



データ宝石箱を用いたウェブアクセスログの視覚化

山口 裕美¹⁾, 池端 裕子¹⁾, 伊藤 貴之¹⁾, 梶永 泰正¹⁾

Visualization of Web access logs using Data Jewelry Box

Yumi Yamaguchi, Yuko Ikehata, Takayuki Itoh and Yasumasa Kajinaga

ABSTRACT

This paper presents a tool for visualizing Web access logs, which has two views. The left view represents Web sitemaps by a set of webpage icons. It places the icons onto display spaces by data jewelry box algorithm. The right view represents Web access statistics by a bar chart. It aggregates the accesses according to user-specified attributes, such as time, client, and status. When a user clicks a bar in the right view, the left view highlights icons that have corresponding accesses. When a user clicks an icon in the left view, the right view shows the access statistics of the clicked webpage. These combinations of the two views help users to find problems of websites, and discover trends of accesses.

Keywords: Web access log, Visualization, Data Jewelry Box

1. はじめに

ウェブサイトを通して簡単に情報発信ができるようになり、現在ではそれを介したオンラインビジネスも増えている。これに伴い、ウェブページが増加し、ウェブサイトの構成が複雑になってきた。ウェブサイトの運営は、主にサイト管理者、設計者が行っており、つねにウェブサーバーに蓄積された膨大なアクセスログの解析結果を反映させて行っている。管理者はサイト内で発生している通信状況を把握するために、設計者はサイト閲覧者が関心のある情報に確実にアクセスできるようなサイトを作るために、アクセスログを解析している。この他、オンラインビジネスの戦略をたてるためにもアクセスログ解析が用いられている。

このようにさまざまな目的でアクセスログが解析されており、そのためのツールが多く開発されている。既存のツールは一画面に簡単な統計グラフを表示するものが多い。この場合、サイト全体のおおまかなアクセス傾向を把握することは可能であるが、個々のページのアクセス状況などのような詳細を把握することは難しい。また、一画面に1つのビューしか表示できないので、複数の結果を比較

するのが困難である。

本報告では、これらの問題点を考慮して、サイトマップとアクセスグラフを同時に表示する視覚化ツールを提案する。サイトマップとは、ウェブサイトを構成するウェブページ群の構成図のことである。また本報告では、アクセスログをユーザーが指定した属性で分類し、その集計数を表した棒グラフを、アクセスグラフと呼ぶ。本ツールは、既存のツールでは見つけられなかったサイト閲覧者のアクセス傾向などを、ユーザーに発見させることを目指している。

以降、次章で近年盛んに研究されているウェブサイトの視覚化技術を概観したあと、3章で本ツールの概要を説明する。4章で、実在するウェブサイトのアクセスログを用いた実験例を示し、5章でまとめと今後の課題を述べる。

2. 従来のウェブサイト視覚化技術

ウェブサイトの視覚化技術の分野では、多くの手法がすでに報告されている。大別すると、サイト構造、リンク構造、アクセス履歴の視覚化などがある。

ウェブサイトの階層構造を表現するサイトマップの典型的な例として、Inxight社が公開しているユーザーインタフェース [Inx] がある。この手法は、放射状に展開された木構造データを双曲面に投影する手法 [Lam96] を用いて、トップページを根にしたウェブサイトの階層構造を表現するとともに、ユーザーの対話操作に連動した局所ズームによる詳細表示を実現している。また、Durandらが提案した MAPA [Dur98] も、ウェブページを階層構造にしたがって、縦横に並べる表現を用いており、ユーザーの

* 原稿受付 ****年**月**日

1) 日本アイ・ビー・エム(株)東京基礎研究所
(〒242-8502 神奈川県大和市下鶴間 1623-14, E-mail :
yyumi@jp.ibm.com)

対話操作にしたがって段階的に上位階層から下位階層を表示している。

ウェブページ間のリンク構造の視覚化手法は、グラフ構造の視覚化手法の応用として多く報告されている。たとえば、力学モデルを適用してノードを適切な分布で配置するグラフ視覚化手法 [Ead84]を、サイトマップの視覚化に適用した手法などが報告されている [Hen95]。

一方、ウェブページの意味や属性から画面空間上の座標値を算出して、その座標値にしたがってノードを配置する手法も知られている。塩澤らは、あらかじめ算出された座標値にしたがって平面上に配置されたノードを持ち上げる操作によって、特定のウェブページのリンク関係を対話的に視覚化する手法を提案している [Shi97]。

ウェブのアクセス情報を視覚化する手法の多くは、個人のアクセス履歴を表現する手法である。具体的にいうと、個人がアクセスしたウェブページ群をアイコン化して、それをアクセス順に線で結んで表示する手法が多く報告されている [Aye95][Fre98][Bed98]。

本報告で提案するツールは、個人のアクセス履歴を視覚化するのではなく、多数のサイト閲覧者によるアクセスの統計や傾向を視覚化するという点、その統計結果とサイトマップを連携させること、という点において従来手法と異なっている。

3. 提案ツールの概要

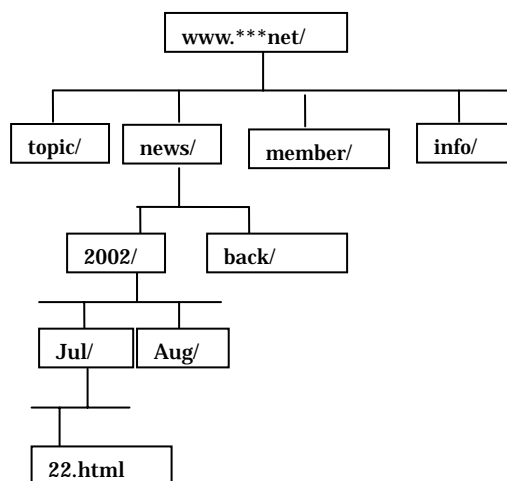
本報告で提案する視覚化ツールは、ウェブサーバーに蓄積されるアクセスログファイルを入力データとする。アクセスログには、サイト閲覧者がウェブページにアクセスしたときの、1)サイト閲覧者の IP アドレス、2)認証を行った際のユーザー名、3)アクセス日時、4)リクエスト内容、5)ステータス、6)転送バイト数、7)アクセスされたページの URL、8)7)のリンク元の URL、などが記録されている。本ツールは、このアクセスログからサイトマップを生成すると同時に、アクセスログのいくつかの属性情報に注目してアクセスグラフを生成する。本ツールはこれらを連携することで、サイト閲覧者によるアクセス傾向を視覚化する。

3.1 データ宝石箱によるサイトマップ表示

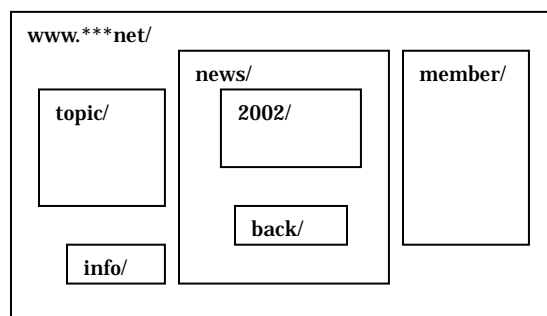
ウェブサイトの階層構造を視覚化するために本ツールでは、アクセスログから URL を抽出してその階層構造を作成し、それを入力データとしてサイトマップを作成する。

まずアクセスログから、アクセスされた全てのウェブページの URL を抽出する。たとえば http://www.***.net/news/2002/Jul/22.html というように、URL は / で区切られたディレクトリ階層からなっているので、この構造をもとに全ての URL から1つの階層構造を作成する(図1(上))。

本ツールでは、この階層構造を入力データとして、著者らが提案したデータ宝石箱アルゴリズム[Ito]を用いてサイトマップを表現する。この手法は、階層構造をもつ多量のデータ全体を、限られた画面空間の中ですべて表示することを目的とした視覚化手法なので、ウェブサイトの全体像を視覚化するのに適している。本ツールでは、ウェブページをアイコンで、ディレクトリを入れ子状の箱で表現して、ウェブサイト全体の階層構造を視覚化する(図1(下))。



(上) URL から作成されたサイトの階層構造



(下) (上)のサイト構造をデータ宝石箱アルゴリズムで配置した例

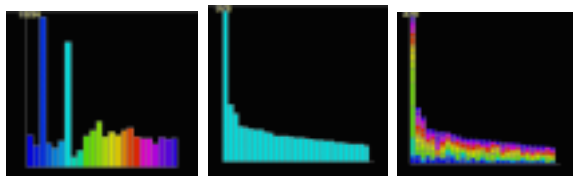
Fig.1 アクセスログから作成された階層構造グラフと配置例

3.2 アクセス数の棒グラフ表示

サイト閲覧者のアクセス傾向を視覚化するために本ツールでは、アクセスグラフを使用する(図2(左))。

アクセスグラフは、横軸に属性の種類、縦軸にアクセス数をとる。アクセス数の絶対数を縦軸にとると、分類別のアクセス数の差が極端に大きい場合は表示が困難になるので、アクセス数の相対数を縦軸にとる。アクセスログの属性のうち、1)、7)、8)に関しては、アクセス数が最も多い30項目を、アクセス数の多い順に表示する(図2(中))。

また、1つの属性で分類したものを、さらに別の属性で分類して集計することも可能である(図2(右))。



(左) (中) (右)

Fig.2 アクセスグラフ

(左) アクセスログを1時間ごとに集計した結果

(中) URLごとに集計した結果

(右) URLごとに集計し、さらに1時間ごとに集計した結果

3.3 2つのビューの連携

本システムは、3.1節で説明したサイトマップを左画面に、3.2節で説明したアクセスグラフを右画面に配置し、これら2つのビューを連携させて用いる。

[連携1] アクセスグラフからサイトマップへの反映

本ツールは、アクセスグラフの中でユーザーが関心のある項目を指定すると、左画面のサイトマップ上で、指定された項目に関するアクセスがあったウェブページのアイコンをハイライトさせる。このときアイコンに高さを与えることで、個々のページのアクセス数を表示する。このようにして、アクセスログの特定の属性に対するアクセス分布を、サイトマップ上で視覚化することができる。

[連携2] サイトマップからアクセスグラフへの反映

左画面のサイトマップにおいて、ユーザーが関心のあるウェブページを指定すると、右画面には、そのページへのアクセスを集計したアクセスグラフが表示される。これによりユーザーは、サイト全体のアクセス傾向だけでなく、特定のページに対するアクセス傾向も知ることができる。

上記のように本ツールでは、サイト管理者による対話的な操作によって、アクセスグラフとサイトマップの連携をとることができる。そしてユーザーは、サイト全体のアクセス傾向だけでなく、特定ページのアクセス傾向までを把握することができる。図3に本システムの様子を示す。ここで使用したデータは1日分のアクセスログで、表示されているアクセスグラフは時間帯で分類したものである。さらに図3では、アクセスグラフの中でユーザーの関心のある時間帯が指定された様子を示しており、左画面のサイトマップでハイライトされたアイコンは高さを持っており、アクセス数を表示していることがわかる。

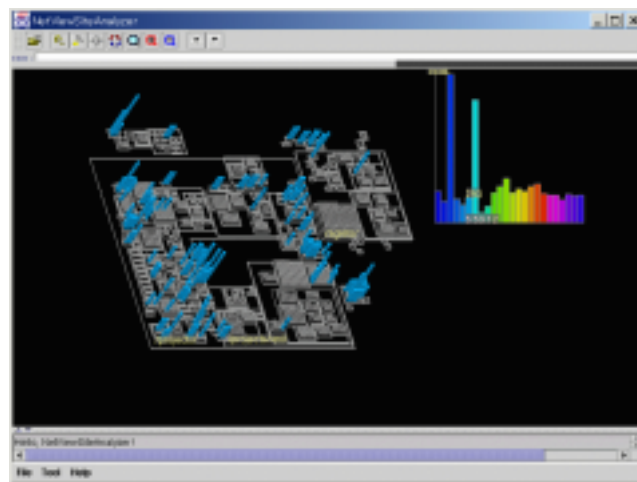


Fig.3 本ツールによる視覚化の例

4. 実験

本ツールを用いて、実在するウェブサイトの1週間のアクセスログを視覚化した実験例を示す。

ウェブページに対するアクセスだけを視覚化の対象にするために、本実験ではまず前処理として、アクセスログからURLが画像ファイルやスタイルファイルなどを示している行を削除した。このアクセスログを日付で分類し、さらに各項目を時間帯で分類した。以上の分類にしたがって、横軸が日付、縦軸がアクセス数を示すアクセスグラフを作成した(図4(左上))。これを見ると、最終日のアクセス数が最も高いことがわかった。

[連携1]を用いて、サイトマップ上で最終日のアクセス分布を表示すると、あるページのアクセス数が午前中に突出していることがわかった(図4(左下))。続いて[連携2]を用いて、そのページへのアクセスを対象として、リンク元のURLで分類したアクセスグラフを表示した。すると、ある新聞会社のURLからのリンクによって訪れているサイト閲覧者が多いことがわかった(図4(右上))。リンク元である新聞会社にアクセスしてみると、オンラインニュースにアクセス数の多かったページが取り上げられていたことがわかった。以上の結果から、午前中に新聞サイトを見て、そのリンクをたどってこのページに来た人が多かったことを推測した。

再び[連携1]を用いて、サイトマップ上で別の時間帯のアクセス分布を表示すると、同一ディレクトリ中のほとんどのページがアクセスされていたことがわかった(図4(右下))。[連携2]を用いて、これらのページへのアクセスをサイト閲覧者のIPアドレスで分類して表示すると、すべてのページに同一IPアドレスからのアクセスがあることがわかった。これらの結果から、あるディレクトリのフ

ファイルをすべて見ている熱心なサイト閲覧者が存在していたことがわかった。

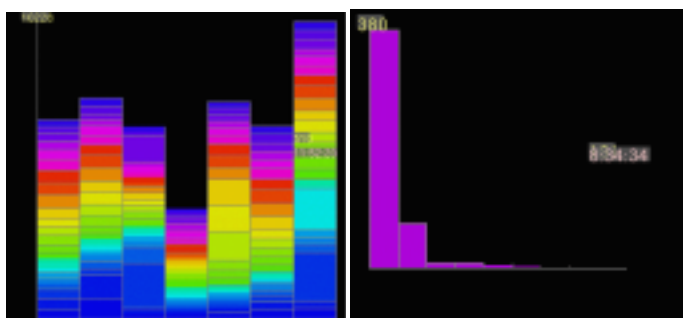


Fig.4 一週間分のアクセスログの視覚化
 (左上) ウェブサイト全体に対するアクセスグラフ
 (右上) あるページに対するアクセスグラフ
 (左下) アクセスグラフ(左上)のある項目に対するアクセス分布をサイトマップ上で表現した例
 (右下) アクセスグラフ(左上)の別の項目に対するアクセス分布をサイトマップ上で表現した例

5. おわりに

本報告では、サイトマップとアクセスグラフを連携させ

る、アクセスログの視覚化ツールを提案した。本ツールを利用することで、ユーザーはサイト全体の大局的なアクセス傾向から、個々のページのアクセス傾向までを把握することができる。

今後の課題として、以下のようなことを検討中である。

アクセスグラフは、属性ごとに独立してアクセス数を集計しているが、複数の属性情報の和や積といった、異なる属性の組み合わせも視覚化できるようにしたい。これにより、ユーザーは異なる属性間の相関から興味深いアクセス傾向を発見することができるであろう。

データ宝石箱アルゴリズムでは、サイトマップ上のウェブページの位置は自動的に決定されるので、ユーザーの意思の入る余地がない。ユーザーが部分的にサイトマップの配置をデザインしたいとき、その要求を満たしながら、その他の部分は自動的に配置させるように、データ宝石箱アルゴリズムを拡張していきたい。

そのほか、サイト閲覧者がサイトに訪問して別のサイトに移動するまでの、リンクをたどる動きの傾向を表示することで、より詳細なアクセス傾向をユーザーに提供できるようなツールを目指したい。

参 考 文 献

- [Aye95] Ayers E., Stasko J., Using Graphic History in Browsing the World Wide Web, 4th International World Wide Web Conference, 1995.
- [Bed98] Bederson, B.B., Hollan, J.D., Stewart, J., Rogers, D., Vick, D., Ring, L.T., Grose, E., Forsythe, C., A Zooming Web Browser, Human Factors in Web Development, pp 255-266, 1998.
- [Dur98] Durand D. G., Kahn P., MAPA: A System for Inducing and Visualizing Hierarchy in Websites, Proceedings of ACM Hypertext '98, pp. 66-78, 1998.
- [Ead84] Eades P., A Heuristic for Graph Drawing, Congressus Numerantium, 42, 149-160, 1984.
- [Fre98] Frecon E., Webpath - A Three Dimensional Web History, IEEE Information Visualization '98, pp. 3-10, 1998.
- [Hen95] Hendley R. J., Drew N. S., Wood A., Beale R., Narcissus: Visualizing Information, Proceedings of the 1995 Information Visualization Symposium, pp. 90-96, 1995.
- [Inx] Inxight Star Tree (TM) SDKs, http://www.inxight.com/products_sp/ht_sdk/index.html
- [Ito] 伊藤, 梶永, 池端, データ宝石箱: 大規模階層型データのグラフィックスショーケース, 情報処理学会グラフィクス&CAD研究会, 2001-CG-104, 2001.
- [Lam96] Lamping J., Rao R., The Hyperbolic Browser: A Focus+context Technique for Visualizing Large Hierarchies, Journal of Visual Languages and Computing, 7, 1, pp. 33-55, 1996.
- [Shi97] 塩澤他, 「納豆ビュー」の対話的な情報視覚化における位置付け, 情報処理学会論文誌, 38, 11, pp. 2331-2342, 1997.