

WWWにおけるナビゲーション履歴の三次元視覚化による 情報検索支援

梅田孝行 垂水浩幸 上林弥彦

tumeda, tarumi, yahiko@kuis.kyoto-u.ac.jp

京都大学大学院情報学研究科 社会情報学専攻

現在 WWW における情報検索ではキーワードなどを用いた検索エンジンが主流となっており、利用者はそこで得られた URL からナビゲーションすることで目的のページを探索する。しかし現状ではナビゲーションする付近のページの情報あまり与えられず試行錯誤で探索するため、情報検索が効率よく行える環境ではない。そこで我々は先人のナビゲーション履歴から情報検索に有用な情報を抽出し、ナビゲーション中の利用者に提示することで探索を支援する手法を提案する。本稿では開発中のナビゲーション共有機構におけるナビゲーション傾向の提示を三次元空間上で行う方法を述べるとともにそのプロトタイプの評価を行う。

Supporting Information Retrieval by 3D-Visualization of Navigation Histories on the WWW

Takayuki Umeda Hiroyuki Tarumi Yahiko Kambayashi

Dept. of Social Informatics, Graduate School of Informatics, Kyoto University

Recently search engines utilizing keywords are popular in information retrieval on the WWW. Users search for their destination resources starting with URLs those engines provide. In the present environment, however, resources are hard to find without information around the given web documents. In this paper, a technique to support information retrieval by extracting some useful piece of information from other users' navigation histories and providing them to the users, is proposed. Furthermore our prototype of 3D-visualization is described and evaluated.

1 はじめに

WWW は世界中の不特定多数の人々に利用されており、WWW サーバの数は日々刻々と増加している。しかし利用者間で情報検索知識を交換しあう機能が十分に提供されていないため、各利用者は単独で情報検索を行うに留まっており、効果的な情報

検索が困難である。

現在この情報検索に関する問題の解決法として各種サーチエンジンや協調フィルタリングなどによる URL 提示手法などが盛んに研究されている。しかしここでは単なる URL と簡単な紹介文だけが提供されることが多く、選択してナビゲーションを始め

た後は何の手がかりもないリンクを辿らなければならず、利用者は試行錯誤の探索を余儀なくされる。そのため現況の情報検索では利用者が独自に培った知識や経験が必要とされる。

このような情報検索で培った知識を利用者間で共有することができれば、利用者はその共有された知識を基にナビゲーションすることで無駄な情報検索を抑えることができ、効率よく目的の情報を得ることができると考えられる。これらの知識は利用者それぞれのナビゲーション傾向に深い関係があるので、このナビゲーション傾向を共有し再利用することが知識共有につながると考えられる。

このような動機に基づき、WWWでの情報検索を支援するナビゲーション共有機構について研究している。本稿ではナビゲーション共有機構の概要について述べ、ナビゲーションデータの管理、利用者へのナビゲーション提示手法、そのプロトタイプの開発について述べる。

2 ナビゲーション共有機構

我々が開発中のナビゲーション共有機構とは利用者がWWWブラウザで辿ったページの履歴（ナビゲーション履歴）を利用者間で様々な形で共有し、再利用することで情報検索を支援するものである。図1にこの機構の仕組みを示す。利用者の過去のナビゲーション履歴をデータベースに蓄積し、それを統合、加工して有用な情報を抽出することで利用者の情報検索にフィードバックさせるというものである。

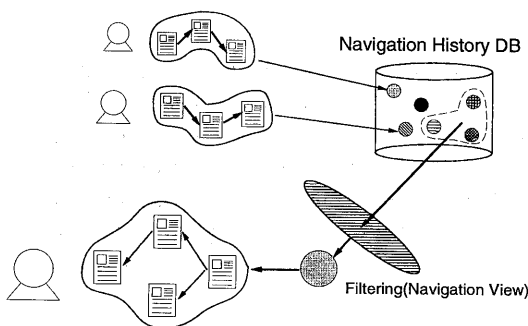


図1: ナビゲーション共有機構の仕組みの概略

2.1 ナビゲーション共有の目的

現在WWWでは非常に多くの利用者が存在しており、年々その数はネットワークの発達などにより

確実に増加している。このような多数の利用者の中には同じような興味をもち、似たような情報を求めているものがあると考えられる。このように興味や目的の近い利用者間で情報交換をシステムの仲介で自動的に行うことで有効な情報が取得しやすくなると考えられる。

また初めて訪れるサイトなど、馴染みのない分野の検索においてもその熟練者のナビゲーションを共有し再現することで、独自に検索能力を高めるよりも簡単にかつ速く検索の技術を学ぶことができると考えられる。

2.2 ナビゲーションビューの概念

このようにナビゲーションを利用者間で共有することで効率よい情報検索ができると考えられるが、問題となるのはWWW利用者の興味の多様性である。単に利用者全員の総体としてのナビゲーション傾向を共有するだけでは、情報が平均化されてしまうので再利用性に乏しい。興味に応じて利用者を分類し、そのグループでの傾向を共有することが望ましい。

このように利用者全体からあるグループを特定し、そのグループ内でのナビゲーション傾向を共有することをナビゲーションビューと呼ぶ。ナビゲーションビューによりある一つのサイトのナビゲーション傾向を、いくつかの特定の興味をもつグループのそれぞれの視点から眺めることができる。ナビゲーションビューはナビゲーション傾向の平均化を避けるために必要不可欠な概念である。

ナビゲーションビューを実現するときに問題となるのが次の二点である。

2.2.1 利用者グループの特定方法

ナビゲーションビューの機能を実現するためには、利用者に対してどのようなグループがあるか、また利用者にグループを自由に設定できるような仕組みが必要である。このための方法として以下の二つの方法が考えられる。

履歴によるフィルタリング 利用者のナビゲーション履歴の相関度を調べ、ある閾値によりグループを特定するという方法がある。このときすべての利用者についてすべてのナビゲーション履歴から相関度を比較しては実時間性からも実用的でないとと言える。処理を軽くするような履歴データ構造の構築が不可欠である。

カテゴリ情報による分類 これはシステム側であらかじめ用意しておいたカテゴリラベル (computer, CSCW, movie など) を利用者が選択することでグループを指定する方法である。このときに問題となるのがシステムが用意できるカテゴリ空間が閉じていることが挙げられる。利用者が目的や興味に応じた独自のカテゴリを作成できることが必要と考えられるが、その場合同じ内容のカテゴリの中でも様々に派生したものが生成される可能性があるので分類機能そのものがうまく働かなくなるおそれがある。

2.2.2 利用者グループに応じて傾向を統合できるような履歴データ管理手法

ナビゲーションビューを考慮しない場合、履歴は利用者全体のもを統合して一つのものとして記録すれば良い。しかしナビゲーションビューでは各利用者の興味などに応じて似通った目的を持つ人のナビゲーションを共有することを目標としており、個人個人の興味を識別するための何らかのプロファイルが必要となる。つまりナビゲーション履歴を利用者ごとに記録する必要が生じる。さらに個人個人のナビゲーション履歴自体が膨大な量となりうるので、それらを簡単に扱えるようなデータ構造も考えなければならない。

3 ナビゲーションデータの管理

ナビゲーション履歴を管理する際に注意すべき点として以下のようなものが挙げられる。

- 利用者のナビゲーションを忠実に再現できる情報を保存する
- 膨大なナビゲーション履歴情報の扱いを効率化する

以下にこれらの条件を満たすようなナビゲーションデータの管理方法について述べる。

3.1 ナビゲーションオブジェクト

利用者がWWWブラウザで行ったナビゲーションの際の履歴を記録するオブジェクトとしてナビゲーションオブジェクトを定義する。ナビゲーションからその傾向を抽出するために必要な情報として、URL、タイトル、閲覧時間の三つを基本に考えた。

ナビゲーションオブジェクトは、図2のように辿ったページの順にこれらの値を配列型変数に格納

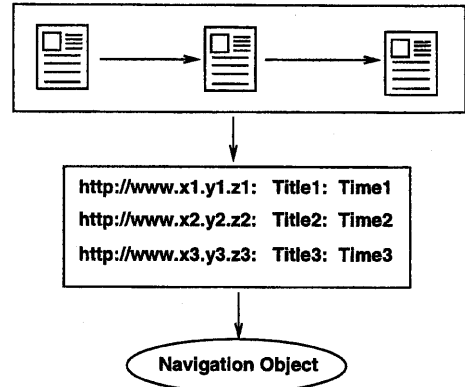


図2: ナビゲーションオブジェクト

したものである。利用者はこのナビゲーションオブジェクトを共有し、再利用する。

3.2 ナビゲーションオブジェクトの拡張

3.2.1 ブラウザ上での操作情報

ナビゲーションで得られる傾向を考えると、ブラウザ上での操作 (マウスポインタ、スクロールバーの動き) が非常に重要である。特に注目度などの尺度においては閲覧時間だけで判断するよりも、これら操作情報を手がかりに本当に利用者がその文書を注視しているかという判定をするほうがよい。¹

3.2.2 時系列情報

一般的に利用者の興味は時間と関係していると考えられる [3][4]。よって閲覧時刻をナビゲーションオブジェクトに登録することで、より細かく利用者の興味を扱えるようになると考えられる。

3.3 統合化ナビゲーションオブジェクト

前節でのべたナビゲーションオブジェクトは利用者がナビゲーションを繰り返すにつれてその数は膨大なものとなる。これらの膨大なデータの中から利用者に有用な傾向を抽出するためには、データ処理の効率を上げるような工夫が必要である。

我々はこの問題の解決策として統合化ナビゲーションオブジェクトというものを実装する。これは

¹ Netscape Communicator 4.x では JavaScript を用いてこれらの情報を取得することができる。

複数のナビゲーションオブジェクトの傾向を一つに統合したものである。

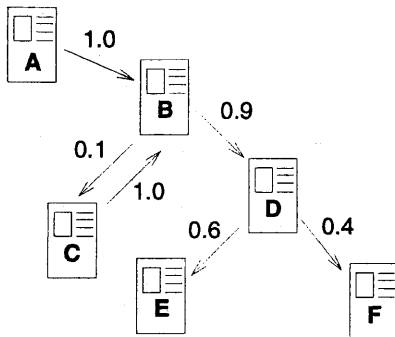


図 3: 統合化ナビゲーションオブジェクトのイメージ図

図 3 に統合化ナビゲーションオブジェクトのイメージを示す。これは意味的には任意の複数のナビゲーションオブジェクトから得られるすべてのページ（ノード）、リンク（有向枝）をマージしたグラフ構造をもつ。各枝はそれに対応するリンクを辿った統計的確率を表す数値をもつ。また各ノードはそれに対応するページに対する利用者の注目度を表すパラメータをもつ。

利用者のナビゲーション履歴ごとにこの統合化ナビゲーションを作成することで、ナビゲーションビューの処理が高速化される。

4 ナビゲーション提示手法

対象となる文書は WWW 文書であり、ハイパーテキスト構造である。複雑なハイパーテキスト構造のナビゲーション傾向を提示する際には、二次元平面より表現力のある三次元空間の方が特徴量の視覚化に向いているといえる [1]。ここでは、この三次元空間を利用したナビゲーション傾向の提示方法について述べる。

ナビゲーション履歴の再利用性を高めるためには、利用者が容易に履歴の統計情報を得られるようなインタフェースが不可欠である。ナビゲーション履歴とは利用者が WWW 空間の中を情報検索しながら歩き回った足跡のようなものであり、その統計情報はいわば「けもの道」のようなものとして表現できる。

4.1 履歴に現れる特徴量

ナビゲーション履歴は、URL、文書のタイトル、注視時間の三つのデータの配列で構成され、このデータから経路情報や閲覧回数などの統計情報を生成する。この履歴の中から得られる情報の視覚化について考察するために、履歴から得られる特徴量について以下に簡単に述べる。

経路情報 WWW 文書間の遷移確率で求められる。

頻繁に辿られる経路を提示したり、めったに辿られない経路を隠すなど、視覚化において重要な意味合いをもつ特徴量である。

注目度 利用者の WWW 文書に対する閲覧時間、回数、ブラウザ上での操作情報などで求める。基本的に閲覧時間、回数が大きければ注目度を高く設定するが、ただ単に表示しただけといった例外的な状態などは操作情報を基に調整する。

分岐度 そのページに存在するリンクの数で求められる。分岐先が多ければそのページは道標のような役割を果たすページであることが推定できる。

4.2 特徴量の視覚化

WWW 文書は三次元空間中に後で述べる手法にしたがって配置される。WWW 文書は空間中の平面で表され、WWW 文書間のリンクはそれらを結ぶ線分で表現される。このような WWW 空間において上で述べた特徴量を視覚化する方法について述べる。

経路情報の視覚化 けもの道のメタファーを用いる。

つまり、その経路を通る利用者の数が増えるにしたがって道幅が広く、目に付きやすいように視覚化する。実際には WWW 文書間のリンクに付随する統計情報に基づいて線分の太さを変えて表示する。なお、リンクの数が膨大で WWW サイトの構造が見づらき場合などは不要な経路（リンク）を非表示にすることも考えられる。

注目度の視覚化 注目度の高いページを利用者にわかりやすく提示するためには、ほかの注目度の低いページと何らかの表示上の区別をつける必要がある。ここでは表示される WWW 文書の色のある注目度のしきい値を越えると変えるという方法を採用する。

分岐度の視覚化 この場合も注目度と同様、あるしきい値を越えるものに何らかの印付けをする

ことが望ましい。ただし、注目度情報と競合しないために別の方法(旗をつけるなど)を採る。

4.3 三次元空間における WWW 文書の配置手法

WWW 文書は世界中の WWW サーバに存在し、その数は膨大なものとなる。しかもこれらが互いにハイパーリンクで結ばれていることも考慮すると、これらすべてを一つの空間に表示することは視認性の面からも処理速度の面からも現実的ではない。よって一つの空間に表示する WWW 文書のある特定の範囲で決定する必要がある。本研究ではこの範囲がある WWW サイト中の文書に限定することにする。つまり一つの空間には一つの WWW サイトの文書だけが表示される。

ある WWW サイトの文書を配置する際には、そのサイトのディレクトリ構造を利用する。なぜなら WWW 文書はおもにこれを管理する人間がディレクトリ構造で意味的に分類した上でおかれることが多いと考えられるので、サイトの意味的構造を理解する上で役に立つと考えられるからである。

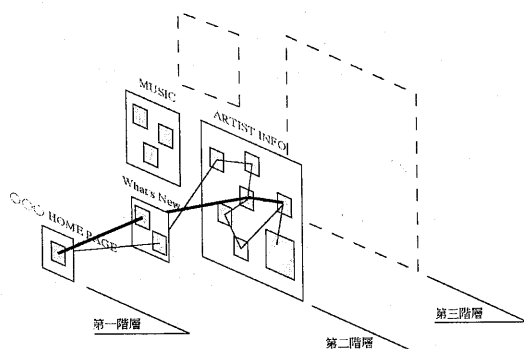


図 4: WWW 文書の配置イメージ図

配置のイメージ図を図4に示す。同じ階層のディレクトリについて、包含する WWW 文書が表示可能なスペースを矩形で確保し、それらが交わらないように適切に配置する。これをすべての階層について行う。階層の深さが増すにしたがって、奥行き方向に配置される。確保した矩形を描画することで WWW 文書のカテゴリをわかりやすく表示することができる。

4.4 配置手法の拡張

前に述べた配置手法では各文書の空間座標をディレクトリごとにグループ化することで意味的分類を行った。つまり階層の水平の関係でクラスタリングを行ったわけである。ここで考えられる拡張方法としては、階層の縦の関係すなわち階層関係に基づくクラスタリングが可能であるということである。

現在の配置手法の実装では第一階層、第二階層それぞれが独立にディレクトリ分類を行い、親ディレクトリ、子ディレクトリとの関係は考慮されていない。そこで子ディレクトリは自分の属する親ディレクトリの近くに配置することが拡張方法として余地がある。

ただしサイトのすべての文書がこれらの条件を満たした上で重なりあうことなく配置されるためには、サイトの文書のディレクトリ構造をすべて把握した上でボトムアップに計算しなければならない。よって前の手法と比べると実用性の点で大きな問題となる。

5 プロトタイプの開発

現在 SUN Ultra1/Solaris2.5 UNIX プラットフォームで JDK1.1 によって三次元視覚化機構を実装中である(図5)。

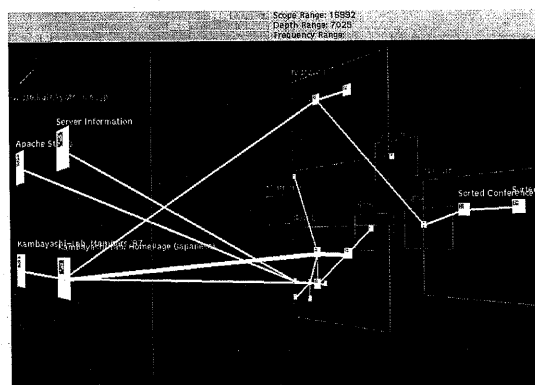


図 5: 視覚化機構の試作プログラム

画面の上部にはスライドバーがあり、これを実行することで WWW 文書の見えかたやリンクの表示の仕方を変化させることができる。画面の下部は WWW 空間を表示する部分であり、ポリゴンで表現された WWW 文書が前述の三次元配置方法にしたがって配置されている。WWW 文書はその属するディレクトリごとに分類されており、ディ

レクトリはグレーの矩形で表示されている。なお WWW 文書やディレクトリには名前も添えて表示される。リンクはポリゴンではなく、経路情報に基づく太さをもった線分で表される。

5.1 視覚化の調整

利用者はマウス一つでこの空間を自由に動き回ることができる。操作を簡単にするために視点は必ず z 平面に対して水平になるように設定してある。利用者は次の三つのスライドバーを調整することで様々な見方ができるようにした。

スコープ範囲設定用 これは利用者の視界を決めるためのものである。近くに密集して WWW 文書が存在する場合は、このスライドバーを調整することで遠くの WWW 文書を見えなくしたり、逆に遠くの WWW 文書を把握できたりする。

階層の深さ設定用 WWW 文書が密集していると、階層の深さの幅を広くとって文書をまばらに表示させたいという欲求が起る。これを実現するのが階層の深さ設定用スライドバーである。

経路の頻度設定用 このスライドバーでは、どれだけ頻繁に辿られている経路かを示す値に対してある閾値を設定することでリンク（経路）の表示 / 非表示を行う。つまりこれを高く設定すれば、それに応じて頻度の高い経路のみが表示されるようになる。

5.2 視覚化のインタフェースの評価

三次元空間を自由に動き回ることにより、さまざまな視点で Web サイトの構造を眺められる。しかし自由度が高すぎるため、好みの視点に合わせるには移動の距離に応じた手間がかかってしまう。ブラウザのナビゲーションと同期して視点をかえる工夫が求められる。

スコープ範囲の設定は、奥行き方向でページが重なる場合などに有効であった。しかし奥行き方向でページが重なって見えることの利点はとくになく、著しくページの視認性が落ちるだけであるので、ディレクトリの矩形を不透明にすることでスコープ範囲の設定を不要にすることが望ましいと考えられる。

階層の深さの設定においては、スライドバーを動かしてやることでページ間のリンクのつながりが把握しやすくなった。

表示する経路の頻度設定により、リンクが多いサイトの文書の構造が把握しやすくなった。

6 おわりに

本稿では、履歴を効果的に再利用することで情報検索を支援するナビゲーション共有機構について述べた。利用者のナビゲーション履歴からその傾向を抽出し、グループで統合する事で履歴の再利用性を高める手法について検討した。また、その時のナビゲーション傾向を利用者にわかりやすく提供する方法についても述べた。

現状ではナビゲーションオブジェクトには利用者のナビゲーションという粗い履歴しか記録できないが、今後ブラウザやプログラミング言語の拡張によって細かい履歴まで取得できると考えられる。これにより WWW を用いた遠隔教育の様子を記録し共有するなど幅広い応用例が考えられる。

今後は統合化ナビゲーションオブジェクトの WWW 文書群のデータ構造を、三次元視覚化との関係をもとにさまざまな視覚化に対応できるように設計する予定である。

謝辞

本研究について御討論頂いた上林研究室の皆様にご感謝致します。なお本研究の一部は情報処理振興事業協会「独創的情報技術育成事業」の支援によるものである。

参考文献

- [1] G.G.Robertson, J.D.Mackinlay, and S.K.Card.Cone Trees:Animated 3D visualizations of hierarchical information. In CHI '91, pages 189-194, 1991.
- [2] Netscape Communications Corporations. "Netscape Object Signing Overview". <http://developer.netscape.com/library/documentation/signedobj/overview.html>
- [3] 朝倉、喜田、垂水：WWW アクセスアドバイザー エージェント-興味に基づく WWW ウォッチング-、情報処理学会第 52 回全国大会 (1996)、4C-1
- [4] 畑島 隆、元田敏浩：時系列情報を考慮したアクセスログ解析、情報処理学会第 54 回全国大会 4S-9