

Webのアクセスパターンとリンク構造の 同時可視化の一手法と適用事例

川本 真規子^{†1} 伊藤 貴之^{†1}

ウェブの可視化手法は、ウェブサイト本体に関する情報（リンク構造や文書内容など）の可視化と、閲覧者のアクセス動向の可視化に大別される。この2種類の可視化を統合することで、ウェブサイト運営に関する有益な知見が得られると考えられる。本報告では、カテゴリ情報を有するネットワークを効果的に画面配置するネットワーク可視化手法「FRUITS Net」を用いて、ウェブサイトのアクセスパターンとリンク構造を可視化する手法を提案する。本手法では、クローラによりリンク構造を、アクセスログファイルによりアクセスパターンを構築し、これらの情報を可視化する。我々は本手法が、ウェブサイトのページ構成やページ内容の再検討に役立つと考えている。本報告では、アクセスパターンとリンク構造を同時に可視化することで、アクセスパターンとリンク構造との関係性、ウェブサイトを訪れる閲覧者の特徴、などを発見した事例を報告する。

A Visualization Technique for Access Patterns and Link Structures of Web Sites and Its Case Study

MAKIKO KAWAMOTO^{†1} and TAKAYUKI ITOH^{†1}

There have been two types of Web visualization techniques: visualization of Web sites themselves based on link structures or lexical contents, and visualization of browsers' behaviors. We think that integration of such two visualization techniques is very useful for Web site management, and therefore we are currently studying on visualization of access patterns and link structures on a single screen. This paper presents a Web visualization technique using 'FRUITS Net', a visualization technique for multiple-category-embedded network data. The presented technique constructs link structures using crawler software, and access patterns from access log files. We expect that users can utilize the knowledge obtained from the visualization results for design and management of Web sites. This paper presents several case studies of relationships between access patterns and link structures, and features of those who access to the Web site and so on.

1. はじめに

ウェブに関する情報可視化の研究は、1990年代中盤から活発に報告されている。ウェブ可視化の対象は、リンク構造や文書内容などウェブサイト本体に関する情報と、アクセス統計をはじめとする閲覧者情報に大別される。これら2種類の情報を一画面に同時に可視化することで、ウェブサイトの構築や管理に関する有用な知見が得られることが期待される。

我々はウェブサイトの閲覧者情報の中でも、アクセスパターンの可視化に着目している。本報告ではアクセスパターンを、複数の閲覧者からアクセスされる同一ウェブページ群と定義する。これをリンク構造と同時可視化することにより、アクセスパターンとリンク構造との関係性、ウェブサイトを訪れる閲覧者の特徴、という知見を視覚的に得られると考えられる。そしてこの知見を、ウェブサイトのページ構成やページ内容の検討などに活用できると考えられる。

ウェブサイトの可視化技術の多くは、ウェブページをノード、ハイパーリンクをエッジに置き換えて、ネットワーク可視化技術を適用することでウェブサイトを表示している¹⁾。これに加えてアクセスパターンも可視化する場合に、リンク構造とアクセスパターンの双方の可読性を高めるための要件は、かなり複雑になる。そのため、汎用的なネットワーク可視化技術をそのまま適用したのでは、リンク構造とアクセスパターンの双方の可読性を高めることは難しいと考えられる。

本報告では、力学モデルと空間充填モデルの2種類の画面配置モデルを組み合わせた「FRUITS Net」²⁾というネットワーク可視化の新しい手法を用いて、アクセスパターンとリンク構造の同時可視化を試みた事例を報告する。本手法では、まずウェブページをノード、ハイパーリンクをエッジ、ウェブサイトのディレクトリ構造をクラスタとした階層型ネットワークを構築する。そして、以下の4つの条件を満たすようにノードを画面配置する。

- 条件 1: 同じアクセスパターンのウェブページを画面上の近い位置に配置する
- 条件 2: ハイパーリンクで接続されたウェブページを画面上の近い位置に配置する
- 条件 3: ウェブページやクラスタの重なりを回避する
- 条件 4: ウェブページやクラスタの画面占有面積を低減する

^{†1} お茶の水女子大学
Ochanomizu University

本報告では、我々の所属研究室のウェブサイト为例として本手法を適用した結果を示し、我々が可視化した典型的なアクセスパターンについて議論する。

2. 関連研究

ウェブサイトにおける閲覧者の行動は、ウェブデザイナーや管理者にとって、とても興味深い情報である。そのため、そのような閲覧者の行動分析や抽出手法についての研究が活発に行われている。アクセスログからのアクセスパターン抽出手法として、文献³⁾では、閲覧者の1セッション内においてアクセスされた一連のURLの推移について、Longest Common Subsequence(LCS) アルゴリズムを用いて頻出のアクセスパターンを抽出するという方法を提案している。また文献⁴⁾では、各ページに閲覧時間の長さに応じて重みを付け、ネットワークマイニングによりアクセスパターンを抽出する方法を提案している。閲覧者の興味遷移の抽出手法として、文献⁵⁾では、ウェブアクセスログデータを解析し、閲覧者の興味やアクセスしている情報が時間と共にどのように変化しているのかを抽出して可視化している。Nasraoui らは、閲覧者の行動に対して類似度を計算し、ファジークラスタリングを適用することでアクセスパターンを求めている⁶⁾。Pitkow らは、マルコフモデルに基づいたユーザ閲覧パターンの予測モデルを提案している⁷⁾。Davison は、ユーザが最近訪れたページの内容からユーザの関心を見つけることによって、アクセスパターンを予測する手法を提案している⁸⁾。

ウェブサイトの可視化手法について、本研究に関連する例として以下があげられる。ウェブサイトのリンク構造を可視化する手法として、ウェブサイトを階層型ネットワークデータとして表現し、力学モデルを用いたネットワークデータの画面配置手法により可視化する手法が挙げられる¹⁾。アクセスログファイルから得られる情報を可視化した例として、文献⁹⁾では、ウェブサイトで提供されるサービスの関連性分析のための可視化方法を提案している。また、ウェブサイトのアクセス分布の可視化手法¹⁰⁾¹¹⁾もいくつか提案されている。このうち文献¹¹⁾では、アクセス統計とリンク構造の同時可視化を試みているが、この手法でのリンク構造は1ページを根とした木構造に限定されている。

ウェブサイトに限定しない汎用的なネットワーク構造の可視化手法は、ネットワーク描画(Graph Drawing) という理論を応用した技術として、現在も活発に研究されている。その研究の最も大きな課題は、ネットワークを構成するノードの効果的な画面配置によって、どのようにネットワーク構造の可読性を高めるか、という点にある。本研究のように、ウェブサイトを構成するネットワークがリンク構造とカテゴリ情報(本報告ではアクセスパター

ン)を有する場合、その可読性を高めるための要件は複雑になる。本報告で採用する可視化手法「FRUITS Net」²⁾は、ノード配置とノード着色の工夫により、リンク構造とカテゴリ情報の両方に対して可読性を高める、という点においてネットワーク可視化の新しい手法であるといえる。^{*1}

3. アクセスパターンとリンク構造の同時可視化手法

本報告で提案する手法の処理手順は以下の通りである。本手法ではまず、前処理として

- アクセスログからのアクセスパターン構築
- クローラを用いたリンク構造構築

により、入力データである階層型ネットワークを生成する。そして、この階層型ネットワークを「FRUITS Net」で可視化する。

本報告では、アクセスログファイルの入手可能な1ドメインを対象として、そのウェブサイトのトップページからクローラによってリンクを辿ることにより、リンク構造を構築する。また本報告では、標準的なアクセスログとして、閲覧者IPアドレス、アクセス日時、アクセスされたファイル名、リンク元ページのURL、使用しているOS名やブラウザ名、などが記録されているアクセスログファイルの使用を前提とする。

ただし、このアクセスログファイルから、閲覧者の全てのページ遷移を抽出できるとは限らない。例えばウェブブラウザの「戻る」ボタンを押した場合などには、キャッシュされたウェブページを再表示するためにサーバへのアクセスが発生せず、結果として閲覧履歴がアクセスログファイルに記録されないことがある。そのため本研究では、以下の3種類の立場を想定するものとする。

立場 a: 閲覧者がどのページからどのページへ辿ったというページ遷移を一切参照しない

立場 b: アクセスログファイルから抽出できるページ遷移のみを参照する

立場 c: 徹底的に閲覧者のページ遷移を記録する

[立場 a]では、アクセスパターンとリンク構造を同時に可視化することで、画面上で閲覧者のページ遷移を想像できるが、その正当性は保証されない。[立場 b]では、視覚的に認識できるページ遷移を可視化することはできるが、キャッシュされたページを表示した場合な

*1 この可視化手法の名称のうち FRUITS は、FRamework and User Interface for Tangled Segments の略である。ネットワーク可視化手法に限らず多くの可視化手法の多くにおいて、画面上で線分群が絡まって可読性が低下する、という問題を解決するために提案されている手法群の総称である。FRUITS Net はその一環として、ネットワーク可視化手法として提案されたものである。

ど、ログファイルに記録されていないため、閲覧者のすべてのページ遷移を正確に可視化することはできない。[立場 c] では、正当性のある形で閲覧者のページ遷移を可視化できる。しかし、そのためにはウェブサイトの各ページにトラッキングコード等を埋め込む、あるいは閲覧者側のパソコンに特定のプログラムをインストールしてデータを採る、などの方法で閲覧者の全ページ遷移を記録する必要がある。だが、これらの方法を使用すると、上述の措置がとられていないウェブサイトへの適用が困難になる、あるいは限られた閲覧者のページ遷移しか記録できない、といった制限が生じる。そのため現時点では、我々は [立場 a] と [立場 b] を前提として研究を進めている。しかし原理的には、[立場 c] を前提として提案手法を適用することも可能である。

3.1 アクセスパターン抽出

アクセスパターン抽出における我々の実装について述べる。図 1 はアクセスパターンを抽出する処理手順である。

本処理ではまず、アクセスログファイルを読み込み、閲覧者と URL の一覧を作成する。ただし我々の実装では、画像や音楽などのコンテンツファイルの URL を削除し、それ以外の URL だけを対象とする。続いて本処理では、閲覧者の IP アドレスの数を n 、アクセスされた URL の数を m として、 $n \times m$ の表を作成する。表の各欄には、各閲覧者から各 URL へのアクセス回数の集計結果を記録する。続いて本処理では、閲覧者のデンドログラムを構築する。このとき 1 閲覧者のアクセス回数を m 次元ベクトルとして、すべての閲覧者ペアについてベクトル間余弦を算出し、これが最大であるペアを併合する。この処理を再帰的に反復することで、デンドログラムを構築する。そして、このデンドログラムを用いて閲覧者を階層的にクラスタリングする。閾値 α を決め、閾値以上のものを 1 つのクラスタとしてクラスタリングを行う。そして、全クラスタの中から構成員数が β 人以上のクラスタを抽出する。続いて各クラスタに対して、クラスタ構成員の γ %以上がアクセスしたページを抽出し、アクセスパターンのデータを構築する。 β 、 γ はそれぞれユーザが定義した値を使用する。また、現時点での我々の実装では、3 ページ以上のページが抽出されたアクセスパターンのみを可視化の対象としている。

3.2 リンク構造構築

リンク構造のデータ構築にはクローラを使用する。本処理では、アクセスログファイルを入力したサイトのトップページを指定し、そこからリンクで繋がっているページを再帰的に探索してリストを作り、得られた URL をもとにリンクのネットワークを構築する。我々の実装では、オープンソースとして提供されている「JSpider」¹²⁾ というクローラを採用して

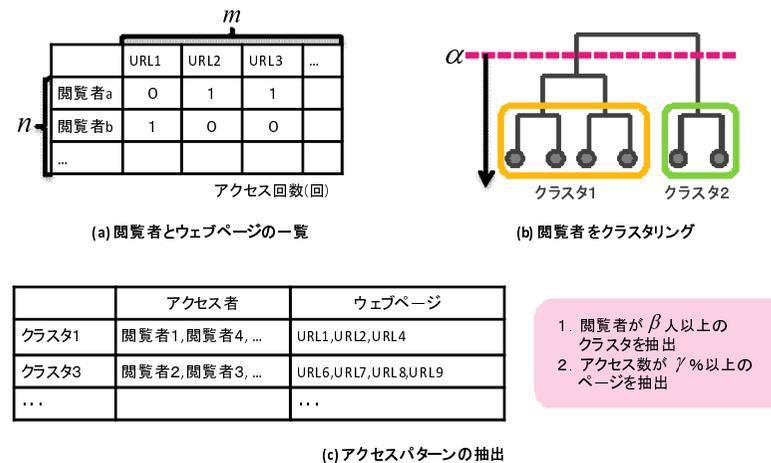


図 1 アクセスパターン抽出の処理手順
Fig. 1 Processing flow of access pattern extraction.

いる。

3.3 データ統合

本手法では 3.1 節および 3.2 節で構築されたデータを統合する。本処理ではまず、3.1 節および 3.2 節で抽出された URL を統合し、ディレクトリ構造に基づいて階層的に格納することで、各 URL を葉ノードとする木構造を生成する。そして各ノードにリンク構造を付加することで、階層型ネットワークを形成する。さらに、各 URL にアクセスパターン情報を付加することで、可視化のための入力データを構築する。

[立場 b] を前提とする場合には、本手法ではもう一度アクセスログファイルを解析し、アクセスされた各々のページに対して、URL とアクセス元 URL (どこのページからアクセスされたか) のペアを抽出する。続いて、ウェブサイトを構成する各リンクに対して、その両端にあるページの URL ペアが何回抽出されたかを集計し、閲覧者の通過頻度としてその集計値の情報を記録しておく。

3.4 可視化

本手法では、力学モデルと空間充填モデルに基づく 2 種類のネットワーク画面配置手法

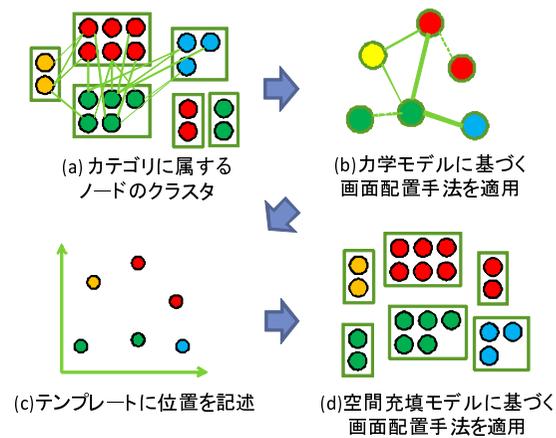


図 2 画面配置処理手順

Fig. 2 Processing flow of the hybrid layout technique.

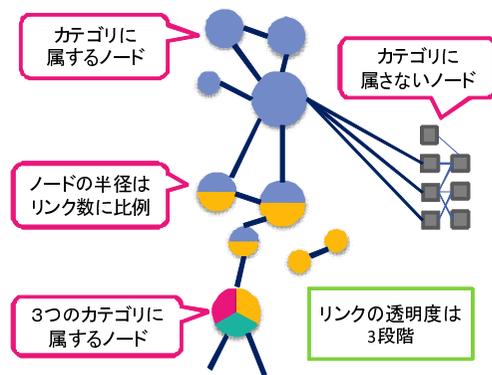


図 3 ノードとリンクの描画

Fig. 3 Drawing style of our implementation.

を組み合わせた可視化手法「FRUITS Net」を、リンク構造とアクセスパターンを統合したデータに適用する。

図 2 はネットワーク可視化手法の処理手順を示す。図 2(a) は色付けされたノードのクラスターの集合の例である。図 2(b) はサブネットワークの例である。ここで、サブネットワークのエッジは、少なくとも 1 つ以上のノードがリンクで結ばれているクラスター間、もしくは、同じ色のノードで構成されるクラスター間を結ぶ。続いて、各クラスターの位置を、力学モデルに基づく画面配置手法を適用することで計算する。具体的には、クラスター間を連結する各エッジにバネを仮想し、反復的解法によってバネ間の力に関する運動方程式を解くことで、クラスターを連結するエッジの長さを適正化する。続いて図 2(c) に示すように、力学モデルによる配置結果をテンプレートと呼ばれるデータとして記述する。最後に、テンプレートに記述されている位置を理想位置として、空間充填モデルに基づく画面配置手法を適用し、力学モデルによる配置結果を修正する。具体的には、クラスターを表現する各長方形が、

空間充填条件 1: 互いに重ならない
 空間充填条件 2: 画面占有面積を最小化する
 空間充填条件 3: テンプレートに記述された理想位置にできるだけ近い位置に配置するという条件をできるだけ満たすような位置に、各長方形を配置する。図 2(d) は空間充填モデルによる配置結果を描いたものである。

以上のクラスター配置手法により、1 章で述べた条件 1~4 をできるだけ満たす画面配置を実現する。具体的には、力学モデルに基づく画面配置手法により [条件 1] と [条件 2] を満たし、空間充填モデルに基づく画面配置手法により [条件 3] と [条件 4] を満たしている。そのため、同じアクセスパターンに属するページや、リンクで接続されたページを近くに配置し、また、重なりを回避しながら、できるだけ多くのページを一画面上に表示することが可能となっている。

図 3 は本手法でどのようにノードやリンクが描かれているかを示している。本手法では、各アクセスパターンに独立した色を割り当てており、アクセスパターンに属するノードは色付けられた円として描かれる。複数のアクセスパターンに属しているノードの場合は、円の内部を分割して 1 つのノードに複数の色を付けられるようにしている。ノードの半径は接続されたリンク数に比例しており、たくさんリンクを持つノードほど大きく描かれる。また、アクセスパターンに属していないノードは、小さな灰色のドットで描かれる。

さらに、リンクの描画に際して、我々は 2 種類の色付け方法を実装している。1 つ目は、両端のノードに着目してリンクを

1. 色づけされたノード同士を連結するリンク
2. 色づけされたノードと色づけされていないノードを連結するリンク
3. 色づけされていないノード同士を連結するリンク

の3種類に分類し, 1. を最も明るい色で描画し, 3. を最も透明度の高い状態で目立たないように描画する, という方法である. 2つ目は, [立場 b][立場 c]に基づいてデータを生成した際に, 閲覧者の通過頻度によってリンクの色を計算する, という方法である. 我々の実装では, 通過頻度の高いものから赤, 緑, 青という順番で色相を段階的に変化させながらリンクに色を割り当てている. 通過頻度がゼロのリンクは薄紫色としている.

4. 適用事例

本章では, 我々の所属研究室のウェブサイトの本手法を適用した事例を報告する. 我々は提案手法を Java JDK1.5.0 を用いて実装し, Apple MacBook (CPU 2GHz, RAM 1GB) および MacOS 10.4.11 を用いて実行した.

我々の所属研究室のウェブサイトの内容は, 図4のようになっている. 研究テーマに関する分野の紹介であるプロジェクトページ, メンバーのホームページへのリンクが張ってあるメンバーページ, 発表論文が置いてあるレポートページ, 大学までの地図が出ているアクセスページ, 研究室に関連するページへのリンク集であるリンクページから成り立っている. メンバーページからは教員, Aさん, Bさん, Cさん... というように個人のウェブサイトへのリンクが張っており, 教員のウェブサイトには, 講義の資料, 経歴, 論文, 海外滞在記などのページがある.

図5は可視化結果画面の一例である. 右端はGUIの操作部分であり, 右下部分に色とボタンの一覧がアクセスパターンに対応している. 色の横のボタンを押すと, 対応するノードとそのノード同士をつなぐリンクがハイライトされる.

4.1 全体図

研究室のウェブサイトを可視化すると, 各ページは図5のように配置された. 色付けられているノードの色に着目してみると, 研究室のトップページと, 教員のトップページに共通のアクセスパターンがないことがわかる. これにより, 同じ研究室のページの中でも研究室のトップページと教員のトップページは一緒にアクセスされることはあまりないということが考察できる. また, 図5の真ん中の白いラインより右側は教員のウェブサイトに含まれるページ, 左側は学生のウェブサイトや研究紹介のページなどその他の研究室のページが配置されている. 各アクセスパターンを見てみると, どれも同じ側のページ内で色付けられて

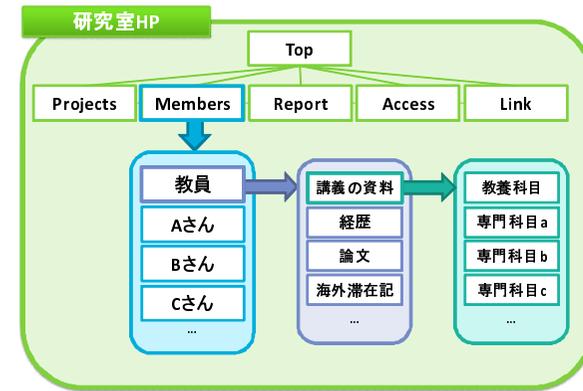


図4 ウェブサイト内容
Fig. 4 Web site contents.



図5 全体図
Fig. 5 Overview

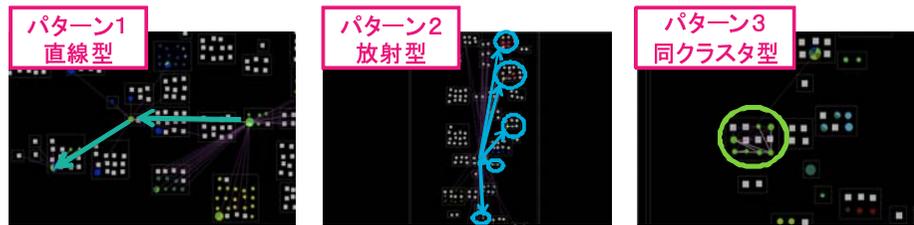


図 6 典型的な 3 パターン

Fig. 6 Three types of Web access patterns

おり、線をまたいで色付けられているアクセスパターンはほぼない。このことから教員のページは教員のページ内で、研究室のページは研究室のページ内で、それぞれ独立してアクセスされていることがわかる。

4.2 典型的なパターン

我々は可視化結果より、図 6 に示すような 3 つの典型的なアクセスパターンが効果的に表現されていることを見つけた。

パターン 1 直線型のアクセスパターン

パターン 2 1 つのノードを中心とした放射型のアクセスパターン

パターン 3 同クラスタ内だけにアクセスが集中している同クラスタ型のアクセスパターン

図 6(左) は、教員の担当講義の資料ページへのアクセスパターンに該当する部分を拡大表示している。この可視化結果から、教員のトップページから担当科目のページへアクセスし、そこから専門科目の資料のページへアクセスした、という閲覧者の軌跡を想像できる。この結果から、目的のページにアクセスするために、トップページから順にリンクを辿ったアクセスが多数あったことがわかる。また、図 6(中央)(右) は研究室のメンバーのページに関連するアクセスパターンに関する可視化結果を示す。(中央) の図は、メンバー一覧のページから、同学年全員のホームページにアクセスする閲覧者が複数いることを示している。(右) の図は、同クラスタ内の大半のページにアクセスされ、特定の個人のホームページだけをアクセスする閲覧者が複数いることを示している。

本事例を通して我々が発見した 3 つの典型的なアクセスパターンのうち、直線型のアクセスパターンをとる閲覧者は、目的の 1 ページにアクセスするためにリンクを辿ってアクセスしているということが想像される。このアクセスパターンをとる場合には、目的のペー

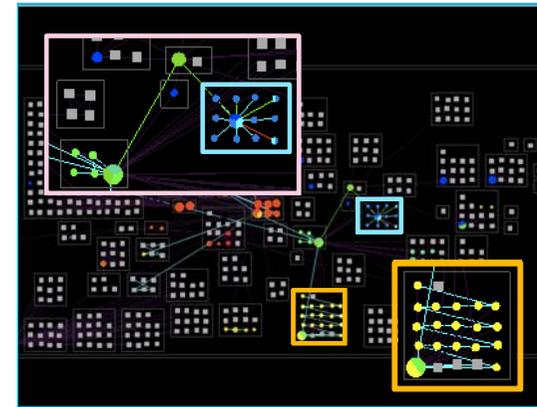


図 7 リンクに色をつけて表示

Fig. 7 Represent frequency of browsers' passes for each link by color

ジまでに、どのくらいのページを経由していったかということが重要であると考えられる。多くのページを経由している場合はリンクの再構築を検討する必要があると考えられる。放射型のアクセスパターンをとる閲覧者は、あるページを中心として紹介されている複数のページに関心があるということが想像される。このアクセスパターンの場合は、一緒にアクセスされたページがどのようなページなのかということを知ることで、閲覧者の興味を把握することができ、ページ内容の充実に役立てられると考えられる。また、同クラスタ型のアクセスパターンをとる閲覧者は、同じディレクトリ内のページにだけ関心があるということが想像される。このアクセスパターンの場合は、同クラスタ内にあるアクセスされなかったページに着目することが重要であると考えられる。アクセスされなかったページの内容は適切か、そのページへのリンクはきちんと張られているかということの再確認に役立つと考えられる。このように、可視化結果に現れたアクセスパターンの分布から、サイトを訪問した閲覧者がどのような意図を持ってアクセスしているのかということを目視的にとらえ、想像することによって、リンクの再構築やページ内容の再考に役立てることができる。

4.3 その他の可視化結果

図 7 は閲覧者の通過人数によってリンクに色をつけて表示した結果である。右下の拡大部分からは、閲覧者がディレクトリ内のページを順番にリンクを辿ってアクセスしているということがわかる。また、左上の拡大部分からは、このディレクトリ内のページにアクセス

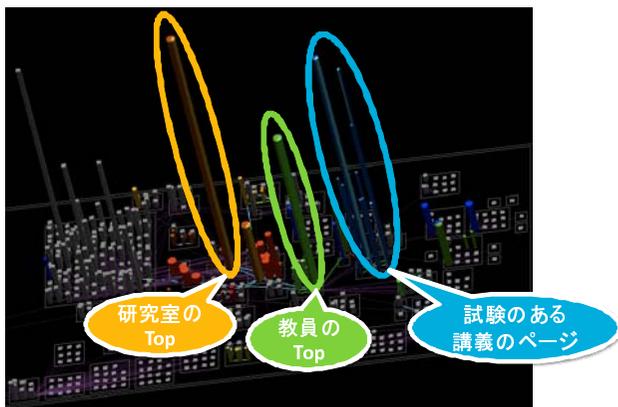


図 8 アクセス数を高さで表示

Fig. 8 Represent frequency of accesses for each Web page by heights

するために多くの閲覧者がリンクを辿ってアクセスしてきているということなどを考察できる。通過頻度の高いリンクを見つけることで、ウェブサイトを訪れる閲覧者にとっての重要なトピックを発見でき、また、リンクの色の有無によって、張られたリンクがきちんと機能しているかという確認に役立つと考えられる。

図 8 は各ページのアクセス数を高さで表示した結果である。この結果からは、研究室のトップページ・教員のトップページに加えて、試験のある講義の資料ページにも同じくらいアクセスがあることがわかる。トップページは他のページへアクセスするための軸となるページなのでアクセスが多いことは自明だが、講義のページにも同じくらいのアクセスがあるということから、このページに掲載されている情報が多くの閲覧者にとって必要であることが、この結果から示唆される。この例のような、多くの閲覧者を集めるページは、トップページから直接リンクを張るなど、少ないアクセスで目的のページまで辿りつけるような位置に置くことで、閲覧者のアクセスがより効率良いものになると期待できる。

5. ま と め

本報告では、「FRUITS Net」を用いたアクセスパターンとリンク構造の同時可視化の一手法を提案し、その適用事例を紹介した。本手法では、クローラを用いてリンク構造のデー

タを、アクセスログファイルからアクセスパターンのデータを得る。そして、それらを統合することによって階層型ネットワークを構築する。続いて、この階層型ネットワークに対して、力学モデルと空間充填モデルに基づく 2 種類のネットワーク画面配置手法を組み合わせた手法により可視化する。

本報告で紹介した適用事例では、3 種類の典型的なアクセスパターンが表現されていることがわかった。また、これらのアクセスパターンがウェブサイトの改良にどのように役立つかということを議論した。

今後は、我々の可視化技術にある問題点の解決に取り組みたいと考えている。現在、アクセスパターン 1 つに対して 1 色の色を割り当てているが、この方法では、非常に多数のアクセスパターンが抽出されたウェブサイトを可視化したときに、色が多すぎて可読性が低下し、アクセスパターンの識別が大変困難になるという問題点がある。そのため、より多くのアクセスパターンを一画面上で表現できるように、色だけに頼らない他のデザインを考える必要がある。また、何万ページもあるような大きなウェブサイトでもスムーズな表示操作を実現できるように、計算速度の向上を図りたい。

上述の問題点に伴って、現時点の我々の実装では、数百個単位のアクセスパターンが抽出されてしまった場合に、抽出されたアクセスパターンの中から手動で 10~20 個を選んで可視化している。一方で、可視化技術の改良によって数百個単位のアクセスパターンを表示できたとしても、ユーザにとって情報提示過多である場合があることも否定できない。よって可視化技術の改良の有無に関わらず、アクセスパターンの抽出個数を制御することは有用であると考えられる。そこで今後の課題として、抽出されたアクセスパターンの優先度を算出し、適切な個数のアクセスパターンを自動選択できるようにしたい。また、現在のアクセスパターン抽出処理は 3.1 節に示した非常に単純な方法をとっており、意味のあるアクセスパターンをもれなく抽出できているのかという点是不確実である。そのため、アクセスパターン抽出処理に関しても再検討し、改善を図りたい。

また、現在の実装では、閲覧者がどこのページからどこのページへ移動したというアクセスの方向については可視化していない。今後の課題として、アクセスログファイルから解析できる範囲だけでもいいから閲覧者のアクセスの方向を可視化する、もしくは第 3 章で想定した 3 つの立場のうち [立場 c] を想定して研究を進める、などの形でアクセス方向の可視化を試みたい。

さらに別の課題として、サイト構成上の問題点を発見しやすくなるように、本可視化手法を改良したいと考えている。具体的には、期待に反してアクセス数の少ないページや、たく

さんのリンクを辿らなければ目的のページまで到達できないページなど、問題のある場所が強調表示されるような可視化機能を設けることで、ウェブサイト上の問題点をより発見しやすくしたいと考えている。

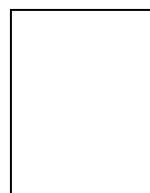
参 考 文 献

- 1) 土井淳, 伊藤貴之, 力学モデルを用いた階層型グラフデータ画面配置手法の改良手法とウェブサイト視覚化への応用, 芸術科学会論文誌, Vol. 3, No. 4, pp. 250-263, 2004.
- 2) T. Itoh, C. Muelder, K.-L. Ma, J. Sese, A Hybrid Space-Filling and Force-Directed Layout Method for Visualizing Multiple-Category Graphs, 2009 IEEE Pacific Visualization Symposium, pp. 121-128, 2009.
- 3) 宇根田純治, 横田治夫, Web ログの共通シーケンス解析, 電子情報通信学会信学技報 DE2002-2, 2002.
- 4) 三原宏一郎, 寺邊正大, 橋本和夫, ページ閲覧時間を考慮した Web ログマイニング手法の提案, 情報処理学会, pp. 39-44, 2007.
- 5) 山田和明, 中小路久美代, 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.
- 6) O. Nasraoui, H. Frigui, A. Joshi, R. Krishnapuram, Mining Web Access Logs Using Relational Competitive Fuzzy Clustering, *Eight International Fuzzy Systems Association World Congress*, 1999.
- 7) J. Pitkow, P. Piroli, Mining Longest Repeating Subsequences to Predict World Wide Web Surfing, *2nd conference on USENIX Symposium on Internet Technologies and Systems*, pp. 139-150, 1999.
- 8) B. D. Davison, Predicting Web Actions from HTML Content, *13th ACM Conference on Hypertext and Hypermedia*, pp. 159-168, 2002.
- 9) 末永高志, 岡田崇, 石打智美, Web アクセスログデータの系列情報を利用したサービスの関連性の分析, 電子情報通信学会信学技報 DE2005-17, 2005.

- 10) 山口裕美, 伊藤貴之, 池端裕子, 梶永泰正, 階層型データ視覚化手法「データ宝石箱」とウェブサイトの視覚化, 画像電子学会誌, Vol. 32, No. 4, pp. 407-417, 2003.
- 11) 山縣修, 中村泰明, アクセス確率による Web サイトのリンク構造可視化ツール, 可視化情報学会論文集, Vol. 26, No. 6, pp. 43-50, 2006.
- 12) 「JSpider」<http://j-spider.sourceforge.net/>

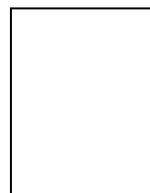
(平成?年?月?日受付)

(平成?年?月?日採録)



川本真規子

2010 年お茶の水女子大学理学部情報科学科卒業。2010 年お茶の水女子大学大学院人間文化創成科学研究科理学専攻情報科学コース博士前期課程在学中。



伊藤 貴之 (正会員)

1990 年早稲田大学工学部電子通信学科卒業。1992 年早稲田大学大学院理工学研究科電気工学専攻修士課程修了。同年日本アイ・ビー・エム(株)入社。1997 年博士(工学)。2000 年米国カーネギーメロン大学客員研究員。2003 年から 2005 年まで京都大学大学院情報学研究科 COE 研究員(客員助教授相当)。2005 年日本アイ・ビー・エム(株)退職, 2005 年よりお茶の水女子大学理学部情報科学科助教授(現在准教授)。ACM, IEEE Computer Society, 情報処理学会, 芸術科学会, 画像電子学会, 可視化情報学会, 他会員。