

# 大規模サイトのナビゲーションを支援するパーソナライズ サイトマップシステム

島村 栄 高野 元 神場 知成  
NEC ヒューマンメディア研究所

本稿ではポータルサイトをはじめとする大規模な WWW サイト内でのユーザのナビゲーションを支援する WWW サイトマップナビゲーションシステムについて述べる。このサイトマップナビゲーションシステムは、大規模な WWW サイト内のドキュメント構造を自動解析し、サイト内コンテンツの目次にあたるサイトマップを自動的に生成する。サイト管理者は部分的にドキュメント構造を補助ファイルにて指定することによって意図した構造を作ることができる。また、自動生成されるマップはユーザの利便性を考慮し、インタラクティブに操作可能なコンパクトなユーザインタフェースと、ユーザのアクセス履歴を元にしたパーソナリゼーション機能を持つ。

## A Personalized Sitemap System for Navigation in Large WWW Sites

Hisashi Shimamura Hajime Takano Tomonari Kamba  
Human Media Research Laboratories, NEC

In this paper, we describe a development of a WWW site map navigation system which helps users to navigate in a large WWW site such as portal sites. This sitemap navigation system automatically explores a WWW site and creates a site map which is like a table of contents in the WWW site. Site administrators can construct a desired structure specifying partial structures of the WWW site on optional files. And the created site map has a compact user interface which enables interactive manipulations, and some functions for personalization based on a user's access history.

### 1. はじめに

World Wide Web (WWW)のアクセスの中心はポータルサイトと呼ばれる大規模な WWW サイトである。ポータルは入り口を意味し、これはユーザが WWW を利用する際に最初に訪れるページであることを意味している。ポータルサイトとしては、検索サービス、プロバイダのトップページ、ショッピングモールといった大量の WWW ドキュメントに対するアクセスを提供するサイトが当てはまる。

多くの大規模 WWW サイトでは、ユーザのサイト内ナビゲーションをスムーズにするためにサイト内のドキュメント構造（サイト構造）を目次のようにマップ化して提供している。しかし、このようなサイトマップを人手で生成するのは手間のかかる作業である。かといってマップの更新を怠ると、マップのリンク先の消滅や、マップからたどることのできないドキュメントの発生によるサービスの低下につながる。

また、マップの提示方法に問題のある場合も多く、ポータルサイトのトップページで目次的に表示されるだけのサイトマップや、同じポータル内であるにもかかわらずコンテンツごとに異なる表示方式のサイトマップが提供されているサイトが多く見受けられる。

本稿では、ポータルサイト提供者によるマップ作成の工数削減とサイトを訪れるユーザの利便性を向上することを目的に開発したサイトマップシステムについて説明する。

## 2. サイトマップの要件

大規模なコンテンツを持つサイトでは、サイトで管理されるドキュメントをそのカテゴリにしたがって目次の形式で表示している。本稿ではこのような大量のドキュメント群をカテゴリ化し、見やすい形式で表示したものをサイトマップと呼ぶことにする。

### 2.1. 課題

従来のサイトマップには以下のような課題がある。

1. 多くの場合人手で作成されており、サイト構造とマップユーザインタフェースの作成には非常に手間がかかる。その一方で全自動で利便性の高いサイト構造を抽出することは非常に困難である。
2. サイトマップがコンテンツと同一ウィンドウ上や、同一ウィンドウの別フレーム上に表示されることが多い。このため、コンテンツを閲覧するとマップ表示が消えてしまったり、マップの表示領域が極端な縦長、もしくは横長に制限されることが多い。
3. ポータルサイトのような大規模サイトのドキュメントを構造化しても構造そのものが大きく複雑になりがちである。こういった構造全体を一度に表示した場合、非常に大きな画面領域が表示に必要となる一方でユーザにとって不必要な情報も詳細に表示されてしまう。

4. ユーザのポータルサイトへのアクセスを考えた場合、まず、新着情報の有無を調べ、その後、決まったページを巡回するというパターンが多いと考えられる。しかし、画一的なサイトマップ表示では、あるユーザの閲覧頻度が高いドキュメントがマップ上でアクセスしづらい場所にあたり、既知の情報を繰り返し新着情報として表示する可能性がある。

### 2.2. 解決の方針

2.1節の課題に対して以下のような解決の方針をとることにした。

1. サイト内ドキュメントの自動探索と、サイト管理者による部分的なサイト構造指定を組み合わせることでサイト構造を自動生成し、それをもとにサイトマップを生成する。
2. サイトマップをサブウィンドウへ表示する。これによって、常にマップが表示されているためユーザの利便性が高い。また、必要に応じてサブウィンドウの大きさ、形状を変えることができるので形状の制限も小さい。
3. サイト構造を階層化し、初期表示は概要構造のみを表示する。ユーザの要求に応じて必要な部分の下位構造を詳細化して表示する機能を実装する。
4. 個々のユーザのアクセス履歴を元にマップ表示を調整するパーソナライズ機能を実装する。

以上から、本サイトマップシステムの要件は次のようになる。

- サイトマップメンテナンスの自動化：  
サイト管理者によるサイト構造の部分指定とプログラムによる自動解析を組み合わせたサイトマップ生成
- サイトマップユーザインタフェース
  - サブウィンドウによるマップ表示
  - サイト構造の詳細度可変表示
  - アクセス履歴によるパーソナライズ機能

### 3. サイトマップ・ユーザインタフェース

#### 3.1. サブウインドウ表示

図 1はサイトマップの概観である。大きいブラウザウインドウが閲覧用ウインドウ、左下のサブウインドウがサイトマップである。ユーザはサブウインドウに表示されたサイトから、閲覧用のウインドウ上へ必要なページを表示することができる。

このサブウインドウ表示でポータルサイトのマップを画面中に常駐させることができ、サイト内ナビゲーションをスムーズに進めることができる。

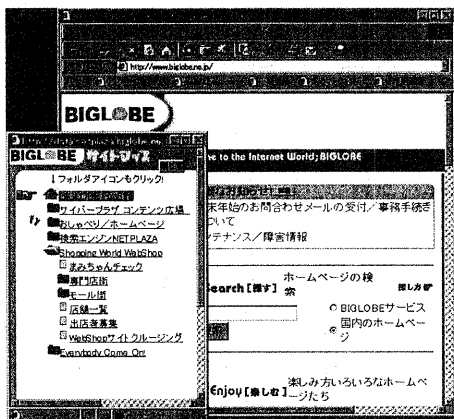


図 1：サイトマップ概観

#### 3.2. 階層構造表示

図 2はサイトマップユーザインタフェースの例である。本サイトマップは階層構造をなすサイト構造をフォルダツリーの形式で表示する。ユーザはフォルダアイコンをクリックすることによって、そのディレクトリの下位構造の表示を展開、縮退することができる。これによって、サイトマップを比較的小さい画面領域に表示することができる。

下位構造を持たないエントリにはフォルダアイコンではなく、ファイルアイコンが付与される。アイコンの右側のテキストはそのディレクトリのコンテンツを示すタイトルであり、ユー

ザはこのテキストをクリックすることによってそのコンテンツの代表ページを閲覧ウインドウ上へ表示することができる。

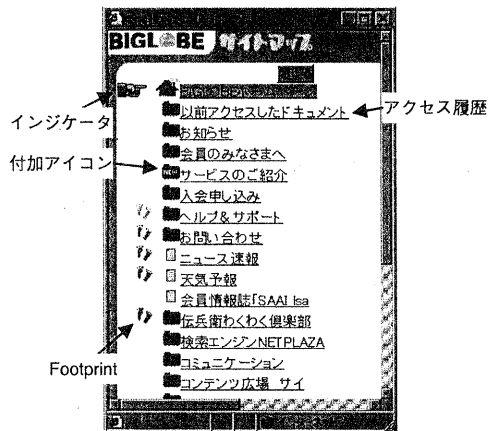


図 2：サイトマップの構成

また、本ユーザインタフェースは以下の機能によりユーザのサイト内ナビゲーションを支援する。

#### ■ 付加アイコン：

サイトマップ生成時にそれぞれのドキュメントに付加情報としてアイコンを貼り付けることができる。図の例では新着情報を示すアイコンが貼られている。これ以外にも推薦ドキュメント、人気コンテンツといった付加情報のアイコン表示が可能である。

#### ■ インジケータ：

図中の指アイコンがインジケータである。サイトマップからドキュメントを取得すると、マップ上のそのドキュメントの位置までインジケータが移動し、取得したドキュメントを示す。

#### ■ Footprint：

サイトマップからアクセスしたドキュメントに足跡アイコンが残る。この足跡アイコンはユーザが次々と新たなドキュメントへアクセスして新しい足跡アイコンが付くにつれてかすれていき、やがて古いものから順に消える。短いアクセス履歴機能として利用することができる。

### 3.3. パーソナライズ機能

本サイトマップは、以下のパーソナライズ機能を持つ。

#### ■ アクセス履歴の表示：

ユーザが前回サイトマップを利用した際に閲覧したドキュメントの履歴を図 2のアクセス履歴用のフォルダにまとめて格納する。サイト内の巡回コースがパターン化しているユーザのナビゲーション支援に有効であると考えられる。

#### ■ 新着情報の表示調整：

サイトマップは新着情報のうち、ユーザが前回サイトマップを利用したとき以降に付加された新着情報のみにアイコンをつける。これによって、前回のアクセス時以降に更新された情報のみを新着情報として扱うことができる。

#### ■ マップ表示位置の記憶：

サーバからサイトマップ UI のダウンロード後、初期設定の際にサイトマップを前回表示した位置へ復元する。本サイトマップユーザインタフェースは画面中に複数のウィンドウを開くため、この機能によってユーザのサブウィンドウ移動の手間を省くことを目的としている。

このパーソナライズ機能によって利用者ごとの利用状況に応じたサイトマップを提供することができる。

## 4. サイトマップの生成

図 3は本サイトマップナビゲーションシステムの構成を示すブロック図である。以下ではサイトマップ生成までの手順について説明する。

### 4.1. サイト構造の生成

本サイトマップシステムにおいて、サイト構造はドキュメント URL のディレクトリ構造から生成する。これは、最終的な人手によるチェックと修正を考えた場合、より複雑な方法で時間をかけて生成するよりも簡便な方法による高速な自動生成の方がよいと考えたからである。

まず図 3のサイト構造情報生成部ではクローラ等で収集されたサイト内の WWW ドキュメントアーカイブから、ドキュメントの URL ディレクトリを単位とした階層構造を生成する。それぞれのディレクトリの内容を表すファイルとして、そのディレクトリ内のファイルから代表ファイルの一つを選ぶ。マップ上では、この代表ファイルのタイトルをそのディレクトリのコンテンツタイトルとして表示する。

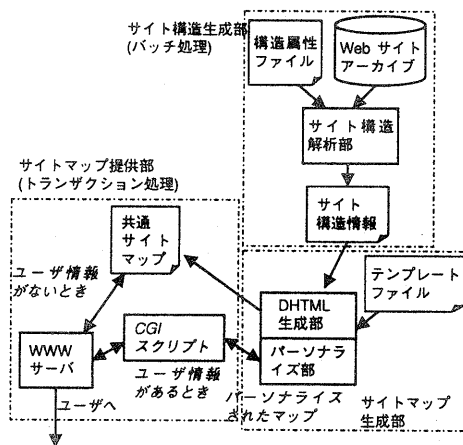


図 3：システム構成図

しかし、単純に URL のディレクトリ構造のみからでは利便性の高いマップは生成できない。そこで、サイト管理者は構造属性ファイルによって自動的に生成されるサイト構造に手を加えることができる。このファイルでは以下の構造属性を指定できる。

- 表示しないディレクトリ名、サーバ名
- 複数のディレクトリ、サーバのクラスタ化
- ディレクトリタイトル、代表ファイルの指定

サイト構造情報の生成方法を図 4のリストを例に説明する。各行はこのサイト内の URL ディレクトリとその代表ファイルおよびタイトルの組を示している。このドキュメントリストに対して以下のような構造属性がサイト管理者によって定義されていると仮定する。

- /Bookmark/Entertainment のタイトルは「娯楽」とする。

- 「サービス案内」というクラスタを定義し、その下位構造に/AP、/ryokin ディレクトリをクラスタリングする

```

http://www.foobar.ne.jp/, FoobarWWW サイト
  whatsnew/index.html, 新着情報
  whatsnew/Jul99/index.html, 7月分
  whatsnew/Jun99/index.html, 6月分
  AP/index.html, アクセスポイント情報
  ryokin/index.html, 料金体系
  Bmark/Bus/index.html, ビジネス
  Bmark/Ent/index.html, エンターテイメント
  Bmark/Com/index.html, コンピュータ
http://search.foobar.ne.jp/, Foobar 検索サービス
  
```

図 4：サイト内ドキュメントリストの例

この構造属性を元に生成したサイト構造情報が図 5 である。

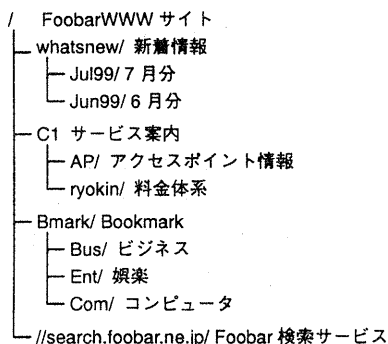


図 5：生成された構造

## 4.2. サイトマップの生成

4.1節の手順で生成されたサイト構造情報をもとにサイトマップユーザインタフェースを生成する。今回の実装では、インタラクティブな展開と縮退を実現するために Dynamic HTML を利用している。

しかし、DHTML でサイト全体のマップをあらかじめ生成するとマップ自体のサイズが非常に大きくなってしまふ。そこで、本サイトマップではサイト構造データと Javascript のプログラムのみをクライアントへ転送し、受信後、Javascript で実際に表示する DHTML をクライアントのブラウザ上で生成している。

## 5. 評価

この節では本サイトマップシステムを用いた実験の結果と今後の課題について考察する。

### 5.1. 利用実験

#### 5.1.1. 実験条件

実験では、ポータルサイト BIGLOBE[5]のサイトマップを本システムにより生成し、11名のユーザに BIGLOBE 内の指定コンテンツを取得する設問を 5 問与えた。それぞれの設問についてサイトマップを使用した場合と不使用の場合を比較した主観的な評価、指定コンテンツ発見までにネットワークを介して取得したドキュメントページ数を比較した。

与えた設問の特徴は以下のとおりである。

1. 単一のページの発見
2. 複数のページの発見
3. 同一コンテンツ内のナビゲーション：サイト内で提供される同一ウインドウ型のコンテンツ内マップの利用が可能
4. サイト構造的に離れた別コンテンツのナビゲーション
5. 既にアクセスしたページの利用：既にアクセスしたコンテンツへの再アクセス

#### 5.1.2. 実験結果

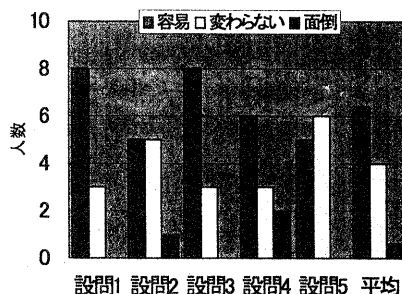


図 6：使用感主観評価

図 6はそれぞれの設問でのサイトマップに対する主観評価のグラフである。3種類の棒グラフはそれぞれ、サイトマップを使うことでナビゲーションが「容易になった」「変わらない」「面倒になった」人数を示している。

設問5を除いて、サイトマップによりナビゲーションが容易になったという主観評価が最も多かった。設問5では変わらないという意見が多かったが、マップ内のユーザ履歴を格納する機能に気づいた人が少なかったことが一因であると考えられる。

図7はそれぞれの設問において、サイトマップを使用した場合としなかった場合に設問終了までに取得した総ページ数の平均である。

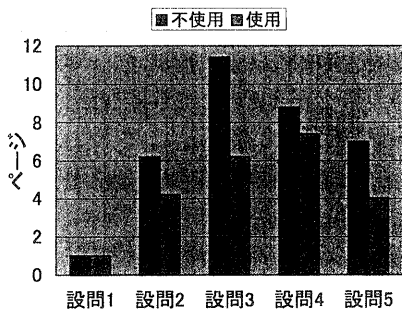


図7: 総転送ページ数

図から、本サイトマップを使用することによって、実際に取得するページが少なくなっている。これは、マップのアクセスのみで求めるページを探索することができるが要因であると考えられる。これによって、ナビゲーションの際のダウンロード待ち時間やネットワーク負荷が減ると考えられる。

実験の結果、本サイトマップUIによってポータルサイト利用者のサイト内ナビゲーションを効率化することが可能であると考えられる。

## 5.2. サイト構造の自動抽出

現状のシステムにおいて、簡便なサイト構造の自動抽出機構では、サイト管理者がかなりの量の構造属性を指定する必要がある。

この理由としては、大規模サイトの場合多くのドキュメントが様々な管理者によって追加されているため、同一コンテンツの内容が複数のディレクトリに渡って格納されていたり、他サーバへのリンクになっていたりするケースがしばしば見られた。

このため、サイト管理者のマップ管理にかかる手間を十分小さくできたとは言えない。サイト管理者の構造指定の手間を減らすために、リンクの参照関係[4]を利用したり、他ドキュメントとの関係記述をコンテンツ作成時にドキュメント内へ埋め込むなどの方法によって利便性の高いドキュメント構造の自動抽出機構の導入が必要であると考えられる。

## 5.3. その他

実験の感想の自由記述やサイトマップ試用者からの意見として以下のものがあった。

- マップウインドウが閲覧ウインドウの裏側に隠れてしまって使いづらいことがある。
- 詳細構造表示を展開しつづけると表示が複雑になるので、ずっと前に展開した部分を自動的に縮退させる機能が欲しい。
- タイトル表示だけでは内容がわかりづらい。

## 6. おわりに

我々はポータル利用者が大量のコンテンツに容易にアクセスするためのサイトマップナビゲーションシステムを開発した。この特徴は以下のとおりである。

- サイト構造の自動解析
- グラフィカルなサイトマップの提供
- パーソナリゼーション

また、本サイトマップによるナビゲーション実験を行い、サイト内ナビゲーションを効率化することが可能であることがわかった。

## 参考文献

- [1] 島村 栄, 神場知成, 古関義幸, 「WWW サイトマップナビゲーションシステム」, 第59回情報処理学会全国大会講演論文集3-89, 1999年9月
- [2] 島村 栄, 高野 元, 「WWWドキュメント検索におけるドメイン名クラスタリングの利用」, 情報処理学会研究報告98-HI-77, 1998年3月
- [3] Sougata Mukherjea, Yoshinori Hara, "Focus+Context Views of World-Wide Web Nodes", Proceedings of ACM Hypertext'97, April 1997.
- [4] Loren Terveen, Will Hill, "Finding and Visualizing Inter-site Clan Graphs", CHI98, April 1998.
- [5] "BIGLOBE", NEC, <http://www.biglobe.ne.jp/>