

アクセス回数を利用した WWWの人気ホームページ道案内システム

杉井俊彦 北英彦 林照峯

三重大学工学部

インターネットの普及に伴い、インターネットを利用するユーザが急増している。インターネットを利用すると、ユーザは自分の気に入った情報を自由に得ることができ、さらに手作りの情報を多くの人に公開することができる。特にWWW (World Wide Web) においては、ユーザはワープロの知識さえあれば簡単に情報を公開できるため、WWWのホームページで与えられる情報はインターネット上にあふれている。その結果、自分の関心がある情報を効率的に手に入れることが困難になっている。

本研究では、他のユーザが見たホームページのアクセス回数を利用して、人気のあるホームページの一覧の提供することによって、自分の興味あるホームページを今までよりも少ない労力で見つけることができるシステムを提案する。

The system of guiding popular homepages by using access counts on World Wide Web

Toshihiko Sugii Hidehiko Kita Terumine Hayashi

Faculty of Engineering, Mie University

On the Internet, users can easily obtain their favorite information and they can publish their information to many other people. Particularly as WWW(World Wide Web), users can easily publish their hand-made homepages, so a huge amount of homepages are published and users should spend much time to find their favorite homepage among them.

Our system provide a list of popular homepages ranking by access counts through a proxy WWW server in order to find favorite homepages easily.

1. はじめに

1960年代末に米国で誕生したインターネットは、その後全世界に広がっている。日本においても、パーソナルコンピュータやモデムの低価格化、パソコン通信サービスとインターネットの相互接続、インターネット接続サービスの一般化などの種々の要因によって、そのユーザ数を急速に伸ばしている。このことはユーザが自由に情報をインターネットに公開するための仕組みとして作られたWWW (World Wide Web) のホームページの数からも留意に推測することができる。

WWWでは普通のユーザでも、ワープロの知識さえあれば簡単に文字や画像といった情報を提供できる。そして、新しいサーバが次々と立ち上がり、無秩序に、かつ、急激にホームページの数が増加している。さらに、WWWのサーバが互いにリンクでつながることにより、総体として巨大な情報空間を形成している。そして、この情報空間は時々刻々と変動しており、その全貌を完全に把握することは不可能である。そのために、現在無数に存在するホームページの中から、自分の興味のある情報が載っているホームページを探し出すことが困難となってきている。

本研究では、この問題を解決するための1つの方法を提案する。それは、自分自身でホームページを見つけるのではなく、他の人がよく行くホームページつまり人気のあるホームページを参考にし、その中から自分の興味のあるホームページを見つけだそうとするものである。それによって、本来無数にあるホームページから自分の興味のあるホームページを見つけだす労力を少なくすることができる。

2. 情報フィルタリング

いろいろな種類の情報が大量に存在するとき、その中から自分にとって有用であると考えられる情報を選別することは非常に困難なことである。おそらく、自分にとって有用であると考えられる情報を入手するための多くの時間がこのような選別のためだけに費やされ、情報を手に入れ、読んで理解するために費やすことのできる時間は少な

くなってしまう。また、苦勞して見つけた情報であるにも関わらず、実際に読んでみるとつまらない内容でしかなかった、ということはよく経験することである。つまり、情報量が多くなればなるほど、情報を選別するためにかかる時間の重要性は多くなってくる。

この大量にある情報の中から、自分にとって有用であると考えられる情報を選別する情報フィルタリング[1][2]は、以下の4つに分けることができる。

1. *Cognitive filtering*

メッセージの内容と、ユーザの情報に対するニーズを比べて、ニーズにあったメッセージを選び出す

2. *Social filtering*

メッセージの内容ではなく、そのメッセージの送り主と受取人との関係に基づいて、ユーザにあったメッセージを選び出す

3. *Economic filtering*

情報を得ることによる利益と、情報を得るために必要なメッセージの長さやその他の心理的な要因を含む対価の比に基づいて選び出す

4. *Collaborative filtering*

他のユーザの主観的な評価を元にして、メッセージを選び出す

本研究では、WWWのホームページのフィルタリングを行う際、ユーザ個人と他のユーザの持つ情報の関連性に基づいて行っているため、特に *Collaborative filtering* と考えることができる。そして、この *Collaborative filtering* を用いることによって、無数のホームページの中から、自分の興味あるホームページを見つけだそうとしている。

3. 検索ホームページとの比較

無数のホームページの中から自分に興味のあるホームページを見つけだす方法として、検索ホームページを利用する方法がある。有名な検索ホームページとしてyahoo[3]や千里眼[4]などがある。これらの検索ホームページでは、適切なキーワードを入力することによってあらかじめ用意されたデータベースにより目的のホームページを見つけ

ることができる。また、最近では検索をかける際にジャンルを指定したり論理演算を利用するなどにより、的を絞って検索効率を上げるようになっているものもある。

ところで、WWWのユーザの中には特定の事柄を調べるためにWWWを使用しているわけではなく、何か面白いホームページがないかと気が向くままにリンクをたどっている人がいる。そういう人たちにとって、検索ホームページを利用する方法が必ずしも向いているとは考えにくい。

4. 人気ホームページ道案内システム「やじうまくん」の概要

一般にユーザがホームページを見る際に、特定の事柄を調べているときを除くと、ほとんどが自分の興味に合うホームページがないかと探していると思われる。このことはホームページを見たことのある誰しもが経験することである。そして、その興味のあるホームページを検索ホームページを使ったり、いくつかのリンクをたどるなどによって見つけても良いが、この無数に存在しているホームページの中から自分に興味のあるホームページを見つけるとなると、かなりの労力が必要となる。このことは精神的に疲れる上、時間の無駄にもつながる。

誰かに面白い、または、役に立つホームページの情報を直接教えてもらったとすると、その労力は必要なくなる。しかし、教えてもらわないとその情報が得られないし、教える側もメールなどを

利用して教えなければならない。

本研究では、他の人から間接的に役に立つホームページに関する情報を得るための方法を提案する。大勢の人が行くホームページや同じ人が何度も行くホームページには何らかの価値があると考えられ、そのホームページに関する情報を知ることができる自分もそのホームページを見に行くことができる。そこで、他のユーザがホームページを見たアクセス回数に注目し、それを人気と定義している。つまりアクセス回数の多いホームページは、人気があり何らかの価値があるホームページと解釈して、自分の興味のあるホームページを探し出す道しるべとしている。

5. 人気ホームページ道案内システム「やじうまくん」の実装

ローカルなネットワークのセキュリティを確保するために、ファイアウォールを設けて外部との接続を制限することがある。ファイアウォールを乗り越えてWWWを使うための仕組みとしてproxyサーバ（代理サーバ）が存在する。ローカルなネットワークからの外部のWWWサーバへのアクセスはproxyサーバを中継することによって行われる（図1）。

このproxyサーバには、ユーザがいつ、どこのホームページを見に行ったのか、などといった情報をログとしてファイルに残すことができる。このアクセスの記録を残すという機能はファイアウォールの利用の有無に関わらず使用できる。

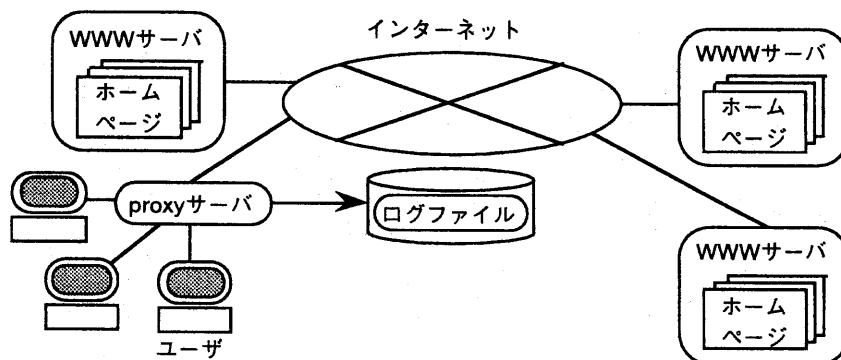


図1 従来のシステム構成図

本システムでは、proxyサーバで取られたログファイルを解析し、アクセス回数の多い順にホームページの所在を示すURLを並べ替え、その結果をホームページの1つとして提供している。本システムの構成図を図2に示す。

一般に1枚のホームページは、全体の構成を示すテキストファイルとその中に取り込まれる図、写真、音声などのファイルから構成されている。WWWでは各々のファイルが別々にサーバからブラウザへ転送され、ブラウザ上で再構成される。そのため、proxyサーバ上では各々のファイルへのアクセスが別々のエントリーとして記録される。本システムでは、proxyサーバでとられたログファイルの内、ホームページと対応するHTMLファイルのログのみを残して利用している。

また、データとして使われるログファイルは、人気という点を考慮して最近過去3カ月間のログを使用している。さらに、ログというものは毎日新しくできるものであるため、本システムでもデータの更新を自動的に毎日行っている。

「やじうまくん」のホームページを図3、図4に示す。図3は、「やじうまくん」でのインデックスになるホームページであり、順位別に分けられたホームページの一覧を表示している。そして、ログファイル及び、データの更新日時をユーザに分かりやすいように上の方に表示している。図4では、図3において”top_50”というのをクリックした時に表示されるホームページである。ここでは、各ホームページの順位、アクセス回数、一

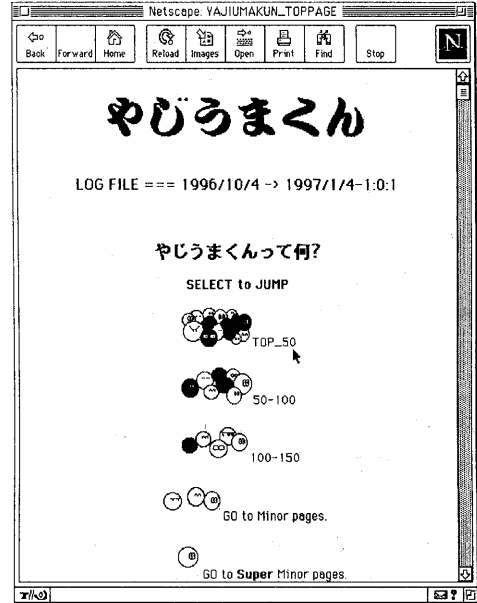


図3 やじうまくんのホームページ1

番最後にアクセスされた最終アクセス日、およびそのホームページのURLを表示している。ホームページのURLでは、すべてにそれぞれのホームページへリンクを張ってあるため、気に入ったホームページが見つければそこをクリックすることによって、すぐにそのホームページに飛ぶことができるようになっている。さらに、人気という点を考慮すると、単にアクセス回数の多いホームページが人気があるとはいいにくい。つまり、いくらアクセス回数が多いホームページであっても、最近そのホームページにアクセスがなければ人気のある

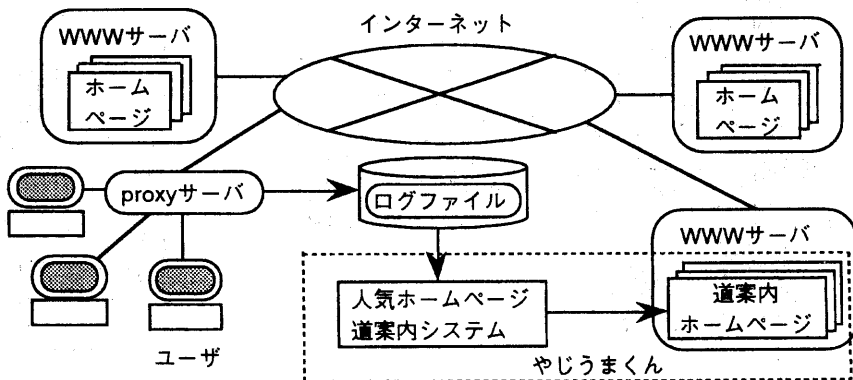


図2 本システムの構成図

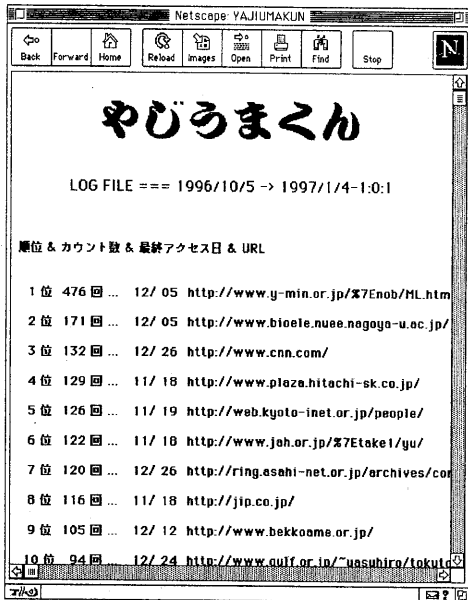


図4 やじうまくんのホームページ2

ホームページとは言いにくい。そのため、本システムでは最近の動向を見るため、付加的に最終アクセス日を表示してある。そして、人気のあるホームページかどうかを判断する基準の1つとしている。

6. 試行実験

6.1 実験方法

本システムを評価するため、試行実験を行った。対象は、本研究室の学生を対象に行った。今回の試行期間は、1995年8月から1996年12月までとした。使用した端末はPC-UNIXが5台、Macintosh 6台で、のべ14人に使用してもらった。

6.2 実験結果

試行実験を行った1995年8月から1996年12月の17カ月間と、過去最近の3カ月間(1996年9月7日～1996年12月8日)においてアクセスされた総アクセス数と有効アクセス数を表1に示す。ここで有効アクセス数とは、ログファイルの中から画像ファイルなどの不要のデータを取り除き、必要なHTMLファイルだけを

選び出し、「やじうまくん」で使用するアクセス数のことをいう。

表1 総アクセス数と有効アクセス数

	総アクセス数 (件)	有効アクセス数 (件)
1995.8～ 1997.1.3	205,777	35,963
1日平均 (平日)	約550.2	約96.2
1日一人 当たり平均 (平日)	約39.3	約6.9
1996.10.4～ 1997.1.3	42,544	7,228
1日平均 (平日)	約644.6	約109.5
1日一人 当たり平均 (平日)	約46.0	約7.8

(1997年1月4日現在)

この試行実験後にアンケートを行ったところ、研究室でのホームページの人気が分かるようになった、という意見が多かった。

また、WWWブラウザとして有名なネットスケープでは、自分の興味のあるホームページを記録しておくBookmarksというものがある。これは、今行ったホームページが自分の興味に合えば簡単に登録しておくことができ、かつそのホームページに行きたいときにはいつでも行けるようになっている。一方、「やじうまくん」では、アクセス回数の多いホームページは人気があるホームページと判断され上位に表示される。つまり、本システムを利用することによって、あたかもグループ内でBookmarksを共有しているのと同じ機能があることが分かった。

問題点としては、WWWのホームページを見る際に、proxyサーバを通すとアクセスが遅くなるという意見や各ホームページがURLで表示されているためホームページの内容が分かりにくいとい

う意見もあった。

7. 考察

本システムを使用することによって、かなり自分に興味のあるホームページを見つける手助けになることが分かった。

しかし、今回の試行実験では対象者が本研究室の学生というある意味では共通した興味のあるグループの中で行ったため、良い結果につながったのかもしれない。これが、興味の異なるグループや規模の大きなグループになると、その結果はまた別のものになると考えられる。そして、その時においてはまた新たな長所、短所があると思われる。

また、現在のシステムにおいても、ホームページのURLの表示では内容までが分かりにくい、などといったシステムでの不十分な点を指摘する意見も多かった。まだまだ試作のシステムであるが、多くの課題があることが分かった。

さらに、ログファイルを使う本研究と似たシステムとして、WebTrends[5]がある。これは、「ホームページ・アクセス状況解析ツール」であり、自分のサイトにどのくらいアクセスしているのか、どのような企業や組織からアクセスしてくるのか、といった情報をログファイルから瞬時に解析し、見やすいレポートをホームページ上で見ることができるものである。WebTrendsでは、ログファイルを解析し、その結果を棒グラフや折れ線グラフなどを立体的にかつカラーで表示しているためとても視覚的に見やすくなっている。

これを「やじうまくん」と比較すると、WebTrendsでは主にどこからのアクセスがあるかというサーバの利用状況の調査に向いている。それに対して、「やじうまくん」ではどこへアクセスしたかというグループ内の利用状況を調べることができるという点が異なる。

8. おわりに

本システムではユーザがホームページをアクセスしたアクセス回数を人気と定義して、人気のあるホームページの一覧を提供することによって、

自分に興味のあるホームページを見つける手助けをするシステムを提案した。しかし、本来人気とはあくまで漠然としたものであり、アクセス回数はその目安の一つにしかすぎない。それにもかかわらず、本システムではそれなりの成果があったことは、アクセス回数が人気の目安として正しかったことを示している。そして、まだまだ不十分な点が多いが更なる工夫によって本来の漠然とした人気に近づくことができ、自分に興味のあるホームページを見つける手助けがしやすくなると考えられる。

参考文献

- [1] 森田昌宏: 情報フィルタリングに関する研究動向. JAIST Reserch Report IS-RR-93-9I, 北陸先端科学技術大学院大学情報科学研究科, 1993
- [2] Paul, R., Neophytos, I., Mitesh, S., Peter, B., John, R.: GroupLens: An Open Architecture for Collaborative Filtering of NetNews, Proc. of CSCW'94, 1994, pp 175-186.
- [3] yahooのホームページ:
<http://www.yahoo.co.jp/index.html>
- [4] 千里眼のホームページ:
<http://www.info.waseda.ac.jp/search.html>
- [5] WebTrendsのホームページ:
<http://www.jp.webtrends.com/>