

湘北短期大学ウェブサイトのアクセスログから見られる 閲覧者の各学科への興味の違い

小棹 理子^a 石田 英弥^b

^a湘北短期大学 ^b早稲田大学大学院博士前期課程

【抄録】

湘北短期大学のWebサイトで2007年7月の1か月間に得られたアクセスログを解析し、訪問者のサイト内ページ遷移をネットワーク図で可視化した。これにより、訪問者が学科によって異なる閲覧動向を示すことが明らかになった。

【キーワード】

ウェブサイト アクセスログ PR マイニング ネットワーク図

1. はじめに

少子高齢化による18歳人口減少の背景の中で各大学では差別化を図り、それぞれの持つ特色を広く一般に伝える媒体としてウェブサイト積極的に利用している^[1]。ウェブサイトは、いわゆるパンフレットや新聞広告、電車内広告などの紙媒体とは異なる。第一に、リアルタイム性がある。ここで言う「リアルタイム性」とは、1日もあれば公開されている情報を修正・更新できる、という意味である。第二に、閲覧者から情報を入手できる。ウェブ掲示板や、ブログ形式でコメントを求めることもできるが、より広報戦略に有益なのはアクセスログの活用であろう。アクセスログとは、ウェブサイトへアクセスしてきた閲覧者の記録である。

アクセスログから得られる情報としては、①ページビュー(PV)(アクセスされたページ数の総計)、②ユニーク訪問者数(セッションIDの数)、③平均訪問ページ数(①/②)、④アクセス元(閲覧者がどのネットワークからアクセスしてきたか)、⑤ページ遷移の状況(サイト内でユーザーがどのページをどのような経路でアクセスしたか)、などがある。たとえば、湘北短期大学のホームページ(トップページ)(<http://www.shohoku.ac.jp/index.html>)にアクセスすると、セッションIDと呼ばれる単一IDが付与され、閲覧ページ、時間、アクセス元などの情報が残る。このIDは、他サイトへ移動するかブラウザを閉じるまで、つまり一湘北短期大学サイト内を閲覧している間—は残るので、同一セッションIDを持つデータから、閲覧者の興味や動向を読み解くことができる。

一日に何千件、何万件とアクセスのあるサイトでは、アクセスログデータの分析の巧拙によって得られる情報も異なってくる。膨大なデータから

<連絡先>

小棹 理子 ozao@shohoku.ac.jp

有用な情報を抜き出すプロセスが必要である。このようなプロセスをデータマイニング^[2,3]という。データマイニングの目的によって用いる手法は異なるが、たとえば意思決定をする際に参考となる知識を得るために傾向や特徴を見出す場合には、分類をしたり、特性ルールを見つけたりする手法が用いられる。たとえばアソシエーション分析(バスケット分析)の適用例では、伝説となっている話、「缶ビール」と「紙おむつ」がある。“スーパーマーケットの膨大なPOSデータを分析したところ、金曜日には「缶ビール」と「紙おむつ」が同時に売れていることが判明した。つまり、お父さんが子供のために買い物に来るついでに自分のビールを買う、という行動パターンがイメージされたわけだ。このことに気づいたスーパーは、「紙おむつ」売り場の近くに「缶ビール」を置き、売上げをさらに伸ばしたそうである。”というものである。後半は「尾ひれ」だそうであるが、「缶ビール」と「紙おむつ」の意外な並売傾向が発見されたことは大きな話題となった。

筆者らはアソシエーション分析をエコマテリアルデータベースのアクセスログ解析に適用し、エコマテリアルの指標の選定や、エコマテリアルの一般関心度調査において有意義な知見を得ることができた^[4]。

ここでは湘北短期大学総務部より同学のウェブサイトへのアクセスログ(2007年7月分、各学科)の提供を受け、ウェブサイト閲覧者のサイト内遷移を可視化し、ウェブサイト訪問者の動きを推察してみた。その結果を報告する。

2. 方法

元データとして、湘北短期大学総務部がGoogle Analyticsによって取得し蓄積したデータをいただいた。データは2007年7月1日から7月31日の1

か月間のものである。

Google Analyticsでは解析対象とするページにJavaScriptによるコードを埋め込み、それを通じてグーグルのサーバーにアクセス情報を蓄る。Google Analyticsのユーザーにはアクセスした生データは提示されないが、各ページ毎のアクセス数、解析対象ページを訪問する直前のページとそのランキング、または解析対象ページの次の訪問先とそのランキングなどの形でアクセス情報が提示される。

今回は、「上記解析対象ページを訪問する直前のページとそのランキング」、さらに「解析対象ページの次の訪問先とそのランキング」の中から、情報メディア学科、総合ビジネス学科、生活プロデュース学科、保育学科の各トップページについて、「ナビゲーションサマリー」に表示される「前のURL」と「次のURL」を取得して解析に用いた。なお、いずれの学科に関しても上位10ページに限定している。ネットワークグラフの作成には「Visual Mining Studio」(数理システム(株))を用いた。

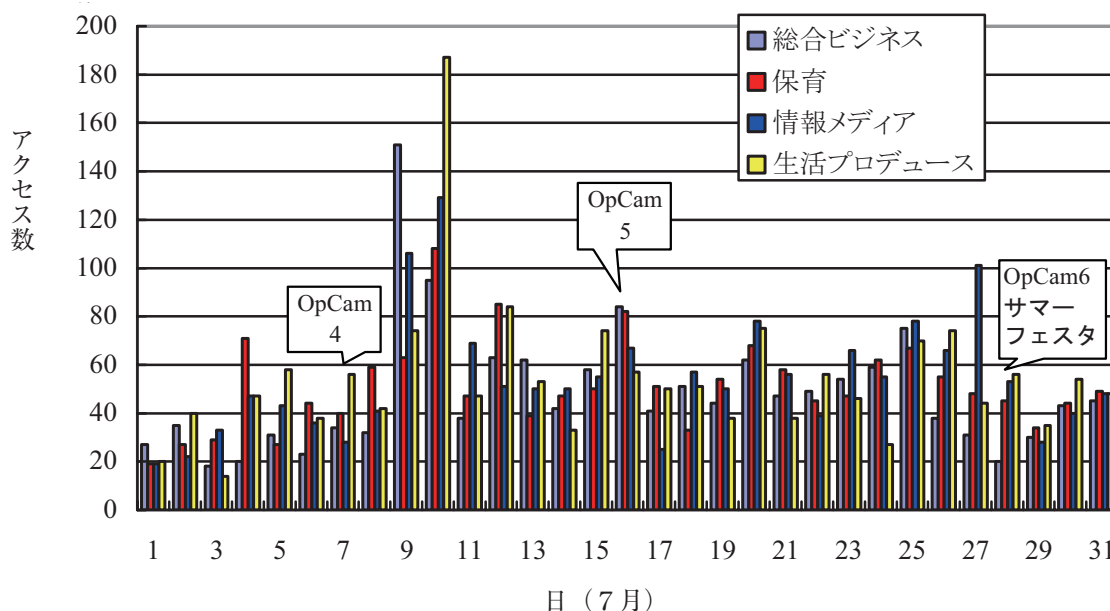
3. 結果と考察

表1に全PV(ページビュー)、図1に全学科のトップページへのアクセス状況を示す。7月9-10日に突出しているところがあり、とくに総合ビジネス学科と生活プロデュース学科に顕著である。その影響か、全学科の閲覧数が増えている。27日には情報メディア学科にややアクセスが集中しているが、理由は不明である。いずれにしても、1ヶ月間で見ると、偏りなく全学科へアクセスされていることがわかる。

表1 各学科の7月の全アクセス数

学科	2007年7月1～31日 全ページビュー (PV)
情報メディア (/guidance/infomedia/index.html)	1686
総合ビジネス (/guidance/business/index.html)	1502
生活プロデュース (/guidance/lifepro/index.html)	1686
保育 (/guidance/hoiku/index.html)	1597

図1 2007年7月の各学科ホームページページビュー



今回は、「上記解析対象ページを訪問する直前のページとそのランキング」、さらに「解析対象ページの次の訪問先とそのランキング」の中から、情報メディア学科、総合ビジネス学科、生活プロデュース学科、保育学科の各トップページについて、「ナビゲーションサマリー」に表示される「前のURL」と「次のURL」を取得して解析に用いた。なお、いずれの学科に関しても上位10ページに限定している。

表2に取得データ例（情報メディア学科）を示す。ここでクリック率とは、/guidance/infomedia

/index.htmlを訪問する前（後）の訪問ページがそれぞれどれだけの割合を占めるかをあらわす。つまり、表2の例では、/guidance/infomedia/index.htmlを訪問する前に、約1/3が/index.html（湘北短大のトップページ）を見ていたことをあらわす。また、/guidance/infomedia/index.htmlの訪問者の18%が次に/index.htmlを見に行ったことがわかる。さらに、このように得られたデータを適宜形式を修正したのち、「Visual Mining Studio」（数理システム（株））を用いて、ネットワークグラフとして可視化した。

表2 取得データ例 (情報メディア学科の例)

前の URL	クリック率	次の URL	クリック率
コンテンツ		コンテンツ	
/index.html	0.345196	/index.html	0.179122
/guidance/index.html	0.166667	/guidance/index.html	0.106168
/guidance/infomedia/index.html	0.090747	/guidance/infomedia/index.html	0.090747
/person/index.html	0.038553	/guidance/business/index.html	0.079478
/guidance/business/index.html	0.036773	/person/index.html	0.070581
/exam/index.html	0.027284	/guidance/teacher.html	0.041518
/guidance/lifepro/index.html	0.02669	/guidance/kamoku.html	0.037367
/guidance/kamoku.html	0.024318	/guidance/lifepro/index.html	0.031435
/guidance/teacher.html	0.020759	/exam/index.html	0.027284
/examinfo/complement.html	0.016014	/examinfo/complement.html	0.022539

図2にネットワークグラフを示す。矢印が遷移の方向を、線の太さがクリック率を表す。線が太いほど、クリック率が高い。たとえば、図の左下、トップページ (/index.html) と各学科のトップページへは太い両矢印 (⇔) で結ばれており、湘北のトップページから各学科トップページの往来が頻繁であることがわかる。保育学科では、学科トップページを訪問したのち、オープンキャンパス情報をアクセスする例が多く、また、「こんな職業を目指そう」を確認して離脱することが推察される。7月当時、総合ビジネス学科では、「こんな職業を目指そう」が人気で、総合ビジネス学科訪問者の2割が同ページを参照し、同学科内にとどまったことを表している。また、その後、実際の就職進路情報を参照する訪問者が多いのも、総合ビジネス学科閲覧者の特徴である。

図2において、特筆すべきは「募集要項・入試制度」(図2左なかほど)と「受験生の方へ」(図2右なかほど)のページ閲覧者の動向である。情報メディア学科あるいは保育学科と同ページ間は両方向の矢印であるが、生活プロデュース学科と総合ビジネス学科へは一方方向の矢印である。つまり、情報メディア学科あるいは保育学科のトップ

ページを閲覧したのち「募集要項・入試制度」を閲覧した場合、再度両学科へ戻る場合もあるが、「募集要項・入試制度」を見てから、生活プロデュース学科と総合ビジネス学科を訪問した場合は、同ページへ戻ってくる確率が低い。「受験生の方へ」の場合も同様である。同ページから総合ビジネス学科向きの矢印が出ていることから、同ページを見て総合ビジネス学科を確認して戻ってこないことが示唆される。以上から、あくまでも各学科トップページを中心としたページ遷移上位10位の結果、それも1ヶ月(7月)のみのデータからではあるが、次のような湘北のウェブサイト訪問者の動向を推察できる。

- ① 保育学科に興味のある訪問者は、他学科にはあまり興味がなく、直接オープンキャンパス情報にアクセスする。
- ② 総合ビジネス学科のトップページ閲覧者は、「こんな職業をめざそう」ページでとどまり、他学科へ流れない。
- ③ 情報メディア、保育学科希望者は、「受験生の方へ」を閲覧する機会が多い。「受験生の方へ」を閲覧して総合ビジネス学科のトップへ遷移した場合は同ページを再閲覧する確率

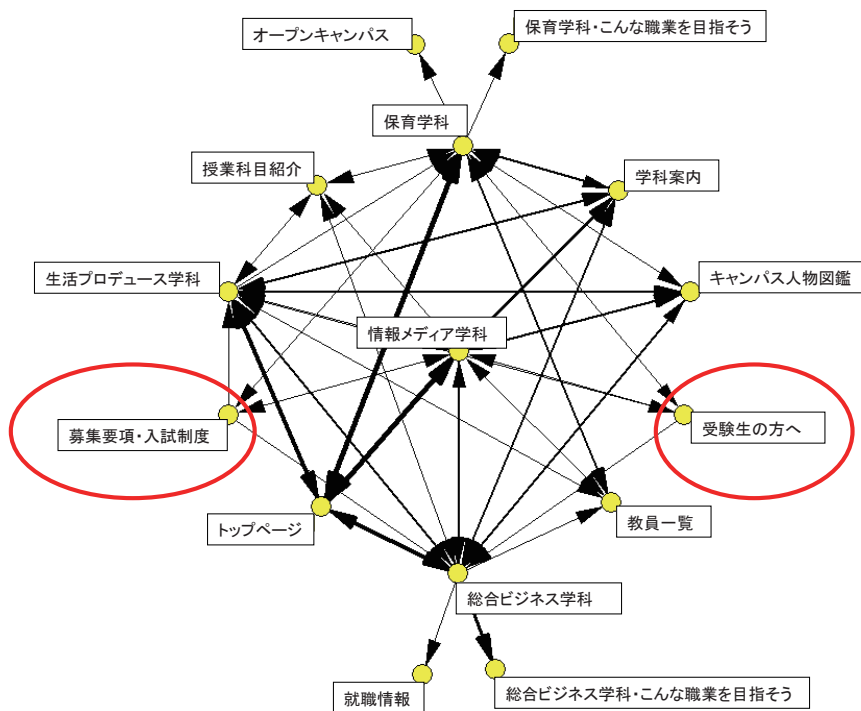
湘北短期大学ウェブサイトのアクセスログから見られる閲覧者の各学科への興味の違い

は低い。

- ④ 情報メディア, 保育学科希望者は「募集要項・入試情報」を閲覧するが多い。「募集要項・

入試情報」を閲覧して総合ビジネス学科と生活プロデュース学科のトップへ遷移した場合は、同ページを再閲覧する確率は低い。

図2 VMSを用いて作成したネットワーク図



4. おわりに

湘北短期大学ウェブサイトのアクセスログの分析を試みた。2007年7月1か月間の限られた期間ではあったが、訪問者が学科によって異なる閲覧動向を示すことが明らかになった。今後、より長い期間でログ解析を行うことで、ウェブサイトの構成や内容など、本学の広報活動に益する知見が得られるのではないかと期待できる。また、今後、アンケートで蓄積された自由記述回答のマイニングを行うことによって、さらに広報にとって有益なナレッジ（知的資産）が得られる可能性がある。

謝辞

本報告は平成18年度湘北短期大学助成金を得て作成されたものです。ここに謝意を表します。また、ご多忙中にもかかわらず、アクセスログデータをご提供いただいた飯塚総務部長に感謝申し上げます。

参考文献

- 1 倉部史記, 「ウェブサイトただ今進化中」, Between 2006 秋, pp54-56 (2006).
- 2 福田剛志, 森本康彦, 森下真一, 徳山豪, 「データマイニングの最新動向—巨大データからの知識発見技術—」, 情報処理 37 (7), pp597-603 (1996)
- 3 Peter Cabena, Pablo Hадjnian, Rolf Stadler, Jaap Verhees, Alessandro Zanasi; “Discovering Data Mining: From Concept to Implementation (IBM Books) ”, Prentice Hall (1997)
- 4 R. Ozao, T. Sawaguchi, H. Ishida, M. Iji, T. Furuyama, Y. Shinohara, K. Yamada, K. Halada, “Eco-MCPS: A newly developed web-based database for eco-materials in Japan”, Mater. Trans, 48 (12) pp. 3043-3049 (2007).

湘北短期大学ウェブサイトのアクセスログから見られる閲覧者の各学科への興味の違い

Access Log Analysis for Public Relations

OZAO Riko ISHIDA Hideya

[abstract]

Access log obtained on the Web site of Shohoku College was analysed, and the page transition of the site visitors was visualized in the form of a network diagram. Differences among the interesting areas for four departments were clearly observed.

[key words]

Web site, access log, data mining, network diagram, public relations

