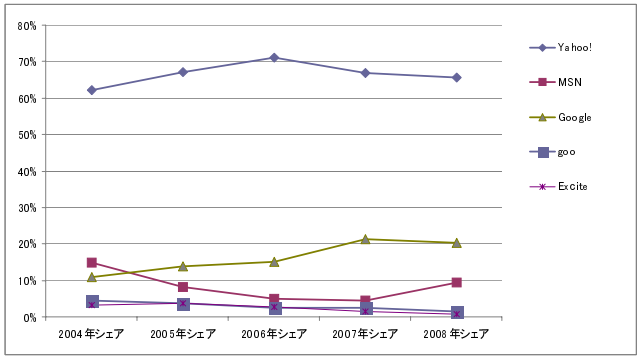

図 14　ポータルサイトのシェア

れらの合計値を 100% とした時の割合である(注8)．解析結果から 2006 年をピークに減少しているものの Yahoo!の利用率が圧倒的であり，また，MSN の健闘も目立つ．2008 年の MSN については全体の 7.6% が Live Search であり，このサイトのユーザ層にも浸透しつつあることがうかがえる．一般の調査では Yahoo!が 56%，Google が 21% という結果も公表されている(注9)が，Yahoo!と Google の差は年々少なくなっているものの，このサイトを閲覧しているユーザ層では Yahoo!がまだ根強い人気を持っていることがわかる．また，上位 5 件以外のサイトについては頻度が少なく，特徴的な傾向は見られなかった．

3.5　検索語の解析

最後にサイト内とサイトに流入するために用いられた検索語の一覧を図 15,16 に示す．図 15 からサイト内で入力された検索語の上位は場所に関する語であった．これはフリーマガジンそのものが地域に密着した情報を発信しているため，ユーザも場所に興味を持つ傾向があると推測される．

対照的に，図 16 からサイトに流入するために使われた検索語の上位を見ると,「温泉」「モテ髪」「よか石けん」「小顔矯正」などこのサイトが対象としているユーザ層が興味を持つキーワードが多い．また，1 位の「長寿の里」について調べたところ，九州で健康食品や化粧品の販売を行っている会社であり「つかってみんしゃいよか石けん」を販売している．推測ではあるが，フリーマガジン掲載時にはまだ知名度が低くこの検索エンジンでこのキーワードを入力すると，このサイトが上位にランキングされていたと推測される．

4.　おわりに

本稿では，30 代前後の働く女性をターゲットにフリーマガジンと連動するサイトのアクセスログの解析を行うことで，ある特定のユーザ層を対象とした，実世界と Web 上における行動の関連性の検証を行った．解析結果から，ユーザの実世界の行動と Web ページのアクセスが密に連動していることを示し，さらに特徴的な行動についても抽出することができた．今後は個々のユーザの行動解析や検索語の時間的推移の解析について行う予定である．

謝辞　本研究を進めるにあたり，貴重なデータを提供していただいた，株式会社 ライフウエル 白神志郎氏に心より感謝いたします．

Nifty, Biglobe, Fresheye, AOL, Livedoor, mixi を対象とした．

(注8)：MSN については http://jp.msn.com/と Live Search の合計としている．

(注9)：朝日新聞 2008 年 6 月 28 日付けの記事 (ニールセン・オンライン調査による)

文　献

[1] D. Beeferman and A. Berger. Agglomerative clustering of search engine query log. In *The 6th ACM SIGKDD International Conference on Knowledge Discovery and Data*

rank	検索語	入力ユーザ数	入力回数
1	新宿	172	230
2	銀座	128	193
3	池袋	112	146
4	渋谷	103	127
5	横浜	75	94
6	吉祥寺	74	100
7	恵比寿	70	90
8	上野	58	65
9	表参道	55	69
10	(整体の店舗名)	54	96
11	品川	46	62
12	六本木	46	56
13	新橋	38	45
14	立川	38	43
15	岩盤浴	36	41
16	ランチ	34	38
17	原宿	33	39
18	大宮	32	38
19	自由が丘	28	32
20	東京	27	38
21	浅草	27	32
22	有楽町	26	34
23	北千住	26	31
24	腸内洗浄	26	30
25	町田	25	28
26	大阪	25	28
27	日本橋	23	27
28	下北沢	23	27
29	高田馬場	22	25
30	目黒	21	25

図 15　サイト内で入力された検索語 (TOP30)

Mining (KDD2000), August 2000.

[2] T. Murata. Discovery of user communities from web audience measurement data. In *The 2004 IEEE/WIC/ACM International Conference on Web Intelligence (WI2004)*, September 2004.

[3] 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, 2003/12.

[4] 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, 2005/6.

[5] 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, 2006/3.

[6] 柴田康徳, 本村陽一, 西田 佳史 山中龍宏, 溝口博. 日常モノデータベースとライフログとの統合による危険の可視化. 第 21 回人工知能学会全国大会, 2007/6.

[7] The internet traffic archive　. http://ita.ee.lbl.gov/.

[8] 宇陀則彦, 伊藤宏美, 松村敦. アクセスログに見る電子図書館利用の傾向. 情報知識学会誌, Vol. 18, No. 2, pp. 161–168, 2008/9.

[9] J. Wen, J. Nie, and H. Zhang. Query clustering using user logs. *ACM Transactions on Information Systems (ACM TOIS)*, Vol. 20, No. 1, pp. 59–81, January 2002.

[10] 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, 2005/9.

[11] 安川美智子, 横尾英俊. クエリログから獲得した関連語のクラスタリングに基づく web 検索. 電子情報通信学会論文誌 D, Vol. J90-D, No. 2, pp. 269–280, 2007/2.

[12] H. Zeng, Z. Chen, and W. Ma. A unified framework for clustering heterogeneous web objects. In *The Third International Conference on Web Information Systems Engineering (WISE'02)*, December 2002.

rank	検索語	入力ユーザ数	入力回数
1	長寿の里	7,384	9,196
2	温泉	3,470	3,897
3	well	3,434	4,991
4	モテ髪	3,235	3,481
5	(脱毛サロン店舗名)	3,165	3,908
6	習い事	3,156	3,534
7	つかってみんしゃい	2,614	3,448
8	目の下のくま	2,521	2,805
9	グルメ	2,290	2,630
10	美容院	1,850	2,019
11	(美容･形成外科医院名)	1,849	2,521
12	長寿の里 つかってみんしゃい	1,675	1,749
13	よか石けん	1,413	1,828
14	上海康茶	1,392	1,850
15	Well	1,365	2,034
16	WELL	1,365	1,897
17	極選上海康茶	1,360	1,885
18	(美容･形成外科医院名)	1,324	1,795
19	ヘアーアレンジ まとめ髪 おだんごヘアー	1,009	1,019
20	つかってみんしゃいよか石けん	997	1,123
21	(脱毛サロン店舗名)	956	1,192
22	目の下のクマ	928	1,056
23	search	919	3,530
24	小顔矯正	796	971
25	下半身痩せ	780	919
26	ファイテン	753	881
27	イベント	722	799
28	通販	672	719
29	浴衣　髪	665	674
30	ひげガール	658	797

図 16　サイトに流入するために入力された検索語 (TOP30)