

# WWW情報空間ナビゲーション手法について

1 L-5

畑島 隆 元田 敏浩 川崎 隆二 長岡 満夫

日本電信電話株式会社 ソフトウェア研究所

(takashi,motoda,kawasaki)@canary.sl.cae.ntt.co.jp,nagaoka@slab.ntt.co.jp

## 1. はじめに

本稿では、インターネット上のコンテンツに対して、利用者の興味度を道標としてユーザをナビゲーションする一手段を提案する。従来のシステムでは、ナビゲーション手段としてページビュー数上位のコンテンツや、関連サイトの一覧表示により行われていた。本稿では情報空間ナビゲーションのために、アクセスログの多角分析と、利用度空間上におけるコンテンツ位置の視覚的提示を一貫して行うシステム“サイバーランキング”について紹介する。

## 2. 背景

WWW のコンテンツからなる情報空間で有用なコンテンツを発見することは困難である。このため、(1)カテゴリ分類した情報空間を提供するディレクトリを作成する、アクセスログを分析し(2)利用度の高いページを推奨表示する、(3)個人属性に即したページを自動設定することにより、利用者のアクセスコスト低減を支援するナビゲーションが行われている。(1)は NTT DIRECTORY や Yahoo! などのサービスが古くから構築されているが、ディレクトリの構成を人手で行っているため、コンテンツ利用傾向が変化するたびに必要となるディレクトリ再構成のコスト問題が指摘されている。(2)は利用度の高いサイトへのリンクを提示することによりWWW情報空間のハブとなるサイト[1] [2]や、クライアントで参照しているサイトの関連サイトを提示するシステム[3]などがある。利用度の指標がページビューのみである点や、利用者の選択自由度が小さい点が問題となる場合がある。(3)は利用者の個人情報の取得が必要なため、プライバシーの問題があることや、個人属性とコンテンツ属性の的確なマッチングが困難であること問題となっている。本稿で提案するサイバーランキングでは、主に WWW サーバ管理者を対象とした以下の項目による情報空間ナビゲーションを用いて問題の解決を狙った。

- ・アクセスログの多角分析による特徴的な利用傾向を持つコンテンツの抽出
- ・WWW 情報空間の視覚的探索可能となる分析結果グラフの提供

## 3. 多角分析と探索可能なグラフ表示による情報空間ナビゲーション

サイバーランキングでは、ページビューや visit 数、アクセスの時系列傾向などの多角的な視点からアクセスログを分析する。その結果を利用者が操作可能なグラフとしてブラウザに表示されたWWW情報空間を探索することにより、一瞥で情報空間全体の利用傾向把握を実現した。同時に、個々のコンテンツのうちアクセス傾向に特徴があるものはグラフでも目立つ位置にプロットされるため、直感的な特徴量抽出も可能である。

また、グラフ表現に VRML を用いることにより、人気の高いコンテンツや関連の強いコンテンツを利用者が自由に探索することが出来る。

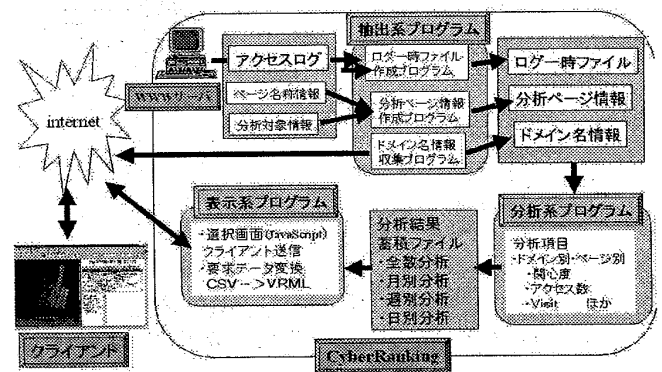


図 1: サイバーランキング構成図

## 4. サイバーランキングの構成と機能

サイバーランキングは抽出系・分析系・表示系のプログラムで実現している(図1)。分析結果を参照には WWW ブラウザを使用し、分析結果の選択画面に JavaScript を、結果の表示には VRML を使用している。サイバーランキングサーバとクライアント WWW ブラウザ間の通信に SSL を使用することによりセキュアな情報閲覧も可能である。以下に本システムを構成する各機能を説明する。

### 4.1. 抽出系機能

抽出系機能では WWW サーバが出力するアクセスログを定期的に収集している。分析対象とする Web ページの URL を設定し、これに従って分析系機能へ渡すログデータを生成する。このときフィルタを使用することにより WWW サーバにより出力形式が異なるアクセスログに対応した。

### 4.2. 分析系機能

分析系機能は、表 1 に挙げたそれぞれの分析項目について複数の分析単位を設定することにより多角的な分析を実現した。分析項目として、ページを構成する

HTML ファイルへのアクセス数 (ページビュー) だけでなくどれくらいの人がページにアクセスしたかを数える visit[4]や、アクセスの時系列を考慮して重み付けをした指標・関心度[5]を算出した。また、1visit のアクセスでのサーバ内の移動経路も算出した。分析単位として、コンテンツとアクセス元のそれぞれから見たアクセス傾向の分析を行い、さらに曜日別・日別・時間帯別の分析も行った。

アクセス数	ページ単位	全体
		ドメインごと
		曜日ごと
		日付ごと
		時刻 (1 時間) ごと
	ドメイン単位	全体
		ページごと
		曜日ごと
		日付ごと
		時刻 (1 時間) ごと
visit	ページ単位	全体
		ドメインごと
		曜日ごと
		日付ごと
		時刻 (1 時間) ごと
	ドメイン単位	全体
		ページごと
		曜日ごと
		日付ごと
		時刻 (1 時間) ごと
関心度	ページ単位	全体
		ドメインごと
		曜日ごと
		日付ごと
		時刻 (1 時間) ごと
	ドメイン単位	全体
		ページごと
		曜日ごと
		日付ごと
		時刻 (1 時間) ごと
関心度 (x 軸) アクセス数 (y 軸)	ページ単位	全体
		ドメインごと
		曜日ごと
		日付ごと
		時刻 (1 時間) ごと
1 visit での サイト内連続参照ページ	ページ単位	全体
		ドメインごと
1 visit での サイト内同時参照ページ	ページ単位	全体
		ドメインごと

表 1 : サイバーランキングの分析項目

(左からグラフ種別, 分析単位, 分析解析単位)

4.3. 表示系機能

クライアントからの要求に従って表 1 の項目を 2 次元もしくは 3 次元グラフで表示する。一画面のグラフに表現された WWW 情報空間の探索による情報空間ナビゲーションを以下のように実現した。

○ 視点変更と部分表示が可能なグラフによる全体傾向の把握と特徴量抽出

グラフ表示に VRML を使用することにより、多次元データを図形の形・大きさ・色などで表現するとともに、利用者の操作によるグラフの部分拡大・縮小や視点方向を変えた表示を実現した。

○ クリックオブジェクトによる情報空間探索

それぞれの画面内のオブジェクトには対応するページの URL を埋め込んでいるため、オブジェクトをクリックすることにより当該 Web ページを開覧することが出来る。またブラウザ画面内に設けたボタンをクリックすることにより、別視点からの分析結果を容易に参照できる。

5. サイバーランキングによる情報空間探索例

クライアントに表示されるナビゲーション画面例を図 2, 図 3 に示す。図 2 は関心度を x 軸に、Web ページに対するアクセス数 (ページビュー) を y 軸にしたグラフである。グラフの各オブジェクトが 1 つ

の Web ページに対応している。グラフ上でオブジェクトが右上にあるほど、常に高い関心を持たれている Web ページであることを示している。このグラフでは 2 つの Web ページが突出して高い関心を集めていることが判る。ディレクトリサーバであれば、これらの Web ページをより目立たせる構成に再構成することにより、利用者の利便と、アクセス数の向上を狙う事が出来る。また、図 2 左下のようにオブジェクトが集積する部分も拡大表示し個々のオブジェクトを操作することが出来る。

図 3 は 1 時間ごとのアクセス数 (ページビュー) を 10 Web ページ同時表示した 3 D グラフである。1 ページに対する時系列推移は図 3 左方から右方への一列で表されている。複数のページに対するアクセス傾向を同一画面にプロットすることにより、個々のページに対するアクセス傾向の特徴が把握できる。図 3 では Web ページごとにピーク時刻がそれぞれ異なることがわかる。この結果を利用し、時刻にあわせて人気のあるコンテンツを入れ替え表示とするサービスを提供することが出来る。

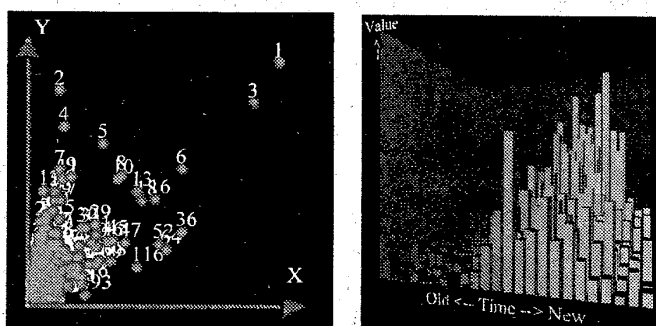


図 2 (左) : ナビゲーション画面例 (2D)

図 3 (右) : ナビゲーション画面例 (3D)

6. おわりに

WWW サーバのアクセスログの分析結果を単純に提示するだけでなくビジュアル表示することによりサーバ空間探索を支援するシステムを提案した。今後は Web サイトのコンサルテーションツールとしての拡充や、より効果的なナビゲーション方式の検討を進める。

参考文献

(1) 100hot.com, <http://www.100hot.com>  
 (2) Super NTT DIRECTORY, <http://super.navi.ntt.co.jp>  
 (3) Alexa, <http://www.alexa.com>  
 (4) visit, <http://www.ipro.com/faq.html>  
 (4) 畑島・元田, "時系列情報を考慮したアクセスログ解析" 情報処理, Vol.38, No.9, pp.848-849, 1997