

WWW アクセスログの有効な解析法について*

4 T-5

畑島 隆† 元田 敏浩‡

日本電信電話株式会社 ソフトウェア研究所§

takashi@canary.sl.cae.ntt.jp

motoda@canary.sl.cae.ntt.jp

1 はじめに

WWWのアクセスログ解析では、一定時間あたりのアクセス数を評価指標としている。これはコンテンツへのアクセス率を指標としているTVの視聴率調査と同様の手法をとっているものと考えられる。しかし、両者にはコンテンツの提供形態やコンテンツの選択のされ方に相違が見られるため、WWWというメディアにあった調査手法を確立する必要がある。本研究では、TVとWWWの相違について検討し、WWWのアクセスログからユーザの動向をより効果的に解析するための手法への要求条件を示す。

2 WWWの商用利用と動向調査

インターネットの普及が進み、主にWWW(World Wide Web)を利用した形での商用利用が企業を中心として始まっている。利用形態は、企業の会社案内や製品紹介といった自社の広告媒体としてだけでなく、オンラインショッピングに代表される「電子モール」や、Webのページを広告スペースとして賃貸するなど、商用利用が進んでいる[1]。商業活動として企業がWWWサイトを設置することになると、その効果の予測や確認といったマーケティング調査の要求が発生する。WWW上でのマーケティング調査は、ユーザへのアンケートやWWWサーバに残されたアクセスログを解析することにより行なわれている。現在提供されているアクセスログ解析サービスでは、WWWサーバのログから時間あたりのページ単位のアクセス数の算出が基本であり、付加価値としてどのような地域のどのような属性のユーザのアクセスが多いかを人口統計学的に解析されているのみである。この手法は

TVの視聴率調査として行なわれているサービスをそのままWWWサーバの動向調査に持ち込んだものであると思われる。しかし、WWWとTVではメディアの性質が異なるため、従来の手法では、調査結果がWWWアクセス者の動向を表しているとはいえない。本研究では、WWWのマーケティング調査手法を確立するために必要と思われる事項を検討する。まず、WWWのログから何が分かり、現状ではどのように利用されているかを述べる。それを踏まえてWWWとTVを利用動向調査の観点から見た時の、メディアの相違について意見を述べ、WWWに適した動向調査の手法を提案をする。

3 アクセスログ解析調査

WWWサーバに残されるアクセスログには自分のWWWサーバに対してどこから、いつ、どのファイルへのアクセスをされたかを示したものが記録されている。このため、ログファイルからWWWサーバの使用状況を知ることが出来ると同時に、アクセスしてきた世界中のユーザがどのようなことに興味事項があるのかを知ることが出来るのである[2]。以下に実際に提供されているWWWのアクセス数調査サービスを例示し、問題点を考察する。Internet Profiles Corporationが提供しているI/PROサービスのうち、I/COUNTサービスは、WWWサイトのアクセスログを夜間にI/PRO社に転送して統計処理を行ない、1日あたりのアクセス頻度をドメイン別等に集計したものを提供している。

アクセス数を調査する場合には、同一ユーザからの多重アクセスの取り扱いが問題となるが、上述のサービスでは、同一ファイルへのリクエストのうち、一定時間以上のアイドルタイムの後で出されたものを‘Visits’としてカウントしている[3]。同様にNetCount[4]、WEB Counter[5]などのサービスが行なわれているが、いずれのサービスも1日、1週間といった一定期

* A Study of effective analysing method for the Access Logs of the WWW Server

† Hatashima Takashi, NTT Software Laboratories

‡ Motoda Toshihiro, NTT Software Laboratories

§ 3-9-11 Midori-cho, Musashino, Tokyo 180, Japan

間のアクセス数を指標としている。

4 WWWのための調査指標の提案

以下に従来の調査手法の問題点を述べ、WWWのメディアの性質をTVと比較により説明することにより提案の根拠を述べる。表1にTVの視聴率調査と従来のWWWの利用動向調査との比較を示す。

表1：TVとWWWの調査手法の比較

	TV	WWW
コンテンツ	番組(時間単位)	ページ(サーバ空間単位)
区分	時間単位	サーバ空間単位
調査指標	時間単位の視聴率	時間単位のアクセス数
調査手法	アクセスのログ解析	アクセスのログ解析
ログがもつ情報	選択先と視聴期間	アクセス先と時刻

4.1 リアルタイムな動向調査が可能なこと

従来の調査指標では、いずれもアクセスログに含まれている時系列の頻度変化が平均化されている。つまり、一定期間のアクセス頻度を1日当たりのヒット数という形で平均化して算出しているが、これはTVの視聴率調査と同様の方式であると考えられる。WWWとTVそれぞれの調査手法について比較すると、TVの調査ではサンプル世帯に設置した装置に1日分のTVチャンネルのアクセスをロギングしておき、これを集計して1分単位や番組単位の視聴率として提供されている。TVでは1つの番組が時間という単位で区切られているため、その番組の放映中のアクセス数の平均が視聴率指標という指標として意味を持つ。しかしWWWの場合は、コンテンツの提供時間の区切りは一般的に存在せず、それぞれのコンテンツに付与されたURL(Uniform Resource Locator)によってサーバ空間上に配置されているので、ユーザはURLを指定すれば何時でもコンテンツへのアクセスが可能である。つまり、コンテンツがWWWサーバ上に存在する限り、時間的に継続してユーザのアクセス動向の観測が可能であることから、調査指標もリアルタイムな動向の調査が可能であるべきである。

4.2 コンテンツ単位の調査が可能なこと

また、TVではどれだけの時間そのチャンネルを選択していたかを計測可能なのに対して、WWWの場合にはサーバにアクセスした時刻はログから知ることが出来るものの、いつまで見ていたかを知ることは困難である。よって調査の主体は時間によって区分する

のではなくURLで指定されたコンテンツを単位とすべきである。

4.3 アクセスの時系列的な傾向を考慮していること

WWWにおいてユーザによるアクセスが持つ意味について、TVの場合の「ザッピング」と同じようにWWWでも様々な場所へ短時間でアクセスしてまわるネットサーフィンを例として考察する。TVの場合は、選択可能なチャンネル数は高々10程度であるが、WWWでは世界中の数千万件にもおよぶコンテンツにアクセス可能である。コンテンツにアクセスするためには、膨大なコンテンツの中から特定のURLを指定するのであるから、ユーザのアクセスという行動はTVのチャンネル選択と比べてより「能動的なアクセス」であると考えられる。言い替えれば、ユーザのアクセスという行動はそのコンテンツへの関心の現れであるから、関心の集まり具合を評価するためには一定時間内のアクセス回数よりも、個々のアクセス傾向の時系列情報を考慮すべきである。以上の議論したように、WWWのための調査指標は以下の条件を満たす必要がある。

- リアルタイムな動向が調査できること
- 時間単位ではなく、コンテンツ単位の調査が可能であること
- アクセスの時系列的な傾向を考慮していること

5 まとめ

本論文では、WWWのアクセスログを利用したWWWの利用動向調査に要求される条件について提案した。現在これらの要求条件を満足する調査指標の関数形式について具体化を進めている。

参考文献

- [1] Electronic Commerce Network,
<http://www.commerce.or.jp/>
- [2] 堀田, 三宅, "HTML リファレンス", pp31-33,
OPEN DESIGN ,CQ 出版, Apr., 1996
- [3] Internet Profiles Corporation,
<http://www.ipro.com/>
- [4] Net Count, <http://www.netcount.com/>
- [5] Computer Networking Service, Inc. ,
<http://www.digits.com/>